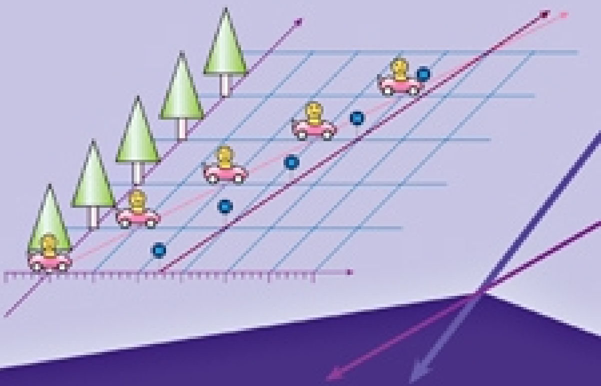


An Illustrated Guide to Relativity

Tatsu Takeuchi

CAMBRIDGE

This page intentionally left blank

An Illustrated Guide to Relativity

Aimed at both physics students and non-science majors, this unique book explains Einstein's Special Theory of Relativity pictorially, using diagrams rather than equations. The diagrams guide the reader, step-by-step, from the basics of relativity to advanced topics including the addition of velocities, Lorentz contraction, time dilation, twin paradox, Doppler shift, and Einstein's famous equation $E = mc^2$. The distinctive figures throughout the book enable the reader to visualize the theory in a way that cannot be fully conveyed through equations alone.

The illustrative explanations in this book maintain the logic and rigor necessary for physics students, yet are simple enough to be understood by non-scientists. The book also contains entertaining problems which challenge the reader's understanding of the materials covered.

TATSU TAKEUCHI is an Associate Professor in the Department of Physics at Virginia Tech. This book grew from the "Highlights of Contemporary Physics" course he taught for many years. Primarily aimed at non-physics majors, it has been highly popular among physics students as well.

An Illustrated Guide to Relativity

TATSU TAKEUCHI
Department of Physics, Virginia Tech

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521763943

First published in print format 2010

ISBN-13 978-0-511-90058-7 eBook (EBL)

ISBN-13 978-0-521-76394-3 Hardback

ISBN-13 978-0-521-14100-0 Paperback

Contents

Preface to English edition *page* viii

Preface to Japanese edition ix

Part I Kinematics: Relativity without any equations 1

1 Welcome to the world of relativity 2

2 Basics 4
- 2.1 Questions about motion 4
- 2.2 Frames of reference 5
- 2.3 Relativity of motion 12
- 2.4 The Law of Inertia 14
- 2.5 Inertial and non-inertial frames 18
- 2.6 What's so "special" about Special Relativity? 26

3 Galilean relativity 30
- 3.1 Basic questions 30
- 3.2 Spacetime diagrams 34
- 3.3 The Galilei transformation 40
- 3.4 Addition of velocities 54
- 3.5 Acceleration and Newton's Second Law 56

4 Einsteinian relativity 59
- 4.1 The mystery of the speed of light 59
- 4.2 Modification to the spacetime diagram 64
- 4.3 The problem 66
- 4.4 The solution 72
- 4.5 Einstein's argument 80
- 4.6 The solution, continued 85
- 4.7 Conservation of spacetime volume 88
- 4.8 The Lorentz transformation 94
- 4.9 The low velocity limit of the Lorentz transformation 108

4.10 Addition of velocities 110
4.11 Dependence of inertia on speed 116

5 Causality 120
5.1 Before and after 120
5.2 Paradox? 122
5.3 Instantaneous communication? 124
5.4 Impossibility of faster than light travel 126
5.5 The light-cone 128

6 Consequences 130
6.1 Synchronization of clocks 130
6.2 Time dilation 132
6.3 What time dilation DOES NOT mean 138
6.4 Lorentz contraction 140
6.5 What Lorentz contraction DOES NOT mean 146
6.6 Twin paradox 148
6.7 Doppler effect 154
6.7.1 Red shift 156
6.7.2 Blue shift 158
6.7.3 Red shift and the expansion of the universe 160

7 Summary of Part I 162

Part II Problems 165

8 Qualitative problems 166
8.1 Reading the spacetime diagram 166
8.1.1 Street lamps 166
8.1.2 Supernovae 168
8.2 Questions on before and after 170
8.2.1 The hare and the tortoise 1 170
8.2.2 The hare and the tortoise 2 172
8.2.3 The hare and the tortoise 3 174
8.2.4 The starship and the supernova 176
8.3 Relativistic sports 178
8.3.1 Tagging up in baseball 1 178
8.3.2 Tagging up in baseball 2 180
8.3.3 The offside rule in soccer 182
8.4 Lorentz contraction 184
8.4.1 Train and tunnel 184
8.4.2 The starship and the enemy space cruiser 1 186

8.4.3 The starship and the enemy space cruiser 2 188
8.4.4 The duel of the space cruisers 190
8.4.5 Trains in a tunnel 192
Solutions to Chapter 8 problems 194

9 Quantitative problems 200
9.1 Addition of velocities 200

Part III Dynamics: Relativity with a few equations 207

10 The world's most famous equation 209

11 The problem 210

12 Newtonian dynamics 212
12.1 The mass–momentum vector 212
12.2 The impulse vector 218
12.3 Inertial mass 220
12.4 Newton's Second Law 222
12.5 Newton's Third Law and the conservation of mass–momentum 224

13 Relativistic dynamics 230
13.1 The energy–momentum vector 230
13.2 The energy–momentum vector of a photon 236
13.3 The work–impulse vector 240
13.4 Conservation of energy–momentum 244
13.5 $E = mc^2$ 246
13.6 Common misconception about $E = mc^2$ 248

14 Summary of Part III 251
Afterword 252
References 254
Index 255

Preface to English edition

This book explains Einstein's Special Theory of Relativity (SR) using diagrams only. Readers who are used to thinking of physics as a vast labyrinth of equations may feel somewhat uneasy about this unconventional approach and fear that it risks losing important information about SR that can only be conveyed via equations. However, this fear is not only unfounded but actually reversed: it is the equations that fail to convey the essence of SR that diagrams can easily display right in front of your eyes. After all, SR, and also the General Theory of Relativity (GR), are about the *geometry* of the spacetime that we inhabit, and what can best describe geometry if not diagrams? Equations are simply inadequate, to wit, one diagram is worth a thousand equations.

So if you are a reader for whom equations are anathema, rest assured that you will get much more out of this book than any physics student will get out of a textbook full of equations. If you are a physics student, this book will provide you with a deep enough understanding of SR that will enable you to reproduce any equation you may need from scratch, if such a need ever arises, and also prepare you for GR as well.

I would like to thank the readers of the Japanese [1] and Chinese [2] editions of this book who have provided precious feedback and encouragement through their reviews, blogs, and email, and have given me added confidence that the approach of this book is the right one. I would also like to thank my students at Virginia Tech whose constant desire to read this book in English motivated me to translate part III from Japanese. (I wrote parts I and II in English to begin with.)

Special thanks are due to my editors at Cambridge University Press: John Fowler for his enthusiastic support of this project, Lindsay Barnes for her meticulous attention to detail in making sure all the text and figures were in order, and Abigail Jones for guiding me through the production process. It was a great joy working with them all.

January, 2010
Tatsu Takeuchi

Preface to Japanese edition

All physical theories, their mathematical expressions notwithstanding, ought to lend themselves to so simple a description that even a child could understand them.

Albert Einstein [3]

Einstein's celebrated Theory of Relativity is one of those scientific theories whose name is so famous that most people have heard of it, but very few people actually know what the theory says, or even what the theory is about. You, too, have probably heard the name, perhaps referred to in a science fiction novel or movie, even if you do not know much about it. And you may have received the impression that it is a very esoteric and difficult theory that could only be understood and appreciated by a select few.

The aim of this book is to show you that that impression is wrong. The Theory of Relativity comes in two flavors, the Special and the General, and if we limit our attention to the Special Theory of Relativity (SR), which is a theory of motion, it is not a particularly difficult theory at all and can be understood by anyone, perhaps "even a child." By "be understood" here, I do not mean that anyone can develop a vague idea of what the theory is saying, but that anyone can understand it in its full glory beginning from its basic tenets to all of its logical consequences. And furthermore, it can be understood without using ANY equations! In fact, *one can develop a deeper understanding of the theory by avoiding the use of equations altogether*. At least, this author thinks so.

Then, why is it that I do not declare that SR is an easy theory outright? The reason is that SR makes some statements about the concept of simultaneity which do not agree with our common sense based on everyday experience, and this is where a slight difficulty lies: we must listen to what SR is telling us with an open mind and not let our common sense obstruct our understanding.

This book is an attempt to explain Einstein's Special Theory of Relativity (SR) without using equations. Instead, we will use drawings called *spacetime diagrams* in a way that will let you "*see*" the essence of the theory. This book has three parts. Part I explains why SR was constructed, what it is telling us, and why SR had to be the way it is.

Part II is a collection of problems. The problems are designed so that by thinking about them you will confront the common sense that hampers our understanding of SR, and know where the pitfalls are that may mislead us to think that there must be something wrong with SR. Part III deals with the famous equation $E = mc^2$. Since $E = mc^2$ itself is an equation, the appearance of some equations in part III could not be avoided. However, the basic argument proceeds via diagrams and can be followed without following the equations. The General Theory of Relativity, which is a theory of gravity, can, in principle, also be explained using drawings only. However, since it is a somewhat complicated theory, we will not be discussing it in this book.

Part I of this book is based on lecture notes I prepared for the course "Highlights of Contemporary Physics" which I have taught at Virginia Tech for many years. Part II consists of the problems I have used for the exams in this course. I thank all the students who have taken this course and have served as guinea pigs to see if the approach of this book works. I would also like to thank Djordje Minic, Joy Rosenthal, Simone Paterson, and Masako Saitō Koike for their critical reading of this manuscript and their many helpful suggestions.

Special thanks are due to Tōru Kawahara of Iwanami Shoten for his enthusiasm in bringing this book to publication. Without his help in digitizing my hand-drawn diagrams, and translating the characters' words in the drawings into Japanese, this book would never have been completed.

August, 2005
Tatsu Takeuchi

PART I

Kinematics: Relativity without any equations

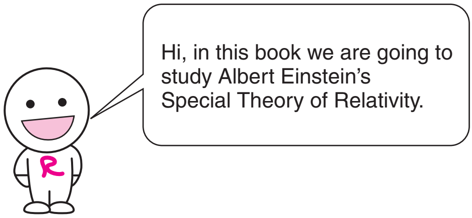
1

Welcome to the world of relativity

Albert Einstein's Special Theory of Relativity, or Special Relativity for short, came into being in 1905 in a paper with the unassuming title of "On the electrodynamics of moving bodies."[1] As the title suggests, Special Relativity is a theory of "moving bodies,"[2] that is: motion. In particular, it is a theory of how motion is perceived differently by different observers. Since motion is the process in which an object's location in *space* changes with *time*, any theory of motion is also a theory of *space* and *time*. Therefore, Special Relativity can be said to be a theory of how *space* and *time* are perceived differently by different observers. The "electrodynamics" part of the paper title refers to the fact that the theory has something to do with *light*, which is an electromagnetic wave. As we will learn in this book, the speed of light in vacuum, which we will call c,[3] plays a very special role in the theory of relativity.

Einstein (1879–1955) was not the first to construct a successful theory of motion. Building upon pioneering work by Galileo Galilei (1564–1642),[4] Sir Isaac Newton (1642–1727) had constructed theories of motion and gravity which were spelled out in his famous book *Philosophiae Naturalis Principia Mathematica*,[5] which is so famous that when people say *the Principia*,[6] it is understood that they are referring to Newton's book. First published in 1687, English translations and commentaries are still widely available in print [5]. Newton's theory worked perfectly well for over 200 years (and still does today in most applications) and succeeded in explaining the motions of objects both in Heaven (the planets, moons, comets, etc.) and on Earth (everything you see around you). However, toward the end of the nineteenth century, a certain mystery was discovered concerning the speed of light which could not be understood within the Galilei–Newton theory. The theory that provided a clear and illuminating resolution to the mystery, and became the theory to supersede that of Galilei–Newton, was Einstein's Special Theory of Relativity.

In this book, we will look at what that mystery was, why the Galilei–Newton theory was in trouble, how Einstein solved the mystery, and what the consequences of Einstein's discovery were. But before we can do that, we need to learn the basics of the study of motion.



Notes

1 "Zur Elektrodynamik bewegter Körper" in the original German. The German word "Körper" shares the same etymology as the English "corpse." They both come from the Latin word "corpus," which means "body." For an English translation of the paper, see [4].
2 Note that the word "bodies" here does not refer to human bodies, but to objects in general. So the expression "moving bodies" should be interpreted simply as "moving objects."
3 The letter c has been traditionally used as the symbol for the speed of light in vacuum because it is the first letter in *celeritas*, which is the Latin word for speed.
4 Galileo's discoveries concerning motion are described in his book *Dialogues Concerning Two Sciences*, the English translation of which can also be found in [4].
5 In English, the title translates to *Mathematical Principles of Natural Philosophy*. Both the title and content were in Latin.
6 The "c" in "Principia" is hard and should be pronounced like a "k."

2

Basics

2.1 Questions about motion

When studying the motion of an object, the most basic questions we would like to ask are things like:

Q1: Is it moving or is it at rest (not moving)?
Q2: If it is moving, what is its direction of motion?
Q3: What is its speed in that direction?

In order to answer these questions, we need to know the object's *location in space* at each *instant in time*, so that we can keep track of how it is changing as *time* progresses. If we find that the object's location in space is not changing with time, that is, if it stays *at the same place*, then we can say that it is "at rest," while otherwise we can say that it is "moving." If the object is "moving," we can specify its direction of motion by saying things like "it is moving to the left" or "it is moving to the right," and we can figure out its speed by determining by how much its location in space is changing per unit time.

Once we have these basic questions under control, we can then start to ask more advanced questions like:

Q4: Is the direction of motion changing with time?
Q5: Is the speed changing with time?
Q6: If they are changing, what is causing it?

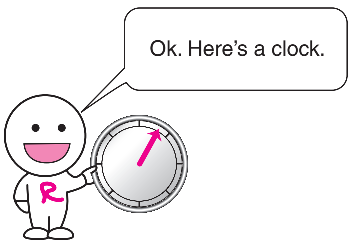
and so on. Questions 2 and 3, and questions 4 and 5, are often combined into the questions

Q2+Q3: If it is moving, what is its *velocity*?
Q4+Q5: Is the *velocity* changing with time?

The term *velocity* means something like "speed and/or direction of motion." When we say "the *velocities* are the same" we mean that both the speeds and the directions of motion are the same. When we say "the *velocities* are different" we mean that either the speeds or the directions of motion, or both, are different.

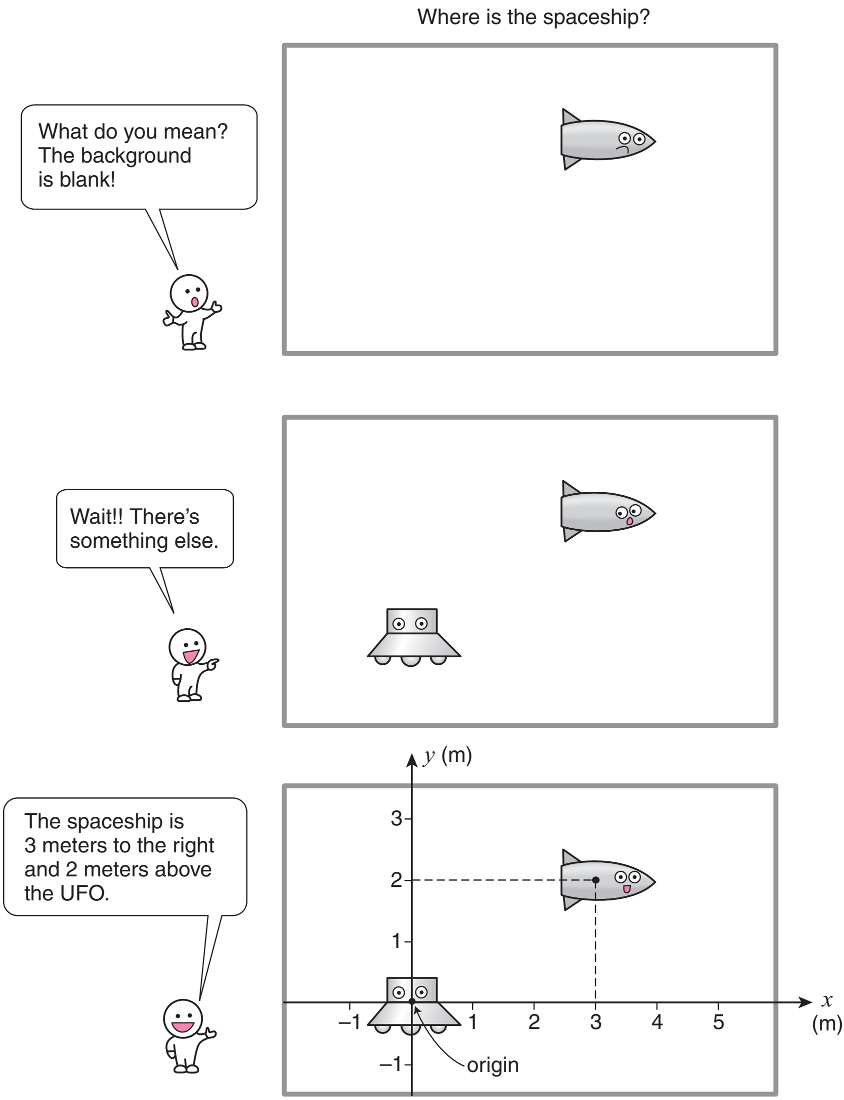
2.2 Frames of reference

Now, what are the tools we need to actually keep track of how an object's *location in space* is changing with *time*? First, and foremost, we obviously need a *clock* to keep track of the *time*. The reading of the clock will give us a number t, which labels each *instant in time*. If we are using SI units,[1] t will be given in seconds. We will be using a different unit later on.



Next, we need to specify the object's *location in space* at each *instant in time*. But the only way to do this is to specify it *relative* to something else. To see this, imagine a spaceship traveling in an empty universe which is devoid of anything else but the spaceship. I am sure you would agree that it would be impossible to state *where* the spaceship is. In fact, the concept of *location* itself does not make much sense in that situation. It is only when there are stars or planets or some other object, like another spaceship, present in the universe that we can specify the location of the spaceship *relative* to the other object.

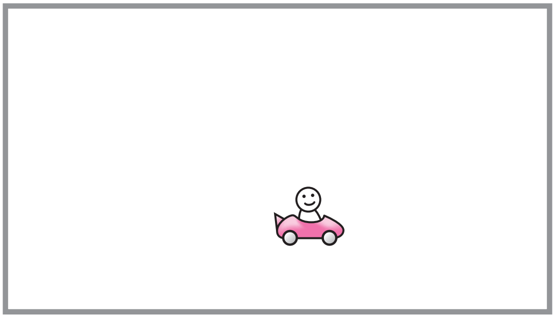
To be more specific, what we need is *a reference point* fixed to some object to define our positions from. Once we have chosen such a reference point, which we will call the *origin*, we can specify the location of the object by saying, for instance, that the object is x meters to the east, y meters to the north, and z meters up from the origin. In other words, we can attach a set of numbers (x, y, z) as a label to each point in space. Of course, we need a ruler to measure these distances, and we also need to specify (or be able to specify) what we mean by east–west, north–south, and up–down.


Where is the spaceship?
What do you mean? The background is blank!
Wait!! There's something else.
The spaceship is 3 meters to the right and 2 meters above the UFO.
y (m)
3
2
1
–1
–1
1
2
3
4
5
x (m)
origin

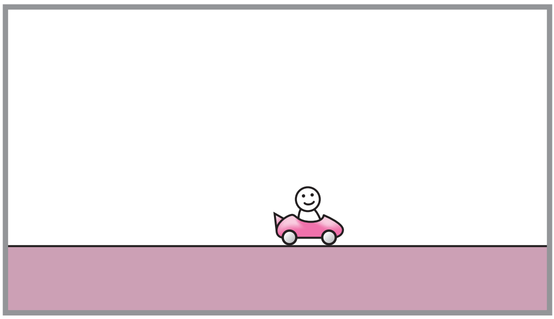
As an example, consider the motion of a car moving along a straight horizontal road as shown in the figure. To specify the position of the car, we can use its distance from some fixed object along the road, say a tree, and say, for instance, that "the car is 5 meters to the right of the tree," or "the car is 3 meters to the left of the tree." To simplify things, we can assign positive numbers to the distance from the tree when the car is to the right of the tree, and negative numbers to the distance when the car is to the left of the tree and say "the car is at $x = +5$ meters" instead of "the car is 5 meters to the right of the tree," and "the car is at $x = -3$ meters" instead of "the car is 3 meters to the left of the tree." This frees us from the need to specify "left" or "right." This procedure allows us to assign a number x (in meters) to the position of the car. The tree, in this case, is the *origin*.

When we have a clock to specify the time, and the origin and the directions in which to measure the distance from the origin set up, we say that we have a *frame of reference*, or simply a *frame*. A *frame of reference* allows us to specify any point in *space* and *time* with a set of numbers. In the *frame* fixed to the tree, it is the reading of the clock t and the distance from the tree x.

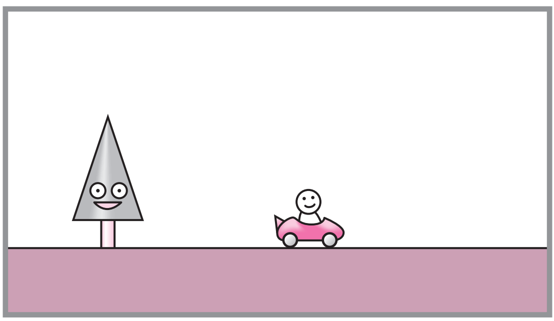
A car...



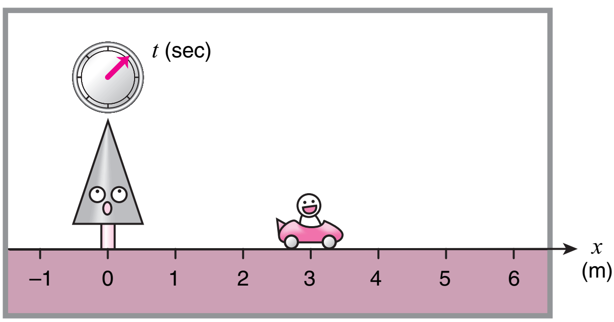
on a straight
horizontal road...



with a tree...



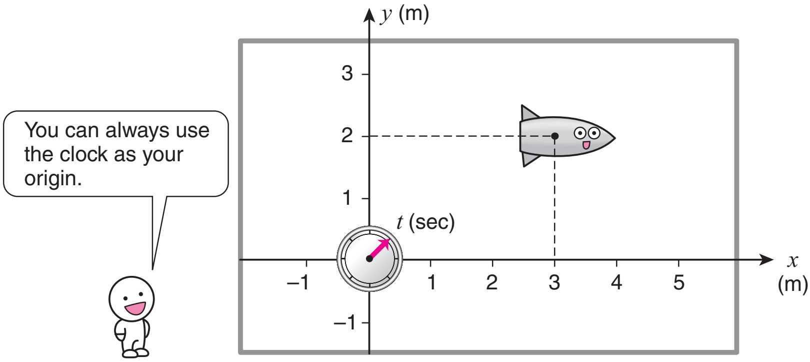
The tree-frame!



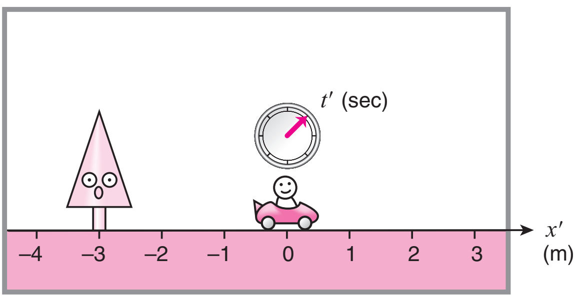
Actually, even in the lone-spaceship-in-the-universe case, there is an *origin* we can use to set up a *frame of reference* since the universe must contain in it, in addition to the spaceship, the clock we use to specify the time (with ourselves to read it, for that matter). It could be that the clock is fixed to the spaceship, in which case we have a *frame* in which the spaceship's position does not change at all with time. It's a boring choice of *frame*, but a *frame* nevertheless.

In effect, the choice of *origin* is equivalent to choosing which object we are going to consider to be at rest, that is "not moving," in that frame. Anything that maintains a constant distance and direction from the origin is also at rest. In other words, the choice of *origin* defines what we mean for an object to be *at the same place*.

Since a clock is always necessary to specify the time, the clock itself is the logical choice for the origin in any frame. We will assume throughout the rest of the book, without comment, and even if the clock is not drawn explicitly in my drawings, that the clock is always fixed at the origin, that is, the clock is always at rest in its frame. In the tree-frame we discussed above to describe the motion of the car, the clock is assumed to be fixed to the tree. If we fix the clock to the car instead, it will move with the car. In that case, we will specify the location of objects relative to the car, and we will have a different *frame of reference* from the tree-frame, which we will call the car-frame.



The car-frame



2.3 Relativity of motion

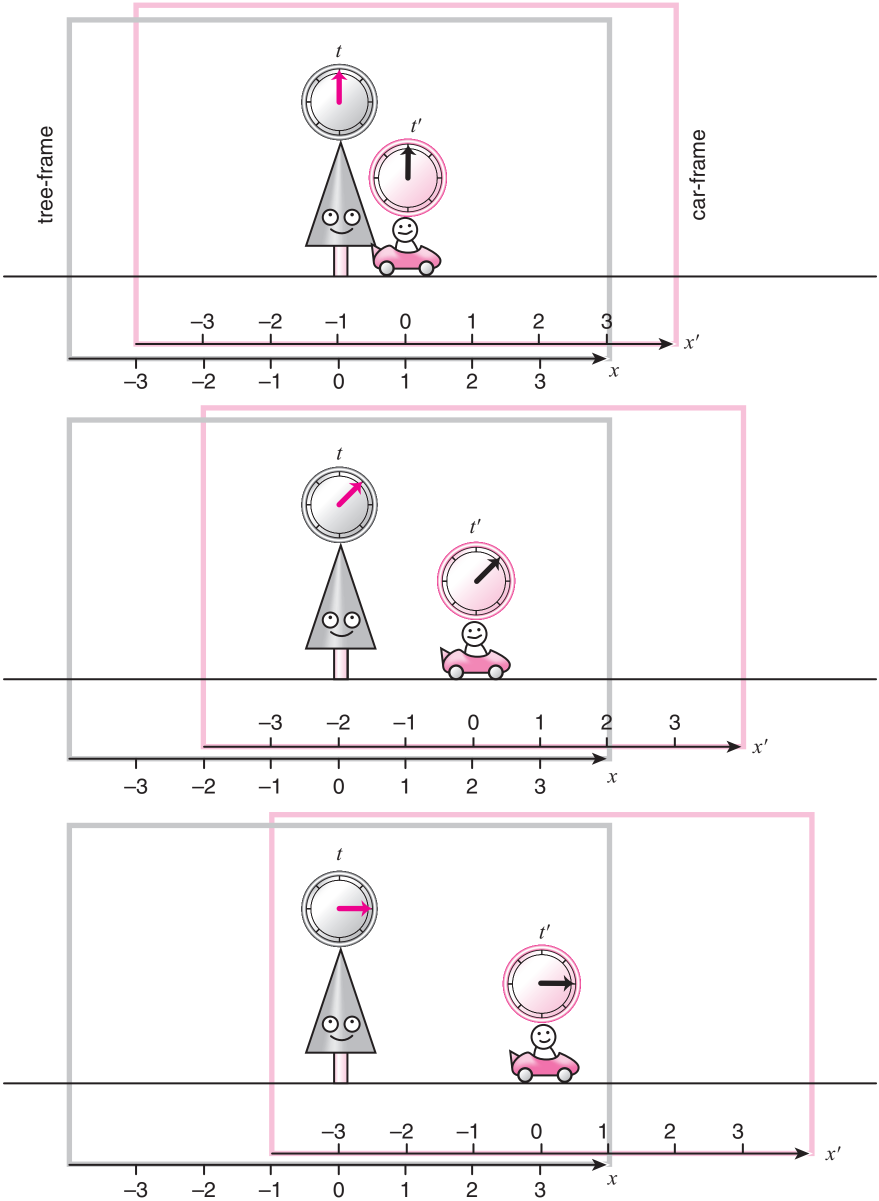
The choice of *frame* to describe the motion of any object is not unique, and, depending on which frame you choose, the answers to the basic questions we listed in section 2.1 would be different, namely:

Q1: Is it moving or is it at rest?
Q2+Q3: If it is moving, what is its velocity?

The answers we come up with in a particular frame will only be correct in that frame. In other words, motion is always *relative*. For instance, in the example of the car driving along a horizontal road, the car is moving but the tree is at rest if observed in the tree-frame, but it is the other way around if observed in the car-frame. This *relativity of motion* can be traced back to the *relativity of position*, that is, to the fact that being *at the same place* is a concept that depends on the observer.

Of course, it should not matter which frame is chosen to describe the motion since what is actually happening to the object is independent of who is doing the observing, or whether there is anyone there to observe it at all.[2,3] In particular, the laws of physics that govern motion must be independent of the choice of frame.[4,5]

However, it can happen that the laws of physics may seem simpler in certain frames than others. This is because the motion that is observed in a particular frame is the motion of the object *relative to the origin*, so, depending on what the origin is doing, the motion can seem more complicated than it really is, or overly simplified.


tree-frame
car-frame
t
t′
−3 −2 −1 0 1 2 3
x′
−3 −2 −1 0 1 2 3
x

2.4 The Law of Inertia

If you ever take physics in school, one of the first laws you learn is the *Law of Inertia*. This law was discovered by Galileo Galilei (1564–1642) and was later incorporated into the theory of motion as one of the guiding principles by Isaac Newton (1642–1727). It is often called Newton's First Law. What the law says is the following:

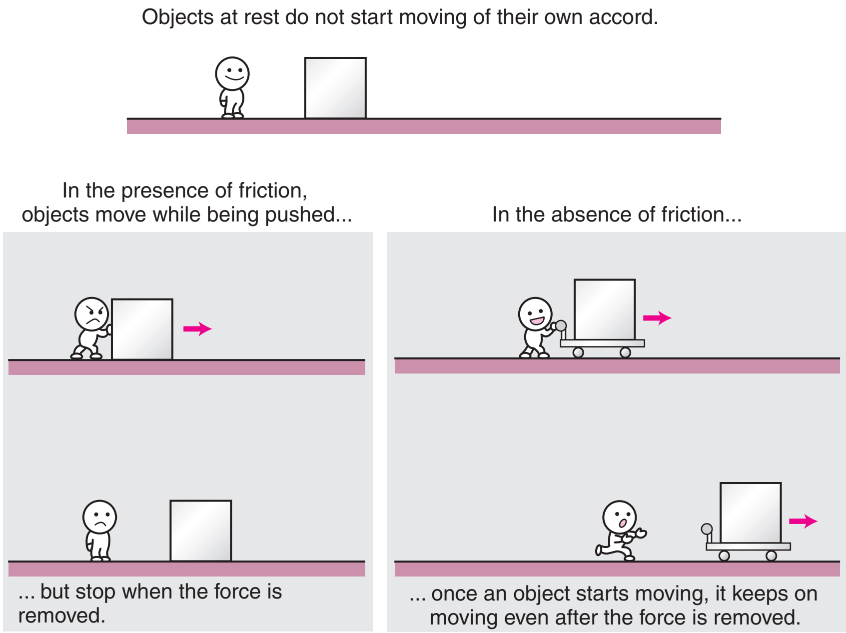
Law of Inertia (aka Newton's First Law):

- If an object is at rest, it will stay at rest unless an external force acts on it.
- If an object is moving, it will keep on moving at the same velocity (that is, in the same direction at the same speed) unless an external force acts on it.

It is a very simple law but it took humans thousands of years to discover it; in particular, the second part.

The first part of the law, that objects at rest will stay at rest unless an external force acts on it, is easy enough to see. Objects do not just suddenly start moving on their own accord.

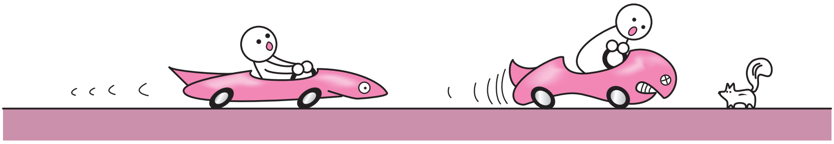
The second part of the law is often obscured by friction, the main reason why it was so difficult to discover. For instance, when you are pushing a heavy object across the floor, it stops moving the moment you stop pushing it instead of continuing to move in the same direction. But this is because friction (an external force) is preventing the object from continuing. If there were no friction between the object and the floor, the object would keep on moving just like the law says even if you stop pushing on it. This can be demonstrated by putting the same heavy object on a cart with well-lubricated wheels. In this case, because the friction between the object and the floor is reduced, once you get the object moving you will find that it is difficult to make it stop or change its direction, confirming the second part of the Law of Inertia. (Make sure no one is in your path before you try this.)


Objects at rest do not start moving of their own accord.
In the presence of friction, objects move while being pushed...
In the absence of friction...
... but stop when the force is removed.
... once an object starts moving, it keeps on moving even after the force is removed.

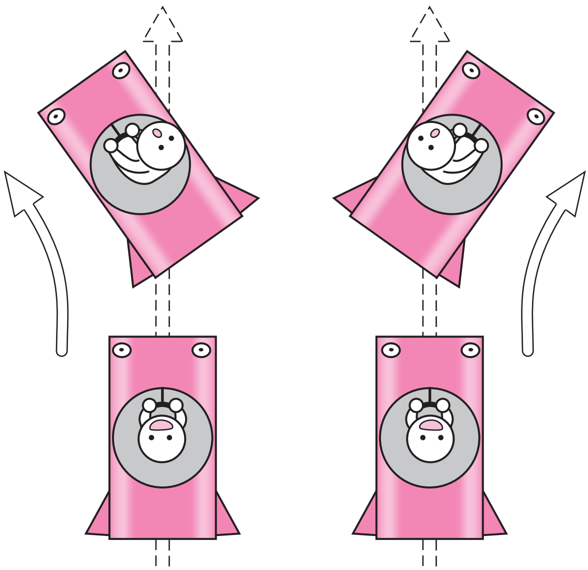
Of course, you do not have to do an experiment like this to confirm this law. You can "feel it" yourself when you are riding a car. When the car is accelerating rapidly you can feel your body being pushed into your seat. This is because your body is resisting the change in its speed. When the car brakes suddenly you can feel your body being thrown forward. This is because your body tries to keep on moving at the same speed as before even though the car under you has slowed down. (And that's why you need to wear a seat belt if you don't want to smash through the windshield.) When the car curves to the left (right), you feel your body being pushed toward the right (left). This is because your body is resisting the change in its direction of motion.

Basically, what the Law of Inertia is saying is that objects resist any change to their velocities. If left alone, an object will continue to move at constant velocity, or, in other words, at a constant speed in the same direction along a straight line. (Being at rest is a special case of this in which the velocity is zero.) This property of objects is called *inertia*.

Your body resists the change in speed.



Your body resists the change in direction.

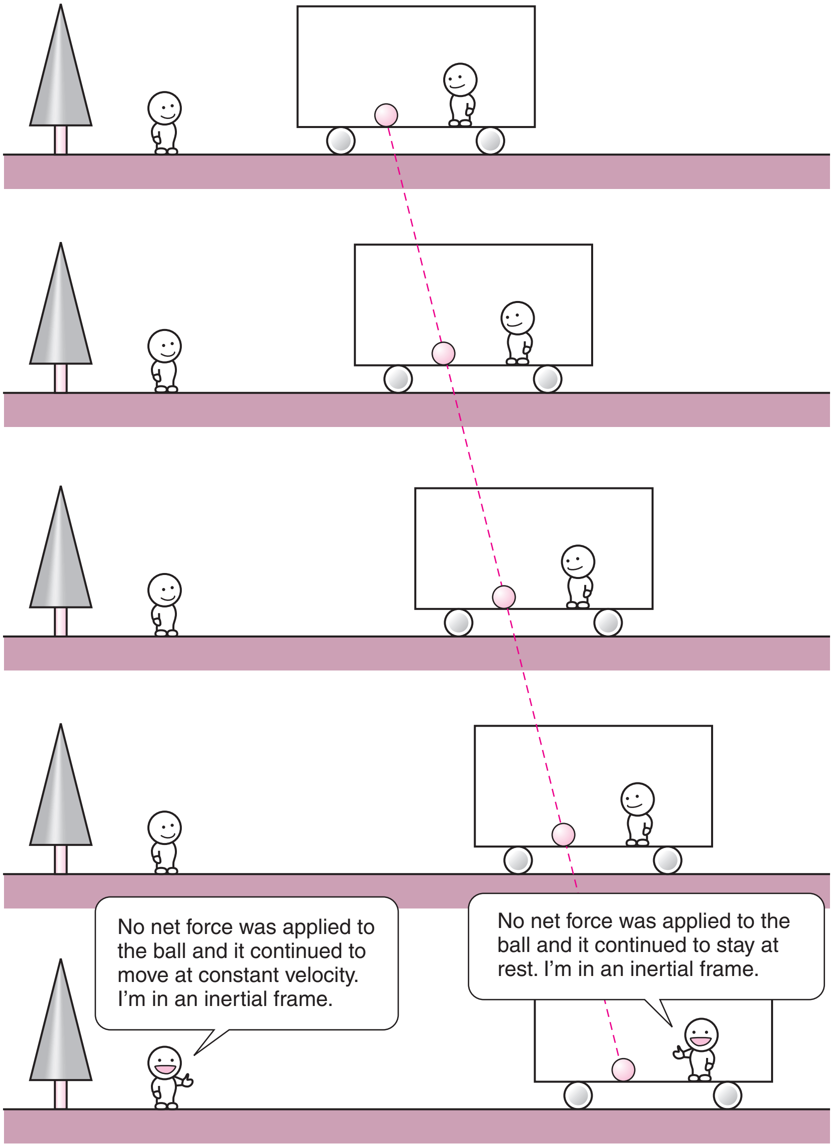


2.5 Inertial and non-inertial frames

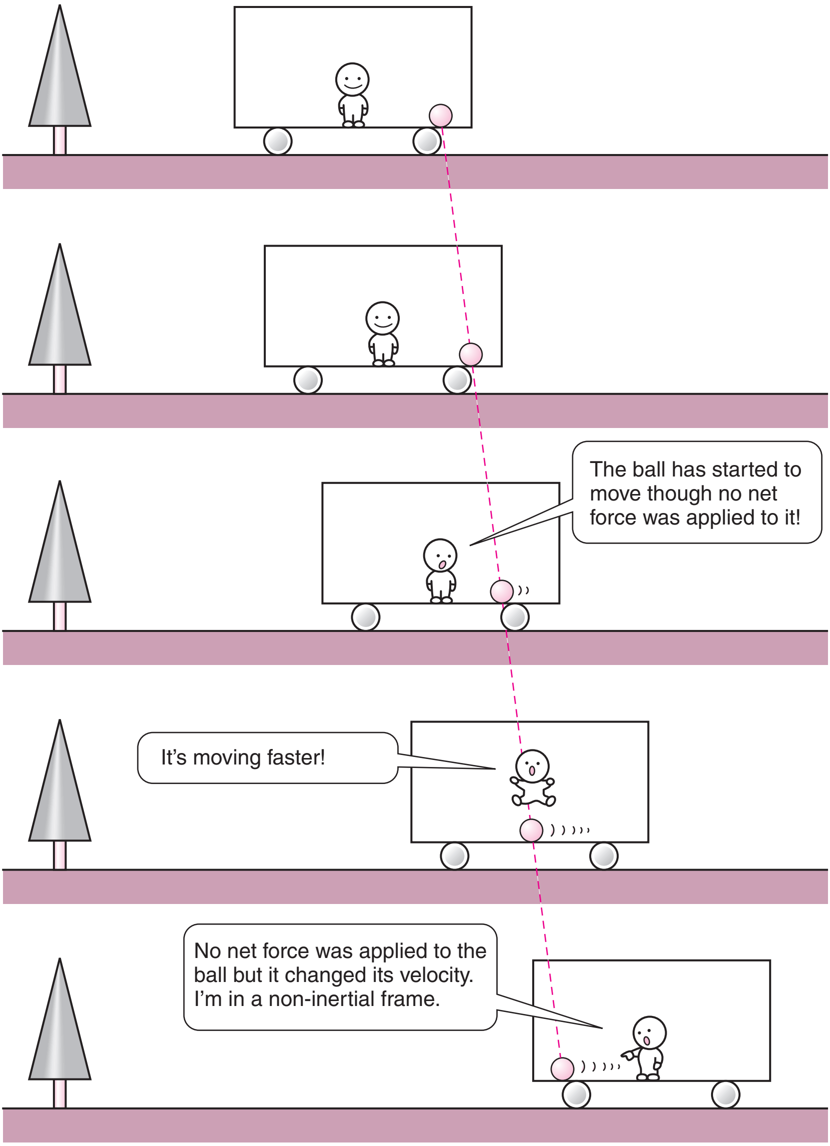
Now, the problem with the Law of Inertia is that it is not correct in all frames. (They probably don't tell you this in school.) For instance, when you are riding your car and using the car-frame to observe the motion of your body, your body will be jerked around when the car accelerates or decelerates or makes a turn even though no forces are acting on your body. Clearly, the Law of Inertia does not hold in the car-frame because the frame itself is changing its speed and/or direction of motion.

The Law of Inertia only holds when the motion is observed from *special* frames called *inertial frames*. If you can find one frame in which the Law of Inertia holds, then any frame which is moving at constant velocity relative to the first is also an inertial frame. This is because if an object is moving at a constant velocity relative to the first frame, then it will be doing so relative to the second frame as well. Therefore, there are an infinite number of inertial frames.

In the example shown in the drawing on the opposite page, the ball is sitting at rest (relative to the train) on the floor of a train which is moving at constant velocity relative to the tree. If the contact between the ball and the train floor is frictionless, the ball receives no net external force. (Gravity is cancelled by the normal force from the floor.) Observed from the tree-frame, it will continue to move at the same velocity as the train. Observed from the train-frame, it will continue to stay at rest. So the Law of Inertia holds in both the tree- and train-frames. They are both inertial frames.


No net force was applied to the ball and it continued to move at constant velocity. I'm in an inertial frame.
No net force was applied to the ball and it continued to stay at rest. I'm in an inertial frame.

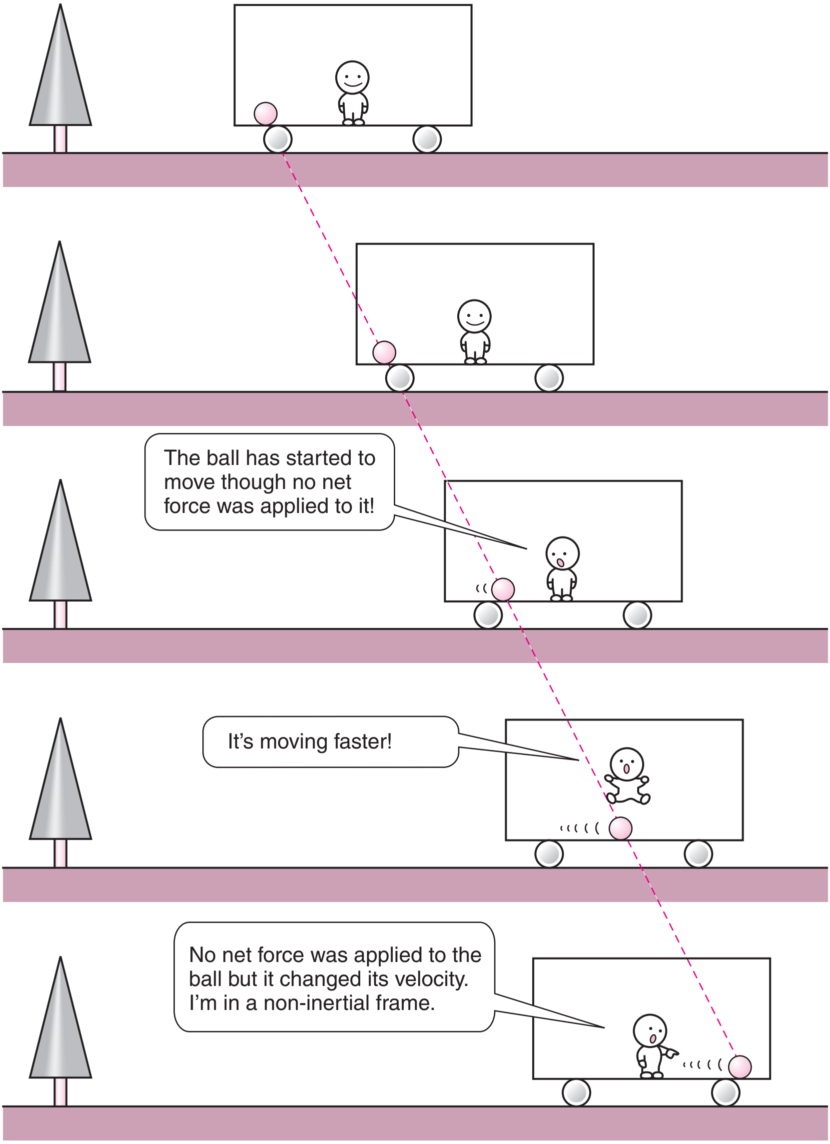
Frames in which the Law of Inertia does not hold are called *non-inertial frames.* They are the ones that are changing speed or changing direction relative to the inertial frames. In the example shown in the drawing on the opposite page, the ball is again sitting on the frictionless floor of the train. But this time, the velocity of the train relative to the tree-frame is not constant; it starts to speed up toward the right in between the second and third drawings from the top. As a result, the ball accelerates to the left in the train-frame even though no net external force is acting on it. This shows that the train-frame is a non-inertial frame.


The ball has started to move though no net force was applied to it!
It's moving faster!
No net force was applied to the ball but it changed its velocity. I'm in a non-inertial frame.

If the train is slowing down, on the other hand, as shown here, the ball accelerates to the right even though no net external force is acting on it. Again, the train-frame is a non-inertial frame.

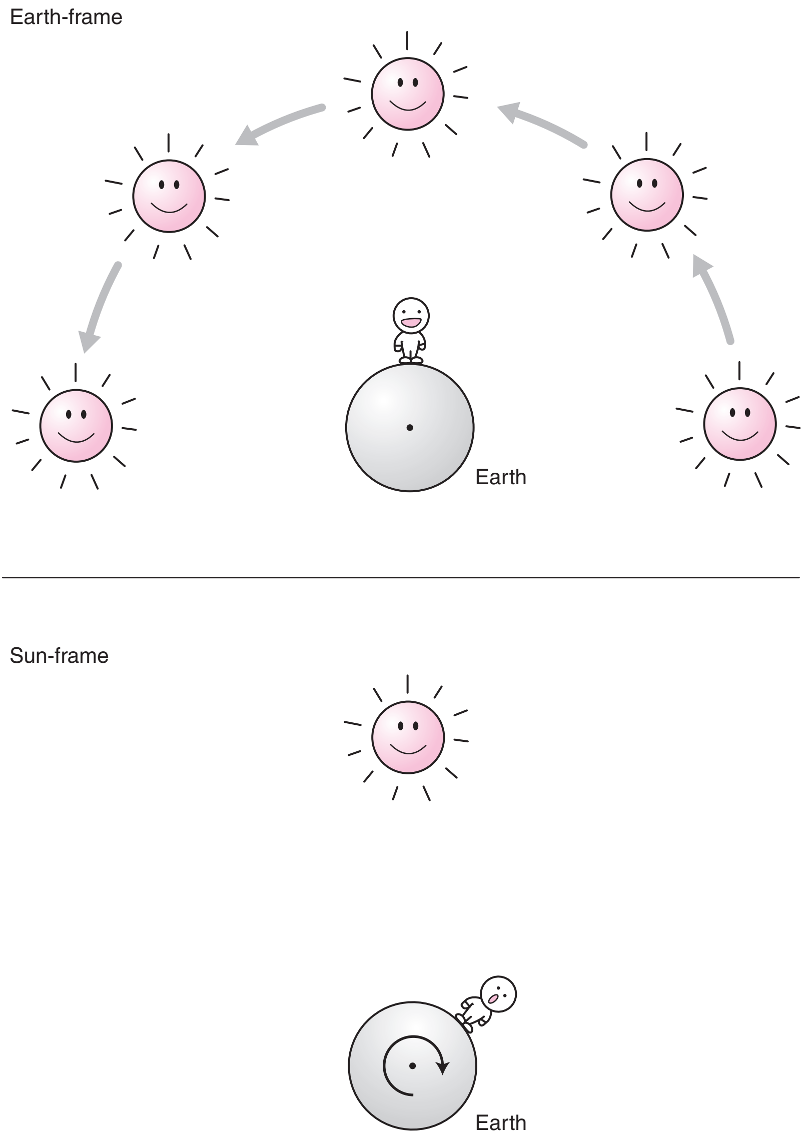
Similarly, the car-frame is an inertial frame as long as the car continues to travel with constant velocity, that is, in a straight line with constant speed, relative to an inertial frame. It will become a non-inertial frame if the car is changing its speed or direction of motion. In an inertial frame, if nothing is happening to the motion of an object it will be observed as nothing happening, but from non-inertial frames it may be observed as something happening because something is actually happening to the frame itself.

The space shuttle in orbit is also in a non-inertial frame. It is in constant free fall towards the Earth.[6] As a result, objects inside the space shuttle look like they are just floating in mid-air. In this non-inertial frame, something is happening to the motion of the objects (they are falling toward the Earth) but it is not observed as such because that same something is happening to the frame also.


The ball has started to move though no net force was applied to it!
It's moving faster!
No net force was applied to the ball but it changed its velocity. I'm in a non-inertial frame.

The frame fixed to the Earth is actually a non-inertial frame because of its daily rotation around its axis, and also its annual rotation around the Sun. As a result of the Earth's rotation, objects on its surface would be flung off were it not for the Earth's gravity holding them down. In fact, this "flinging off" effect is largest near the equator, where the radius of rotation is largest, and can be observed as an apparent reduction in the Earth's gravitational pull compared to that at the poles, though the effect is very small. (About 0.3%.)[7] This is why we know for certain that it is the Earth that is rotating around its axis and not the Heavens rotating around Polaris.[8] The Earth's annual rotation around the Sun leads to a difference in the apparent gravitational pull of the Earth between aphelion and perihelion, but this effect is much much smaller. (About 0.006%.)

However, for the purposes of this book, the Earth-frame is, to a very good approximation, an inertial frame. That is how we discovered the Law of Inertia in the first place by doing experiments on the surface of the Earth, and also why people before Copernicus (1473–1543) thought that the Earth did not move. So we will continue to pretend that the Earth-frame is an inertial frame since it is much easier to talk about trees and cars than about spaceships and stars (at least to me).


Earth-frame
Earth
Sun-frame
Earth

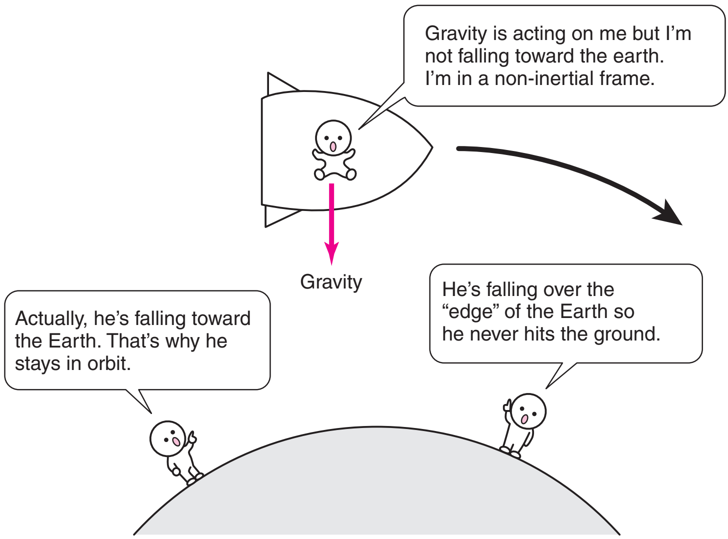
2.6 What's so "special" about Special Relativity?

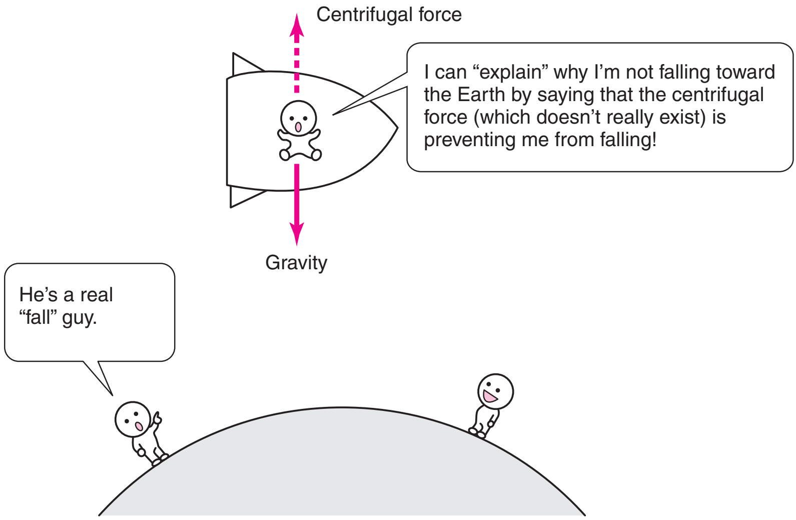
Though it is possible to write the laws of physics so that they are valid in a *general* frame of reference, including both inertial and non-inertial frames, they end up being fairly complicated.

In the case of Newton's theory of motion, its usual formulation you learn in school is only valid in the *special* inertial frames. To make them work in a *general* frame, the theory has to be augmented by *inertial forces*. For instance, in the frame fixed to the space shuttle in orbit, nothing inside the shuttle is falling toward the Earth even though gravity is acting on them. To explain this, you have to introduce a fake force called the *centrifugal force*, which doesn't really exist, to cancel the gravitational force.

Now, as we will see, the Galilei–Newton theory of motion breaks down when the speeds of objects approach the speed of light c. And Einstein's Special Theory of Relativity tells us how it has to be modified in the *special* inertial frames. That is why it is called the "special" theory. Generalizing the theory to *general* frames of reference took Einstein 10 years longer; his General Theory of Relativity was not completed until 1915.[9] As you can imagine, General Relativity is a fairly complicated theory so we are not going to discuss it further in this book.

In the following, we will limit our attention to motion as observed from inertial frames. So when we simply say "frame" from now on, we will mean an inertial frame.


Gravity is acting on me but I'm
not falling toward the earth.
I'm in a non-inertial frame.
Gravity
Actually, he's falling toward the Earth. That's why he stays in orbit.
He's falling over the "edge" of the Earth so he never hits the ground.


Centrifugal force
I can "explain" why I'm not falling toward the Earth by saying that the centrifugal force (which doesn't really exist) is preventing me from falling!
Gravity
He's a real "fall" guy.

Notes

1 The system of units that was developed in France after the French Revolution. It is used all over the world except in the USA. The acronym "SI" stands for *Systeme International d'Unites* = International System of Units (in French, the adjective comes after the noun). It was originally based on the size of the Earth and the mass density of water, and is a very democratic and universal choice of standards compared to yards, feet, and inches, which were based on the sizes of human body parts.

2 If you are a Zen Buddhist or a Quantum Mechanic, you may disagree. I am stating Einstein's point of view here.

3 In Einstein, words: "Physics is an attempt conceptually to grasp reality as something that is considered to be independent of its being observed. In this sense one speaks of *physical reality.*" See [3], page 240.

4 In Judeo-Christian-Islamic terms, what God/Allah does in Her/His wisdom is independent of how we Humans choose to view it in our ignorance.

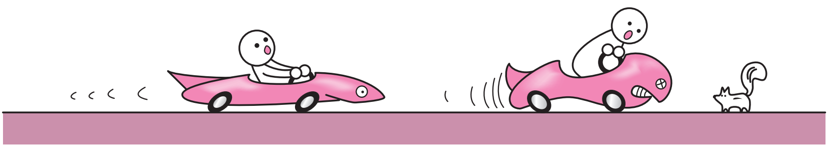
5 If the laws of physics that govern motion are independent of the choice of frame, then it should be possible to express them in a frame-independent way. Of course, you need a frame to describe motion in the first place so this does not mean that you want a theory which does not make any reference to your frame choice. Rather, you want to express your laws in such a way that they are correct in any frame, that is, in a *general frame of reference.* This idea is called the principle of general relativity, or the principle of general coordinate invariance, and was the guiding principle behind Einstein's theories. Einstein's General Theory of Relativity is precisely a theory that satisfies this requirement. But it is a very difficult theory (mathematically) and beyond the scope of this book. The Special Theory of Relativity is less ambitious in that it restricts its attention to *inertial frames of reference.*

6 The space shuttle in orbit is actually falling towards the Earth. As it falls vertically toward the Earth, it travels at a great speed horizontally as well, and as a result it "falls over the edge" of the Earth, so to speak. Because the Earth is round, it keeps on falling along the curvature of the Earth and ends up never hitting the ground.

7 This reduction in apparent gravitational pull may be small, but large enough to save fuel and increase payload when launching rockets into orbit. That is why the space shuttle is launched from Florida and not from Maine, since Florida is closer to the equator.

8 This can also be demonstrated using the Foucault pendulum.

9 The first paper on General Relativity appeared in 1916, in the journal *Annalen der Physik*, volume 49 [4]. Even though General Relativity is a theory of motion, it is also a theory of gravity. Hints that the theories of motion and gravity had to be unified existed as far back as the discovery by Galileo that objects fell at the same rate independently of their masses when released in gravity [4]. Newton incorporated this into his theories by assuming that the *inertial* and *gravitational masses* were the same [5]. General Relativity provides a geometrical explanation why they are the same by associating gravity with the structure of spacetime itself.



3
Galilean relativity

3.1 Basic questions

Let us recall the basic questions we asked about the motion of an object in section 2.1, namely:

Q1: Is it moving or is it at rest?
Q2+Q3: If it is moving, what is its velocity?

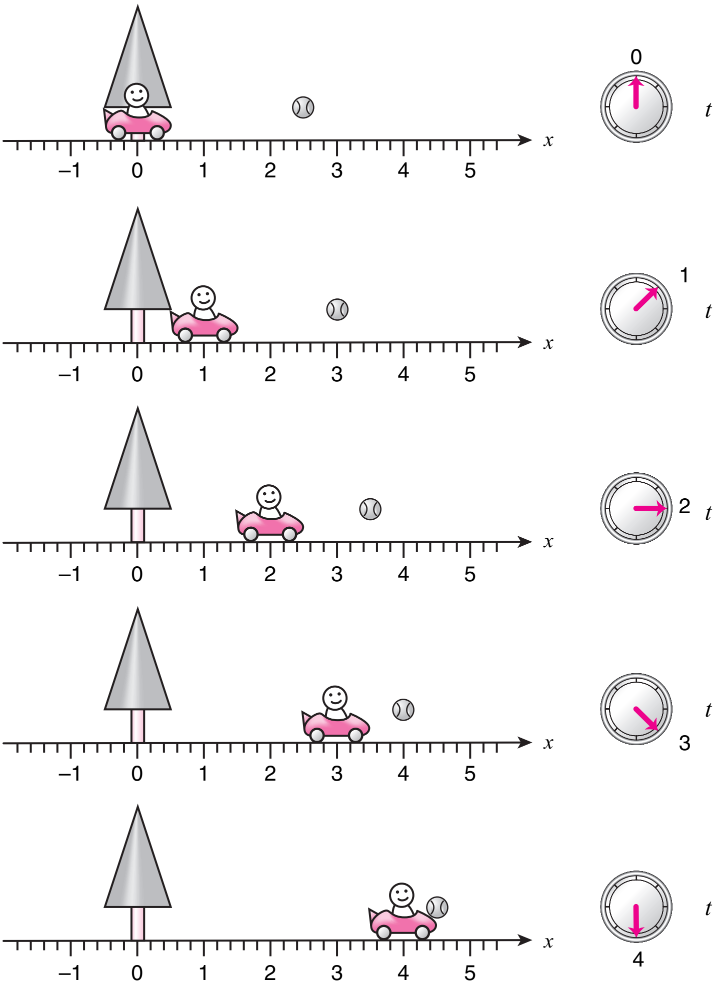
As we have seen, in order to answer these questions we must first choose a frame, and the answers depended on our frame choice.

Let us actually try this out. Consider the car moving along a straight horizontal road as shown in the figure. In the tree-frame, at every *instant in time* the clock fixed to the tree (though it's not drawn on the tree, assume that it is) will give some reading t (in seconds) and *at the same instant* the car will be somewhere along the road at some position x (in meters). The figure shows the sequence of this position from time $t = 0$ seconds to $t = 4$ seconds in 1-second intervals. As you can see, we can tell from the figure that:

A1: the position of the car is changing with time so it is obviously moving, and
A2+A3: the position of the car is changing at a rate of $+1$ meters every second so its velocity is $+1$ meters per second. (The plus sign indicates that the motion is toward the right, so this number tells us not only the speed of the car but also its direction of motion.)

The figure also shows a ball which is also moving horizontally, and we can see that the ball is moving at a velocity of $+0.5$ meters per second. Of course, the tree is at rest in this frame, so we can assign it a velocity of 0 meters per second.

Motion of tree, car, and ball observed in the tree-frame.

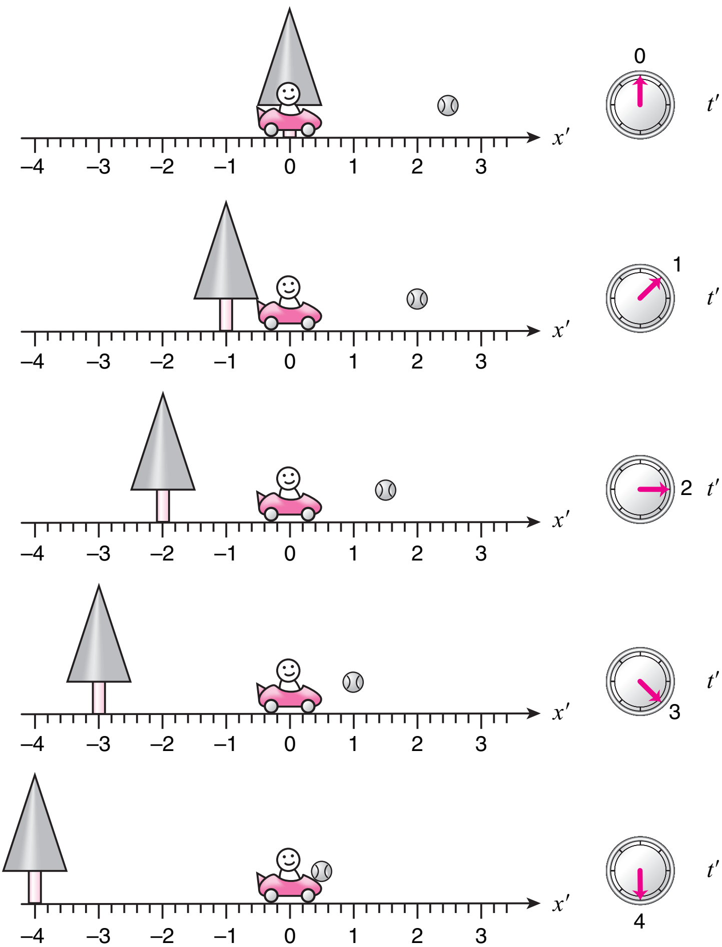


What will the exact same motion look like if observed from the car-frame?[1] In the car-frame, time is specified by the reading of a clock fixed to the car. Let's call this reading t' (in seconds) to distinguish it from the reading of the clock fixed to the tree t. To keep things simple, we will assume that the car-clock is synchronized with the tree-clock so that their readings, t' and t, agree at all times.[2] The location of an object in space in the car-frame is specified by its distance from the car, which we will call x' (in meters) to distinguish it from the location of an object in the tree-frame x which was the distance from the tree.

Then, at each instant labeled by t' in the car-frame, the relative locations of the objects will be the same as that observed in the tree-frame at $t = t'$. At $t' = 1$ second, for instance, the relative locations of the tree, car, and ball are the same as those in the tree-frame at $t = 1$ second, which was shown in the second from top figure on the previous page. The only difference is that we must now use the distance from the car to specify the location of objects. So the tree, which is at $x = 0$ meters in the tree-frame, is at $x' = -1$ meter in the car-frame. The car, which is at $x = +1$ meter in the tree-frame, is at $x' = 0$ meters in the car-frame. And the ball, which is at $x = +3$ meters in the tree-frame, is at $x' = +2$ meters in the car-frame. This is depicted in the second from top figure on the opposite page.

We can figure out the locations of the objects for all other times in a similar manner, and we will obtain the sequence of figures shown here. The clock is actually fixed to the car though it is drawn on the margin. This sequence describes the motions of the tree, the car, and the ball in the car-frame. If we follow how the location of the tree is changing with time, we can see that, after each second, the tree moves to the left by one meter. So the velocity of the tree is -1 meters per second, where the minus sign indicates that the motion is toward the left. Similarly, the ball is moving at a velocity of -0.5 meters per second. The car is at rest so its velocity is 0 meters per second.

Motion of tree, car, and ball observed in the car-frame.



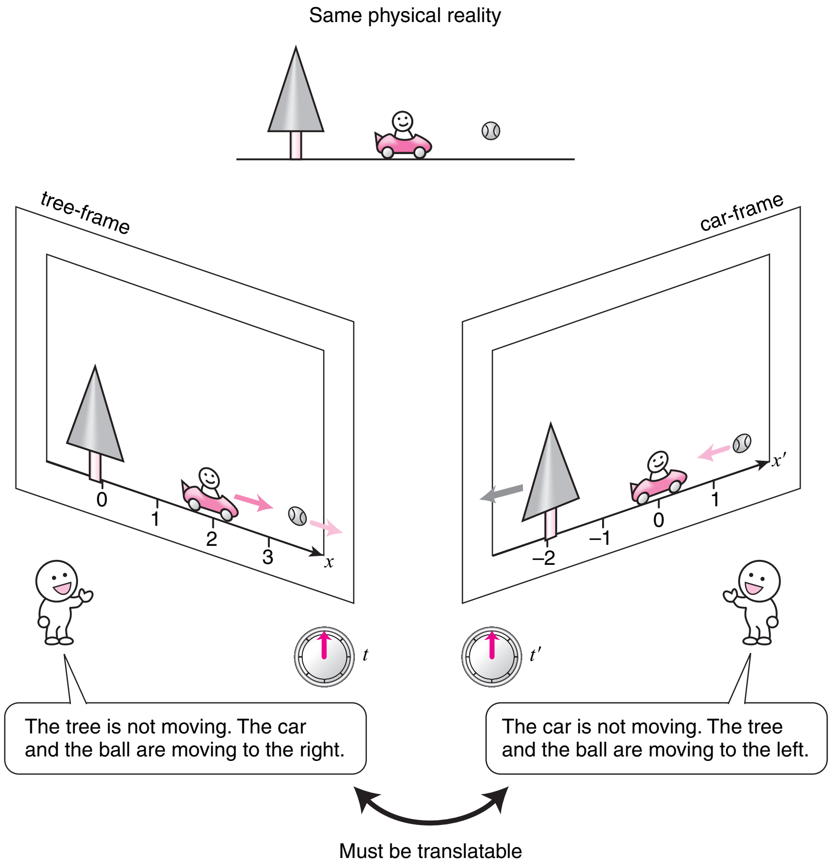
So the answers to the basic questions in the two frames can be tabulated as follows:

object	velocity in tree-frame (m/s)	velocity in car-frame (m/s)
tree	0 (at rest)	-1
car	$+1$	0 (at rest)
ball	$+0.5$	-0.5

Now a natural question to ask is: how are these different observations from different frames related? They are both observations of the exact same objective and physical reality so they cannot be independent of each other. Each observation must contain the exact same information about the reality being observed, since it should not matter which frame you use, which means that the observation in one frame must be translatable into another. What is the rule of translation?

3.2 Spacetime diagrams

Now before we try to answer this question, I would like to introduce a compact way of describing the observed motion of objects, since keeping track of the locations of the objects pictorially, as I have done, is rather bulky. I have indicated the positions of the objects at only five instances, but that already takes up an entire page. If we wanted to study the motion of objects in more detail at smaller chronological increments, things would quickly get out of hand.


Same physical reality
tree-frame
car-frame
0
1
2
3
x
−2
−1
0
1
x′
t
t′
The tree is not moving. The car and the ball are moving to the right.
The car is not moving. The tree and the ball are moving to the left.
Must be translatable

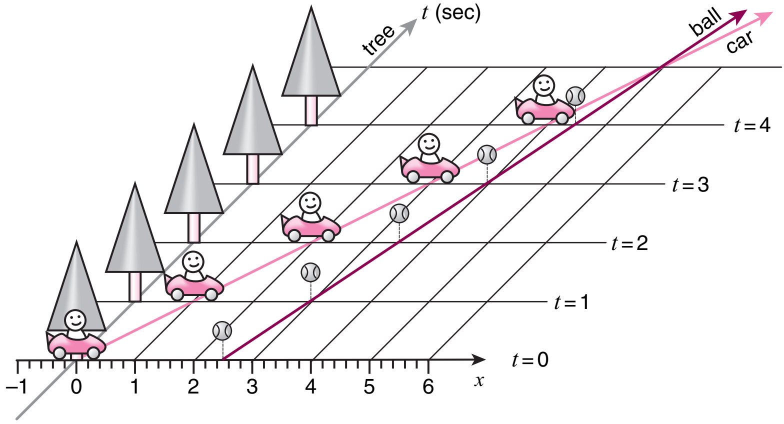
Imagine that we have a movie of the tree, car, and ball taken in the tree-frame. Each frame[3] of the movie captures the positions of the three objects at a particular instant in time as observed in that frame of reference. Assume that the chronological separation of neighboring frames is constant. Cut these frames apart and stack them up in chronological order as shown in the figure so that the tree in each frame overlaps with the tree in the next. If you look at this stack from above, the line connecting the images of the tree in each frame will be a vertical line as shown in gray. The line connecting the images of the car is shown in pink, and the line connecting the images of the ball is shown in crimson.

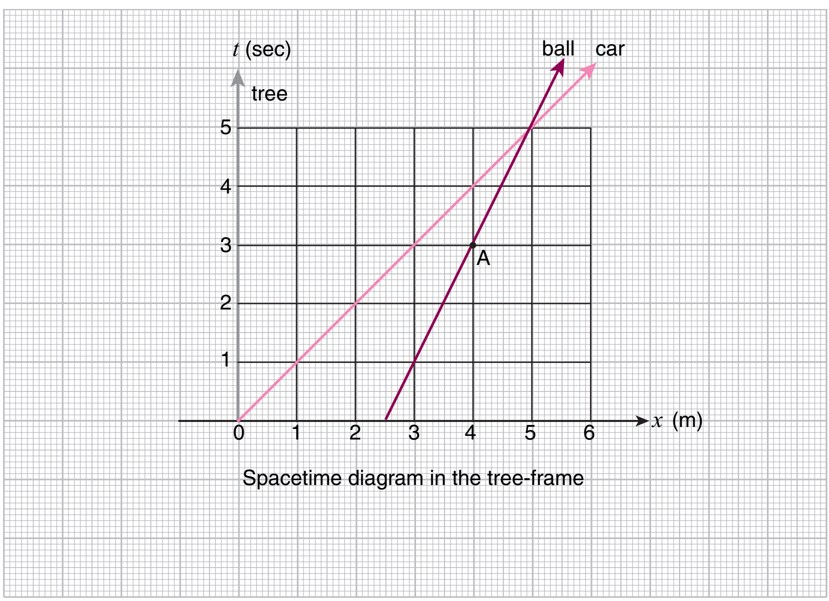
The graph you obtain in this fashion, with the horizontal axis corresponding to *space* and the vertial axis corresponding to *time*, is called a *spacetime diagram.* It is a compact and visual way of describing the motion of objects.

The motion of each object is described by a line on the spacetime diagram. If the object is moving at constant velocity, this line will be straight as in the case of the tree, the car, and the ball shown here. These lines on the spacetime diagram are called *worldlines.*[4] The spacetime diagram represents the *world* in which the motion takes place, and the *worldlines* represent the motions of objects in that *world.*

The position of any object at any given time can be read off of the spacetime diagram by looking at how the object's worldline is positioned against the square grid in the background. The square grid is composed of vertical lines parallel to the time-axis which connect the points that are *at the same place* in the tree-frame, and horizontal lines parallel to the space-axis which connect the points that are *at the same time.* If you want to know the position of the ball, for instance, at some particular time, say at $t = 3$ seconds, all you need to do is slice the diagram with a horizontal line going through the $t = 3$ seconds point on the time-axis, and look at where it crosses the worldline of the ball. This happens at point A on the diagram. Following the vertical line that goes through A down to the space-axis, we find that A is at $x = 4$ meters, which tells us that the ball was at $x = 4$ meters at $t = 3$ seconds.

As I mentioned above, an object moving at a constant velocity is represented by a straight worldline on the spacetime diagram. How steep this worldline is tells us how fast the object is moving. The closer the worldline is to the horizontal the faster the object is, and the closer the worldline is to the vertical the slower the object is. If the worldline is completely vertical, the object is at rest.





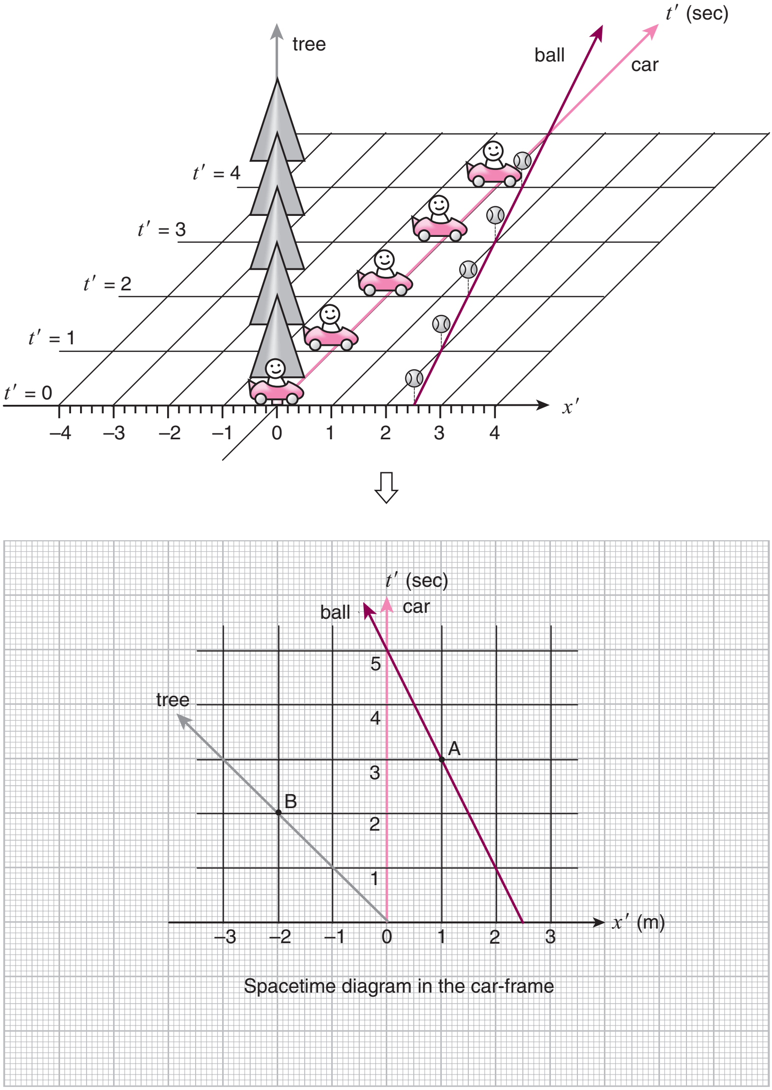
Spacetime diagram in the tree-frame

The spacetime diagram for the same motion observed from the car-frame is shown here. To obtain it, we need a movie taken in the car-frame, but actually we can use the movie taken in the tree-frame as it stands.[5] All we need to do is to stack the frames of the movie so that the images of the car overlap giving a vertical line on the spacetime diagram.

Again, we can read off where any object is at any given time by looking at how the object's worldline is positioned against the square grid in the background. This time, the square grid is composed of vertical lines parallel to the time-axis, labeled t' to distinguish it from the t-axis in the tree-frame, which connect the points that are *at the same place* in the car-frame, and horizontal lines parallel to the space-axis which connect the points that are *at the same time*.

If we want to figure out where the ball is at $t' = 3$ seconds in this frame, we slice the diagram with a horizontal line which goes through the $t' = 3$ seconds point on the t'-axis, and look at where it crosses the worldline of the ball. This is at point A on the diagram. Following the vertical line that goes through A down to the x'-axis, we find that A is at $x' = 1$ meter, which tells us that the ball was at $x' = 1$ meter at $t' = 3$ seconds.

If we want to figure out where the tree is at $t' = 2$ seconds, we slice the diagram with a horizontal line which goes through the $t' = 2$ seconds point on the t'-axis, and look at where it crosses the worldline of the tree. This is at point B on the diagram. Following the vertical line that goes through B down to the x'-axis, we find that B is at $x' = -2$ meters, which tells us that the tree was at $x' = -2$ meters at $t' = 2$ seconds.

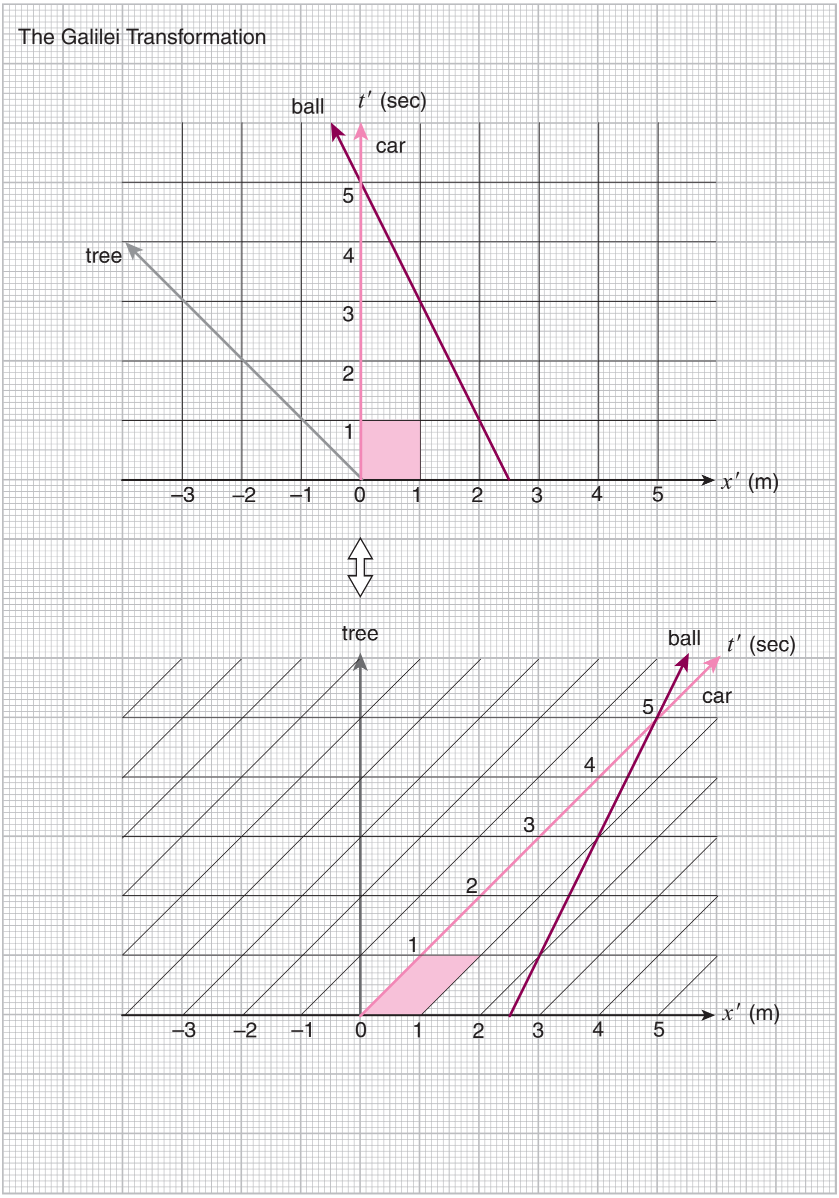

t′ (sec)
tree
ball
car
t′ = 4
t′ = 3
t′ = 2
t′ = 1
t′ = 0
x′
–4 –3 –2 –1 0 1 2 3 4
t′ (sec)
ball
car
tree
5
4
3
2
1
A
B
x′ (m)
–3 –2 –1 0 1 2 3
Spacetime diagram in the car-frame

3.3 The Galilei transformation

Once we have the two spacetime diagrams representing the observations from the tree-frame and the car-frame, we can ask: how are the two diagrams related? How can we obtain one spacetime diagram from the other?

Let's start from the car-frame diagram shown on the upper-half of the opposite page. Since the diagram was obtained by stacking up many movie frames so that the car stays in the same place, to obtain the diagram in the tree-frame all we have to do is shift each movie frame to the right until the image of the tree overlaps with that in the frame immediately below it. This procedure will result in a diagram like the one shown on the lower-half of the opposite page, which is precisely the spacetime diagram in the tree-frame except for the skewed grid inherited from the car-frame diagram.

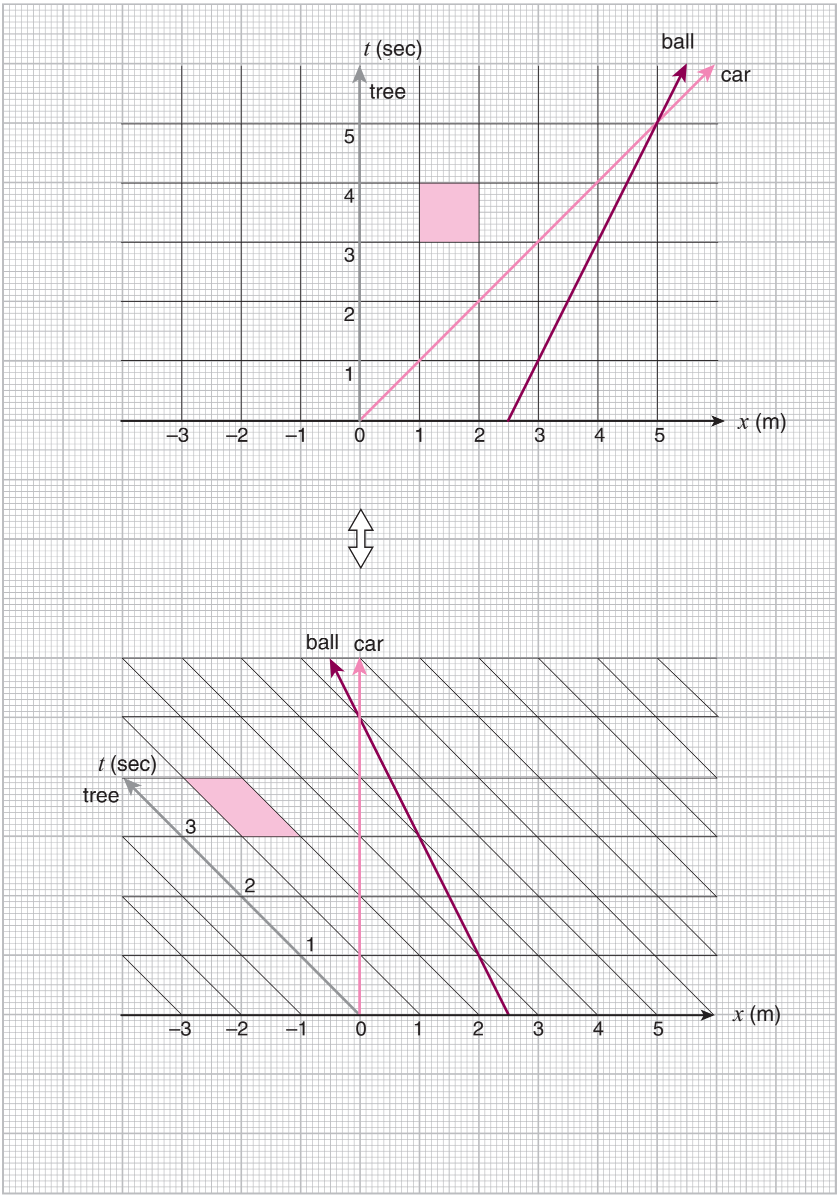
Note that the shaded square in the top figure is mapped onto the shaded parallelogram in the bottom figure, and that the area of the shaded regions are the same. This is because the amount of "film" contained in the unit square, or the "number of events" in it, so to speak, is conserved in the procedure.



The Galilei Transformation
ball
t' (sec)
car
5
4
3
2
1
tree
–3
–2
–1
0
1
2
3
4
5
x' (m)
tree
ball
t' (sec)
car
5
4
3
2
1
–3
–2
–1
0
1
2
3
4
5
x' (m)

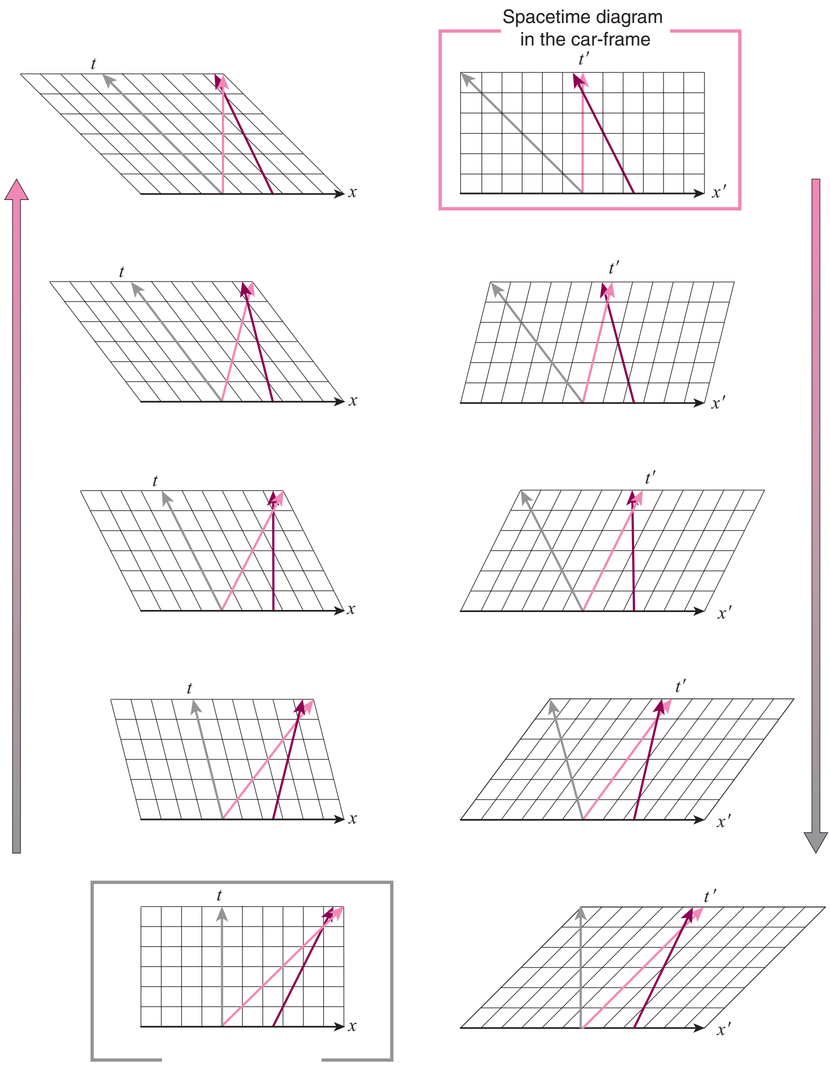
We can similarly start from the tree-frame diagram, shown on the upper-half of the opposite page, and obtain the car-frame diagram by following the reverse procedure. We shift each frame of the movie to the left until the image of the car overlaps with that in the frame immediately below it. This will give us the diagram shown on the lower-half of the opposite page, which is the spacetime diagram in the car-frame except for the skewed grid inherited from the tree-frame diagram.

Again, I have shaded a unit square on the tree-frame grid, and the corresponding parallelogram on the car-frame diagram. The area of the shaded region is conserved.


t (sec)
tree
ball
car
5
4
3
2
1
x (m)
–3
–2
–1
0
1
2
3
4
5
ball
car
t (sec)
tree
3
2
1
x (m)
–3
–2
–1
0
1
2
3
4
5

The intermediate steps of both procedures are shown here. To transform the car-frame diagram into the tree-frame diagram, we start from the car-frame diagram, shown top-right in the figure, and skew its square grid to the right until the worldline of the tree becomes vertical. The worldlines then match those in the tree-frame diagram, shown bottom-left.

Similarly, to transform the tree-frame diagram into the car-frame diagram, start from the tree-frame diagram, shown bottom-left, and skew its square grid to the left until the worldline of the car becomes vertical (start from the bottom-left figure and proceed upwards). The worldlines then match those in the car-frame diagram, shown top-right.



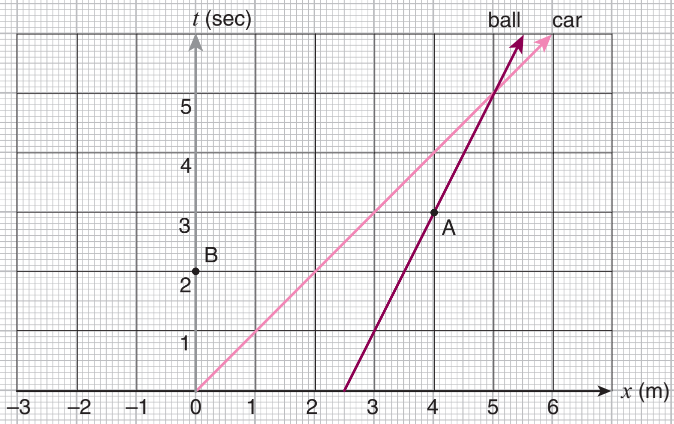
Spacetime diagram
in the tree-frame

This skewing of the grid which takes you from one frame to another is called the *Galilei Transformation*. It lets you figure out what the motions will look like in a different frame when given the spacetime diagram in one frame.

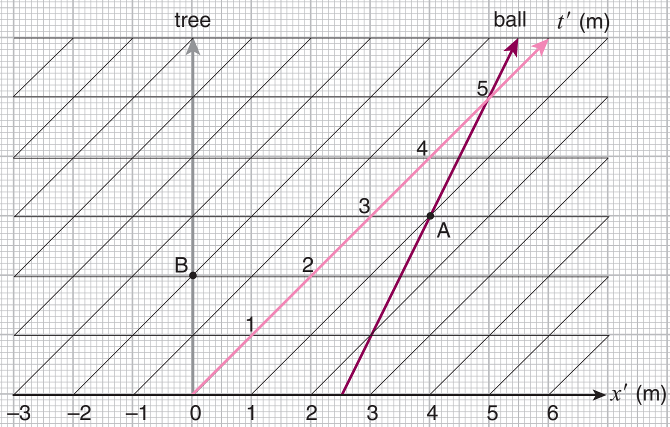
Now if you look at the tree-frame diagram that we obtained by skewing the car-frame diagram (bottom figure), we can still read off the information about the car-frame if we keep the skewed grid in place and don't replace it with the square grid of the tree-frame. In fact, even though the worldlines of the tree, the car, and the ball overlap with those in the tree-frame diagram (top figure), we can still consider this to be a car-frame diagram as long as the car-frame grid is still there.

For instance, at $t' = 3$ seconds, the ball is at point A. If we follow the line parallel to the car-worldline which goes through A down to the space-axis, we can tell that its position in the car-frame is $x' = 1$ meter. At $t' = 2$ seconds, the tree is at point B. If we follow the line parallel to the car-worldline which goes through B down to the space-axis, we can tell that its position in the car-frame is $x' = -2$ meters.

This suggests an alternative way of looking at the *Galilei Transformation*. Instead of skewing the entire diagram to go from one frame to another, we can simply superimpose onto a single diagram different grids, each representing the observation from a particular frame.



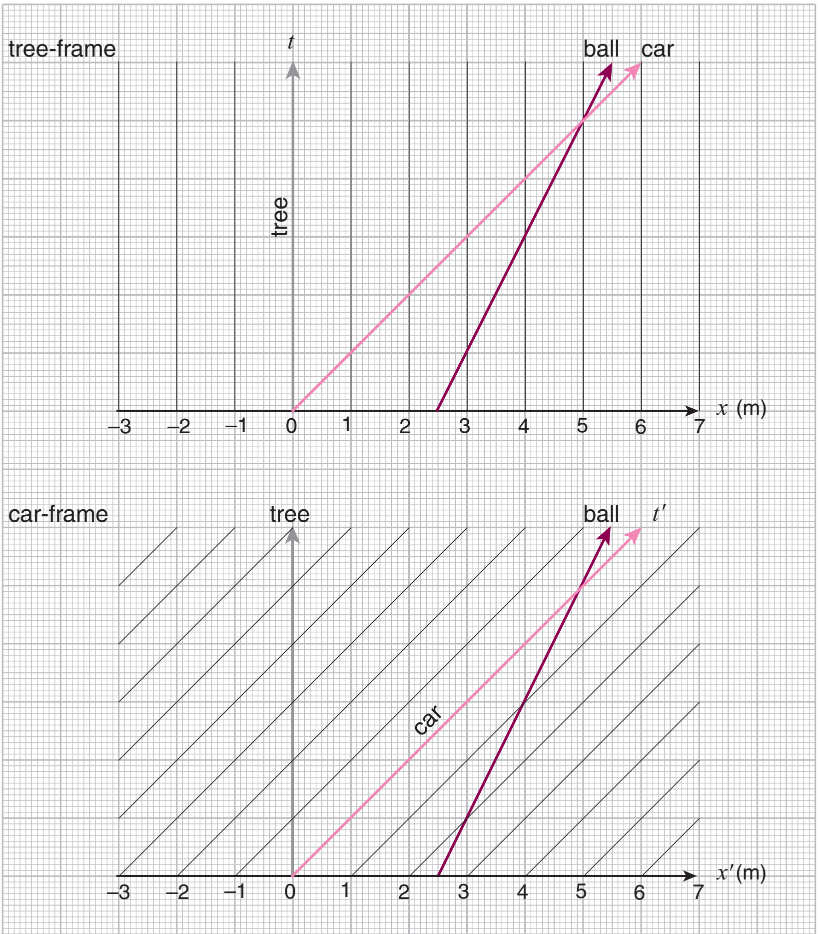
Motion of ball and car in the tree-frame



Motion of ball and tree in the car-frame

So how can we figure out the grid that we are supposed to use? Recall that the "position" of an object in each frame is the distance of the object from the origin. So the lines that connect the points that are *at the same place* in each frame must be parallel to the worldline of the origin (see figure below).

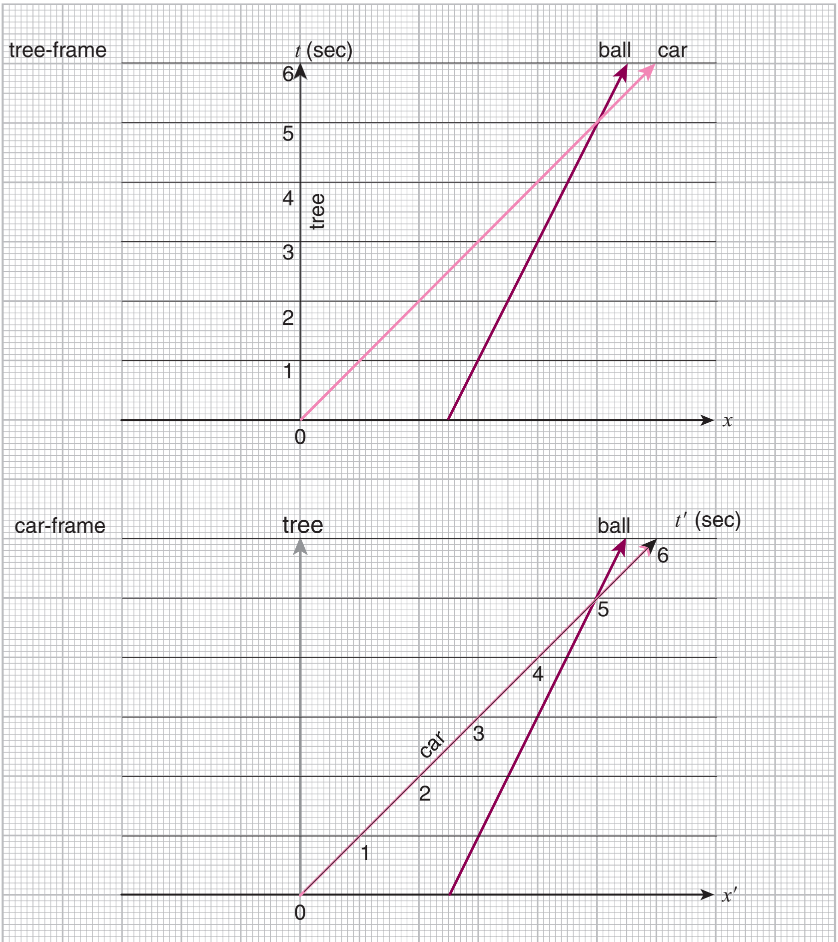
In the tree-frame, the tree is the origin so the lines that connect points



Lines parallel to the time axis connect points that are "at the same place."

that are *at the same place* must be parallel to the tree-worldline. In the car-frame, the car is the origin so the lines that connect points that are *at the same place* must be parallel to the car-worldline.

On the other hand, the lines that connect the points that are *at the same time* are horizontal in any frame (see figure below).[6]

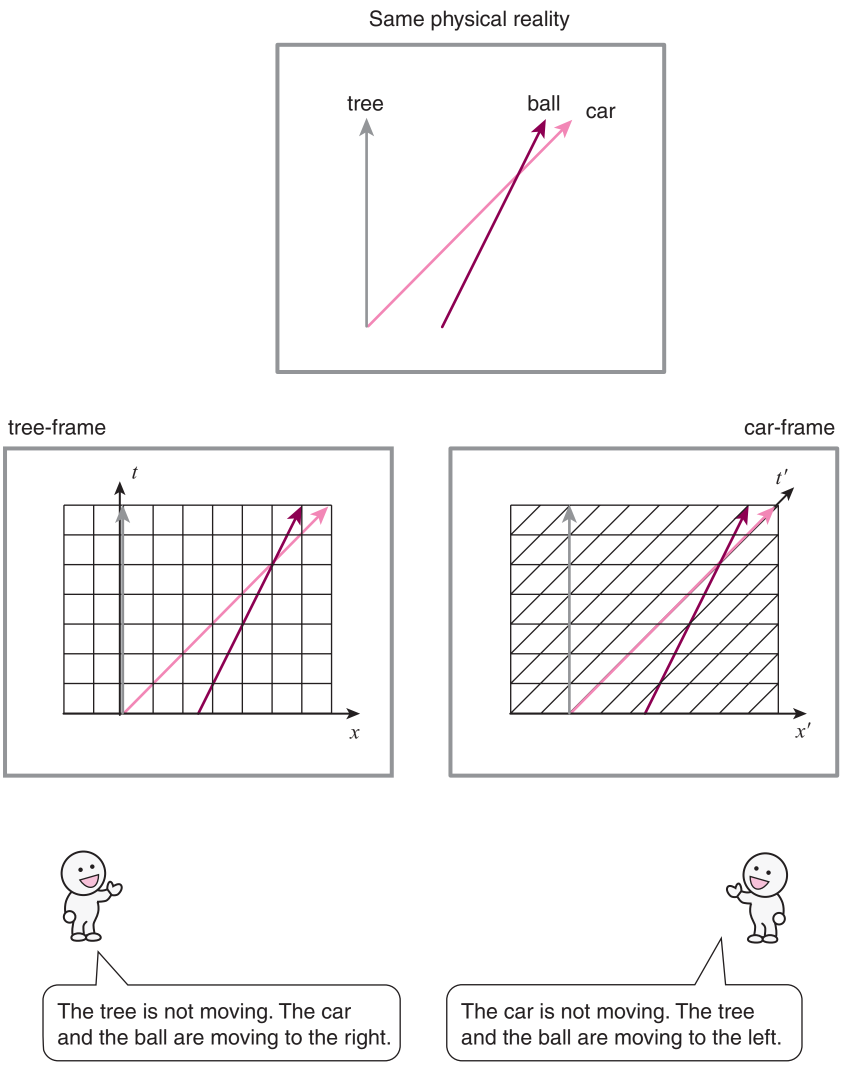


Lines parallel to the space axis connect points that are "at the same time."

So finding the skewed grid which represents the observation from any frame on any spacetime diagram is easy. The horizontal lines that connect the points that are *at the same time* remain the same in all frames. The lines that connect the points that are *at the same place* are drawn parallel to the worldline of the object to which the origin is fixed.

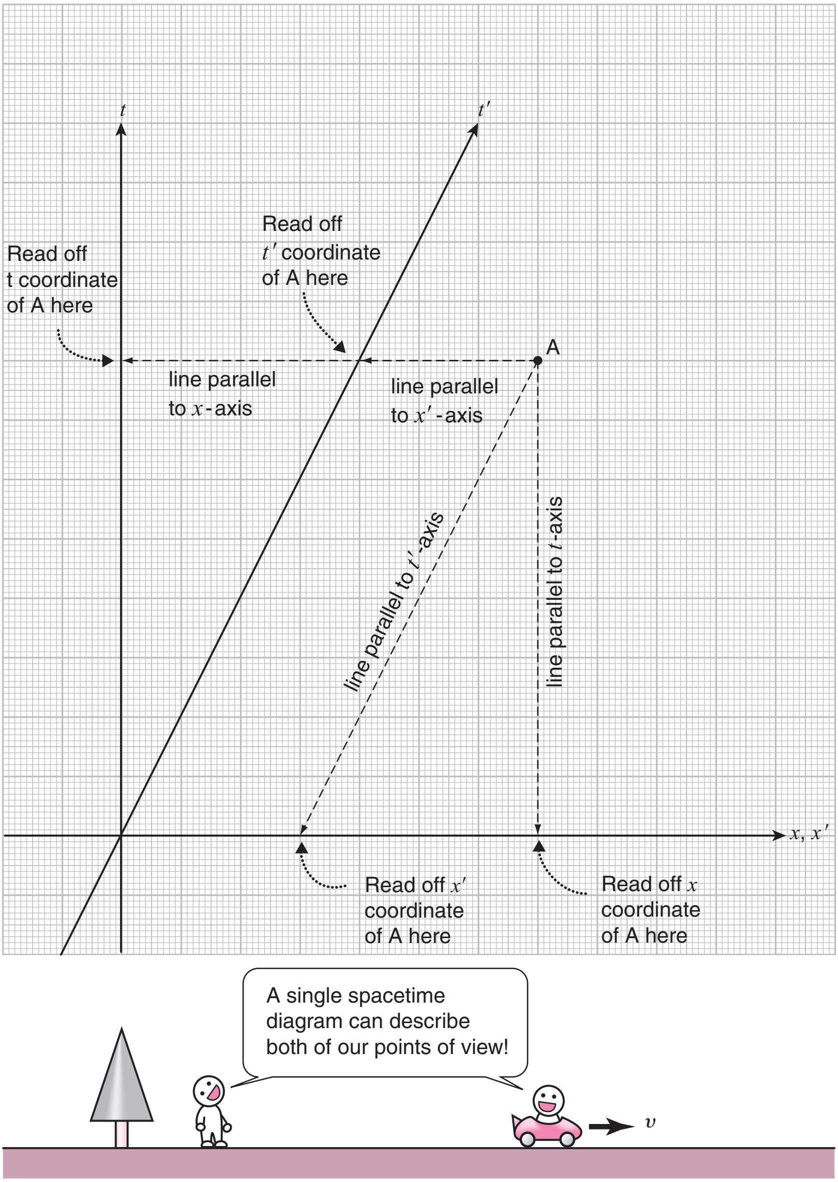
For instance, if, in addition to the tree- and car-frames, we wanted to figure out the grid that describes the observation from a frame of reference fixed to the ball, we would keep the horizontal lines from the tree- and car-frames as they are, and add lines parallel to the ball-worldline which connect the points that are *at the same place* in the ball-frame. (Try this yourself.)

From this point of view, a spacetime diagram without any grid on it can be considered a representation of the objective and physical reality, and superimposing a grid on it corresponds to choosing the frame from which to make the observation.


Same physical reality
tree
ball
car
tree-frame
car-frame
t
x
t′
x′
The tree is not moving. The car and the ball are moving to the right.
The car is not moving. The tree and the ball are moving to the left.

Once we get used to this idea, we can save ourselves the trouble of drawing grids on the spacetime diagram since we will be able to “see” them without actually drawing them. All you need are the space- and time-axes.

To find out *where* a point on the spacetime diagram is in a particular frame, draw a line through the spacetime point which is parallel to the time-axis of that frame (which is the worldline of the origin of that frame) and read off its position from where the line intersects the space-axis. To find *when* the spacetime point is in that frame, draw a line parallel to the space-axis, which is a horizontal line in all frames, and read off the time from where the line intersects the frame’s time-axis. That’s all there is to it.


t
t′
Read off
t coordinate
of A here
Read off
t′ coordinate
of A here
A
line parallel
to x-axis
line parallel
to x′-axis
line parallel to t′-axis
line parallel to t-axis
x, x′
Read off x′
coordinate
of A here
Read off x
coordinate
of A here
A single spacetime
diagram can describe
both of our points of view!
v

3.4 Addition of velocities

The Galilei transformation tells us how observations in one inertial frame are related to observations in another. Let us now ask the following question:

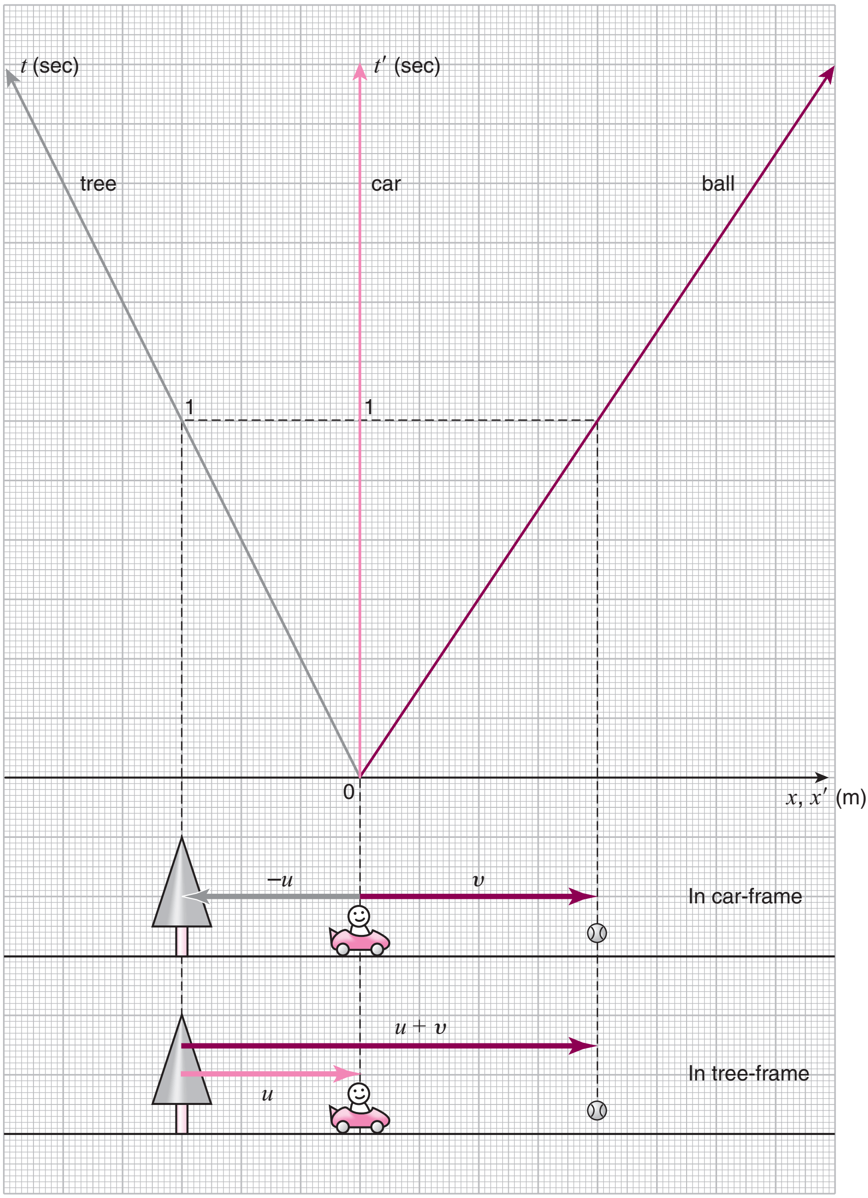
- Assume that the car is moving at a constant velocity of u meters per second when observed from the tree-frame. The velocity of an object, say a ball, is v meters per second when observed from the car-frame. What is the ball's velocity when observed from the tree-frame?

To simplify things, let's assume that the tree, the car, and the ball are all at the same place at time $t = t' = 0$. Then, at a later time, the distance from the origin of each object will be proportional to its speed in either frame. The spacetime diagram which describes the motion of the ball and the tree in the car-frame is shown here. The car is at rest at the origin, the tree is moving with speed u meters per second to the left (velocity $-u$), and the ball is moving with speed v meters per second to the right (velocity $+v$).

If we take a time-slice of the spacetime diagram at $t = t' = 1$ second, then the relative positions of the tree, the car, and the ball will be as shown: the distance between the tree and the car will be u meters, the distance between the car and the ball will be v meters, and the distance between the tree and the ball will be $u + v$ meters. Since the distance between the tree and the ball has increased from zero to $u + v$ meters toward the right in one second, the velocity of the ball in the tree-frame is $u + v$ meters per second.

So according to the Galilei theory, to translate the velocity of an object observed in frame 1 (the velocity of the ball in the car-frame) to that observed in frame 2 (the velocity of the ball in the tree-frame), all you need to do is add the velocity of the origin of frame 1 relative to frame 2 (the velocity of the car in the tree-frame).

This rule can easily be checked by looking at the table on page 34 which gave the velocities of the tree, the car, and the ball in the tree- and car-frames. The velocity of the ball in the car-frame was -0.5 meters per second. The velocity of the car in the tree-frame was $+1$ meters per second. Adding these together, we obtain $+0.5$ meters per second, which was the velocity of the ball in the tree-frame.


t (sec)
t′ (sec)
tree
car
ball
1
1
0
x, x′ (m)
−u
v
In car-frame
u + v
In tree-frame
u

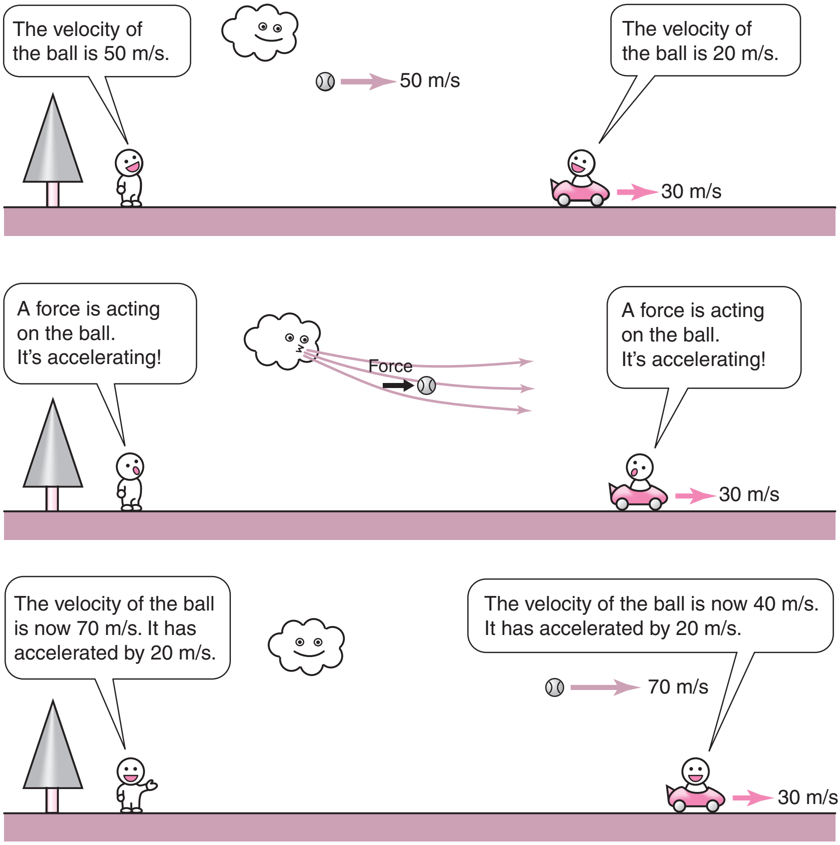
Now I'm sure this is all very easy for you, and you are probably wondering by now why I am spending so much time on such trivialities. But bear with me. Thinking about things that are trivial to you using spacetime diagrams will help us in understanding what is to come.

3.5 Acceleration and Newton's Second Law

As we have seen, the velocity of an object depends on the frame from which the observation is made. However, according to the Galilei transformation, the velocities observed from different frames only differ by the relative velocity of the two frames. This implies that if the velocity of the object changes, both frames will agree by how much it has changed.

To see this, consider a ball traveling at velocity v meters per second when observed from the car-frame. The car, on the other hand, is traveling at velocity u meters per second relative to the tree-frame. The velocity of the ball in the tree-frame is $u + v$ meters per second, as we just discussed in the previous section. Now, let's say that the velocity of the ball increased from v meters per second to $v + \Delta v$ meters per second in the car-frame. The corresponding velocity in the tree-frame is $u + v + \Delta v$. So in both frames the velocity has increased by Δv.

This fact is important for Newton's Second Law, which provides a prediction for the change in velocity when an external force acts on an object. Since the change in velocity is independent of the inertial frame (according to the Galilei transformation), Newton's Second Law applies equally to all inertial frames.[7]



Both observers agree that the ball has accelerated by 20 m/s.

Notes

1 Since the car-frame is moving at a constant velocity relative to the tree-frame, it is also an inertial frame.
2 We will learn later that, surprisingly, clocks in different frames cannot be synchronized. But we will assume that they can be for the moment.
3 Here, the word "frame" is used in a different sense from "frame of reference."
4 Weird term. Must be a direct translation from German.
5 Or at least, we think we can.
6 At least, that is what we think.
7 Newton's Second Law expressed in an equation is

$$F = ma$$

where F is the external force, m is the mass of the object, and a is the acceleration = rate of change of the velocity. When the external force is zero, that is $F = 0$, then this equation tells us that $a = 0$, which means that the velocity of the object does not change (this is Newton's First Law). When the external force is non-zero, then the acceleration is proportional to F but inversely proportional to m. This means that the larger (smaller) the mass is, the smaller (larger) the acceleration. So the mass m is a measure of the object's *inertia*, that is, its ability to resist change in its velocity. In this sense, the mass m is sometimes called the *inertial mass* to distinguish it from the *gravitational mass*, which is a measure of how strongly the object feels the effect of a gravitational field.

4

Einsteinian relativity

4.1 The mystery of the speed of light

Now the surprising thing about the Galilei–Newton theory that we have been discussing so far is that it is *wrong*. It is not *wrong* in the sense that it is *completely wrong*, but *wrong* in the sense that there is a limit to its applicability and in certain cases it does not work.[1] And that case involves the speed of light.

The speed of light in a vacuum is very very fast.[2] It is 299 792 458 meters per second,[3] or roughly 3×10^8 meters per second. Since we do not want to end up writing this big number repeatedly, we will just represent it with the letter c. To give you an idea just how fast this is, it is fast enough to circumnavigate the Earth seven and a half times per second. The time it takes for light to travel 30 centimeters (about a foot) is only 1 nano-second, which is 0.000 000 001 seconds.

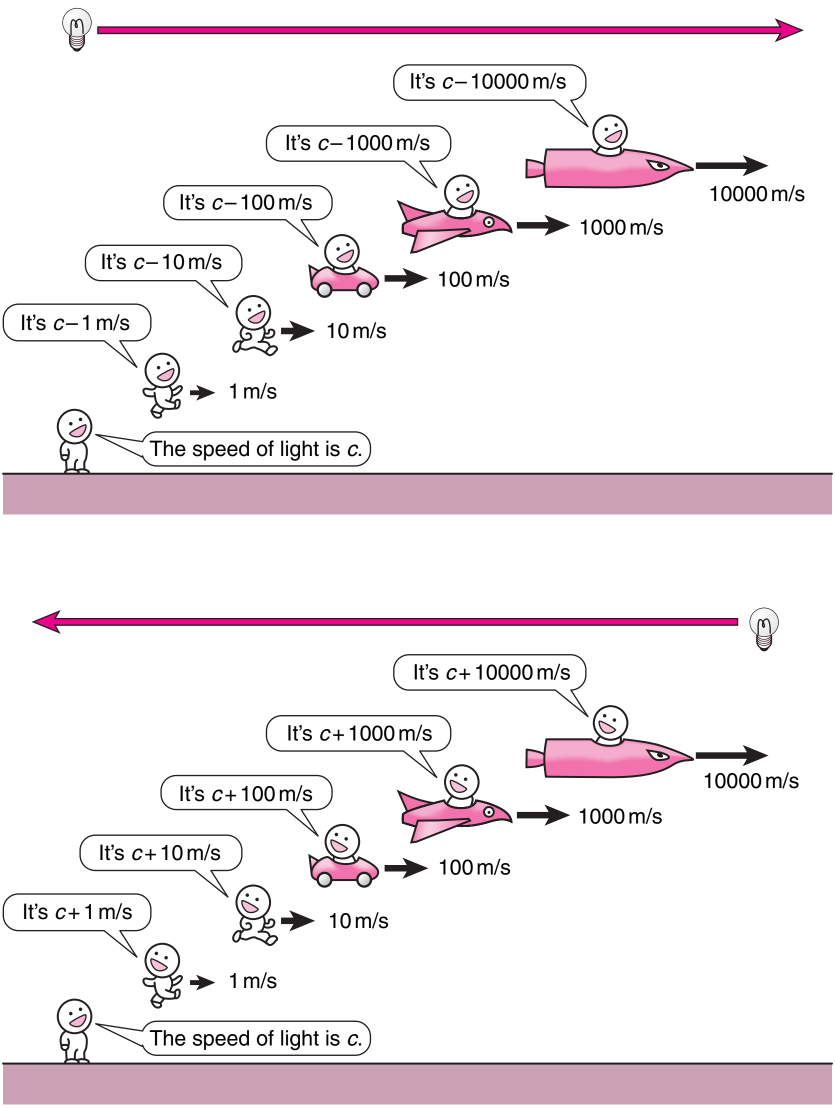
Because c is so large, it was very difficult to measure what it was for a long time. Galileo himself tried it but did not succeed.[4] But by the end of the nineteenth century, the technology had advanced to the point that very accurate measurements of c were possible.

What the physicists at the time were trying to figure out was how the speed of light depended on the motion of the observer, and also on the motion of its source. Now, I am not going to explain why the physicists wanted to know this or how the measurement was actually done since we are not interested in history at the moment. All you need to know is the surprising result:

- The speed of light in a vacuum is always equal to c regardless of the motion of the observer or the motion of the source. It is c in all inertial frames.

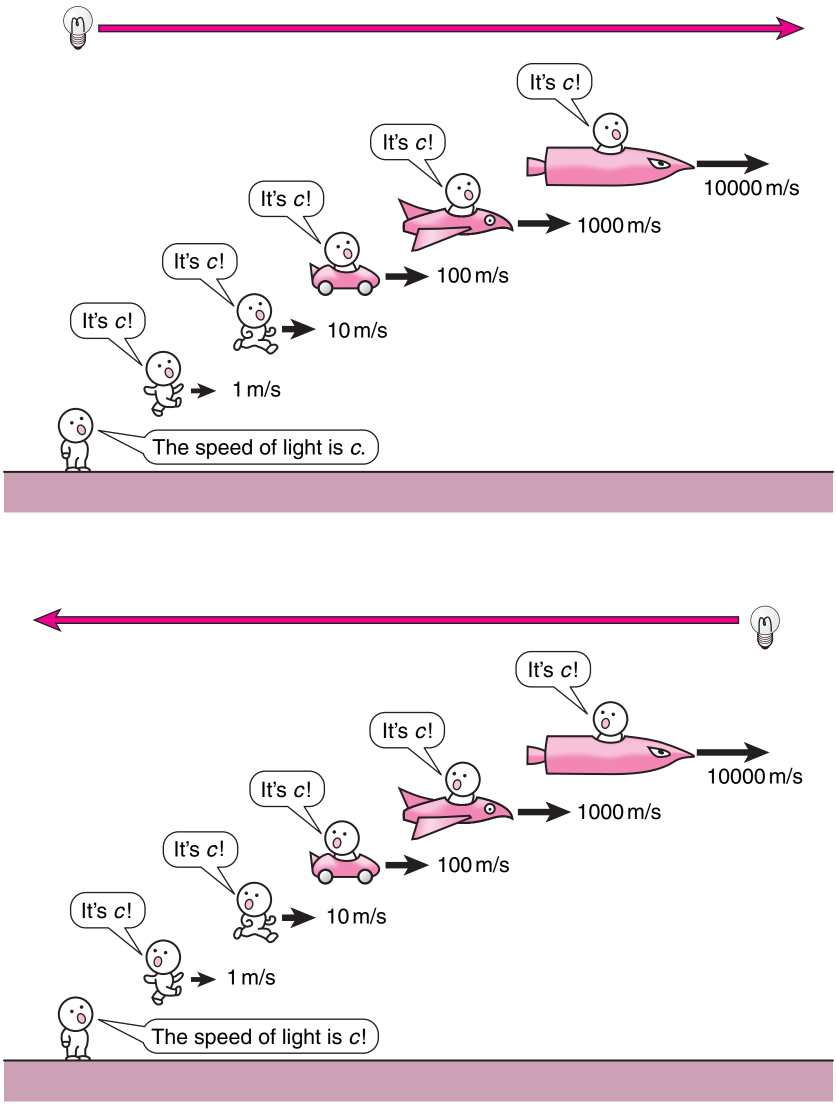
What this statement is saying is this: You could chase after a beam of light as fast as possible, or run away from a beam of light as fast as possible, . . .

What you expect:

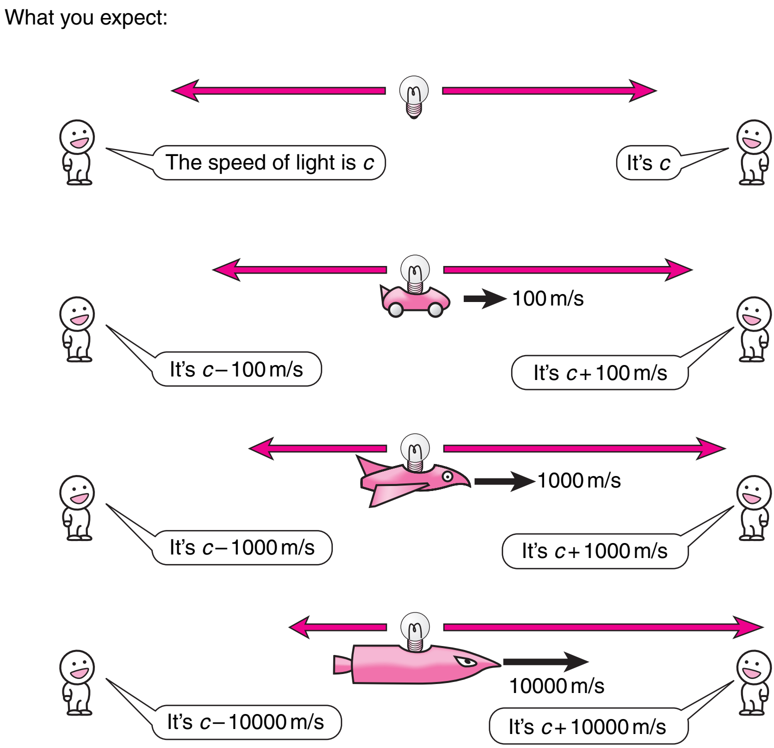


... but the speed of light will always be c.

What really happens:

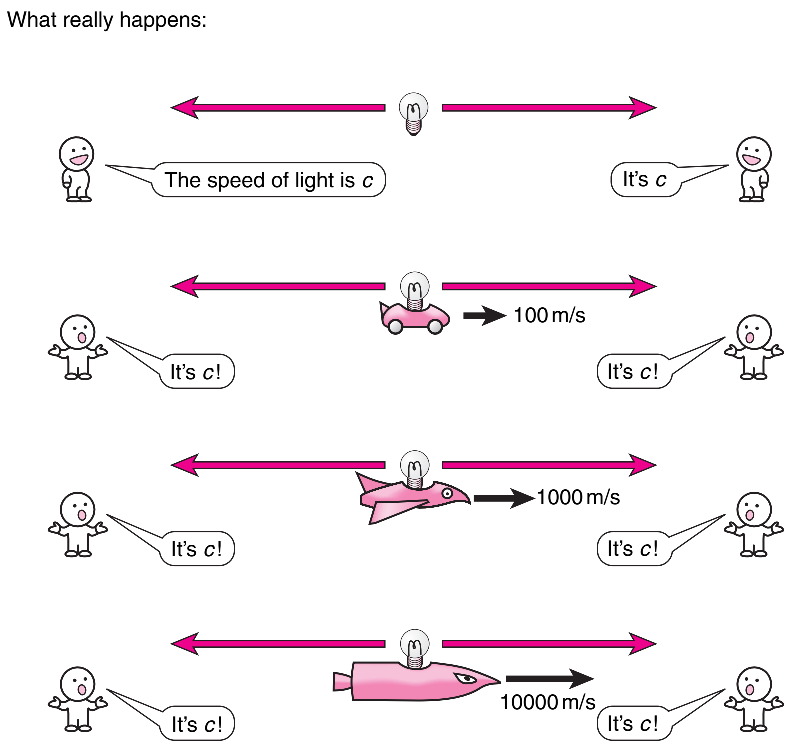


Or the source of the light could be moving away from you at a very fast speed, or moving towards you at a very fast speed, ...



...but the speed of light will always be c.

Clearly the addition-of-velocities rule we derived from the Galilei transformation does not work here!



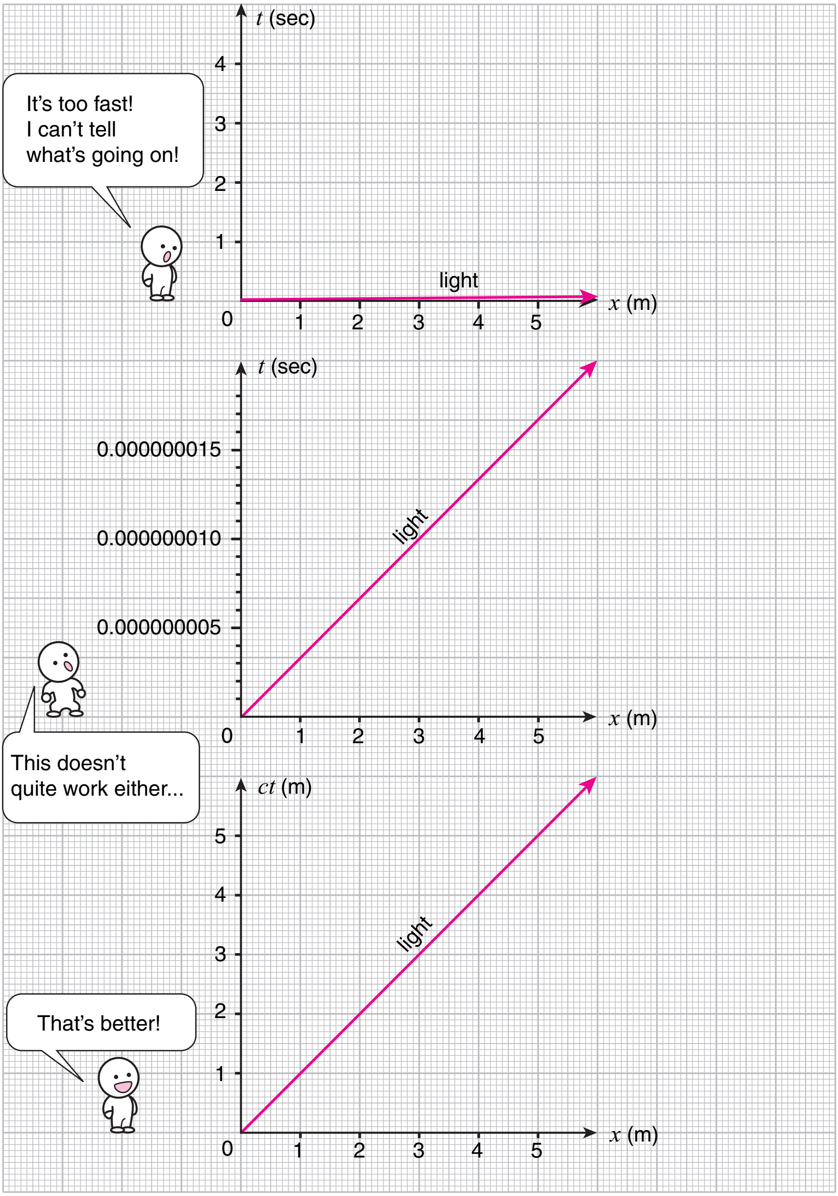
4.2 Modification to the spacetime diagram

Now we are going to analyze this problem using spacetime diagrams, but we need to make a slight modification since the speed of light is so fast. If we keep on using meters to label the space-axis and seconds to label the time-axis, the worldline of a beam of light would look virtually horizontal because of its great speed.

So what can we do? We could stretch the diagram in the vertical direction until the worldline of a beam of light is not so close to the horizontal, but then the units of time that label the time-axis will be so short, as shown in the second graph, that we have no sense of what kind of time intervals are involved in the problem.

A similar problem exists in astronomy, where saying things like "the distance between the Earth and the Sun is about 1.5×10^{11} meters," or "the nearest star system is Alpha Centauri which is about 4.1×10^{16} meters away," doesn't tell us much because the numbers involved are so huge that we have no sense of how large these distances really are. What astronomers do to solve this problem is to convert spatial distances into time using the speed of light c. Instead of giving the distance in meters, they use the amount of time it takes for light to cover that distance to describe how far things are. A light-minute is the distance light travels in a minute (about 1.8×10^{10} meters), a light-day is the distance light travels in a day (about 2.6×10^{13} meters), and a light-year is the distance light travels in a year (about 9.5×10^{15} meters). The distance between the Earth and the Sun is about 8 light-minutes, and the distance to Alpha Centauri is about 4 light-years. Since we have a good sense of how long a minute or a year is, and given the extreme speed of light, this gives us a feel for just how far away these objects really are.

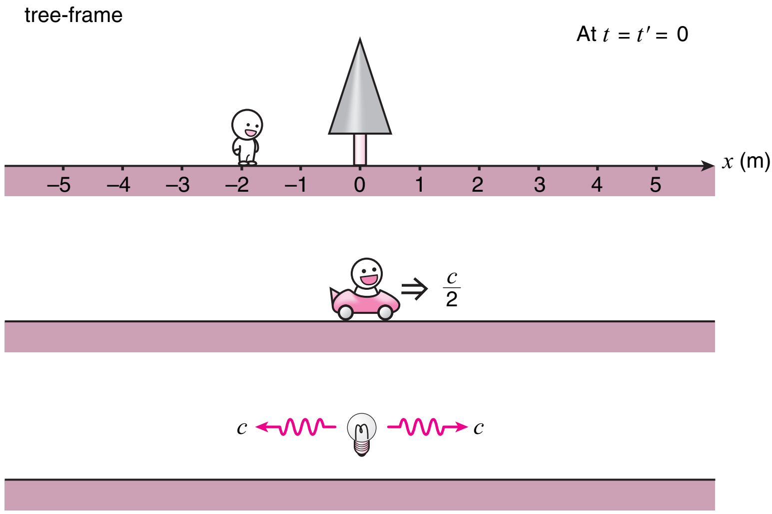
We are going to use a similar trick here and use c to convert very short time intervals into manageable lengths. If we multiply c with t, the product ct gives the distance light travels in the time interval t. For instance, light travels 0.3 meters in 1×10^{-9} seconds, 0.6 meters in 2×10^{-9} seconds, 0.9 meters in 3×10^{-9} seconds, and so on and so forth. The idea is to use this distance travelled by light as the measure of time. So instead of saying 1×10^{-9} seconds of time, we can say 0.3 meters of time, and we definitely have a good idea of how long 0.3 meters is (about a foot) which gives us an idea of how short 1×10^{-9} seconds is, given the extreme speed of light. Furthermore, using ct instead of t to measure time is also convenient since the speed of light will be exactly 1: it takes 1 meter of time for light to travel a distance of 1 meter. The worldline of a beam of light will be at a 45 degree angle from the horizontal on the spacetime diagram, as shown in the final graph below.

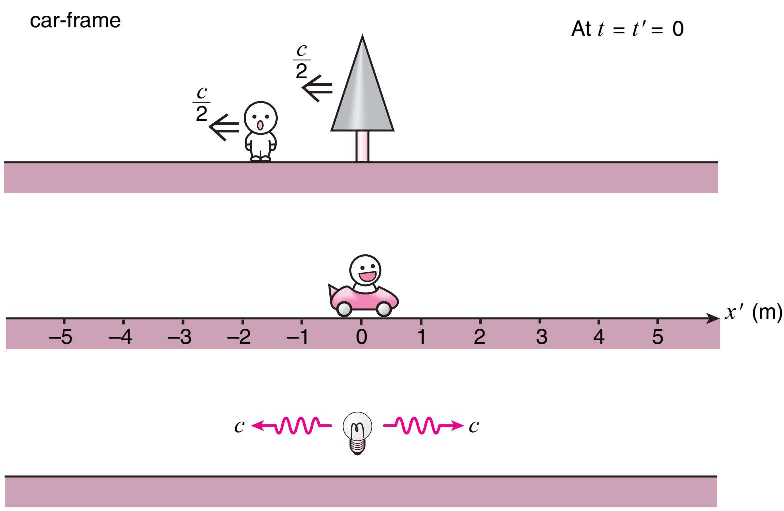


4.3 The problem

Let us now consider the following situation, which exemplifies the problem we are facing.

A tree is at rest on the ground, and a car is moving toward the right at half the speed of light. (It is a very fast car.) Both the tree and the car have clocks on them which are set to agree at time $t = t' = 0$ when they are both *at the same place* as a light bulb which is switched on at that instant. Two photons, that is, particles of light, will be emitted from the light bulb at the speed of light c, one toward the right, and another toward the left. How are the motions of the tree, car, and the photons observed in the tree- and car-frames?


tree-frame
At $t = t' = 0$
x (m)
–5
–4
–3
–2
–1
0
1
2
3
4
5
$\frac{c}{2}$
c
c

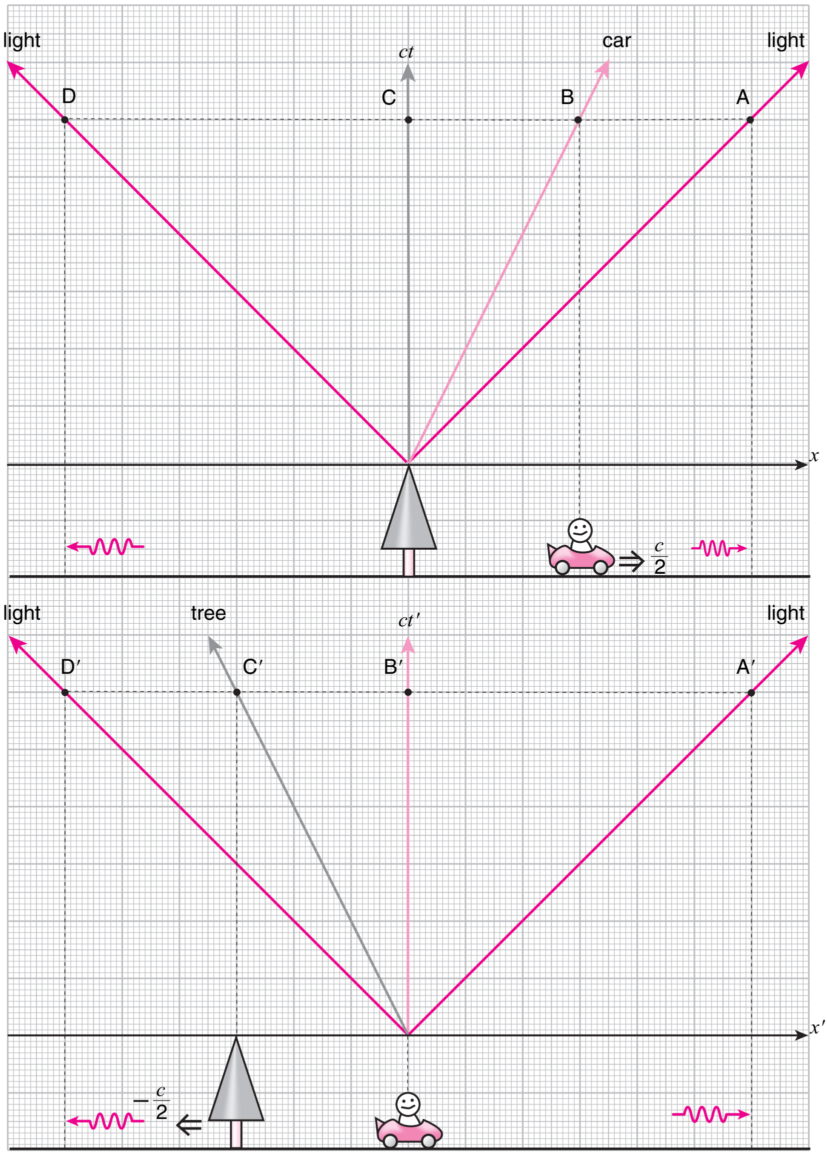

car-frame
At $t = t' = 0$
$\frac{c}{2}$
$\frac{c}{2}$
x' (m)
–5
–4
–3
–2
–1
0
1
2
3
4
5
c
c

The observation in the tree-frame is given by the spacetime diagram shown in the top figure opposite. The tree is at rest at the origin, the car is moving toward the right at speed $\frac{1}{2}c$, and the two photons are moving at speed c to the right and to the left. At any instant in this frame, the two photons are equidistant from the tree, while the car is midway between the tree and the photon on the right.

The exact same motion observed in the car-frame is given by the spacetime diagram shown in the bottom figure. The car is at rest at the origin, and the tree is moving toward the left at speed $\frac{1}{2}c$, and the two photons are again moving at speed c to the right and to the left. At any instant in this frame, the two photons are equidistant from the car, while the tree is midway between the car and the photon on the left.

Now, I am not making this up. The speeds of the photons are exactly the same in both frames. And this is regardless of what the motion of the light bulb was when it was turned on. The light bulb could have been at rest in the tree-frame, or it could have been fixed to the car, or it could have been moving at a totally different velocity all together, but the speeds of the two photons are c in both the tree-frame and the car-frame, and also in any other frame. This is an experimentally established fact and we cannot argue with it even if we don't like it.

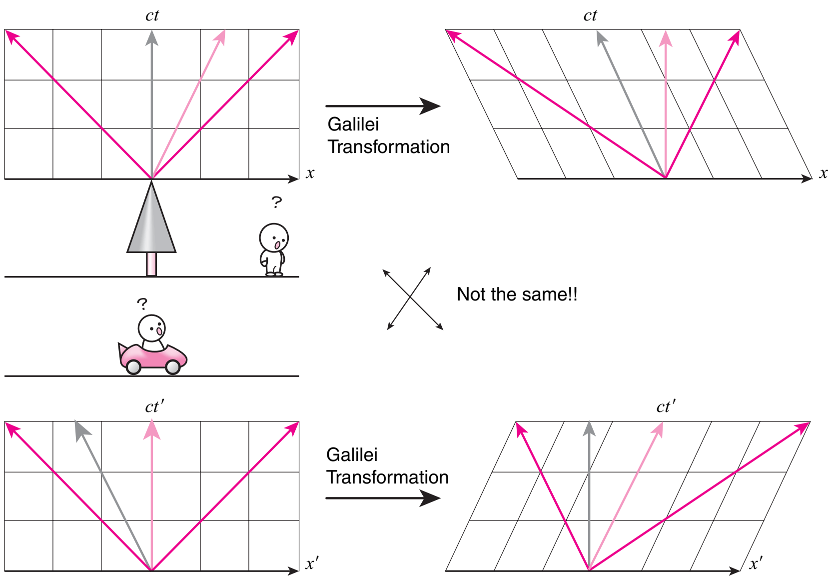
But how can this be?



light
ct
car
light
D
C
B
A
x
$\frac{c}{2}$
light
tree
ct'
light
D′
C′
B′
A′
x'
$-\frac{c}{2}$

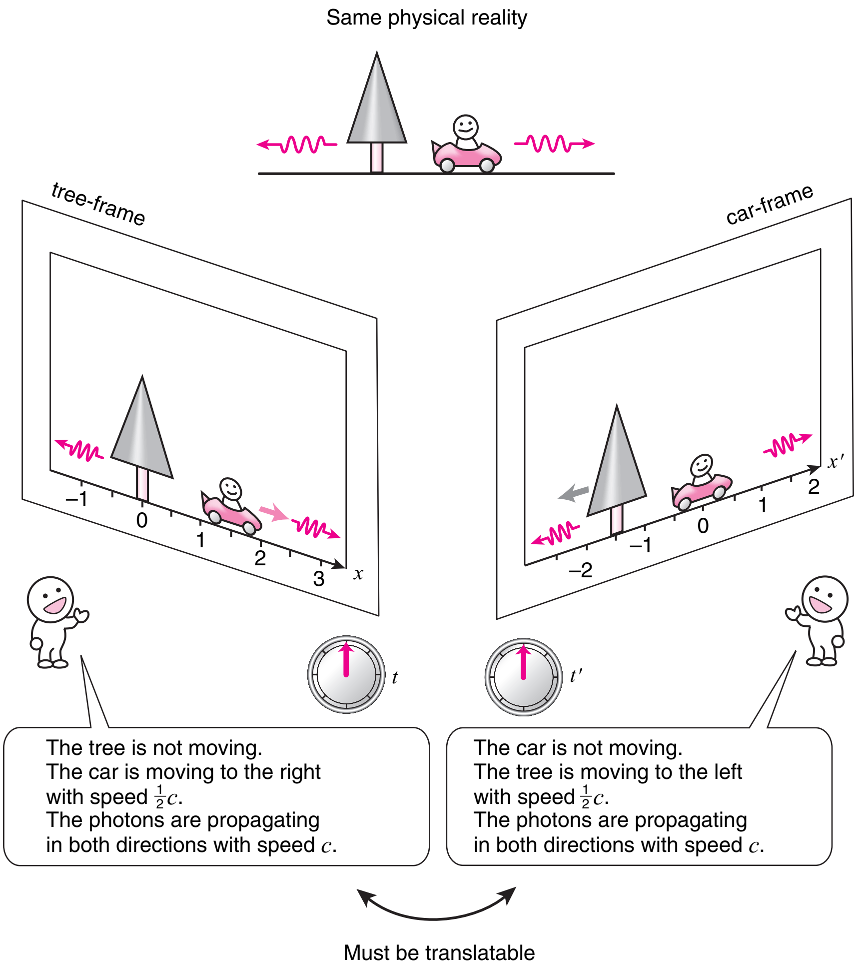
Now, I'm sure you agree that this is not what we expect at all. We expect that if the photon is traveling at speed c toward the right in the tree-frame, it will be observed as traveling at half of c toward the right in the car-frame. If the photon is traveling at speed c toward the left in the tree-frame, we expect it to be observed as traveling at 1.5 times c toward the left in the car-frame. This expectation is shown below in the spacetime diagram on the top-right, which is obtained by a Galilei transformation from the spacetime diagram in the tree-frame (shown top-left).

Similarly, if the photon is traveling at speed c toward the right in the car-frame, we expect it to be observed as traveling at 1.5 times c toward the right in the tree-frame. And if the photon is traveling at speed c toward the left in the car-frame, we expect it to be observed as traveling toward the left at half of c in the tree-frame. This expectation is shown in the spacetime diagram on the bottom-right, which is obtained by a Galilei transformation from the spacetime diagram in the car-frame (shown bottom-left).



What this is telling us is that our expectation, and therefore the Galilei transformation, is wrong since it does not agree with observation.

So what is going on? We have two observers observing the exact same physical reality from two different frames so their observations cannot be independent. The observation in one frame must contain the exact same information as the observation in the other frame, since the choice of frame should not matter, so there must exist a translation which tells us how the observations from different frames are related. What is that translation?



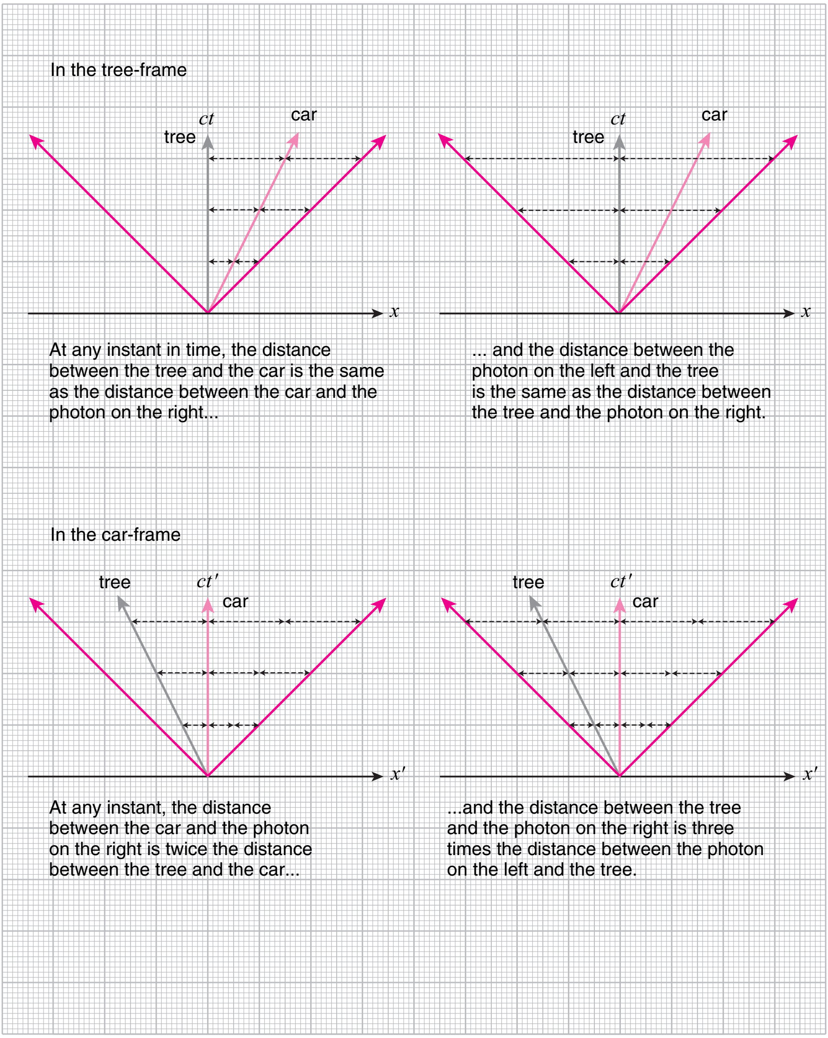
4.4 The solution

How can the observations from the tree- and car-frames be the observation of the exact same physical reality?

At any instant in the tree-frame, the car is midway between the tree and the photon on the right. But at any instant in the car-frame, the car is only one-third of the way between the tree and the photon on the right. How can the car be at two different places *at the same time*?

Or take the photon traveling toward the left. In the tree-frame, at any instant it is at the same distance from the tree as the photon traveling toward the right. But in the car-frame, at any instant it is only one-third of the distance from the tree as the photon traveling toward the right. How can the photon be at two different places *at the same time*?

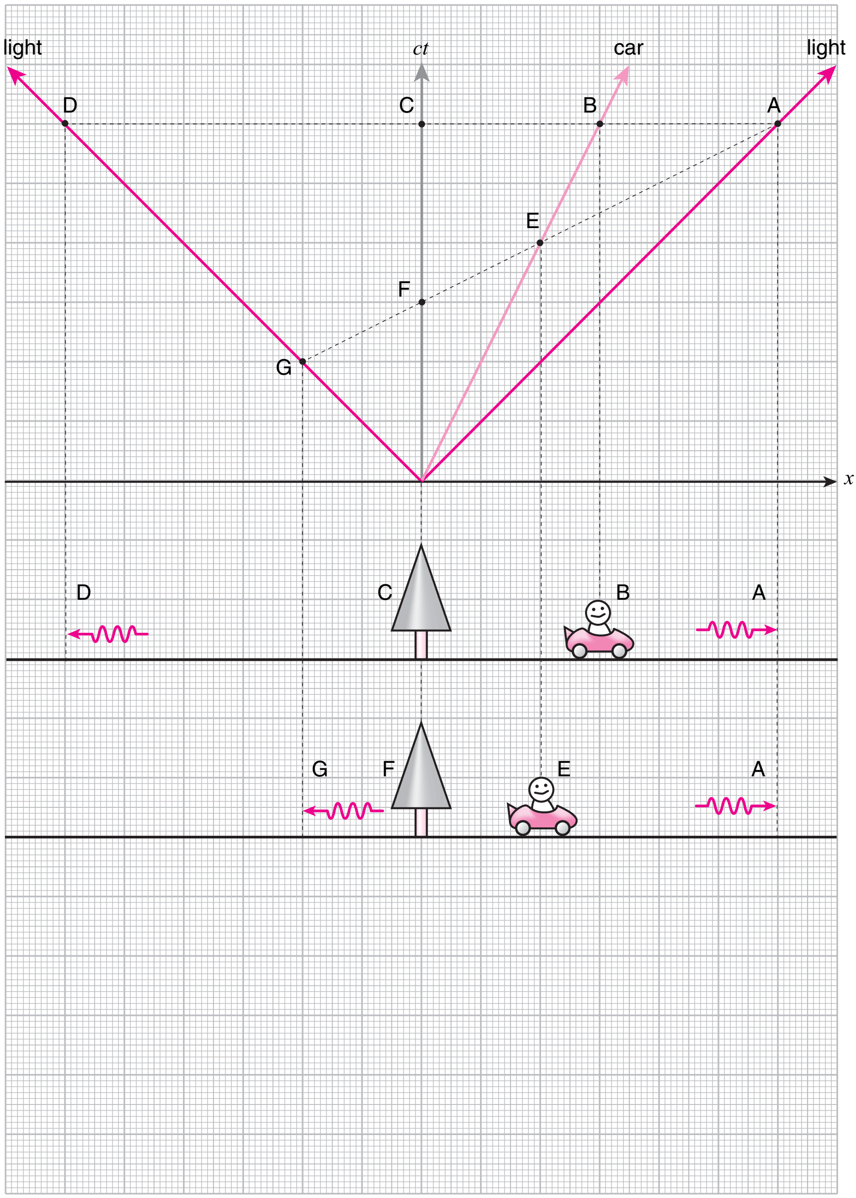
I am sure you would agree that *nothing can be at two different places at the same time*, whether it be a car or a photon. That is just common sense. Then how are we to interpret our current problem? Now, some people may argue that this shows that reality itself will depend on the observer, and that there is no such thing as an objective reality. But Einstein disagrees. He points out that while *nothing can be at two different places the same time*, they can be *at two different places at two different times.*


In the tree-frame
ct
tree
car
x
At any instant in time, the distance between the tree and the car is the same as the distance between the car and the photon on the right...
ct
tree
car
x
... and the distance between the photon on the left and the tree is the same as the distance between the tree and the photon on the right.
In the car-frame
tree
ct′
car
x′
At any instant, the distance between the car and the photon on the right is twice the distance between the tree and the car...
tree
ct′
car
x′
...and the distance between the tree and the photon on the right is three times the distance between the photon on the left and the tree.

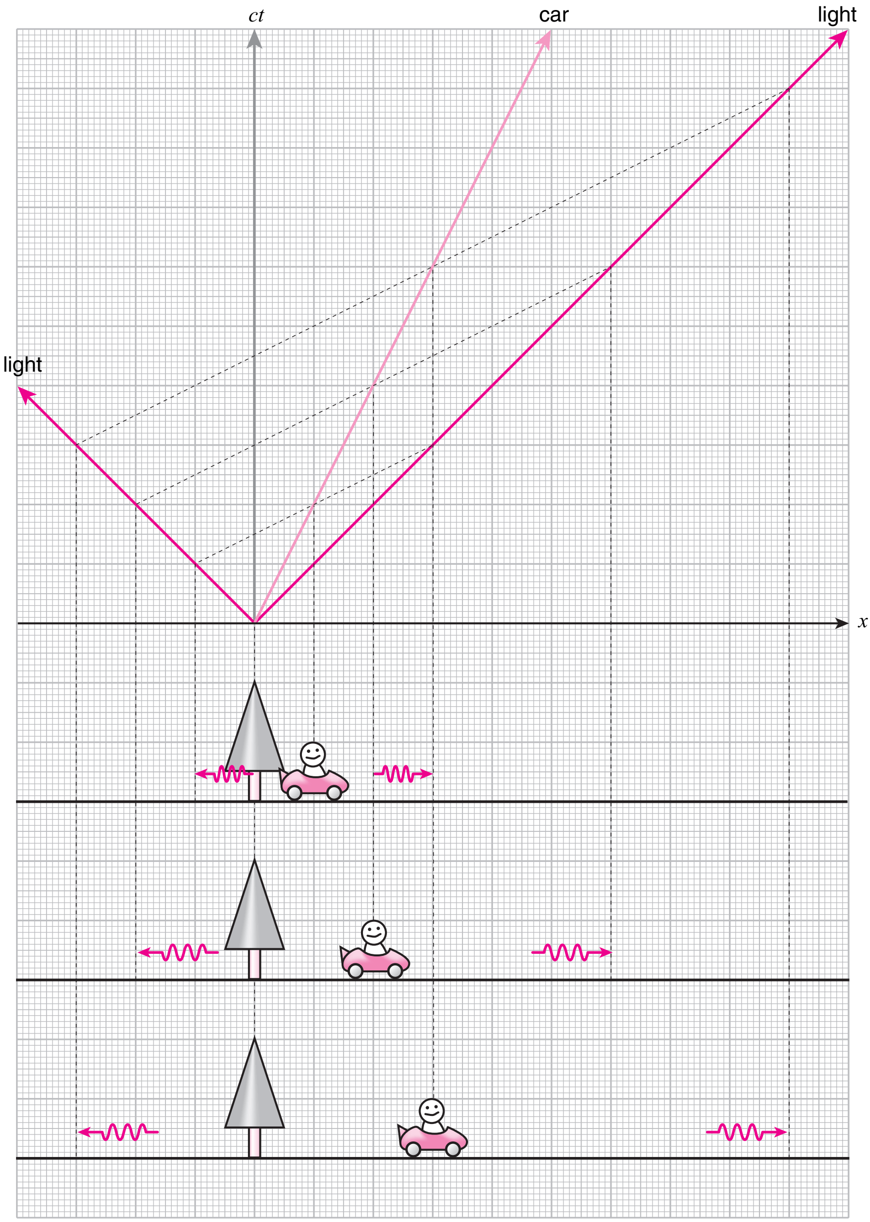
Consider the spacetime diagram in the tree-frame shown here. Slicing the diagram along a horizontal line tells us where the objects are at that particular instant *in the tree-frame*. Such a slice is shown on the spacetime diagram. We can see that the car at B is midway between the tree at C and the photon on the right at A. In the car-frame, the car is supposed to be only one-third of the way between the tree and the photon on the right at any instant, which is clearly not the case here. Now trace the worldline of the car backwards from B toward the origin until you reach point E. This point is at an earlier time (in the tree-frame) than A, but at this point the car is precisely one-third as far from the tree as the photon on the right is at A.

Similarly, the photon on the left at D is at the same distance from the tree as the photon on the right at A. In the car-frame, the photon on the left is supposed to be only one-third as far away from the tree as the photon on the right at any instant. Now, trace the worldline of the photon on the left backwards from D to the origin until you reach point G. This point is at an earlier time (in the tree-frame) than either A or E, but at this point the photon on the left is precisely one-third as far from the tree as the photon on the right is at A.

Notice that the points A, E, and G all fall on a straight line. There seems to be a definite pattern here. Could it be that though the points A, B, C, and D are *at the same time* in the tree-frame, they are not *at the same time* in the car-frame, and the points that are *at the same time* as A in the car-frame are actually points E, F, and G?


light
ct
car
light
D
C
B
A
E
F
G
x
D
C
B
A
G
F
E
A

It is easy to see that if you slice the tree-frame diagram with any tilted line parallel to the line AEFG, then the relative positions of the tree, the car, and the two photons turn out to be precisely what would be expected in the car-frame.

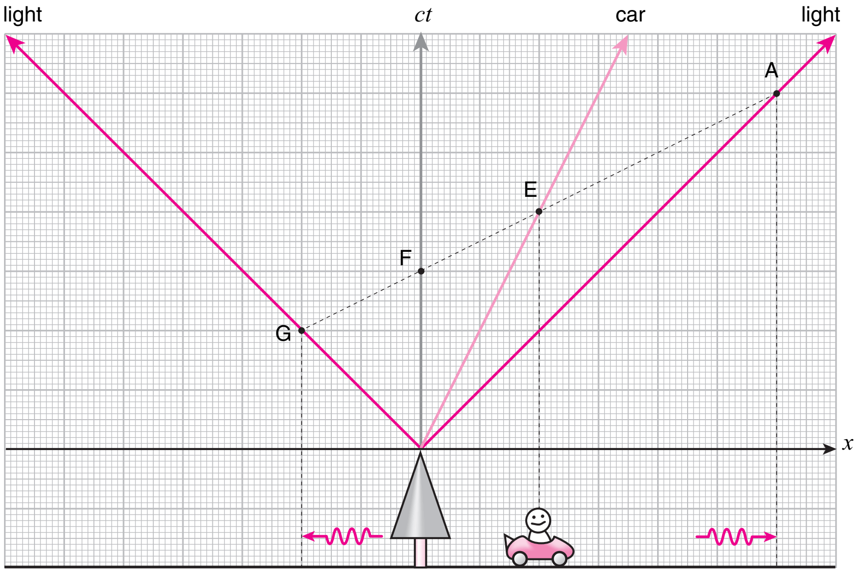


ct
car
light
light
x

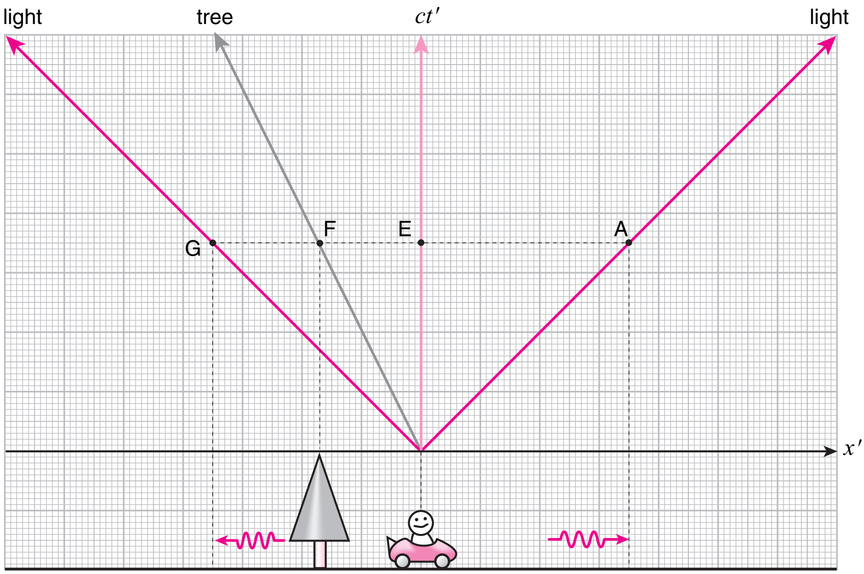
So could it be that though the points A, E, F, and G are not *at the same time* in the tree-frame, they are *at the same time* in the car-frame? Do they correspond to points that fall on a horizontal line on the car-frame diagram?

This is a revolutionary idea indeed, since Einstein is saying that the concept of events happening *at the same time* depends on the frame, just like the concept of *at the same place* depends on the frame. Two events that happen *at the same time* for one observer do not happen *at the same time* for another observer, and vice versa.

Spacetime diagram in the tree-frame



Spacetime diagram in the car-frame

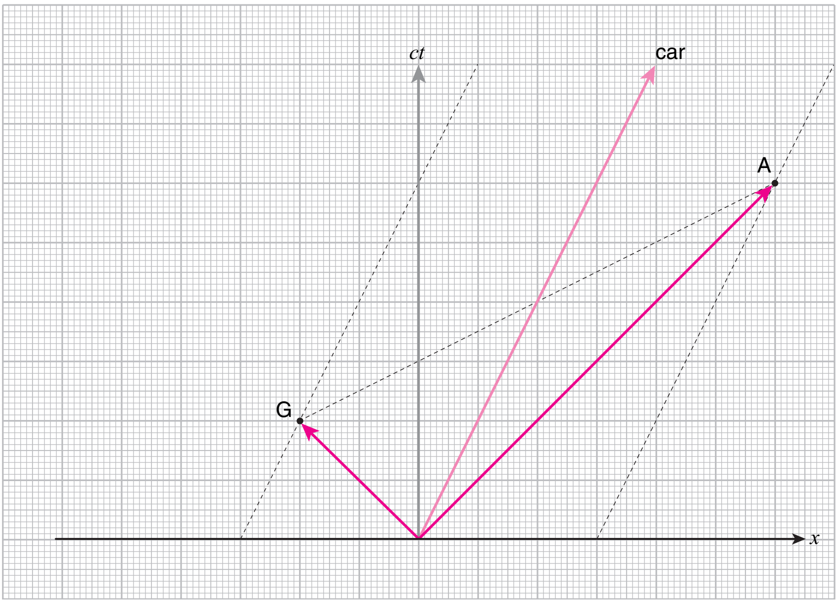


4.5 Einstein's argument

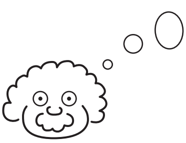
Let's take a look at the argument of Einstein himself presented in his book *Relativity: The Special and the General Theory* [6], which was written for the layman.

Einstein says: *If two photons both travel at the same speed c, then the time it takes for them to cover the same distance must be the same.* I am sure you all agree with this statement. After all, this is what you mean when you say that two objects are traveling at the same speed.[5]

Now notice that on the spacetime diagram on page 75, the points A and G are the same distance away from the car. (The diagram is reproduced below so that you can see this more clearly.) In the car-frame,



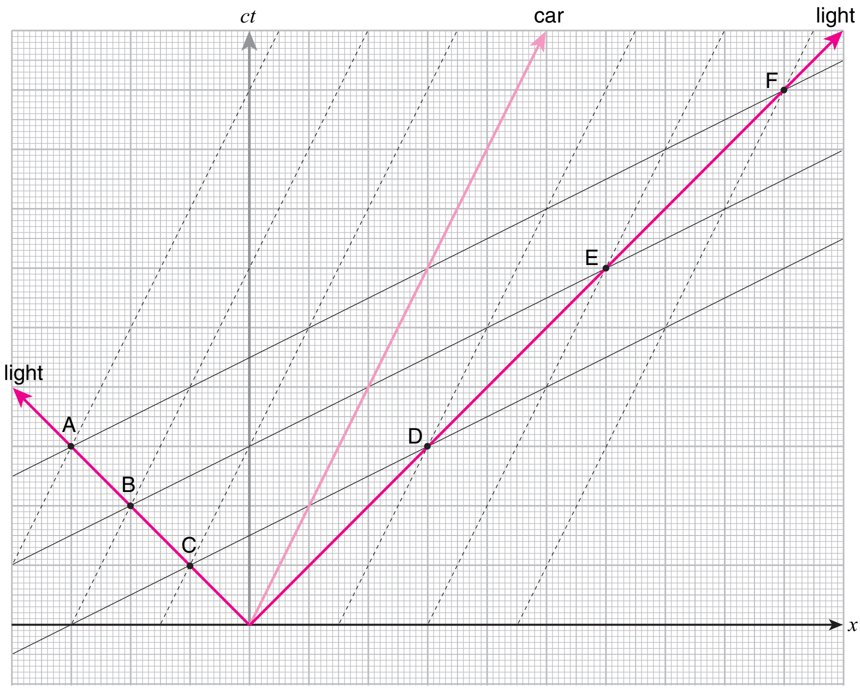
Einstein's argument:



In the car-frame A and G are at the exact same distance from the car. Therefore, in the car-frame the photons travel the exact same distance to reach A and G. Since the speed of light in both directions is c in the car-frame, it takes the exact same amount of time for the photons to reach A and G. Therefore, A and G must be at the same time in the car-frame.

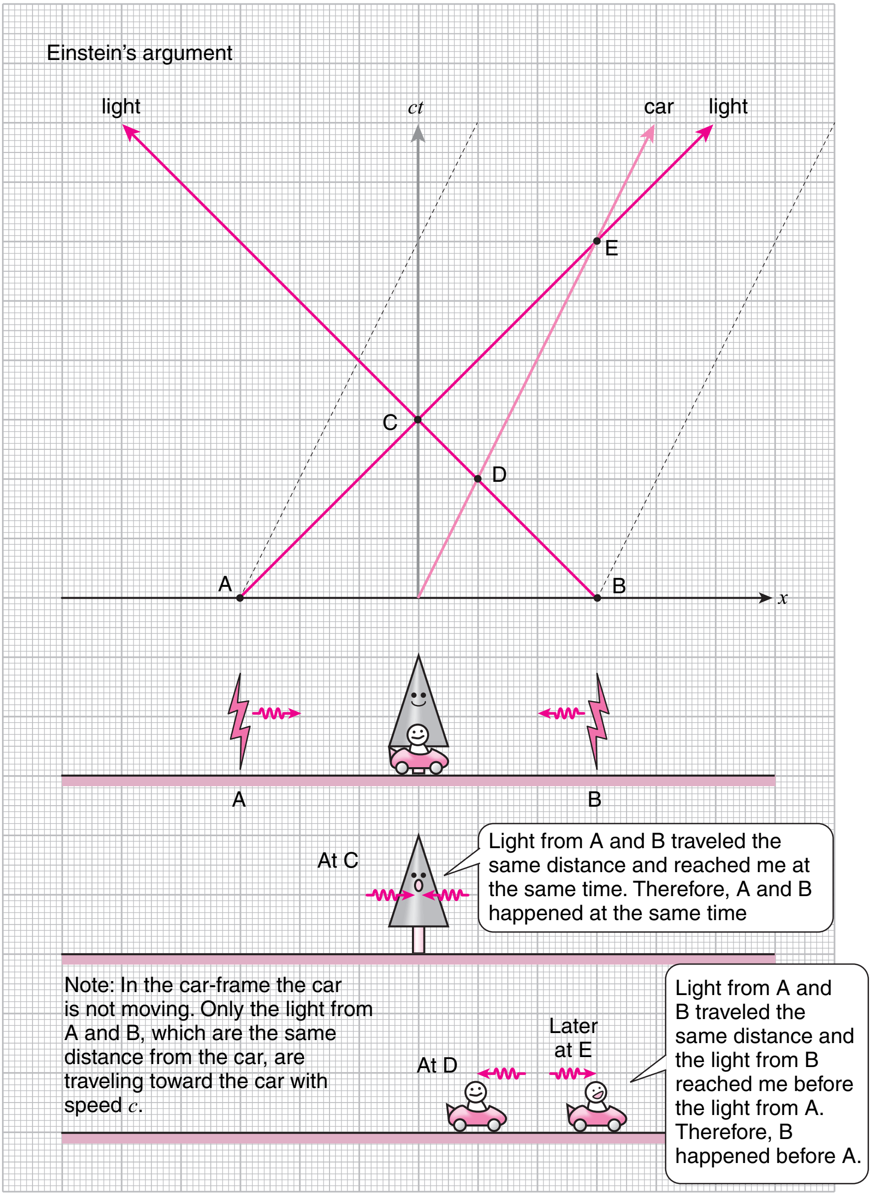
the two photons have traveled the exact same distance to reach points A and G, so they must have taken the exact same amount of time to do so since the speed of light is always c in any frame. And since the two photons started out from the origin at the same time, points A and G must also be *at the same time* in the car-frame, even though they are not *at the same time* in the tree-frame.

Indeed, any two points that are equidistant from the origin in the car-frame must be reached *at the same time* in the car-frame by the two photons. So on the diagram shown here, points A and F are *at the same time*, points B and E are *at the same time*, and points C and D are *at the same time* in the car-frame.

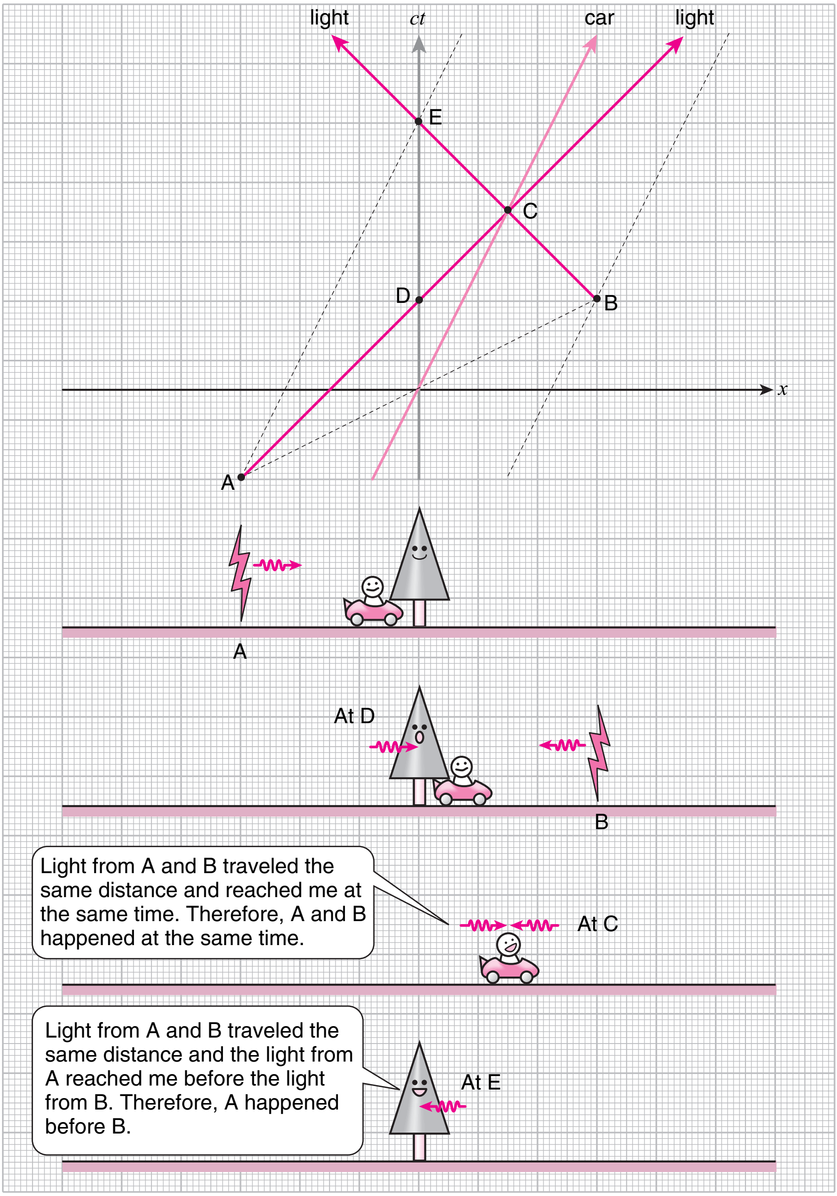


Einstein continues: After all, how do we know whether two events that are spatially separated happened at the same time in the first place? Take two lightning strikes that happen at points A and B on the spacetime diagram shown here. The observer in the tree-frame, who is with the clock fixed to the tree, will see the lightning flashes simultaneously at point C. Since the flashes of light from A and B are both traveling at the same speed c, and they cover the same distance to reach the tree, the observer at the tree will conclude that A and B happened *at the same time* in the tree-frame.

On the other hand, the flash of light from B reaches the car at point D, while the flash of light from A reaches the car at point E. The observer on the car will see the flash from B before the flash from A. Since both flashes covered the same distance in the car-frame also, and since they are again traveling at the same speed c, the observer on the car will have to conclude that B happened before A in the car-frame.


Einstein's argument
light
ct
car
light
E
C
D
A
B
x
A
B
At C
Light from A and B traveled the same distance and reached me at the same time. Therefore, A and B happened at the same time
Note: In the car-frame the car is not moving. Only the light from A and B, which are the same distance from the car, are traveling toward the car with speed c.
At D
Later at E
Light from A and B traveled the same distance and the light from B reached me before the light from A. Therefore, B happened before A.

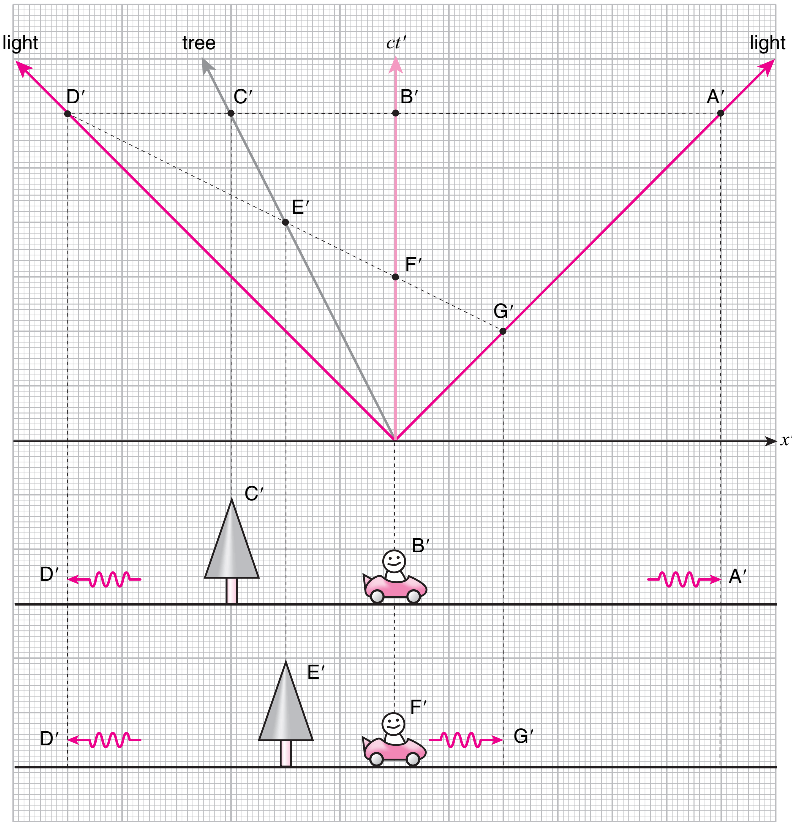
Similarly, if the lightning flashes occur at points A and B shown here, the observer in the car-frame will conclude that A and B happened *at the same time*, but the observer in the tree-frame will conclude that A happened before B.



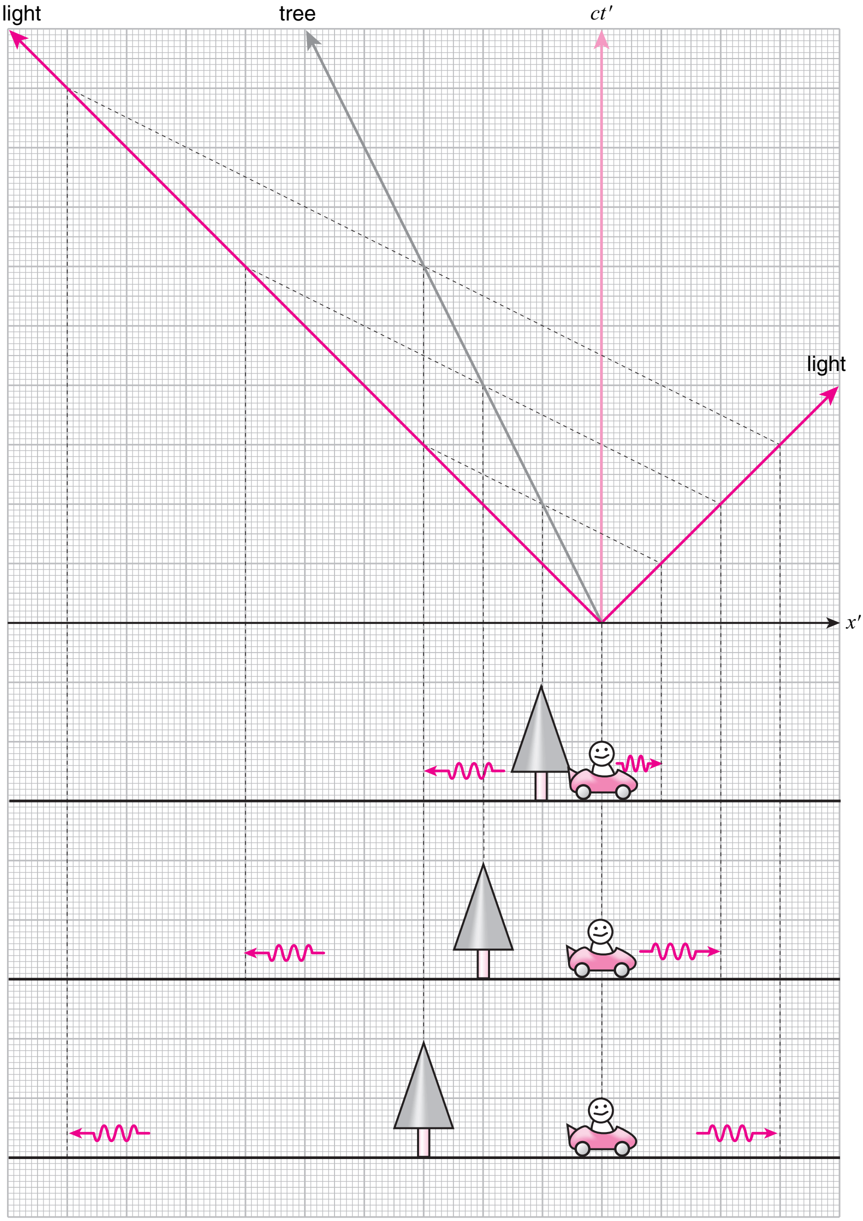
4.6 The solution, continued

We have seen that points that are *at the same time* in the car-frame lie along tilted lines on the spacetime diagram in the tree-frame.

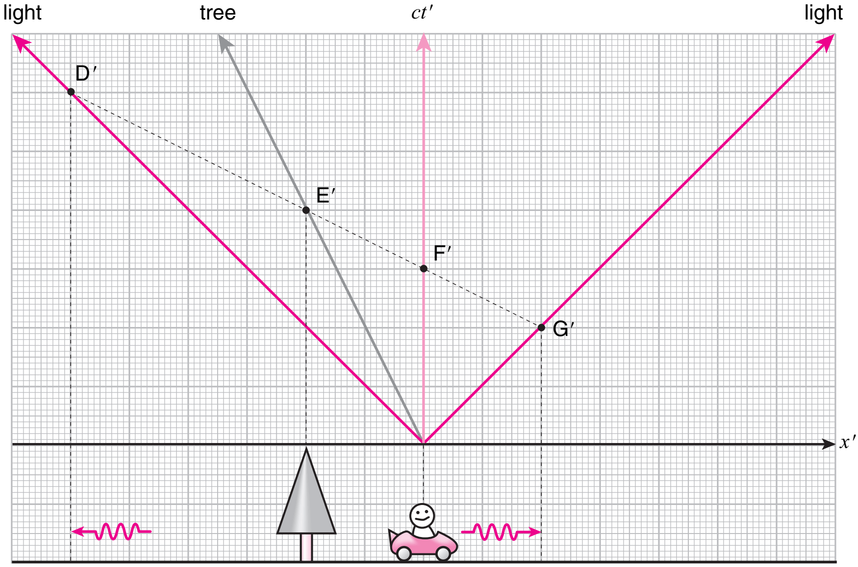
We can go through the exact same argument to conclude that points that are *at the same time* in the tree-frame lie along tilted lines on the spacetime diagram in the car-frame, as shown here.



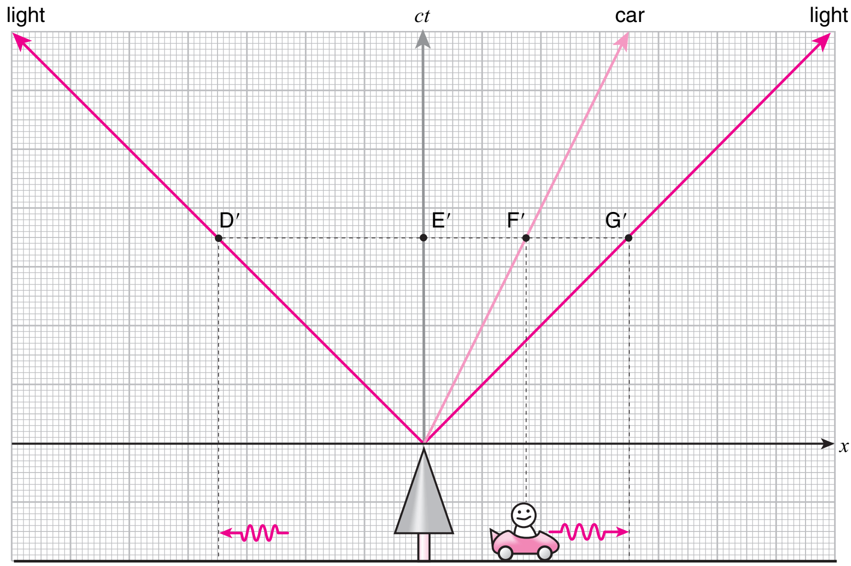
Slice the the car-frame spacetime diagram with any line parallel to the line D′E′F′G′ on the previous page and we will always find the correct relative positions of the tree, the car, and the photons in the tree-frame.



So the points D′, E′, F′, and G′ on the car-frame diagram must correspond to points that fall on a horizontal line on the tree-frame diagram as shown here.



Spacetime diagram in the car-frame

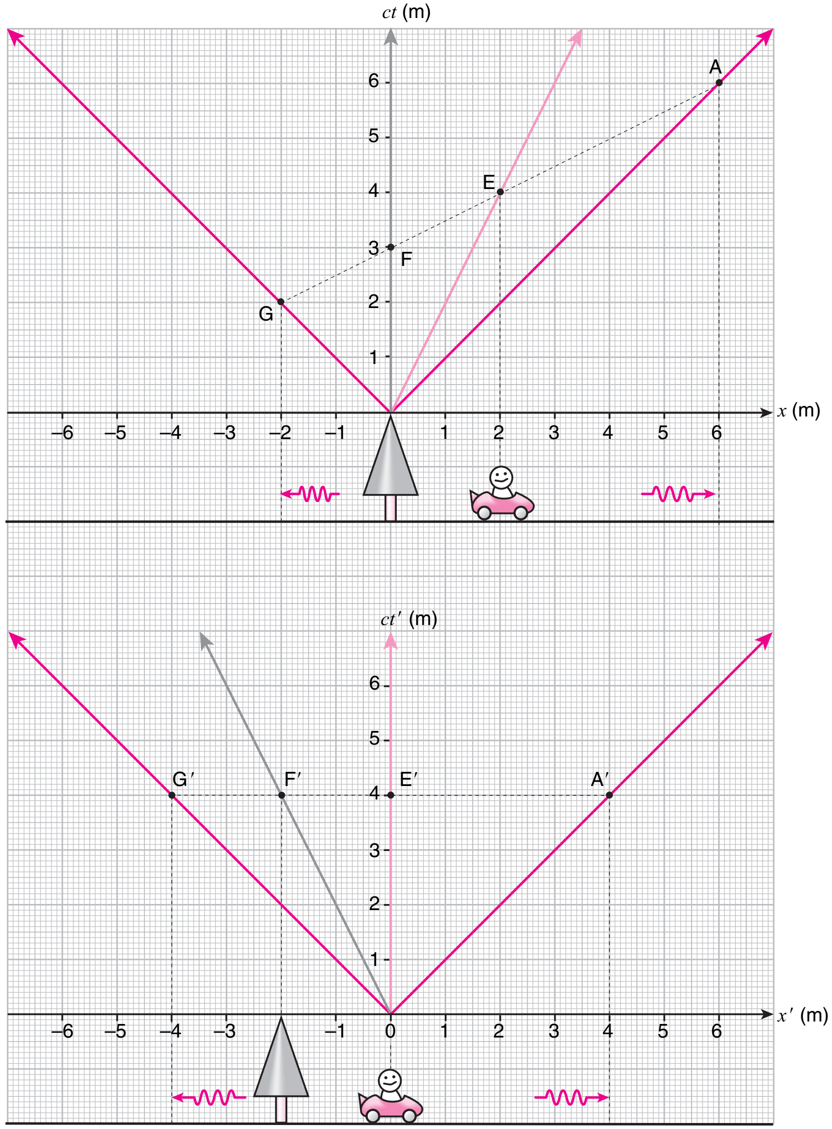


Spacetime diagram in the tree-frame

4.7 Conservation of spacetime volume

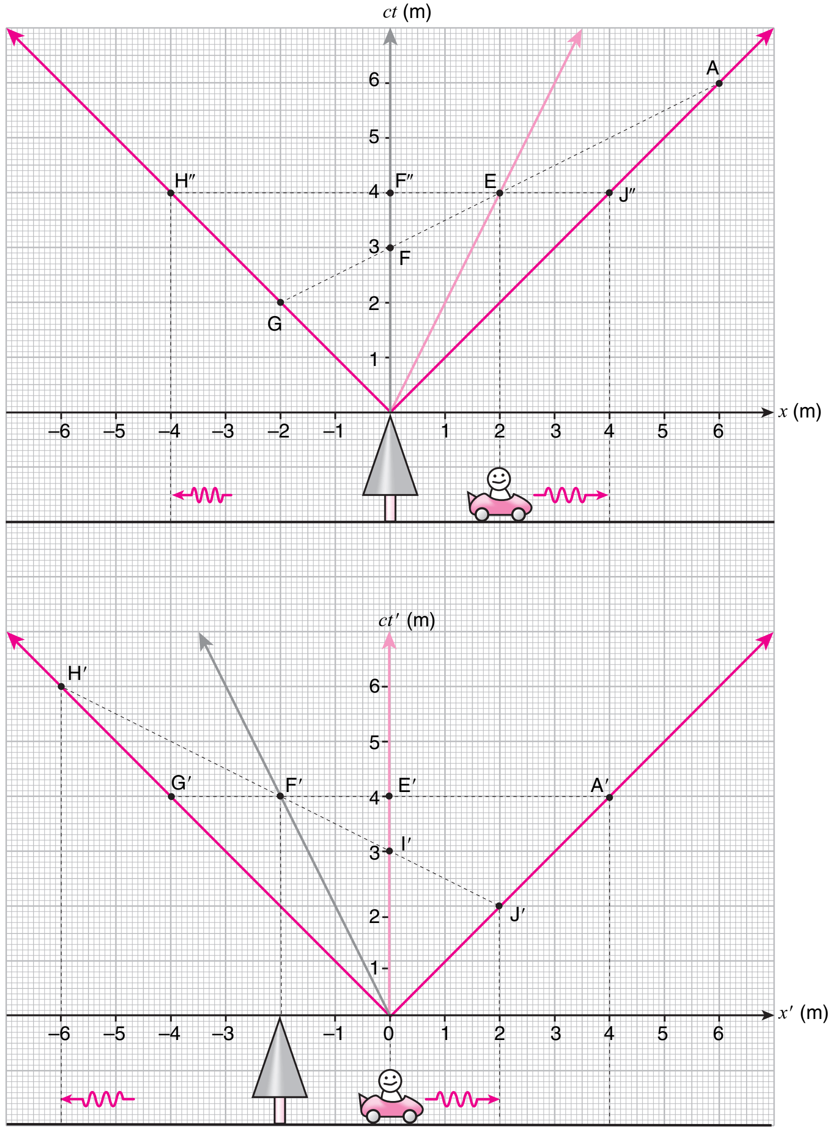
Now, I argued on page 78 that the points AEFG in the tree-frame diagram fall on a horizontal line on the car-frame diagram, and I also just argued on page 87 that points D′E′F′G′ on the car-frame diagram fall on a horizontal line on the tree-frame diagram. But I have not told you in either case which horizontal line it is. So let me do so now.

Take the line AEFG on the tree-frame diagram first. (Reproduced here on the opposite page, top figure.) One is tempted to think that they correspond to the points A′E′F′G′ on the car-frame diagram as shown on the bottom figure since the spatial separations of the tree, the car, and two photons in x' are the same as those in x.



ct (m)
x (m)
A
E
F
G
6
5
4
3
2
1
–6
–5
–4
–3
–2
–1
1
2
3
4
5
6
ct′ (m)
x′ (m)
G′
F′
E′
A′
0

However, this cannot be correct since it implies that the points H′F′I′J′ on the car-frame diagram, shown on the opposite page, bottom figure, correspond to the points H″F″EJ″ on the tree-frame diagram, shown above it. And F″ is not the same point as F even though there must exist a one-to-one correspondence between the points on the two diagrams.



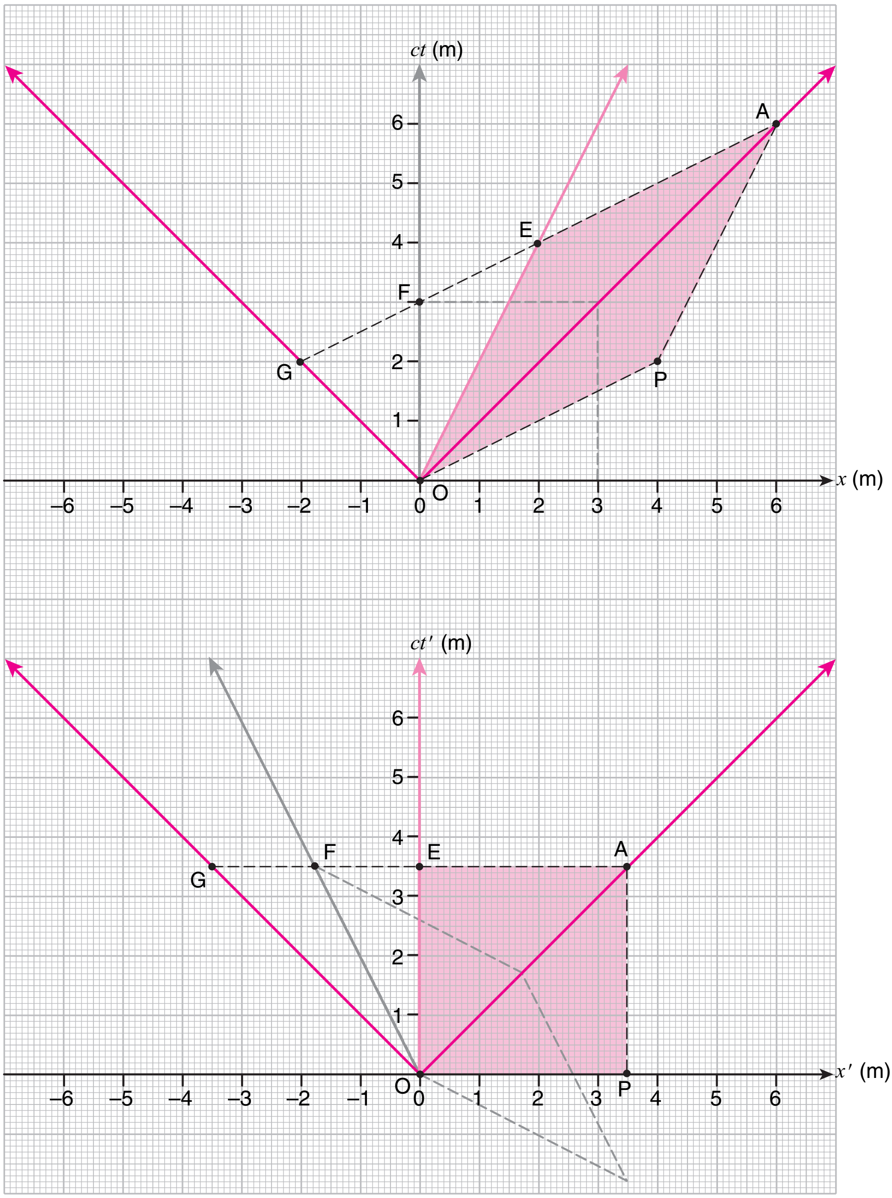
ct (m)
x (m)
A
H″
F″
E
J″
F
G
−6 −5 −4 −3 −2 −1 1 2 3 4 5 6
1 2 3 4 5 6
ct′ (m)
x′ (m)
H′
G′
F′
E′
A′
I′
J′
−6 −5 −4 −3 −1 0 1 2 3 4 5 6
1 2 3 4 5 6

So what is the correct correspondence? On the tree-frame diagram, draw a line parallel to the car worldline that goes through A, and a line parallel to AEFG that goes through the origin O. Call the point where the two lines cross P. Note that P is *at the same time* as O, and *at the same place* as A in the car-frame. So the diamond AEOP on the tree-frame diagram must correspond to a square, like the one shown in the bottom figure, on the car-frame diagram. The corresponding square on the car-frame diagram is the one that has *the same area* as the diamond on the tree-frame diagram.

In the current case we are considering here, the area enclosed in the diamond AEOP on the tree-frame diagram is $12\,\mathrm{m}^2$. (Just count the number of squares inside the diamond.) So the length of the sides of the square on the car-frame diagram must be $\sqrt{12\,\mathrm{m}^2} = 2\sqrt{3}\,\mathrm{m} \approx 3.5\,\mathrm{m}$. Therefore, points AEFG lie on the horizontal line with $ct' \approx 3.5\,\mathrm{m}$, as shown on the car-frame diagram.

Under this rule, the square on the tree-frame diagram which has OF as one of its sides will be transformed to the diamond with the same area on the car-frame diagram as shown. So under the inverse transformation, F will be mapped back to where it originally was.

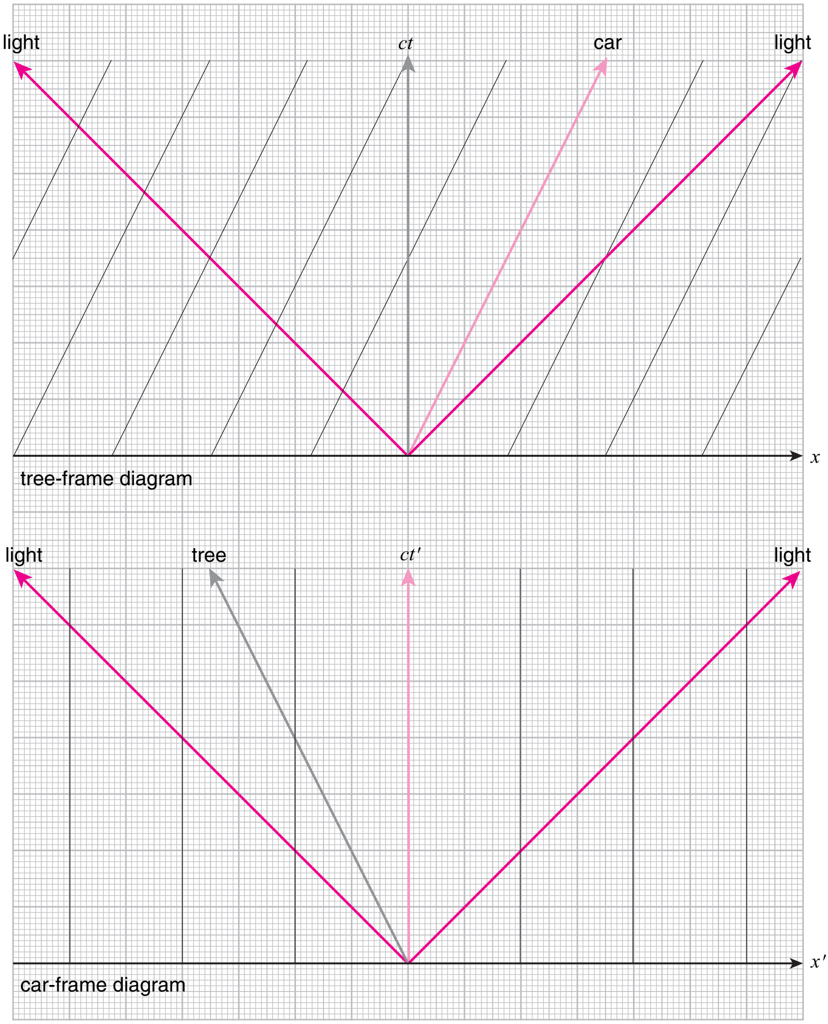
This conservation of spacetime area[6] maintains the symmetry between the tree- and car-frames, since each is moving at the exact same speed when observed from the other frame, and ensures that the correspondence between the points on the two diagrams is one-to-one. It is a consequence of the fact that the "number of events" enclosed in AEOP must be independent of the frame of reference. Recall that a similar situation also existed in the Galilei transformation case. There too, the areas of corresponding regions on the spacetime diagrams of different frames were the same.


ct (m)
x (m)
A
E
F
G
P
O
6
5
4
3
2
1
0
−6
−5
−4
−3
−2
−1
ct′ (m)
x′ (m)

4.8 The Lorentz transformation

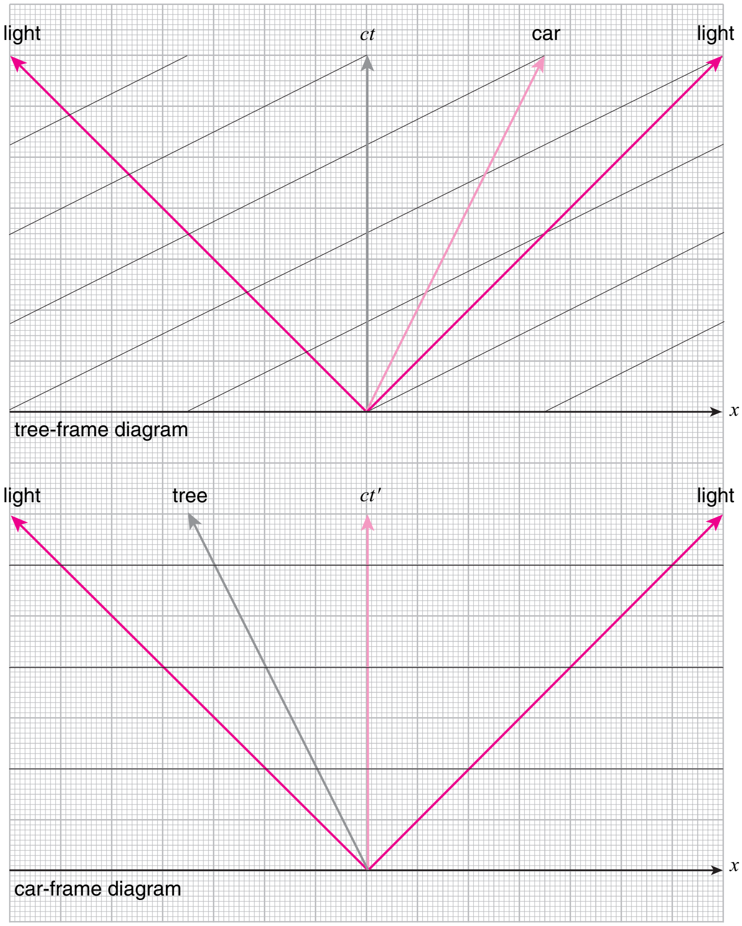
Let us now summarize what we have learned so far.

The points that are *at the same place* in the car-frame fall on vertical lines on the car-frame diagram, but they fall on tilted lines that are parallel to the worldline of the car on the tree-frame diagram. This is nothing new.



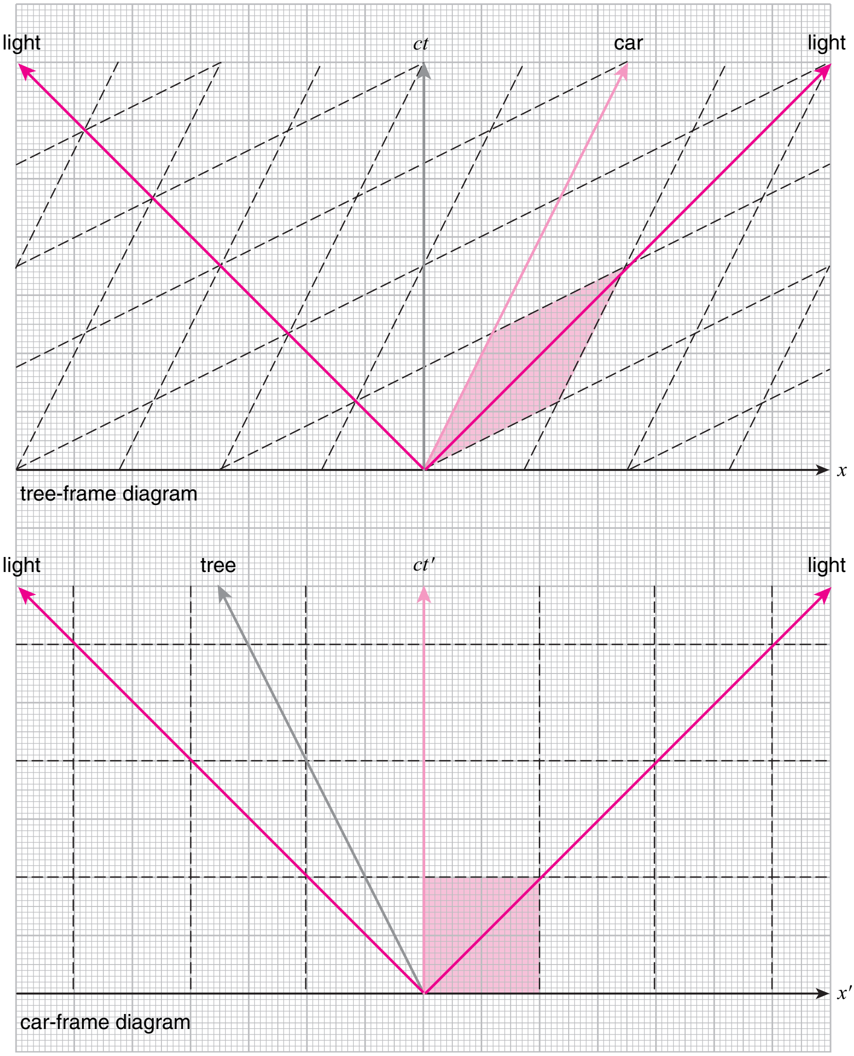
Lines that connect points that are at the same place in the car-frame.

What is new is that the points that are *at the same time* in the car-frame fall on horizontal lines on the car-frame diagram, but they fall on tilted lines on the tree-frame diagram as shown here. The tilt is easily determined since if we take the worldlines of two photons that were emitted from the spacetime origin, then the points along them that are equidistant from the car must be *at the same time*.



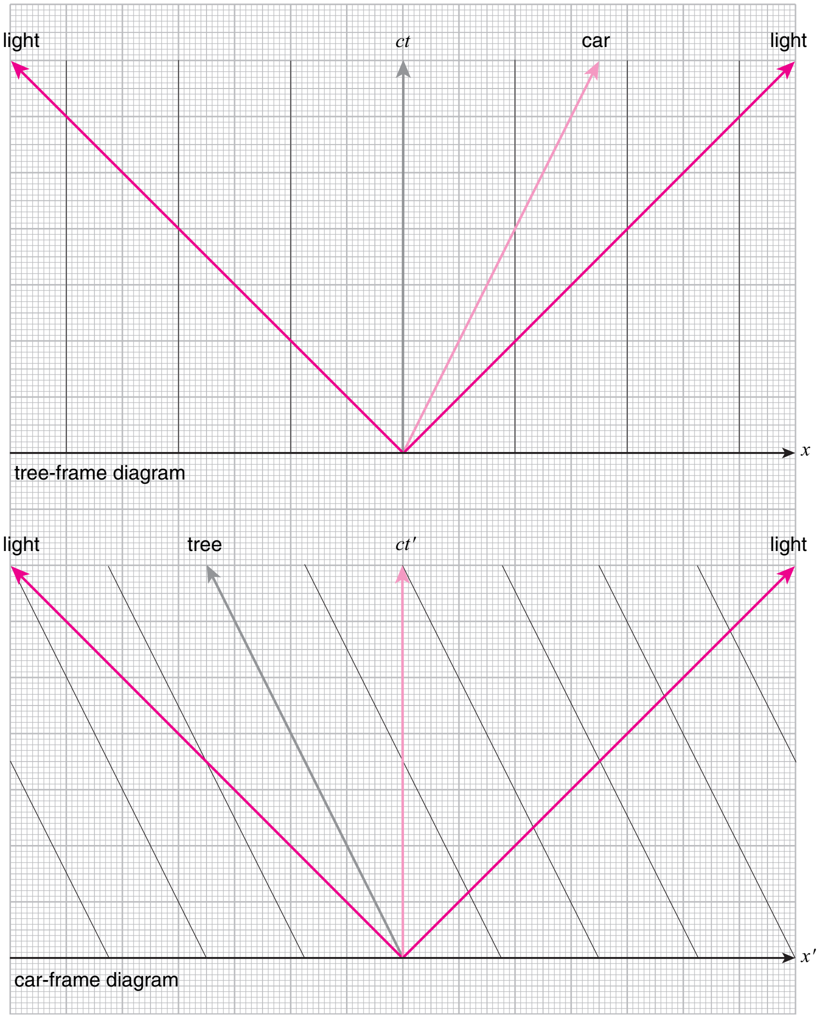
Lines that connect points that are at the same time in the car-frame.

Putting this together, we conclude that the square grid on the car-frame diagram corresponds to the skewed grid on the tree-frame diagram shown here. The spacing of the grid lines must be such that the spacetime area enclosed in a unit cell is the same on both diagrams.



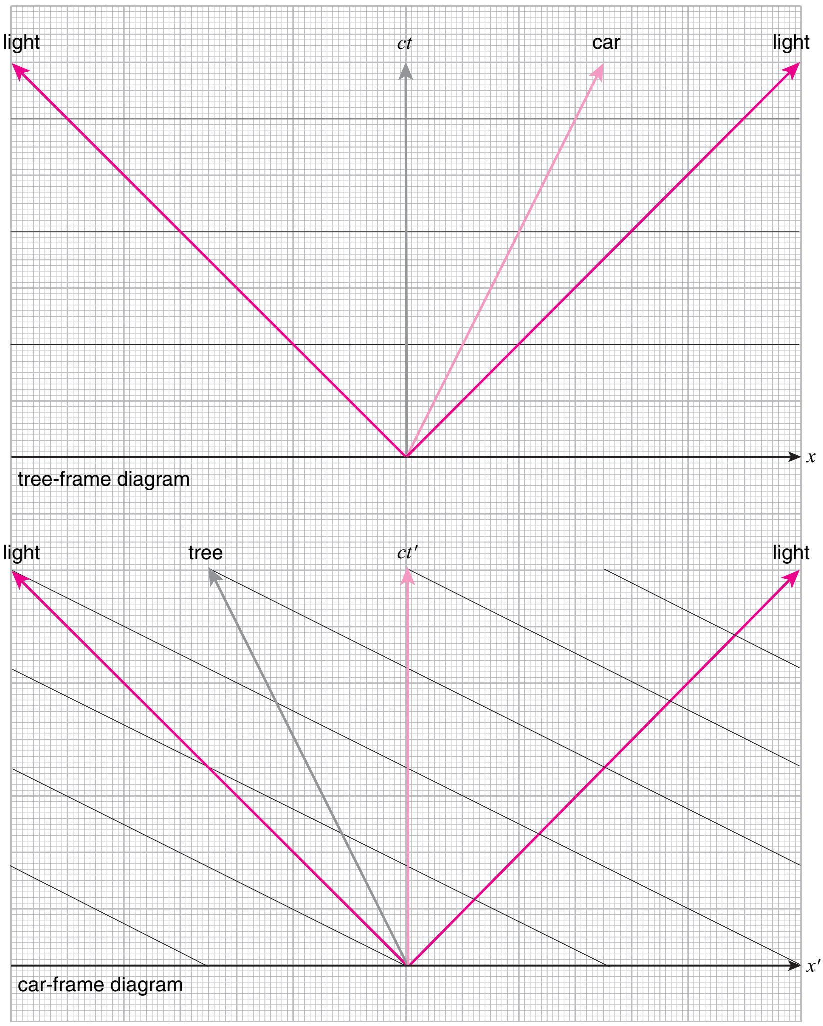
Corresponding grids in the tree- and car-frames.

Similarly, the points that are *at the same place* in the tree-frame fall on vertical lines on the tree-frame diagram, but they fall on tilted lines that are parallel to the worldline of the tree on the car-frame diagram.



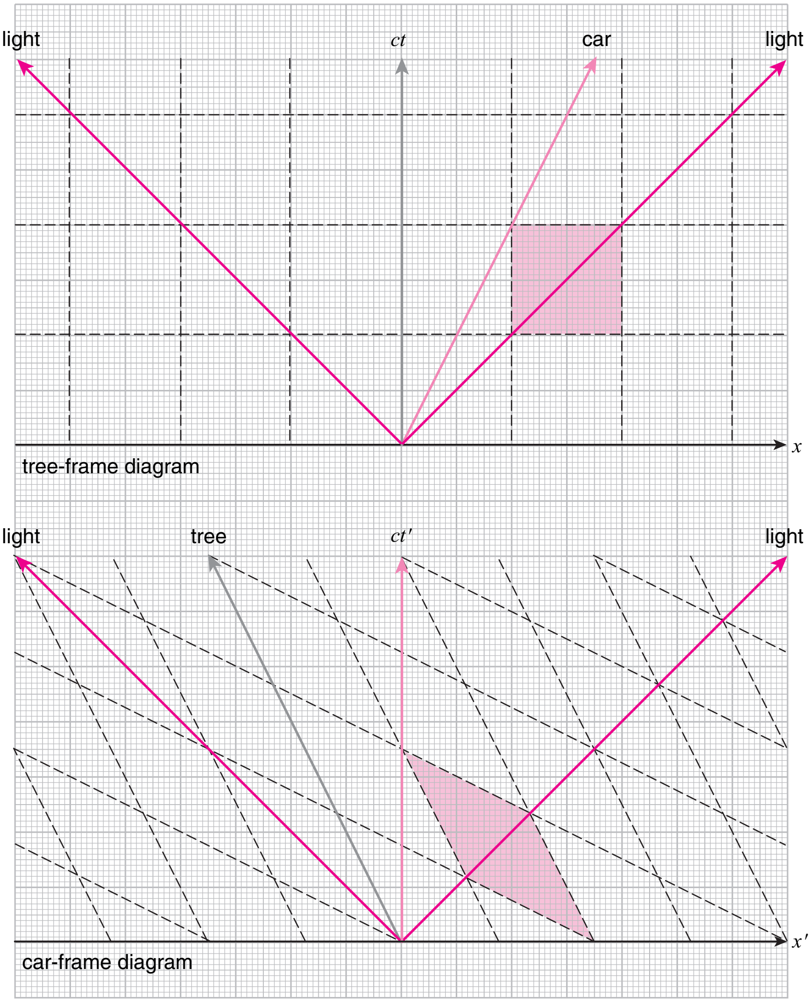
Lines that connect points that are at the same place in the tree-frame.

And the points that are *at the same time* in the tree-frame fall on horizontal lines on the tree-frame diagram, but they fall on tilted lines on the car-frame diagram as shown here. Again, the tilt is easily determined since points along the worldlines of two photons emitted from the spacetime origin that are equidistant from the tree must be *at the same time*.



Lines that connect points that are at the same time in the tree-frame.

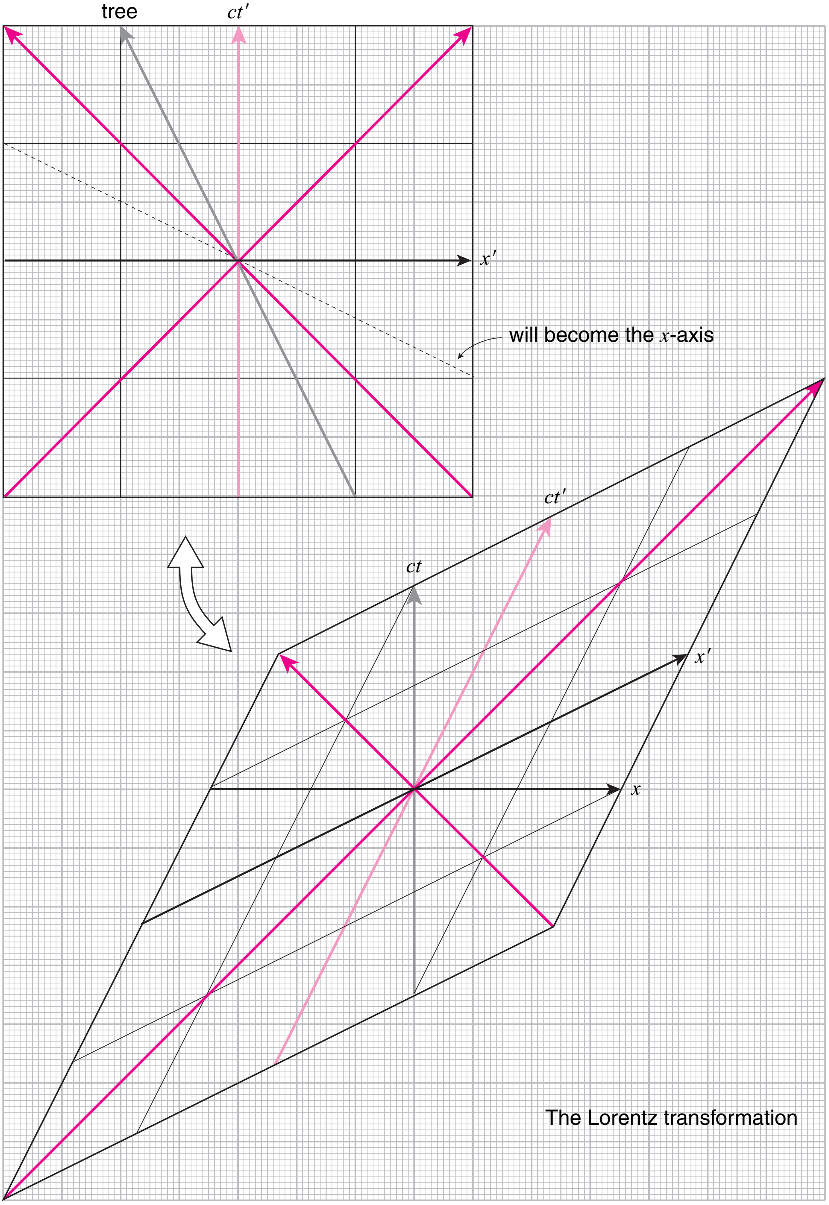
So the square grid on the tree-frame diagram corresponds to the skewed grid on the car-frame diagram shown here. Again, the spacing of the grid lines must be such that the spacetime area enclosed in a unit cell is the same on both diagrams.



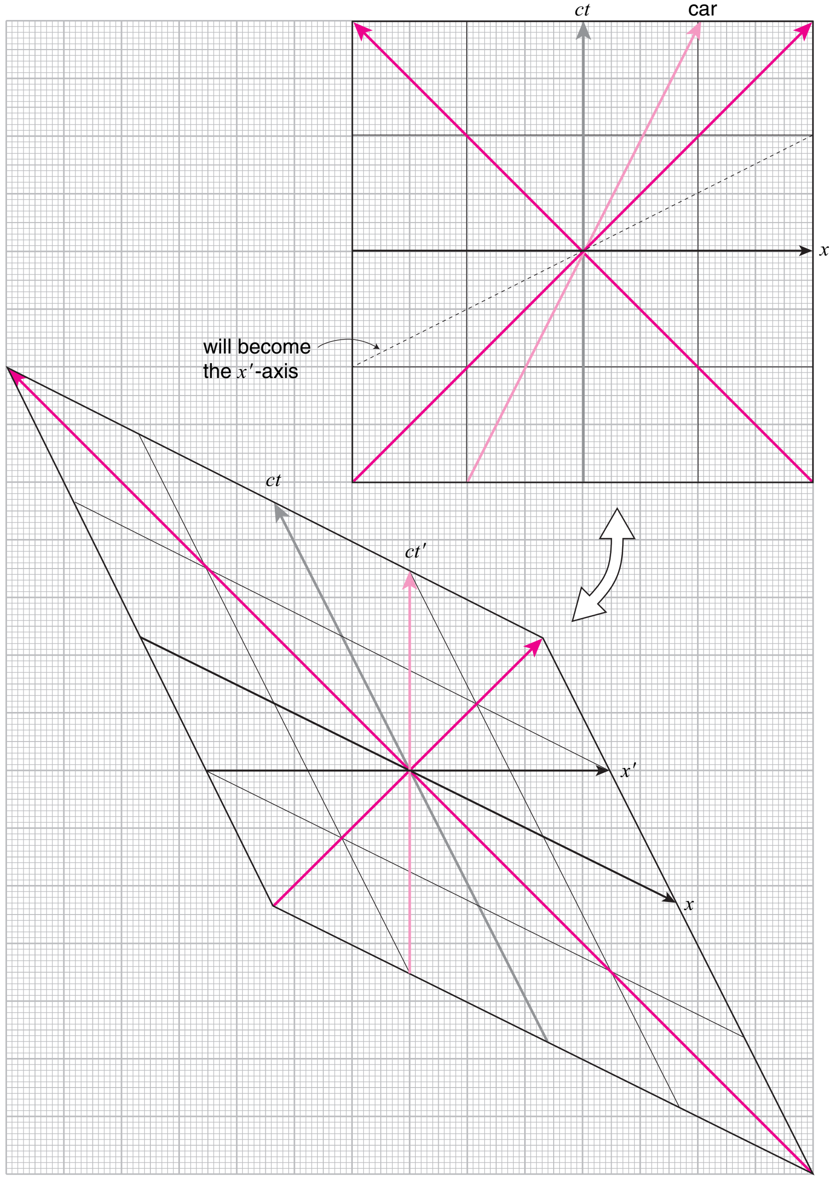
Corresponding grids in the tree- and car-frames

So to go from the car-frame to the tree-frame, the spacetime diagram must be stretched in the northeast and southwest directions, and squeezed from the northwest and southeast directions, until the worldline of the tree becomes vertical while maintaining the worldlines of both photons at an angle of 45 degrees from the horizontal. The amount of squeezing in the northwest–southeast direction must be the reciprocal of the amount of stretching in the northeast–southwest direction in order to keep the spacetime area inside a unit cell invariant.

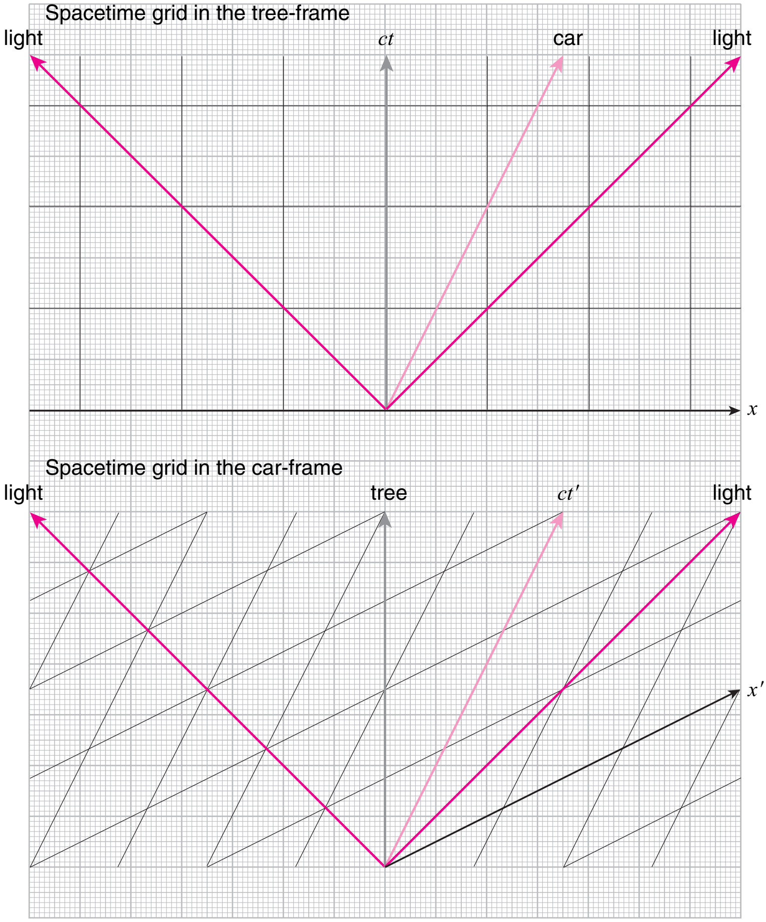
This transformation is known as the *Lorentz transformation.*[7,8]


tree
ct'
x'
will become the x-axis
ct'
ct
x'
x
The Lorentz transformation

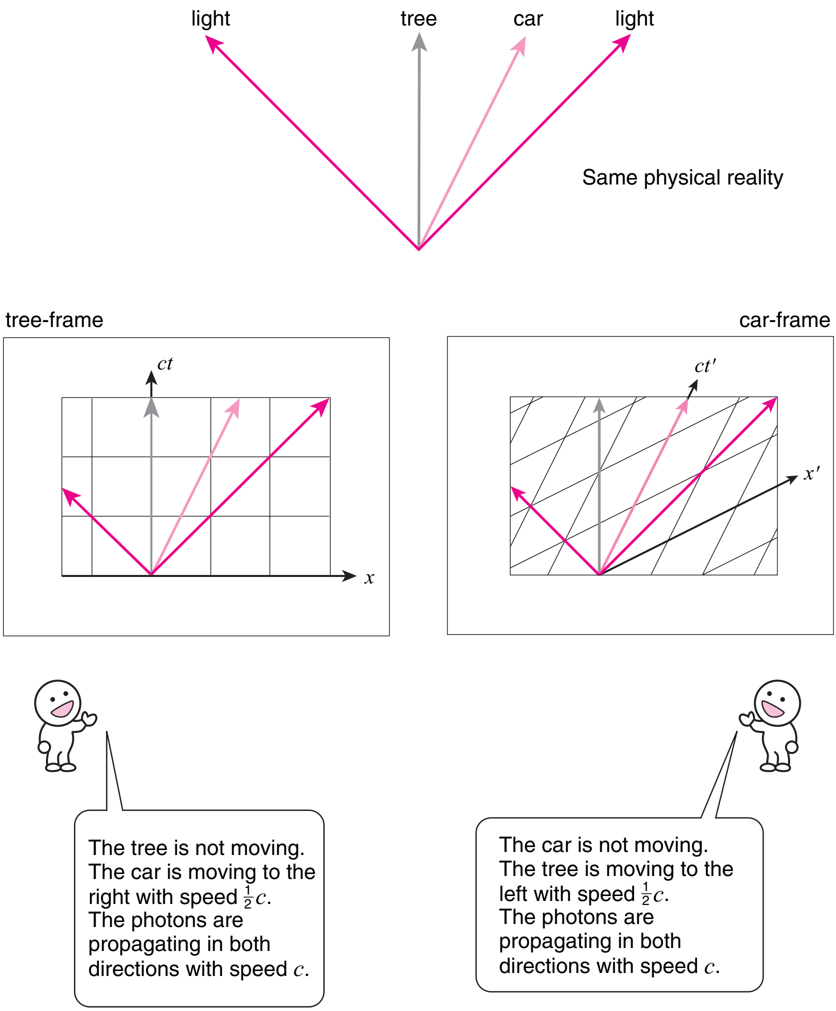
The reverse transformation is shown here. To go from the tree-frame to the car-frame, the spacetime diagram must be stretched in the northwest and southeast directions, and squeezed from the northeast and southwest directions by the same factor, until the worldline of the car becomes vertical while maintaining the worldlines of both photons at an angle of 45 degrees from the horizontal. Again, the amount of squeezing must be the reciprocal of the amount of stretching to keep the spacetime areas invariant.


ct
car
x
will become
the x′-axis
ct
ct′
x′
x

As in the Galilei transformation case, instead of skewing the spacetime diagram, we can use the same diagram but simply superimpose different grids to represent the observation from different frames.



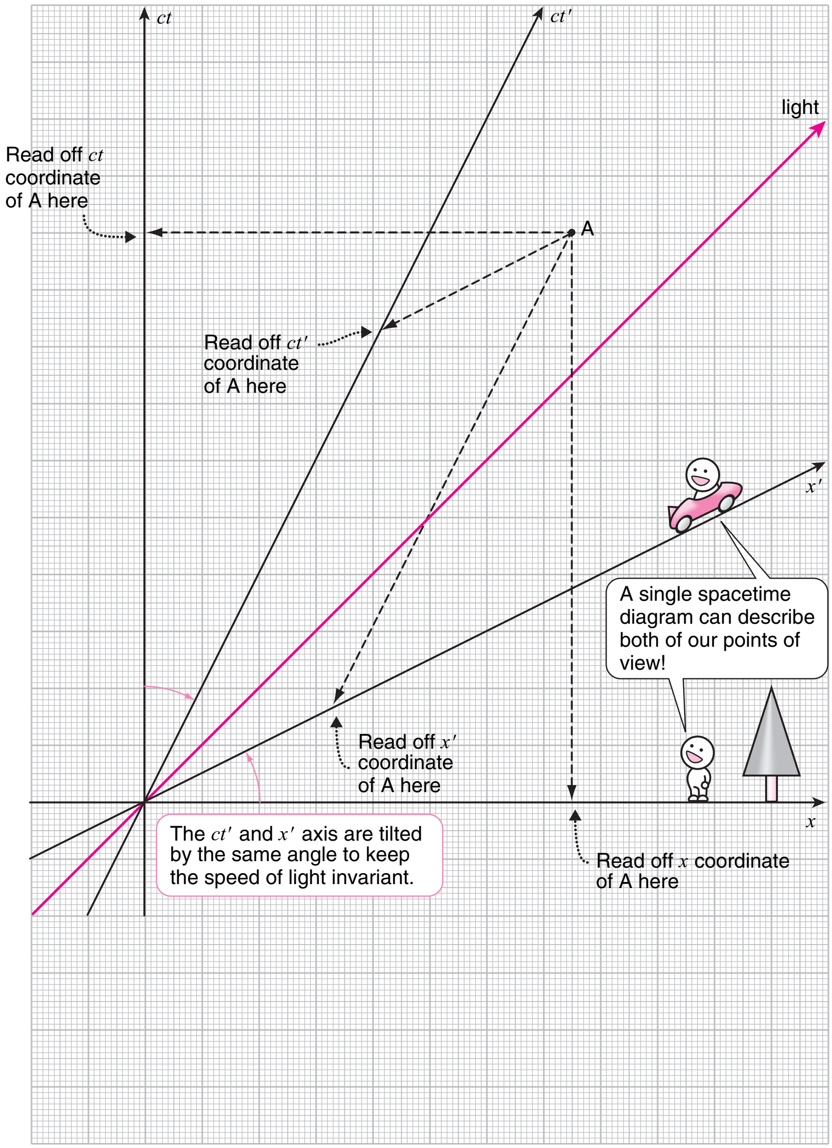
Therefore, we can consider the spacetime diagram without any grid on it as representing the objective physical reality, and superimposing a particular grid as the choice of frame.



Once we get used to the idea, we do not have to superimpose a complete grid. All we have to do is to represent the observation from different frames with different space- and time-axes.

Recall that, in the Galilei transformation case, all frames shared the same space-axis. This was because lines parallel to the space-axis connected points *at the same time*, and these were assumed to be the same for all frames. On the other hand, the time-axis for each frame was given by the worldline of the origin, where the clock was, in each frame.

In the Lorentz transformation case, the space-axis also becomes frame dependent since "*at the same time*" becomes a frame-dependent concept. The angle that the space-axis makes with the horizontal must be the same as the angle that the time-axis makes with the vertical. This ensures that photons will travel at speed 1 (since we are using ct as the time) in that frame.



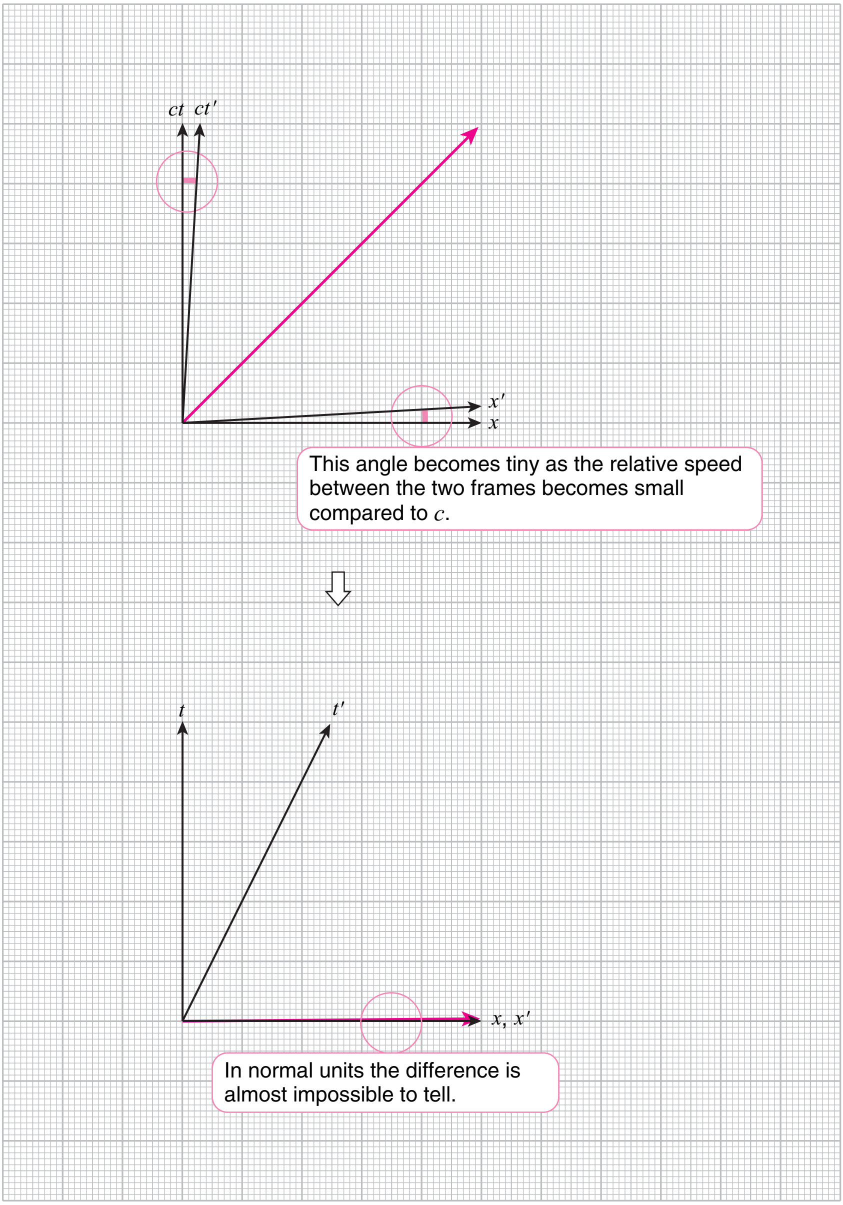
ct
ct′
light
Read off ct coordinate of A here
A
Read off ct′ coordinate of A here
x′
A single spacetime diagram can describe both of our points of view!
Read off x′ coordinate of A here
x
The ct′ and x′ axis are tilted by the same angle to keep the speed of light invariant.
Read off x coordinate of A here

4.9 The low velocity limit of the Lorentz transformation

If the Lorentz transformation is the correct transformation that relates observations in different inertial frames, and the Galilei transformation is wrong, why is it that the Galilei–Newton theory worked so well until people encountered the mystery concerning the speed of light?

Let's think what happens to the Lorentz transformation when the relative speed between the two frames is small compared to the speed of light. On the tree-frame spacetime diagram, the car-frame is represented by tilted time- and space-axes. The angle between the car-frame time-axis and the tree-frame time-axis (the vertical) depends on the relative speed between the tree and the car. In the example we have been considering, the car is moving at half the speed of light so this angle is rather large. But if the car is moving at a much smaller speed, like what we would encounter in everyday life, this angle will be very small. And since the angle between the car-frame space-axis and the tree-frame space-axis (the horizontal) must be the same as the angle between the two time-axes, that angle will also be very small. Furthermore, if we revert back to using seconds to measure time instead of meters, then the spacetime diagram will be crushed in the vertical direction so that the angle between the two space-axes is further diminished while the angle between the two time-axes is enhanced. And we find that, to a very good approximation, the Lorentz transformation is just the Galilei transformation at these low speeds!

So the reason why the Galilei transformation worked (and still does in most situations) is because the speeds involved were so much smaller than the speed of light c. It is only when the speeds of objects approach the speed of light that one has to use the Lorentz transformation to obtain the correct answer.


ct ct′
x′
x
This angle becomes tiny as the relative speed between the two frames becomes small compared to c.
t
t′
x, x′
In normal units the difference is almost impossible to tell.

4.10 Addition of velocities

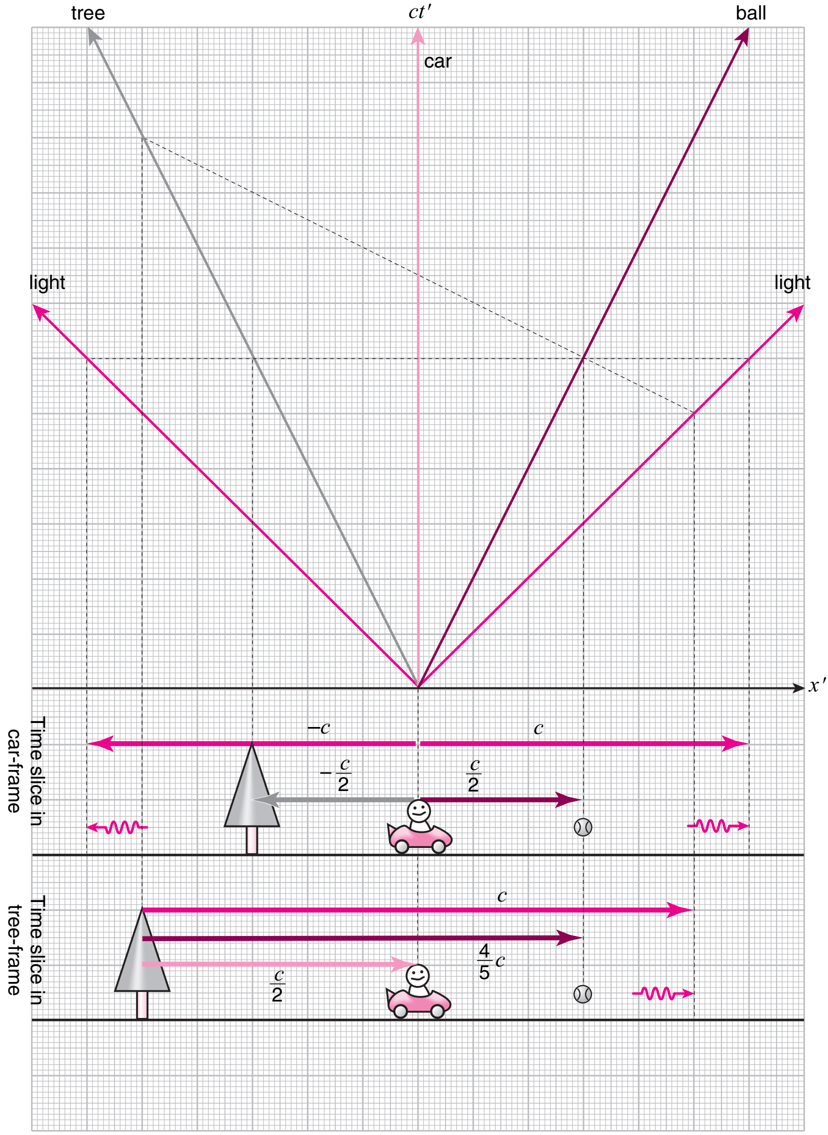
The Lorentz transformation tells us how to relate the observations in one frame to those in another. So let's consider the following problem: a ball is traveling at velocity $\frac{1}{2}c$ in the car-frame. The car, on the other hand, is traveling at velocity $\frac{1}{2}c$ in the tree-frame. What is the velocity of the ball in the tree-frame?

The Galilei transformation would tell us that the velocity of the ball is c in the tree-frame, but, as we have seen, the Galilei transformation is wrong. In particular, the velocity of a photon, which is faster than the ball in the car-frame, is c in both frames. So the velocity of the ball in the tree-frame has to be slower than that.

The spacetime diagram on the opposite page shows the motions of the tree, the car, the ball, and the two photons in the car-frame. Slicing the diagram along a horizontal line shows the relative positions of the objects at a particular instant in the car-frame. As you can see, the car is at rest in the middle, the two photons are equidistant from the car, the ball is midway between the car and the photon on the right, and the tree is midway between the car and the photon on the left. This tells us that the speeds of the tree and the ball in the car-frame are both $\frac{1}{2}c$, since they only cover half the distance the photons cover in the same amount of time.[9]

A time slice in the tree-frame would be tilted on this spacetime diagram as shown. The angle of tilt from the horizontal must be the same as the angle of the worldline of the tree from the vertical. The relative positions of the tree, the car, the ball, and the photon on the right at this instant in the tree-frame are shown at the bottom of the figure. As you can see, the tree is at rest on the left, the car is midway between the tree and the photon, while the ball is four-fifths of the way between the tree and the photon. This shows that the velocity of the car in the tree-frame is $\frac{1}{2}c$, as expected, while the velocity of the ball in the tree-frame is $\frac{4}{5}c$.

What we have found is that when we have two frames, 1 and 2 (the car- and tree-frames), that are moving relative to each other, the velocity of an object in frame 1 (the velocity of the ball in the car-frame) and the velocity of frame 1 as seen from frame 2 (the velocity of the car in the tree-frame) do not simply add up to give you the velocity of the object in frame 2 (the velocity of the ball in the tree-frame). The actual velocity is smaller than the simple sum.[10]

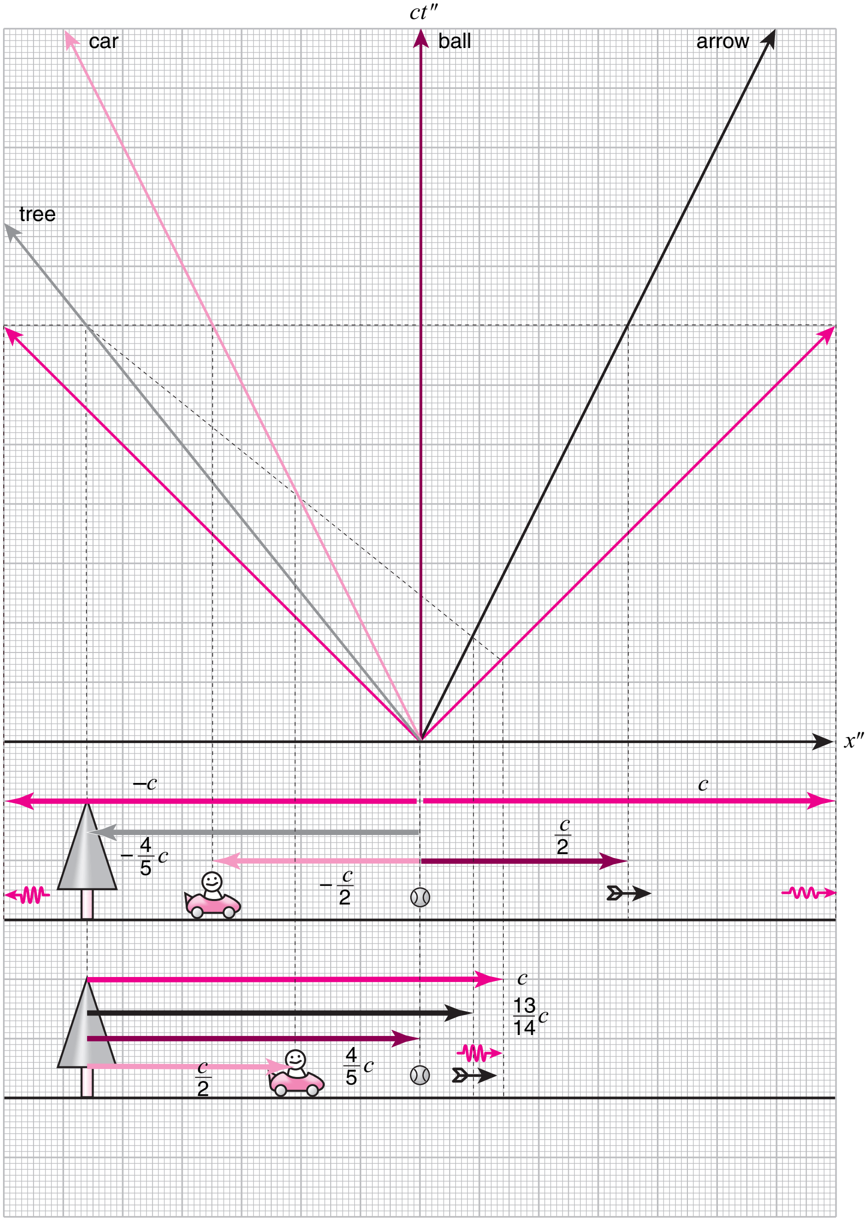

tree
ct′
ball
car
light
light
x′
Time slice in car-frame
−c
c
−c/2
c/2
Time slice in tree-frame
c
4/5 c
c/2

Now let's say that, in addition to the tree, the car, the ball, and the two photons, there is an arrow traveling at velocity $\frac{1}{2}c$ in the ball-frame. We can tell immediately from the result of the previous page that the velocity of the arrow in the car-frame is $\frac{4}{5}c$. But what is the velocity of the arrow in the tree-frame?

The spacetime diagram in the ball-frame is shown here. A time-slice along a horizontal line tells us the relative positions, and thus the velocities, of the objects in the ball-frame. The ball is at rest in the middle, the car is moving to the left with speed $\frac{1}{2}c$, the tree is moving to the left with speed $\frac{4}{5}c$ (since the ball is moving to the right at speed $\frac{4}{5}c$ in the tree-frame), the two photons are moving in their respective directions with speed c (as always), and the arrow is moving to the right with speed $\frac{1}{2}c$.

A time-slice in the tree-frame is along the tilted line as shown. The angle of tilt from the horizontal must be the same as the angle of the worldline of the tree from the vertical. The relative positions of the objects at that instant in the tree-frame are shown on the bottom. The tree is at rest on the left, the car is half way between the tree and the photon, the ball is four-fifths of the way between the tree and the photon, and the arrow is $\frac{13}{14}$ of the way between the tree and the photon.[11] This tells us that the velocities of the car, the ball, and the arrow in the tree-frame are $\frac{1}{2}c$, $\frac{4}{5}c$, and $\frac{13}{14}c$, respectively.[12]

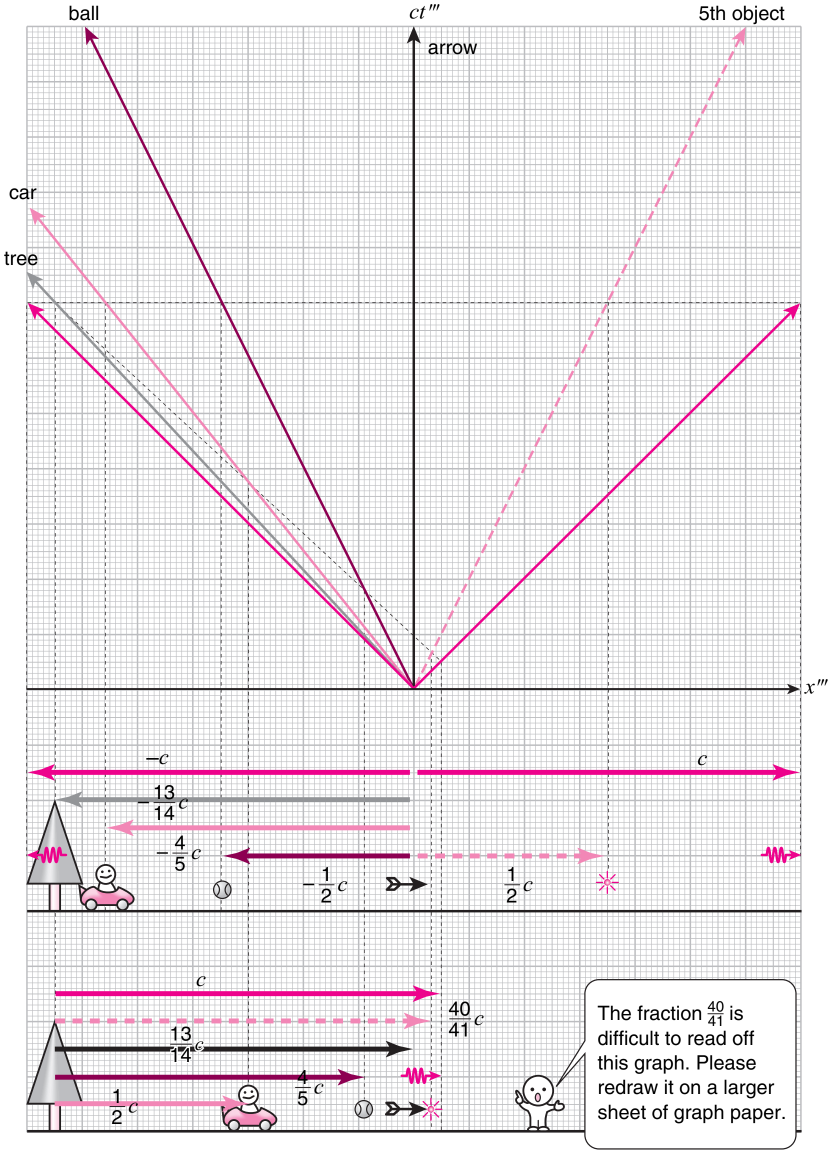
Notice that, in the tree-frame, the difference in the velocities of the arrow and the ball is smaller than the difference in the velocities of the ball and the car (this is clear from the bottom figure), even though the relative speeds of the objects when observed in a frame in which one of the objects is at rest is $\frac{1}{2}c$. The ball is faster than the car by $\frac{1}{2}c$ in the car-frame, but it is only faster than the car by $\frac{4}{5}c - \frac{1}{2}c = \frac{3}{10}c$ in the tree-frame. The arrow is faster than the ball by $\frac{1}{2}c$ in the ball-frame, but it is only faster than the ball by $\frac{13}{14}c - \frac{4}{5}c = \frac{9}{70}c$ in the tree-frame.



ct″
car
ball
arrow
tree
x″
−c
c
−4/5 c
c/2
−c/2
c
13/14 c
4/5 c
c/2

This pattern will continue if we consider another object moving at velocity $\frac{1}{2}c$ in the arrow-frame. I will only show the spacetime diagram on the opposite page and not explain the details, but the velocity of such an object in the tree-frame will be $\frac{40}{41}c$, which is faster than the arrow by $\frac{40}{41}c - \frac{13}{14}c = \frac{27}{574}c$.[13]

What this discussion shows is that, in contrast to the Galilei–Newton theory, you cannot accelerate objects to arbitrarily large speeds. As the speeds of objects approach c, it becomes more and more difficult to accelerate them. You can keep on increasing the speed of an object by $\frac{1}{2}c$ increments relative to the frame the object is in prior to each acceleration, but, when observed from the initial frame, the speed will increase at ever-diminishing increments as you approach the speed of light. And even though the speed will keep on getting closer and closer to c, it will never reach it or exceed it. Of course, this is to be expected: since the speed of light is the same to all observers, any speed that is smaller than the speed of light in one frame must be smaller than the speed of light in all frames.



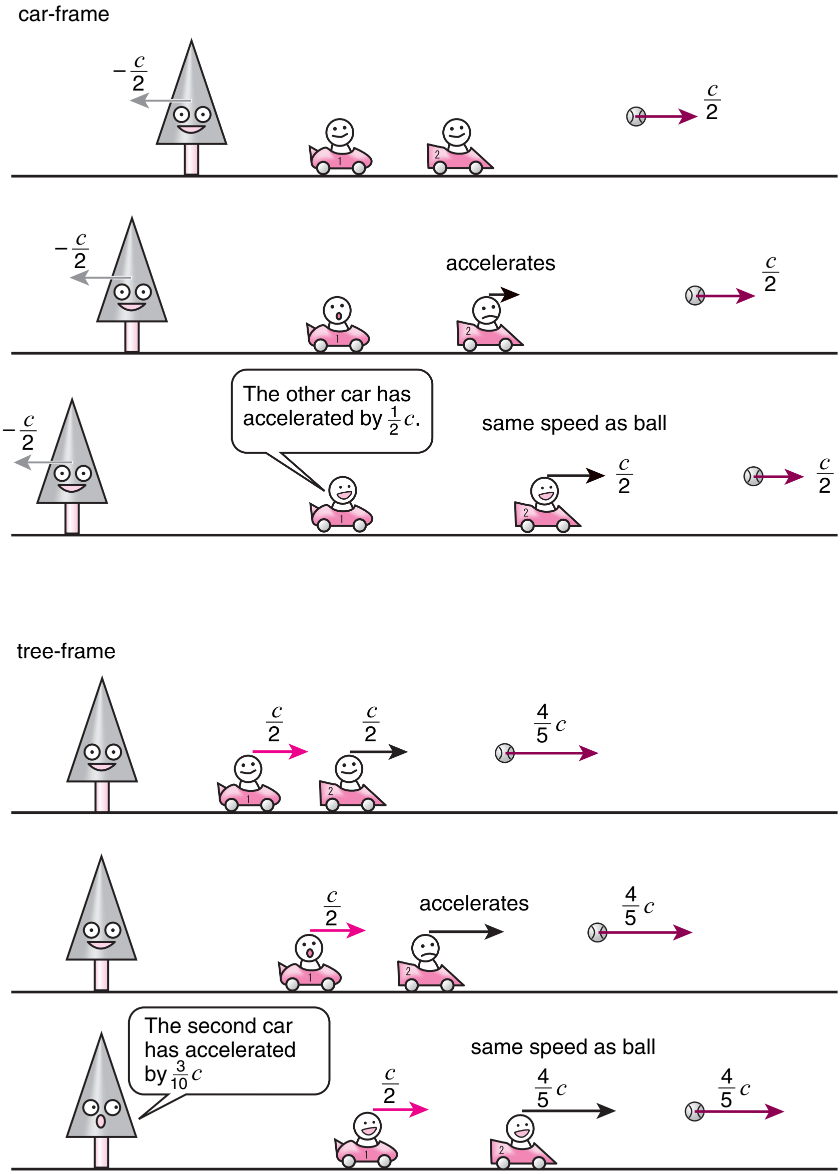
ball
ct'''
5th object
arrow
car
tree
x'''
$-c$
c
$-\frac{13}{14}c$
$-\frac{4}{5}c$
$-\frac{1}{2}c$
$\frac{1}{2}c$
c
$\frac{40}{41}c$
$\frac{13}{14}c$
$\frac{4}{5}c$
$\frac{1}{2}c$
The fraction $\frac{40}{41}$ is difficult to read off this graph. Please redraw it on a larger sheet of graph paper.

4.11 Dependence of inertia on speed

The result of the previous section implies that Newton's Second Law has to be modified since different observers will not agree by how much the velocity of an object has changed. If an object at rest in the car-frame accelerates to the same velocity as the ball, the change in its velocity in the car-frame is $\frac{1}{2}c$, but in the tree-frame it is $\frac{3}{10}c$. If an object at rest in the ball-frame accelerates to the same velocity as the arrow, the change in its velocity in the ball-frame is $\frac{1}{2}c$, but in the car-frame it is $\frac{3}{10}c$, and in the tree-frame it is $\frac{9}{70}c$.

Now, presumably, the same amount of effort on the part of the object is necessary for it to accelerate from rest to $\frac{1}{2}c$ in the car-frame as well as in the ball-frame. But when observed from the tree-frame, the amount of acceleration the same effort accomplishes is smaller when accelerating from the ball-frame to the arrow-frame than when accelerating from the car-frame to the ball-frame. This can be interpreted to mean that the object has more inertia in the ball-frame than in the car-frame when observed from the tree-frame.

So according to the Lorentz–Einstein theory, the inertia of an object becomes a frame-dependent concept. Faster moving objects are observed to have more inertia than when they are moving slower. And as the object's speed approaches c, its inertia becomes infinitely large.


car-frame
$-\frac{c}{2}$
$\frac{c}{2}$
accelerates
The other car has accelerated by $\frac{1}{2}c$.
same speed as ball
tree-frame
$\frac{4}{5}c$
The second car has accelerated by $\frac{3}{10}c$
same speed as ball

Notes

1 By "not work" we mean that it does not agree with experimental observation.

2 In this book, when we say "speed of light" we will always mean its speed in a vacuum. Light is slower when traveling through transparent substances such as water or glass because the photons bump into the atoms in the way, which slows them down. So, in those cases, the speed of light is not equal to "the speed of light" c.

3 This nine-digit number is actually *exact* since a *meter* is defined so that the speed of light ends up being this number.

4 Galileo's experiment is described in his book *Dialogues Concerning Two Sciences*, which we mentioned in endnote 4 in Chapter 1. An English translation can be found in [4]. See the dialogue for the first day.

5 In an equation, what Einstein is saying can be expressed as

$$\text{time it takes} = \frac{\text{distance}}{\text{speed}}.$$

I'm sure everyone knows this and uses it all the time. For instance, Virginia Tech is located in Blacksburg, which is about 300 miles away from Washington DC. If you drive at a speed of 60 miles per hour, you will reach Washington DC in 5 hours. The calculation you just did in your head is expressed by this equation.

6 When we have more than one space dimension, we will have the conservation of spacetime *volume*. In mathematical terms, what this means is that the determinant of the transformation matrix which relates the spacetime coordinates of the two frames must be 1.

7 Named after Hendrik Antoon Lorentz (1853–1928). The Lorentz transformation equations, shown below, were known well before Einstein's 1905 paper.

8 In equations, the Lorentz transformation can be written down as follows. Let (ct, x) be the time and space coordinates in the tree-frame, and (ct', x') be the time and space coordinates in the car-frame. Then (ct, x) and (ct', x') are related as

$$ct' = \frac{ct - \beta x}{\sqrt{1-\beta^2}}, \qquad x' = \frac{x - \beta ct}{\sqrt{1-\beta^2}}, \qquad \beta \equiv \frac{u}{c},$$

where u is the speed of the car as observed in the tree-frame. In matrix notation, this can be written as

$$\begin{bmatrix} ct' \\ x' \end{bmatrix} = \gamma \begin{bmatrix} 1 & -\beta \\ -\beta & 1 \end{bmatrix} \begin{bmatrix} ct \\ x \end{bmatrix}, \qquad \gamma = \frac{1}{\sqrt{1-\beta^2}}.$$

We can check that the determinant of this transformation matrix is indeed 1 as mentioned in endnote 6.

9 The velocity of the ball in the car-frame is $+\frac{1}{2}c$, since it is traveling toward the right, while the velocity of the tree in the car-frame is $-\frac{1}{2}c$, since it is traveling toward the left.

10 If the velocity of an object in frame 1 is v_1, and the velocity of frame 1 as seen from frame 2 is u_{12}, then the velocity of the object in frame 2 is

given by

$$v_2 = \frac{v_1 + u_{12}}{1 + \dfrac{v_1 u_{12}}{c^2}} \; .$$

When $v_1 = \frac{1}{2}c$, and $u_{12} = \frac{1}{2}c$, it is easy to check that $v_2 = \frac{4}{5}c$, according to this formula. This agrees with the velocity we obtained by simply looking at the appropriate time-slice of the spacetime diagram.

11 The last fraction may be a bit difficult to see. You need to count the tiny gridlines in the background.

12 You can also check this result against the formula given in endnote 10.

13 If you are not good at fractions, comparing the sizes of fractions like $\frac{3}{10}$, $\frac{9}{70}$, and $\frac{27}{574}$ may seem difficult. In that case, just use a calculator!

5

Causality

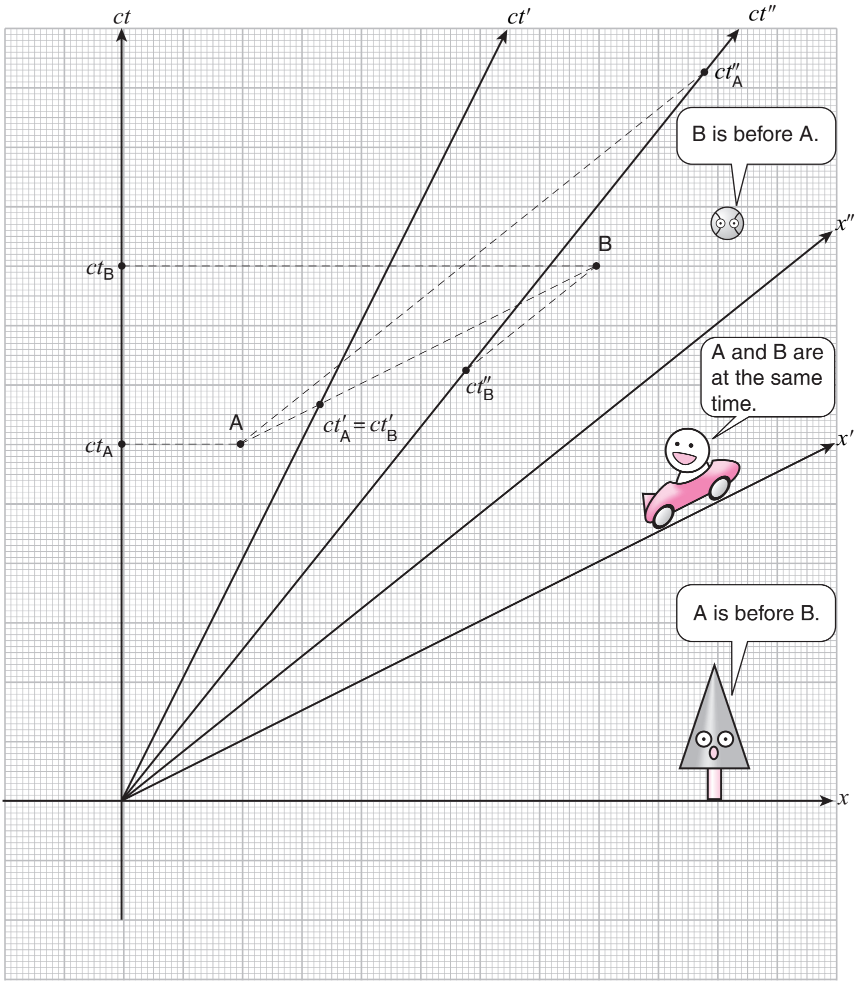
5.1 Before and after

As we have seen, the only way to reconcile the experimentally observed fact that the speed of light does not depend on the inertial frame with our belief in an objective reality, was to abandon the notion that "*at the same time*" meant the same thing for all observers. Two events that are "*at the same time*" in one frame may not be "*at the same time*" in another frame, and vice versa.

Now, some of you may have already realized that this could lead to a problem with the notion of *causality*, namely, the notion of *cause* and *effect*. If an event A is the *cause* of another event B, then A must happen *before* B, and B must happen *after* A. But according to Einstein, the chronological order in which two events happen can depend on the frame of reference!

To make our discussion concrete, consider again the tree planted firmly in the ground, the car moving to the right with speed $\frac{1}{2}c$ in the tree-frame, and the ball moving to the right with speed $\frac{4}{5}c$ in the tree-frame (which corresponds to $\frac{1}{2}c$ in the car-frame). The spacetime diagram on the right shows the space- and time-axes for all three frames: the tree-frame (x, ct), the car-frame (x', ct'), and the ball-frame (x'', ct'').

Now look at the points A and B on the spacetime diagram. I think it is clear that A *happens before* B in the tree-frame. But in the car-frame, the line that connects A and B is parallel to the x'-axis, which means that A and B happen at the same time. And in the ball-frame, if we draw lines that go through A and B that are parallel to the x''-axis, it is easy to see that the line that goes through B intersects the ct''-axis at a point which is earlier than the point at which the line that goes through A intersects the ct''-axis. This means that B happens before A in the ball-frame.



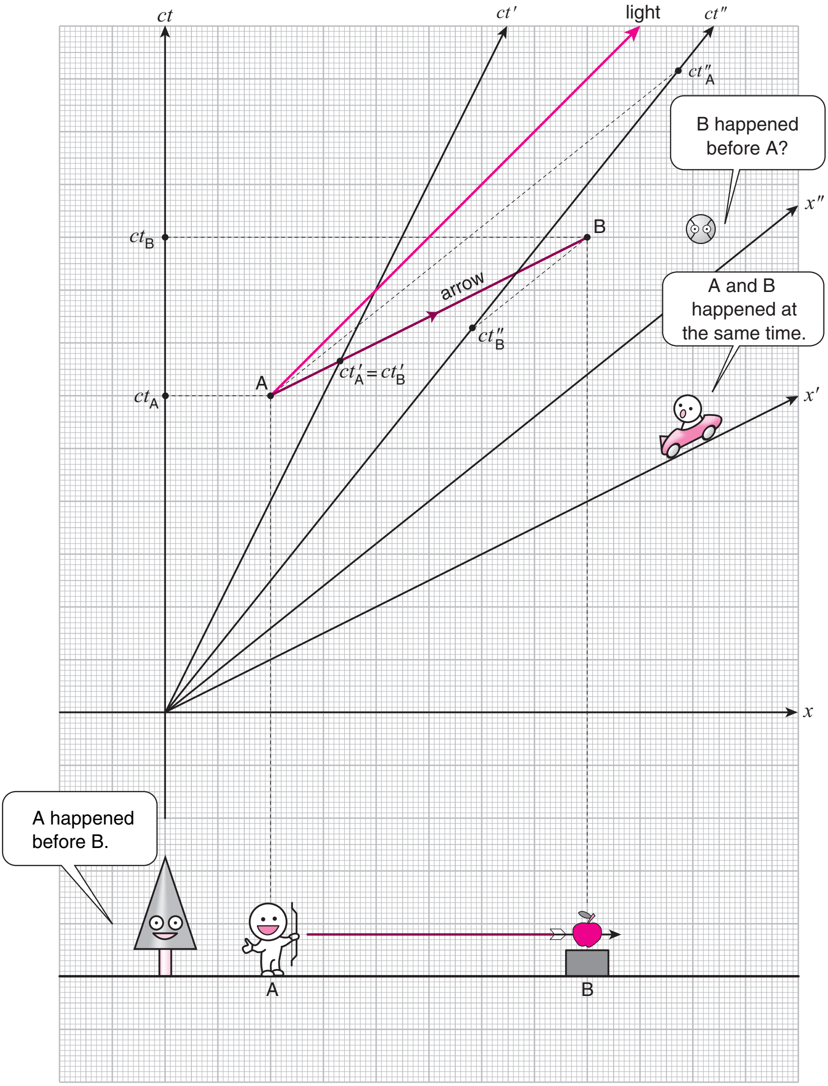
ct
ct′
ct″
ct″A
B is before A.
x″
B
ctB
A and B are at the same time.
ct″B
ct′A = ct′B
A
ctA
x′
A is before B.
x

5.2 Paradox?

Now imagine that you shot an arrow at A and it hit a target at B. In the tree-frame, A would be observed to happen before B, so there is no problem. But in the car-frame, A and B are at the same time so the arrow will have traversed the distance to the target instantaneously. And in the ball-frame, B happens before A, so the target was hit *before* the arrow was released! This does not make any sense at all, does it? The release of the arrow is clearly the *cause*, and the target being hit is clearly the *effect*. The *effect* cannot happen before the *cause*. Does this mean that the Lorentz transformation leads to a logical inconsistency? (Oh, no!)

It would if it were actually possible to shoot an arrow at A to hit a target at B. But if you look at the proposed worldline of the arrow carefully, you can see that it must travel faster than the speed of light to go from A to B. Just compare it to the worldline of a photon emanating from point A which is shown in the diagram. Now we have already discussed in section 4.10 that nothing traveling at speeds slower than c can be accelerated beyond c by incremental increases of its speed. So shooting an arrow from A to B is actually impossible. (Phew! That's a relief.)

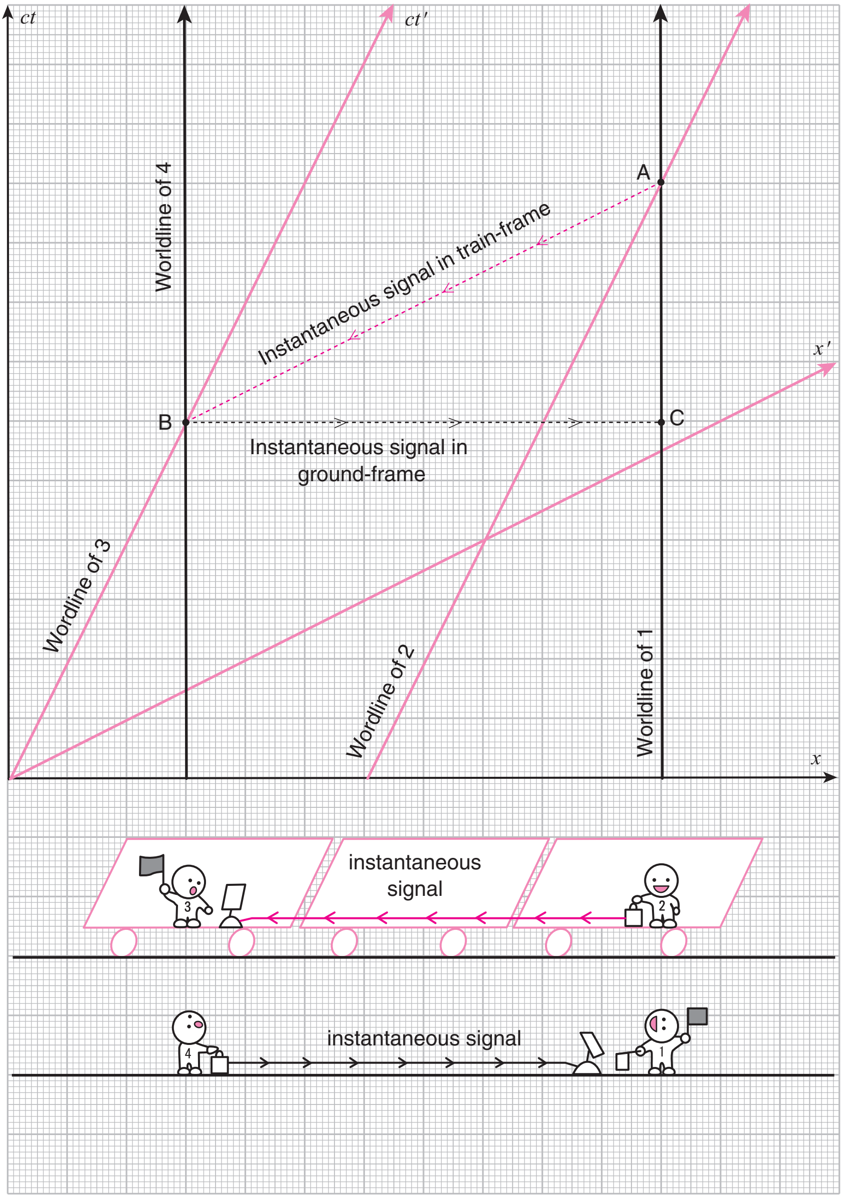
But would it be possible to shoot some kind of "beam," instead of an arrow, from A to B with a speed faster than c from the very beginning so that it doesn't have to be accelerated from slower speeds? Well, if it were, it would lead to the breakdown of causality, so it had better be impossible![1]


ct
ct′
light
ct″
ct″A
B happened before A?
x″
ctB
B
arrow
A and B happened at the same time.
ct″B
A
ct′A = ct′B
ctA
x′
x
A happened before B.
A
B

5.3 Instantaneous communication?

As we have seen, an object traveling at speeds faster than c can lead to the breakdown of causality. In fact, it is not only objects that must not travel faster than c but any kind of signal that conveys information. To see this, I am going to use the following example from the book *The Einstein Paradox* [7] by Colin Bruce.

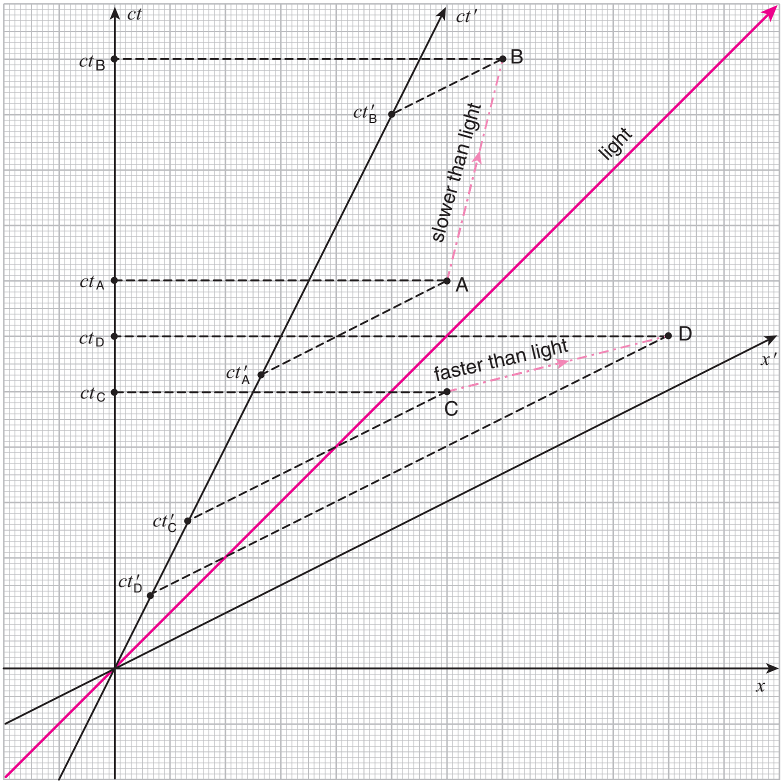
Assume that it is possible to communicate instantaneously with anyone at any distance.[2] Now imagine a long train traveling along straight horizontal tracks. Observer 1, who is standing along the railroad tracks, signals to observer 2, who is riding the locomotive of the train, as the locomotive passes point A on the spacetime diagram. Observer 2 immediately forwards the message received from observer 1 to observer 3, who is riding the caboose, using the instantaneous communication device. Since the signal is, presumably, *instantaneous* in the train-frame, it reaches observer 3 at point B on the spacetime diagram. Observer 3, upon receiving the message, immediately forwards it to observer 4, who happens to be standing just alongside the caboose at that particular instant. Observer 4 then immediately forwards the message to observer 1, again using the instantaneous communication device. Since the signal is, presumably, *instantaneous* in the ground-frame, it reaches observer 1 at point C. But C is before A! So observer 1 receives his own message before he even sent it. A paradox has occurred because we assumed that a signal could be sent at speeds that exceed c.


ct
ct′
x′
x
Worldline of 4
Worldline of 3
Worldline of 2
Worldline of 1
A
B
C
Instantaneous signal in train-frame
Instantaneous signal in ground-frame
instantaneous signal
instantaneous signal
3
2
4
1

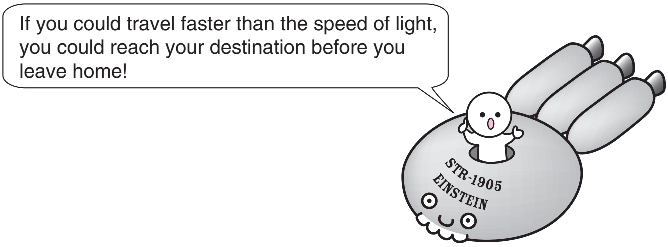
5.4 Impossibility of faster than light travel

As long as no object nor any signal can exceed the speed of light, causality does not break down. This is because the chronological order of two events on the spacetime diagram will depend on the frame only when the line connecting them has a tilt which is smaller than 45 degrees from the horizontal. In such a case, there always exists a frame in which the two events happen at the same time. And when observed from another frame moving to the left (right) relative to that frame, the event on the left happens before (after) the event on the right. And to connect those two events by a worldline, the object or signal must be traveling faster than the speed of light.

On the other hand, the worldline of an object or signal traveling at a speed slower than the speed of light will have a tilt larger than 45 degrees from the horizontal. Any two events on the spacetime diagram that can be connected by such a worldline have a fixed chronological order independent of the frame.



ct
ct'
ct_B
B
ct'_B
slower than light
light
ct_A
A
ct_D
D
x'
ct'_A
faster than light
ct_C
C
ct'_C
ct'_D
x



If you could travel faster than the speed of light, you could reach your destination before you leave home!
STR-1905
EINSTEIN

5.5 The light-cone

Take any point on the spacetime diagram, say A, and draw two lines at 45 degree angles that go through it. These represent the worldlines of photons that go through point A. The two lines separate the spacetime diagram into four regions. Any point in the bottom or top regions can be connected with point A with a worldline of an object or signal whose speed does not exceed c. On the other hand, this cannot be done for any of the points in the left or right regions. So the points in the bottom region are the points that are in the *past* of A. Only the events happening in this region can be the *cause* of anything that happens at A. The points in the top region are the points that are in the *future* of A. Only the events happening in this region can be the *effect* of anything that happens at A. The points in the left and right regions do not have a fixed chronological order with A. They are neither in the past nor in the future of A. They are the points causally disconnected from A.

The two lines that separate these regions is known as the *light-cone*. It may not look like a cone here, but if we consider two space dimensions, then the surface that separates the causally connected and disconnected regions will be a cone.[3] The top and bottom regions of the diagram are called the *inside* of the light-cone, and the left and right regions are the *outside* of the light-cone. Any point inside the light-cone of A is said to be *time-like separated* from A. Any point outside the light-cone of A is said to be *space-like separated* from A.[4]

The light-cone represents the causal structure of spacetime. It tells you that only a limited region of the spacetime diagram can be considered the future, or the past, of any event. And this is because the inside of the light-cone is the only region that all inertial observers agree as happening before, or after, the event in question.

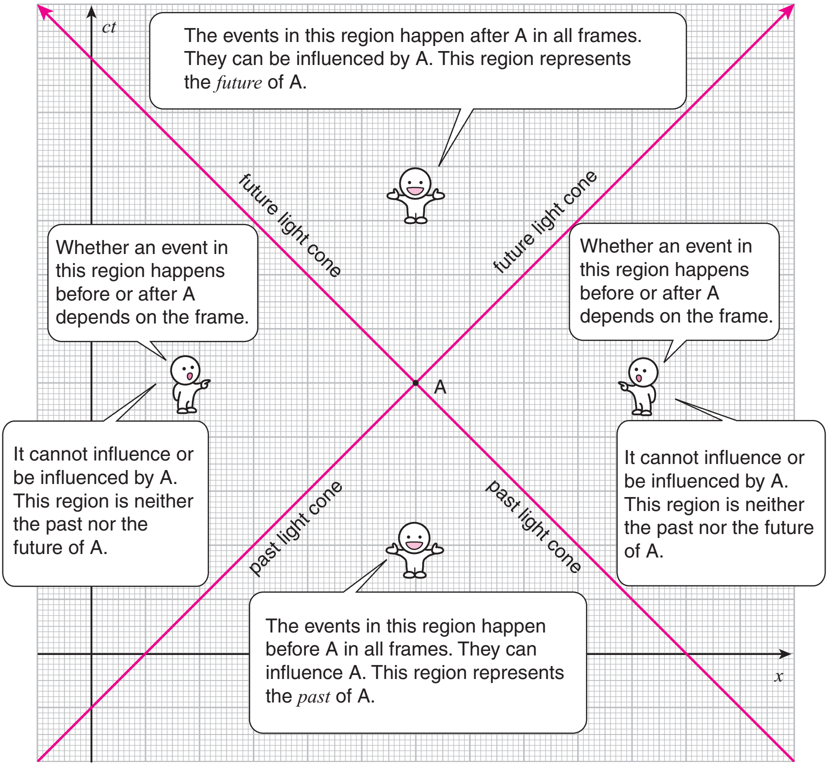
Notes

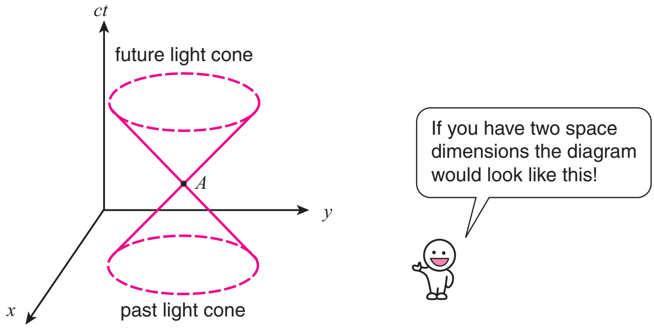
1 Particles that travel faster than the speed of light have been considered theoretically and are called *tachyons* (tachy- from the Greek $\tau\alpha\chi\upsilon\varsigma$, meaning *fast*, with the suffix -on for *particle*). If tachyons existed causality would break down. So any theory that predicts their existence is considered a bad theory.

2 This happens all the time in science fiction movies.

3 With three space dimensions, the light-cone is not exactly a cone in the usual sense of the word, but the terminology is still used due to the lack of a better word.

4 It's nerdy terminology so you don't have to use it, but it's useful to know.


ct
The events in this region happen after A in all frames. They can be influenced by A. This region represents the *future* of A.
future light cone
future light cone
Whether an event in this region happens before or after A depends on the frame.
Whether an event in this region happens before or after A depends on the frame.
A
It cannot influence or be influenced by A. This region is neither the past nor the future of A.
It cannot influence or be influenced by A. This region is neither the past nor the future of A.
past light cone
past light cone
The events in this region happen before A in all frames. They can influence A. This region represents the *past* of A.
x


ct
future light cone
A
y
x
past light cone
If you have two space dimensions the diagram would look like this!

6
Consequences

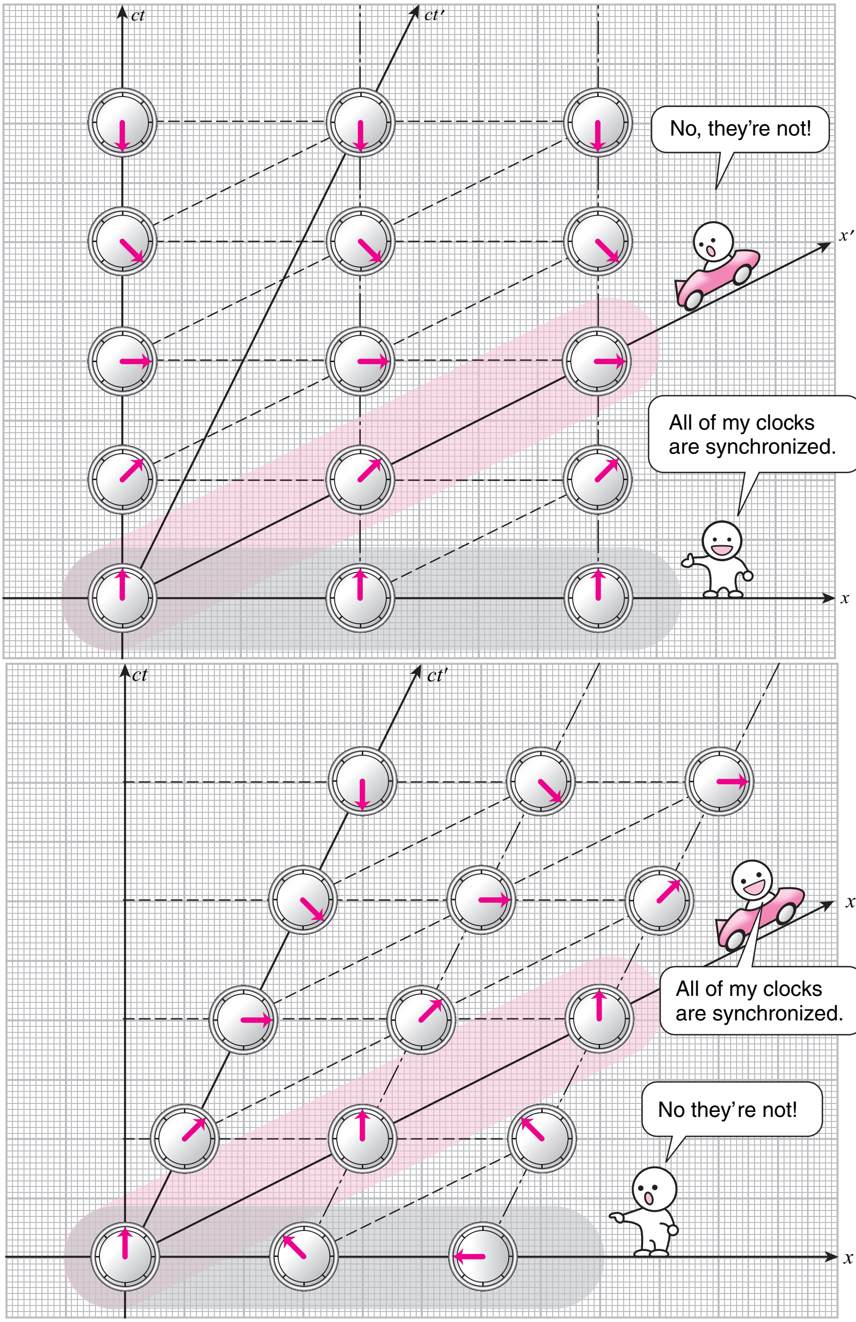
Let us now take a look at some of the consequences of the fact that the concept of simultaneity is relative.

6.1 Synchronization of clocks

In our discussion so far, we have assumed that there is a single clock at the origin which keeps track of the time in each frame. But of course, we can have multiple clocks at fixed distances from the origin, and have them synchronized so that they all give the exact same reading in that particular frame. For instance, in the top diagram shown on the opposite page, three clocks in the ground-frame have been synchronized. However, if the same clocks are observed from the car-frame which is moving relative to the ground, it is clear from the diagram that they are not synchronized at all.

Similarly, clocks that are synchronized in the car-frame are not synchronized in the ground-frame, as shown in the bottom diagram.

So when you have multiple clocks in a single frame, different observers in different frames will never agree on whether they are synchronized or not. This is a rather trivial consequence of the relativity of simultaneity, but something that may not be so obvious until you think about it.



ct
ct′
No, they're not!
x′
All of my clocks are synchronized.
x
ct
ct′
x′
All of my clocks are synchronized.
No they're not!
x

6.2 Time dilation

Now, let's say we wanted to compare the reading of a clock in the ground-frame with the reading of a clock in the car-frame to see if they are running at the same rate.

First, let's consider carefully how we should compare the running of two clocks. I think everyone would agree that the correct procedure is:

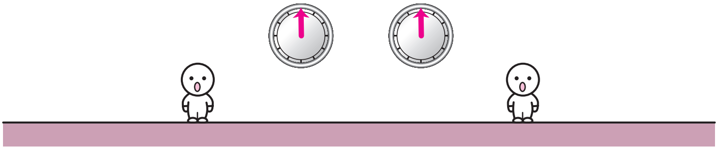
1. Synchronize the two clocks, that is, make sure that initially they give the same reading *at the same time*.
2. Wait a while and then compare the readings of the two clocks *at the same time*.

You normally would not even mention the *at the same time* requirement since it is so obvious. But as we have been discussing, *at the same time* for one observer is not *at the same time* for another.

Now, I think it is clear from the discussion in the previous section that the only situation in which observers in both frames would agree that two clocks are giving the same reading *at the same time* would be if the two clocks were *at the same place*.

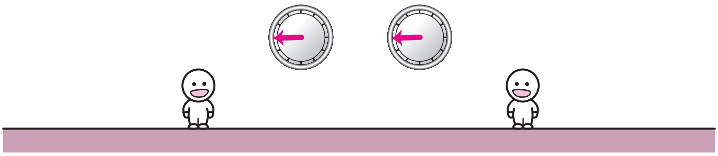
If the two clocks are in the same frame

Step 1 Synchronize the clocks so that they give the same reading "at the same time."



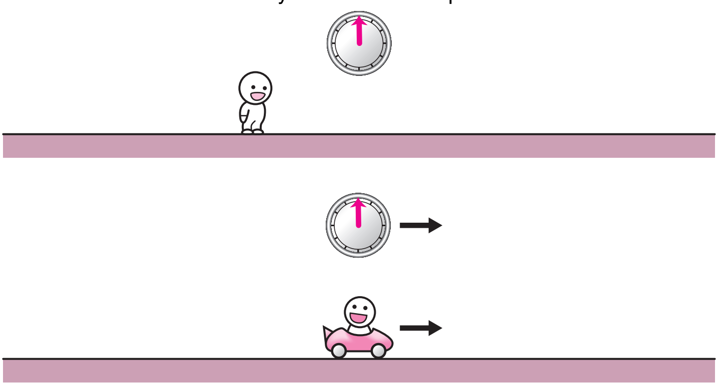
wait a while

Step 2 Compare the readings of the two clocks "at the same time."



If the two clocks are in different frames

Step 1 Synchronize the clocks as they pass each other so that they are at the same place.



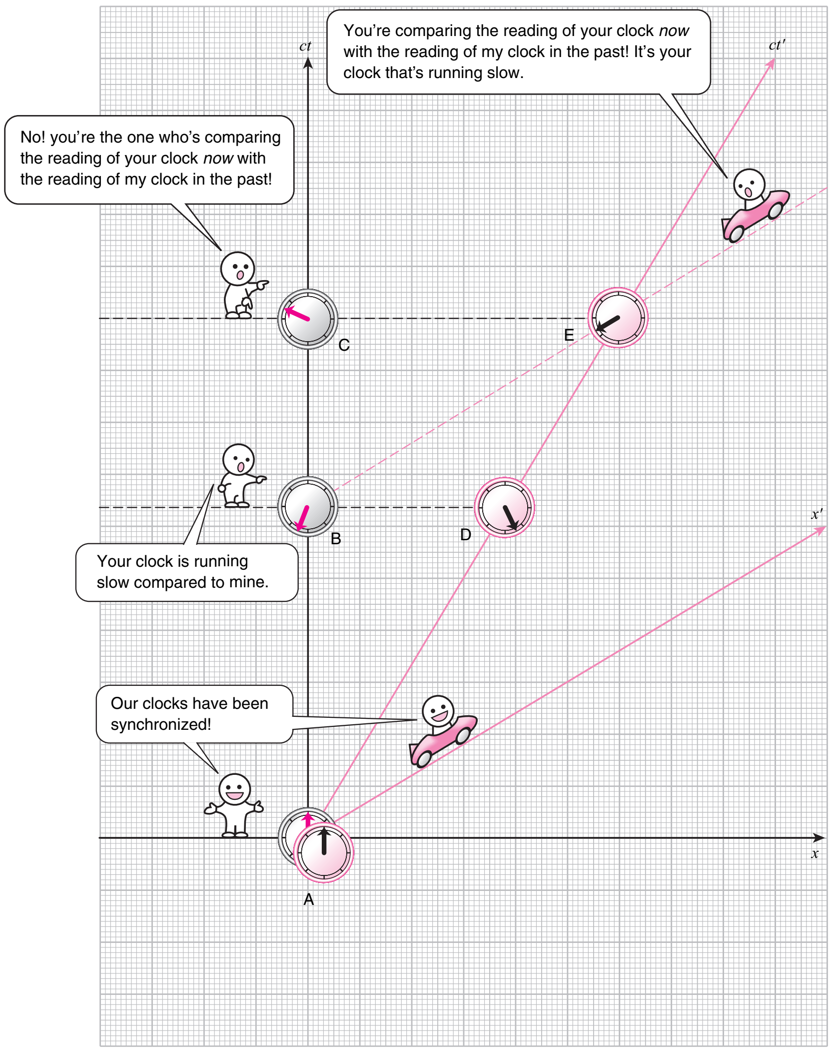
So let's assume that two observers in the ground- and car-frames synchronized their clocks at the origin, labeled A, of the spacetime diagram as shown. Both clocks are at the same place at the same time in both frames so if the clocks give the same reading here, step 1 on page 132 is OK.

Step 2 is the problem. Since both observers are in inertial frames, the clocks will keep on moving away from each other after A. So when the ground-frame observer decides to compare the reading of the ground-frame clock with the reading of the car-frame clock some time later, the two clocks will be some distance apart, and if he compares the readings of the two clocks *at the same time* in the ground-frame, it will NOT be *at the same time* in the car-frame. Similarly, when the car-frame observer compares the reading of the car-frame clock with the reading of the ground-frame clock, he will compare the readings of the two clocks *at the same time* in the car-frame, which is NOT *at the same time* in the ground-frame.

For instance, if the ground-frame observer wants to compare the reading of the ground-frame clock at B with the car-frame clock, he will naturally compare it with the reading of the car-frame clock at D, which is *at the same time* as B in the ground-frame. But B is *at the same time* as E in the car-frame, which is in the future of D. So when the ground-frame observer compares the readings of the clocks at B and D, the car-frame observer at E will think that the other observer is comparing the present reading of the ground-frame clock with the past reading of the car-frame clock.

The car-frame observer will naturally compare the reading of the ground-frame clock at B with the reading of the car-frame clock at E since they are simultaneous in the car-frame. But in the ground-frame, E is in the future of B, and it is actually C that is simultaneous with E. So when the car-frame observer compares the readings of the clocks at B and E, the ground-frame observer at C will think that the car-frame observer is comparing the present reading of the car-frame clock with the past reading of the ground-frame clock.

What this discussion shows is that both observers will think that the other observer is comparing the "present" reading of the other clock with the "past" reading of their own clock. But what is the "present" reading of the other clock compared to the "present" reading of your clock?

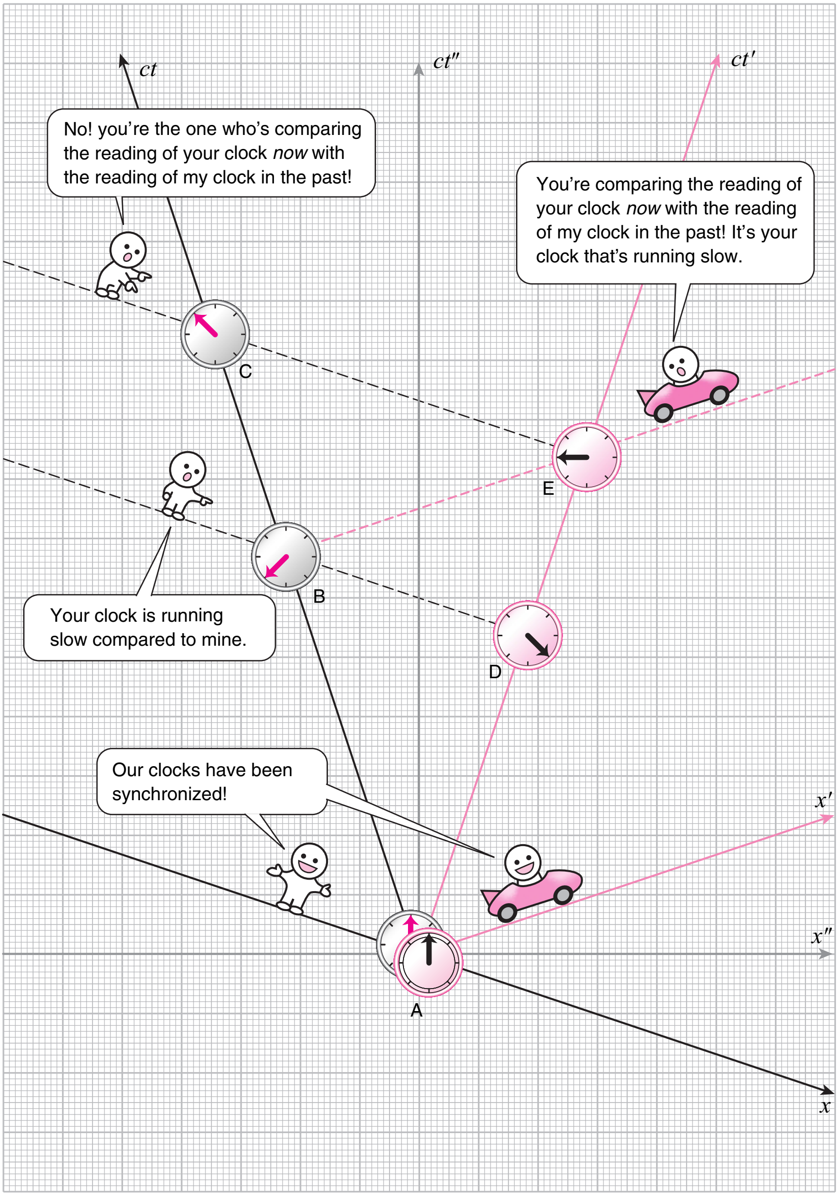

You're comparing the reading of your clock *now* with the reading of my clock in the past! It's your clock that's running slow.
No! you're the one who's comparing the reading of your clock *now* with the reading of my clock in the past!
Your clock is running slow compared to mine.
Our clocks have been synchronized!
ct
ct′
x
x′
A
B
C
D
E

The little characters in the figure on the previous page are both claiming that the other clock is running slow. The easiest way to see this is to Lorentz transform to a frame in which the ground- and car-frames are moving in opposite directions at the same speed. This lets you see the two frames in a completely symmetric fashion. The result of such a transformation is shown here.

Now, the ground-frame observer compares the reading of the ground-frame clock at B with the reading of the car-frame clock at D. Since the two clocks were synchronized at A, this is the same thing as comparing the lengths of AB and AD, and clearly AD is shorter than AB. So the reading of the car-clock at D will be earlier than the reading of the ground-clock at B, and the ground-frame observer at B will conclude that the car-frame clock is running slow compared to the ground-frame clock.

The car-frame observer compares the reading of the car-frame clock at E with the reading of the ground-frame clock at B. AB is shorter than AE. So the reading of the ground-clock at B will be earlier than the reading of the car-clock at E, and the car-frame observer at E will conclude that the ground-frame clock is running slow compared to the car-frame clock.

So what we have discovered is that moving clocks are always observed to be running slower. This effect is known as *time dilation.*



ct
ct''
ct'
No! you're the one who's comparing the reading of your clock *now* with the reading of my clock in the past!
You're comparing the reading of your clock *now* with the reading of my clock in the past! It's your clock that's running slow.
C
E
Your clock is running slow compared to mine.
B
D
Our clocks have been synchronized!
x'
x''
A
x

6.3 What time dilation DOES NOT mean

Now, *time dilation* is often misunderstood to mean that time itself is flowing at a slower rate in a moving frame compared to a frame at rest. And consequently, time intervals measured by the moving clock will always be shorter than those measured by the stationary clock.

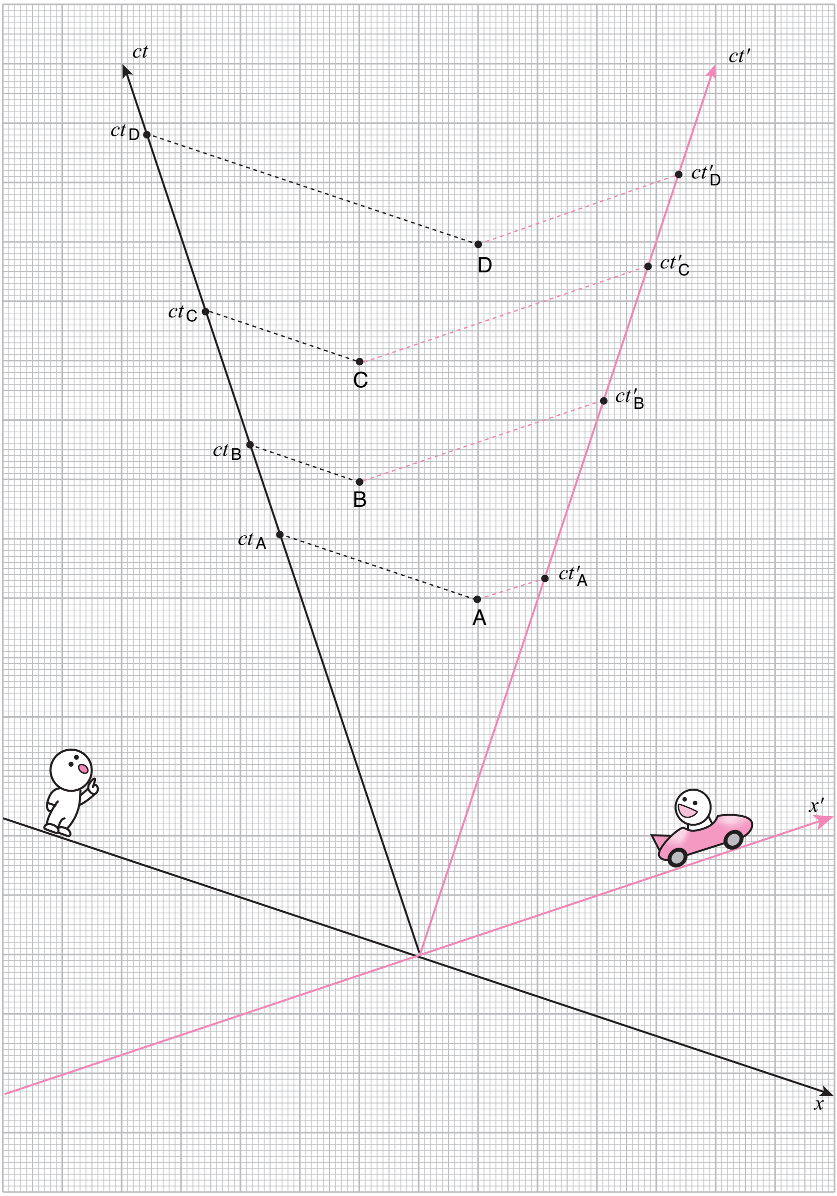
This is not true at all.

First, the relationship between the stationary and moving frames is completely symmetrical. After all, which frame we consider to be moving is a matter of choice. So if time is flowing at a slower rate in the moving frame than the stationary frame, then time has to be flowing at a slower rate in the stationary frame than the moving frame. This is a clear contradiction.

Second, as we have seen, when the two observers are comparing the readings of their clocks, they are not measuring the chronological separation of the same events at all. They are always measuring events along the worldline of the other clock.

If the two observers actually measure the time interval between the same events on the spacetime diagram, then there is no rule that says that one observer will obtain a shorter time than the other. For instance, take the four points A, B, C, and D shown here. The chronological separation between A and B is measured to be shorter in the ground-frame than in the car-frame. However, the time interval between B and C are measured to be the same in both frames, while the time interval between C and D is longer in the ground-frame than in the car-frame.

Time dilation is a phenomenon caused by the relativity of simultaneity, and has nothing to do with how fast time is flowing.



ct
ct'
ct_D
ct'_D
D
ct'_C
ct_C
C
ct'_B
ct_B
B
ct_A
ct'_A
A
x'
x

6.4 Lorentz contraction

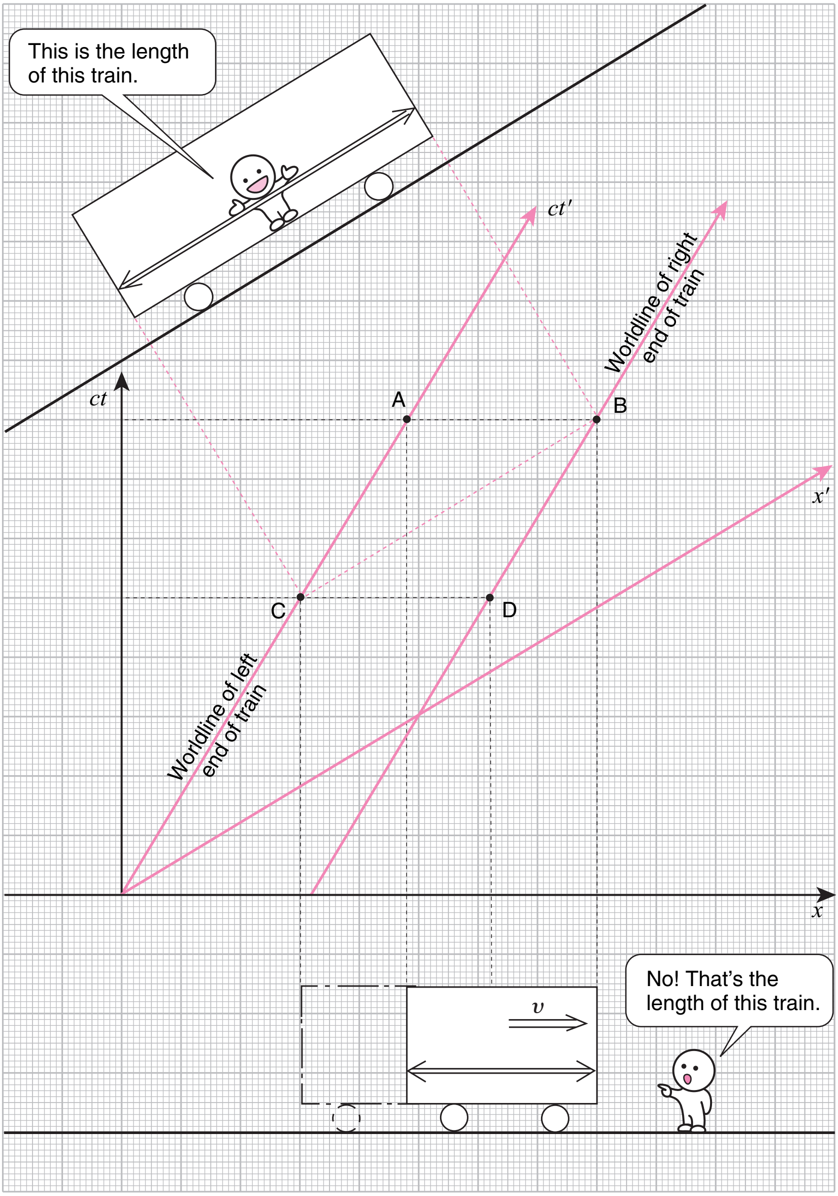
Next, let's think about the measurement of length.

First, ponder for a moment what we mean by "length." We mean the distance separating the two ends of an object *at the same time*, don't we? Say we wanted to measure the length of a moving train. We wouldn't measure the position of the front end of the train at one time, and the position of the back end of the train at another time, and then take the difference, would we? The train would have moved in between the two times and the answer we get would not be what we think of as its length. But we have been discussing that the concept of being *at the same time* depends on the observer!

Consider the spacetime diagram shown here. The observer riding the train would record the position of the front end of the train and the position of the back end of the train *at the same time* in the train-frame and obtain the length of the train by taking the difference. This would be the distance between points B and C. But the observer on the ground would record the position of the front end of the train and the position of the back end of the train *at the same time* in the ground-frame and obtain what he thinks is the length of the train. This would be the distance between points A and B. Clearly AB must be shorter than CB.

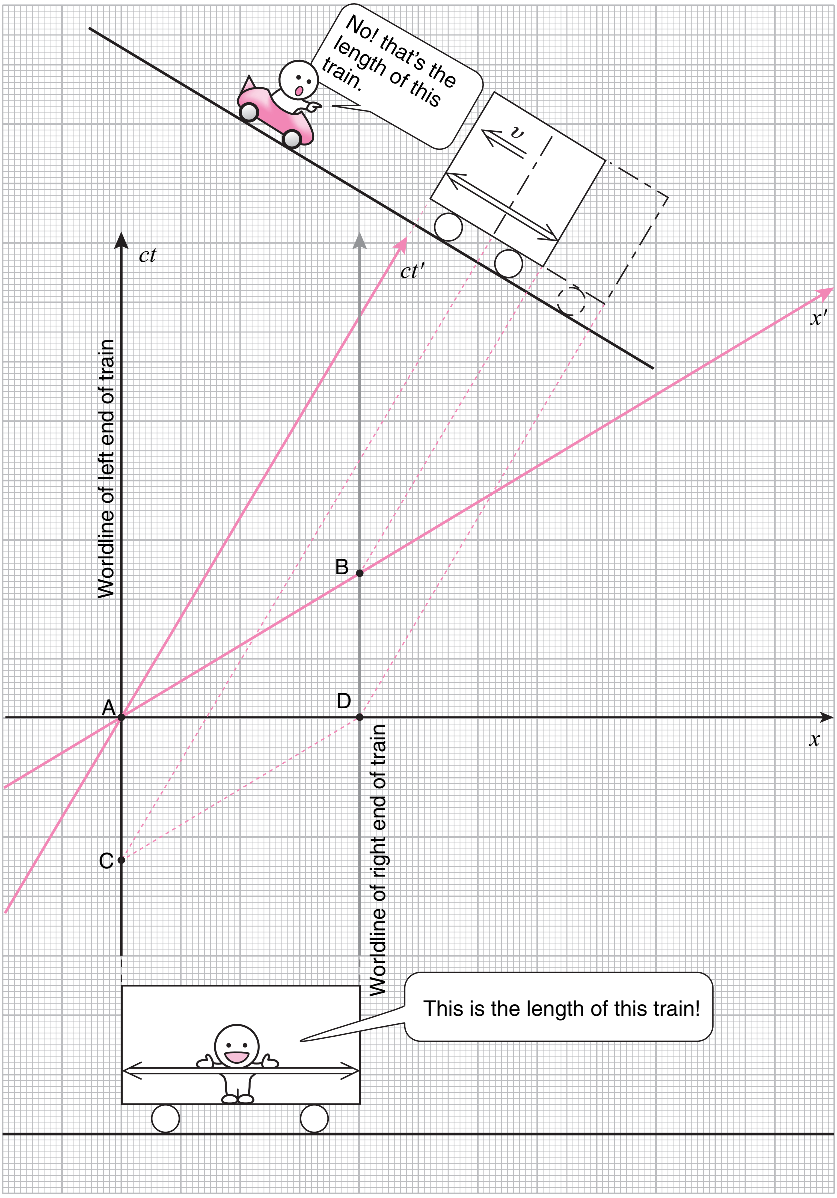
Since what we think of as the natural length of an object is its length at rest, we can conclude that moving objects look shorter than their natural lengths at rest. This is called *Lorentz contraction.*

There is a slight complication in this derivation, however, due to the fact that the Lorentz transformation involves stretching and squeezing of the spacetime grid so it may not be that clear whether the moving frame would associate a shorter distance to CB than the stationary frame would to AB.

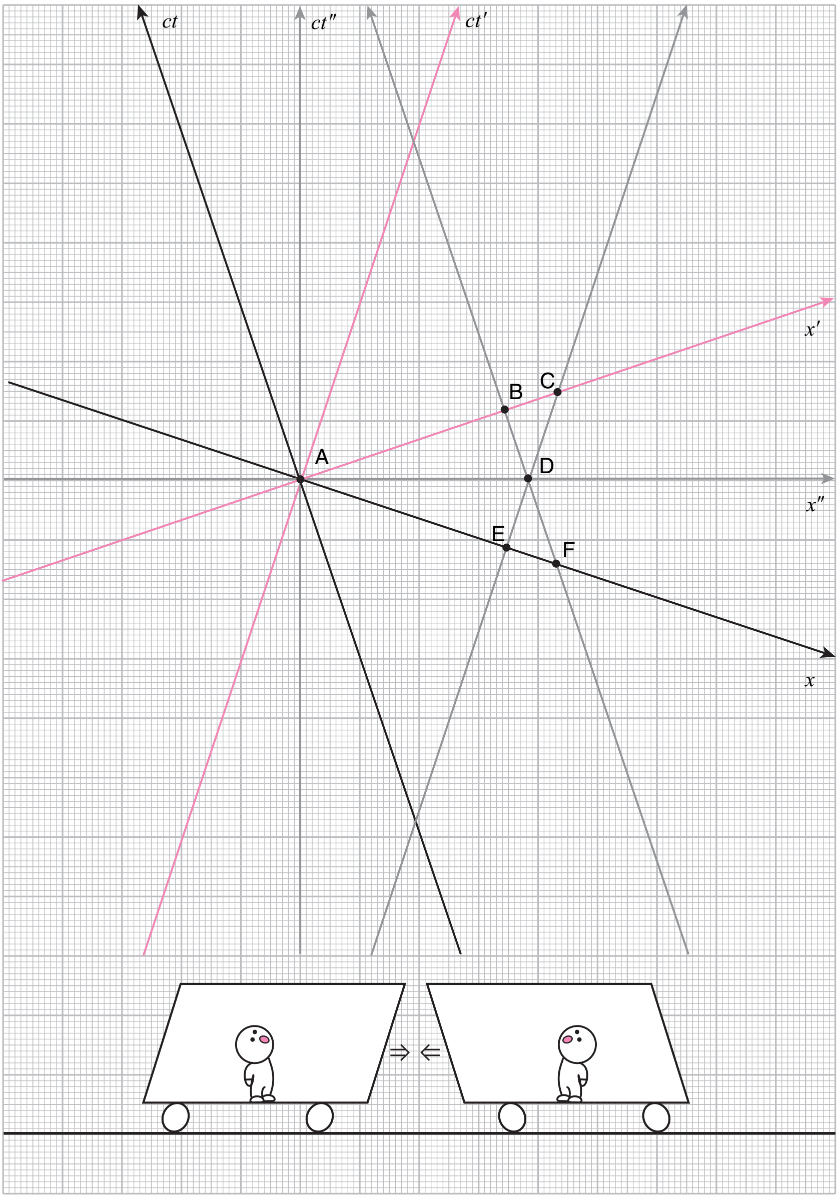

This is the length of this train.
ct'
Worldline of right end of train
ct
A
B
x'
C
D
Worldline of left end of train
x
No! That's the length of this train.
v

If we consider the case where the train is at rest, this complication becomes clear. In this spacetime diagram, the observer on the stationary train will measure the distance between points A and D and call it the length of the train. A moving observer in another train or car in the (x', ct')-frame, on the other hand, will measure the distance between points A and B and call that the length of the train.

Now since the train will move to the left between CD and AB in the (x', ct')-frame, AB measured in the (x', ct')-frame should be shorter than AD measured in the (x, ct)-frame. But this is not clear from this particular spacetime diagram.


No! that's the length of this train.
v
ct
ct'
x'
Worldline of left end of train
B
A
D
x
C
Worldline of right end of train
This is the length of this train!

The way to circumvent this complication is to Lorentz transform to a third frame in which the first two frames are moving at the same speed in opposite directions. Here, we show in the spacetime diagram two trains of equal lengths that are moving in opposite directions at the same speed. The train moving to the left is the (x, ct)-frame train, and the train moving to the right is the (x', ct')-frame train. In the (x, ct)-frame, the length of the left-moving train (which is at rest in that frame) is AF while the length of the right-moving train is AE. Clearly, AE is shorter than AF. In the (x', ct')-frame, the length of the right-moving train (which is at rest in that frame) is AC, while the length of the left-moving train is AB. Clearly, AB is shorter than AC.



ct
ct″
ct′
x′
x″
x
A
B
C
D
E
F
⇒ ⇐

6.5 What Lorentz contraction DOES NOT mean

Now, just like *time dilation*, Lorentz contraction is often misunderstood to mean that space itself shrinks in the direction of motion in the moving frame, and consequently, spatial distances measured in the moving frame will always be longer (since they are using a shorter ruler) than in the stationary frame.

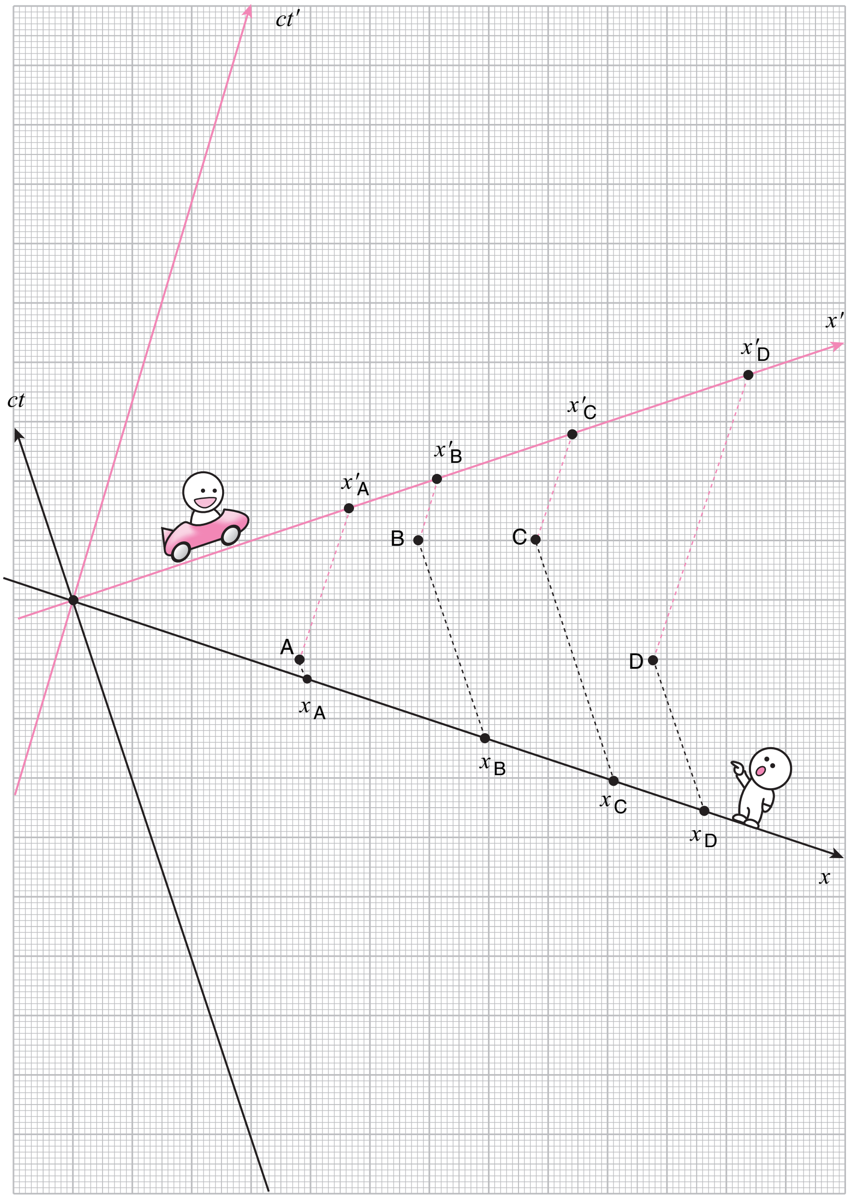
Again, this is not true.

First, as in the *time dilation* case, the relationship between the stationary and moving frames is completely symmetrical, since which frame we consider to be moving is a matter of choice. So if space shrinks in the moving frame compared to the stationary frame, then space has to shrink in the stationary frame compared to the moving frame. A clear contradiction.

Second, when the two observers are measuring lengths they are not measuring the spatial separation of the same events at all. They are always measuring events along a time-slice in their own respective frames.

If the two observers actually measure the spatial separation between the same events on the spacetime diagram, there is no rule that says that one observer will obtain a shorter length than the other. For instance, take the four points A, B, C, and D shown here. The spatial separation between A and B is measured to be longer in the ground-frame than in the car-frame. However, the spatial separation between B and C are measured to be the same in both frames, while the spatial separation between C and D is shorter in the ground-frame than in the car-frame.

Lorentz contraction is a phenomenon caused by the relativity of simultaneity, just like time dilation, and does not imply the stretching or shrinking of space itself in any way.


ct'
x'
x'D
x'C
x'B
x'A
ct
B
C
A
D
xA
xB
xC
xD
x

6.6 Twin paradox

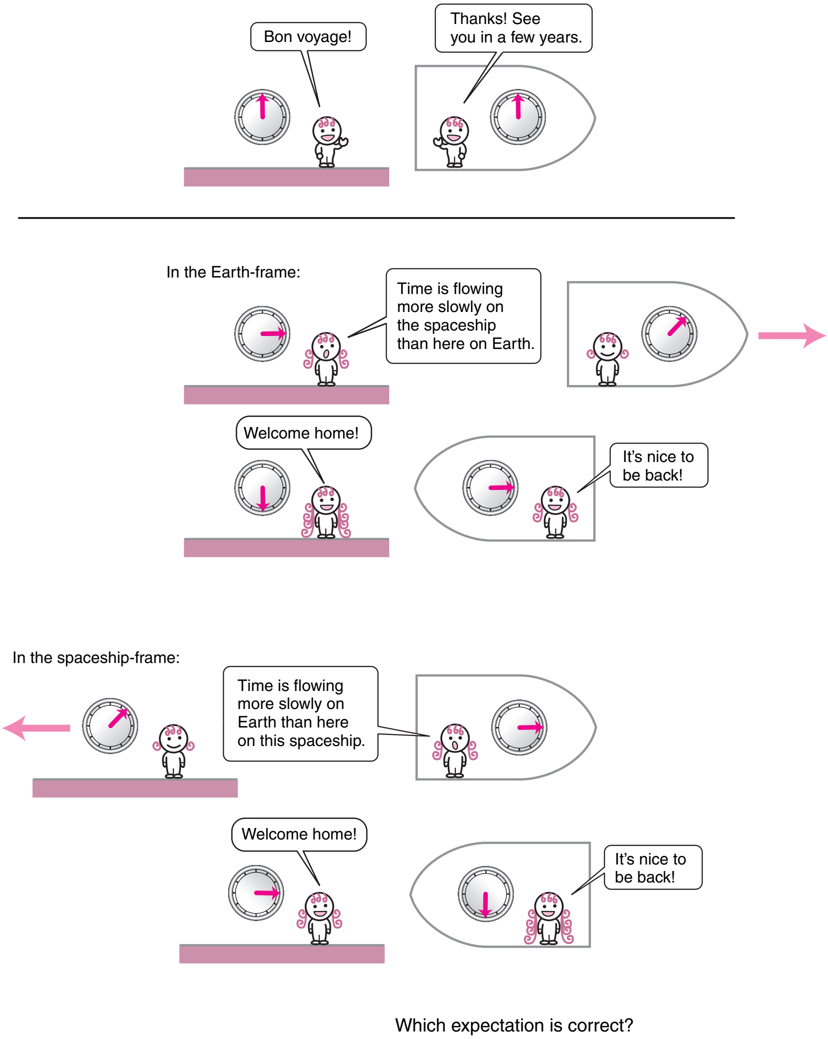
A common argument made against relativity is the following:

Consider sending an astronaut to a distant star in a spaceship at very high speed. Assume that the astronaut has a twin sibling waiting back on Earth.

In the Earth-frame of reference, time on the spaceship will be observed to pass more slowly than on the Earth due to time dilation. It may seem as if only a few years have passed on the ship while decades pass on the Earth. So the twin of the astronaut waiting on the Earth expects the astronaut to be the younger of the two upon return.

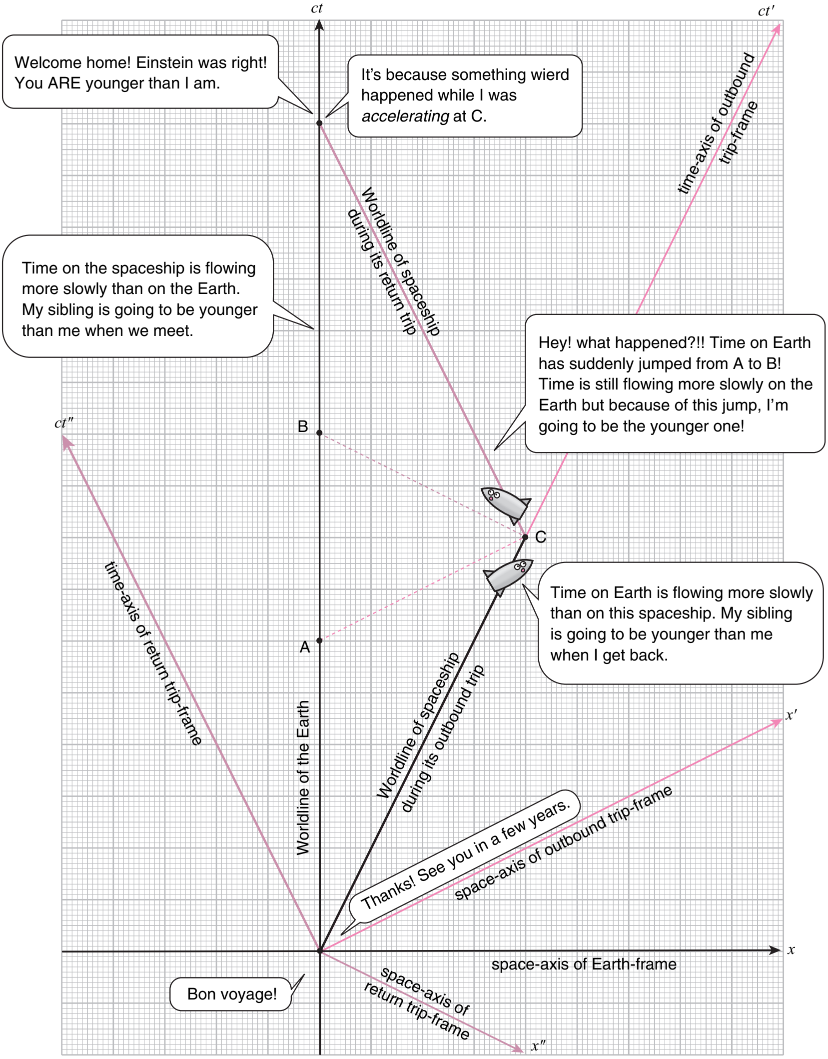
In the spaceship-frame of reference, it is the Earth that is moving at a very high speed so time on Earth will be observed to pass more slowly than on the spaceship. Decades will pass on the ship while only a few years pass on Earth. So the astronaut expects that the twin sibling waiting on the Earth to be the younger of the two upon return.

Isn't this a contradiction? What happens when the astronaut comes back to Earth? Which of the twin siblings will be older? How can both observations be correct? This problem is known as the "twin paradox."


Bon voyage!
Thanks! See you in a few years.
In the Earth-frame:
Time is flowing more slowly on the spaceship than here on Earth.
Welcome home!
It's nice to be back!
In the spaceship-frame:
Time is flowing more slowly on Earth than here on this spaceship.
Welcome home!
It's nice to be back!
Which expectation is correct?

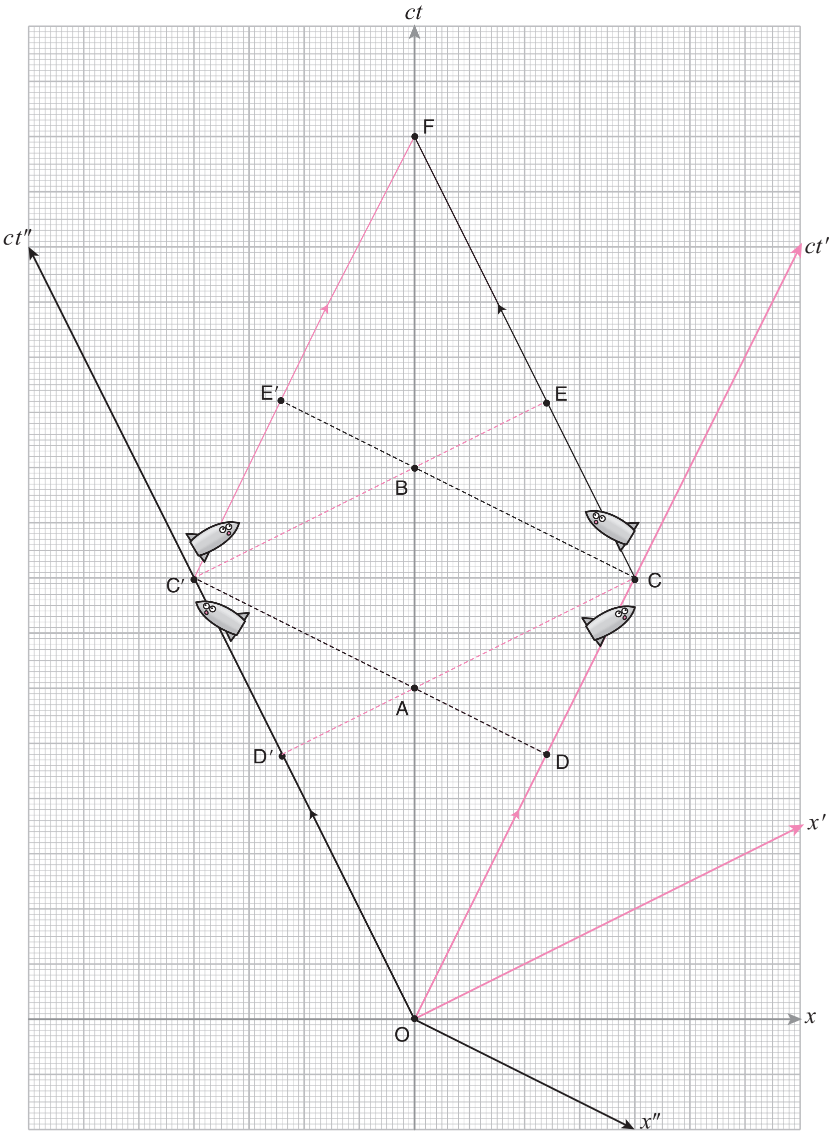
To resolve this question, let us first notice that in order for the astronaut to return to the Earth, she must change direction somewhere. Otherwise, her spaceship will keep on moving at constant velocity, taking her farther and farther away from the Earth. Changing direction means that she must switch from one inertial frame to another, and that breaks the symmetry between the two siblings.

Now changing direction involves acceleration, which puts the astronaut in a non-inertial frame. And as we discussed earlier, special relativity is a theory that only applies to observations made from inertial frames, so we can't really tell what the astronaut observes during the acceleration process. But let's for the sake of argument assume that the acceleration process can be made as short as we like and can be considered an instantaneous jump from one inertial frame to another. In the spacetime diagram shown here, this jump occurs at C. Notice that in the frame the spaceship was in before the jump, C is simultaneous with A on Earth. But in the frame that the spaceship is in after the jump, C is simultaneous with B on Earth. So, after the jump, the astronaut discovers that time on Earth has suddenly jumped from A to B. Because of this effect, the astronaut will be younger than her sibling when she returns to the Earth.


ct
ct′
Welcome home! Einstein was right! You ARE younger than I am.
It's because something wierd happened while I was *accelerating* at C.
time-axis of outbound trip-frame
Worldline of spaceship during its return trip
Time on the spaceship is flowing more slowly than on the Earth. My sibling is going to be younger than me when we meet.
Hey! what happened?!! Time on Earth has suddenly jumped from A to B! Time is still flowing more slowly on the Earth but because of this jump, I'm going to be the younger one!
ct″
B
C
Time on Earth is flowing more slowly than on this spaceship. My sibling is going to be younger than me when I get back.
A
time-axis of return trip-frame
Worldline of the Earth
Worldline of spaceship during its outbound trip
x′
Thanks! See you in a few years.
space-axis of outbound trip-frame
x
space-axis of Earth-frame
Bon voyage!
space-axis of return trip-frame
x″

We can make the process completely symmetric by sending both siblings on separate spaceships in opposite directions at the same speed, have them turn around at a predetermined distance from the Earth, and have them return at the same time (on the Earth). The worldlines of their trips are shown here: they both start at the origin O, sibling 1 travels toward the right and turns back at C, while sibling 2 travels toward the left and turns back at C′.

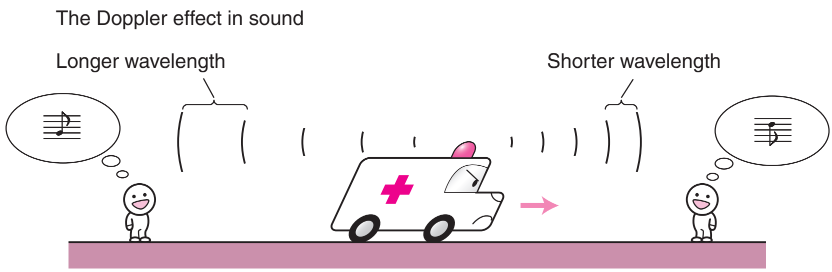
Both siblings will observe time on the other spaceship to be passing at a slower rate due to *time dilation*. However, in sibling 1's frame before the turn-around, C is simultaneous with A and D′, but after the turn-around, C is simultaneous with B and E′. So as sibling 1 turns around at C, time in the other spaceship suddenly jumps from D′ to E′. Similarly, in sibling 2's frame before the turn-around, C′ is simultaneous with A and D, but after the turn-around, C′ is simultaneous with B and E. So as sibling 2 turns around at C′, time in the other spaceship suddenly jumps from D to E. So even though both siblings will observe time in the other spaceship to be passing at a slower rate when they are cruising at constant velocity, time on the other spaceship will suddenly move forward when they turn around. The net effect is that both siblings will be the exact same age when they return to the Earth.

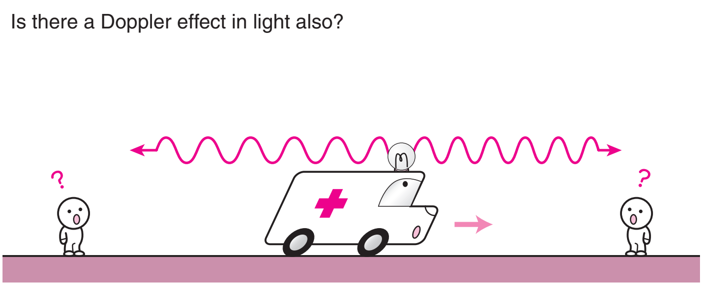

ct
F
ct″
ct′
E′
E
B
C′
C
A
D′
D
x′
x
O
x″

6.7 Doppler effect

As a final example of the consequences of relativity, let us discuss the *Doppler effect*[1] *in light*. Doppler effect refers to the phenomenon in which the observed frequency of a wave changes depending on the motions of the wave source and the observer. For instance, if an ambulance passes by with its sirens blaring, you will notice that the sound is at a higher pitch when the ambulance is moving towards you than when the ambulance is moving away from you. That is an example of the Doppler effect.

Now, if you have learned about the *Doppler effect in sound* in school, you may recall that it was caused by the fact that the speed of sound relative to the sound source and the observer depended on their motions relative to the air through which the sound was propagating. Since the speed of light, on the other hand, is independent of the motion of the light source or that of the observer, you may wonder whether there is a Doppler effect in light at all. The effect does indeed also happen in light, but due to the relativity of simultaneity.


The Doppler effect in sound
Longer wavelength
Shorter wavelength

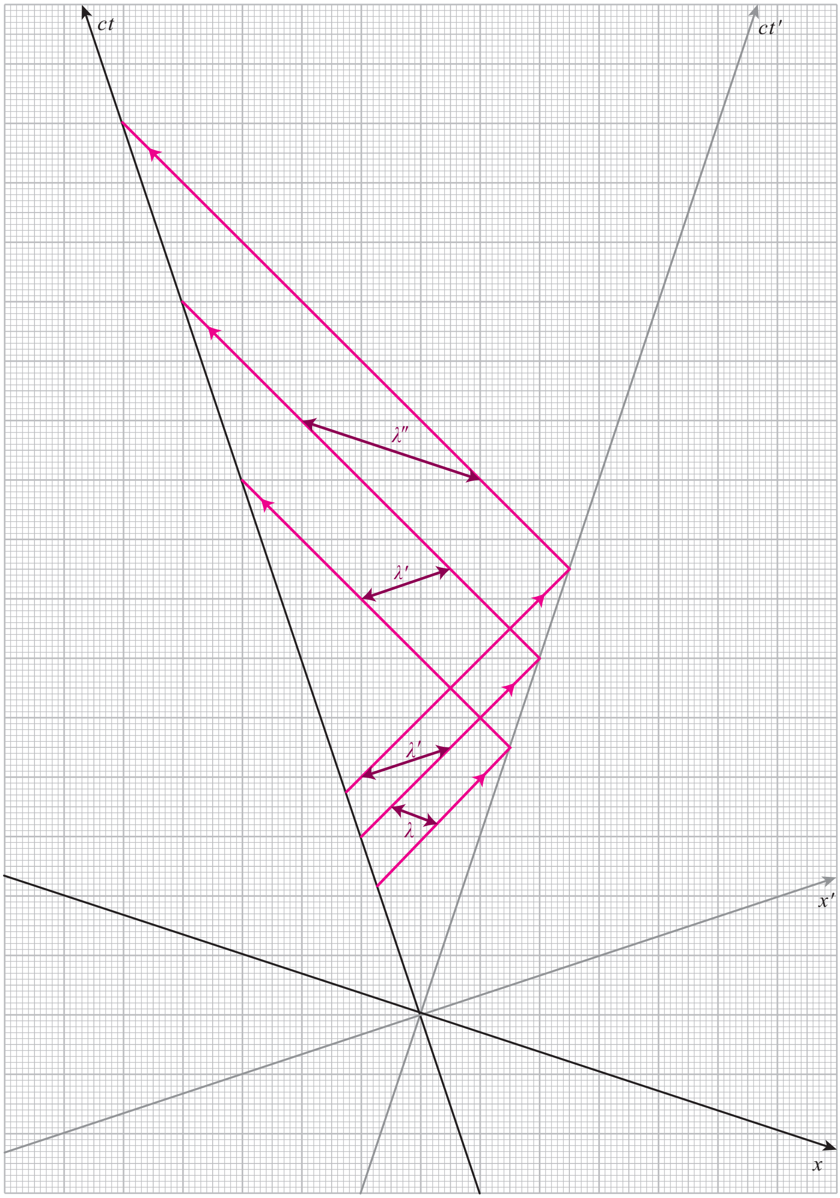

Is there a Doppler effect in light also?
?
?

6.7.1 Red shift

Consider a light source whose worldline is given by the ct-axis of the spacetime diagram shown here. It is moving to the left as it gives off light at a particular frequency. The worldlines of the crests of the light-wave are shown here in red. The spatial separation of the successive crests give the wavelength of the light. In the light-source frame, the initial wavelength is labeled λ in the diagram. This light is received by an observer moving away from the light source to the right whose worldline is given by the ct'-axis. The spacetime diagram is drawn in a frame in which the source and observer are moving away from each other in opposite directions at the same speed so that the scales on the (x, ct) axes and the (x', ct') axes are the same.

It is clear from this diagram, I think, that the time intervals separating the successive wave crests that arrive at the observer are larger than those of the successive wave crests leaving the light source. This means that the frequency of the light that is observed by the observer is smaller than the frequency of the light at the source. This is also clear from the comparison of wavelengths: the wavelength of the incoming light as observed from the observer frame is λ', which is longer than that in the light-source frame λ.[2] If the light is reflected by the observer and travels back to the source, its wavelength in the light-source frame will be λ'', which is again longer than λ'.

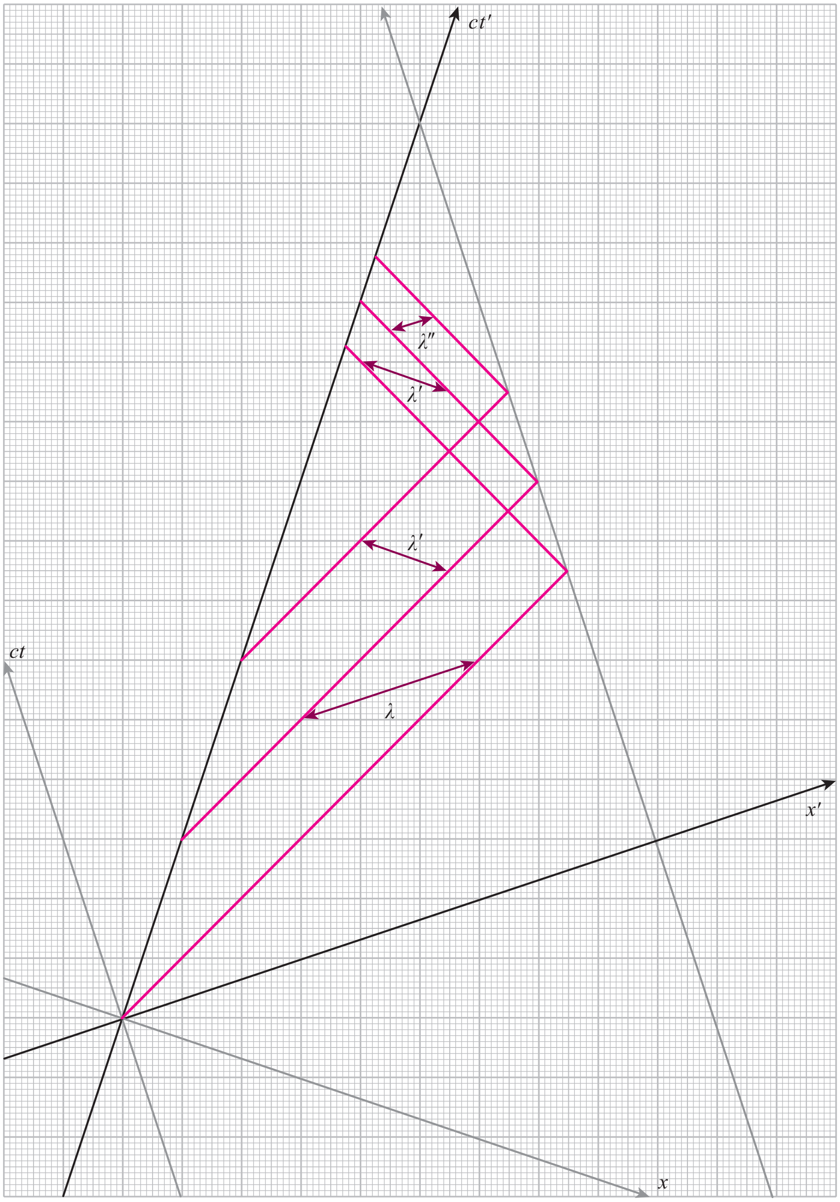
What you can conclude is that if the light source is receding away from you, the frequency will be shifted to a smaller value, or, equivalently, the wavelength will be shifted to a longer value. This effect is known as *red shift*. The terminology comes from the fact that the different colors of the rainbow are light of different frequencies/wavelengths. The red end of the rainbow is light with lower frequencies (longer wavelengths), and the blue/violet end of the rainbow is light with higher frequencies (shorter wavelengths). Since the Doppler effect discussed here will cause all the colors to be shifted in the direction of red, it is called red shift.



ct
ct'
λ''
λ'
λ'
λ
x'
x

6.7.2 Blue shift

What if the light source is coming towards you instead of receding away from you? The spacetime diagram for that situation is shown here. The light source is moving to the right along the ct'-axis, while the observer is moving to the left along the worldline parallel to the ct-axis. The spacetime diagram tells us that, in this case, the frequency (wavelength) of light received by the observer is higher (shorter) than that emitted by the source. This effect is known as *blue shift* since the colors of the rainbow will be shifted towards blue/violet.


ct'
λ''
λ'
λ'
ct
λ
x'
x

6.7.3 Red shift and the expansion of the universe

The amount of shifting that occurs towards the red or the blue depends on how fast the light source is moving relative to the observer. So by measuring by how much the frequency of light has shifted, one can not only determine whether the light source is receding away or coming towards you, but also what its speed is. That is how astronomers know that (1) all the galaxies in the universe on average are receding away from us, and (2) the farther away the galaxy is, the faster the speed by which it is receding, by measuring the amount of red-shift in the light coming from those galaxies. This is what tells us that the universe is expanding. Two natural questions immediately come to mind:

- How do astronomers know what the original frequency of the star-light that we observe on Earth was? If we see yellow light, has it red-shifted from blue, or has it blue-shifted from red?
- How do astronomers know how far away a particular galaxy is?

The second question is actually a very difficult problem in astronomy and beyond the scope of this book. I will have to ask you to study it yourself in a book on astronomy. (See, for instance, [8].)

The answer to the first question is: the astronomers look for light emitted by specific atoms. Atoms give off very specific frequencies of light called *spectral lines*, so-called since they show up as distinct lines if you use a prism to separate the frequencies into a rainbow of colors. The pattern of the lines is unique to each atom. So even if the pattern is red-shifted, astronomers can still identify which atom the spectral lines came from. And since the original frequencies of those spectral lines are known from experiments on Earth, astronomers can determine the amount of red-shift that the light has undergone.

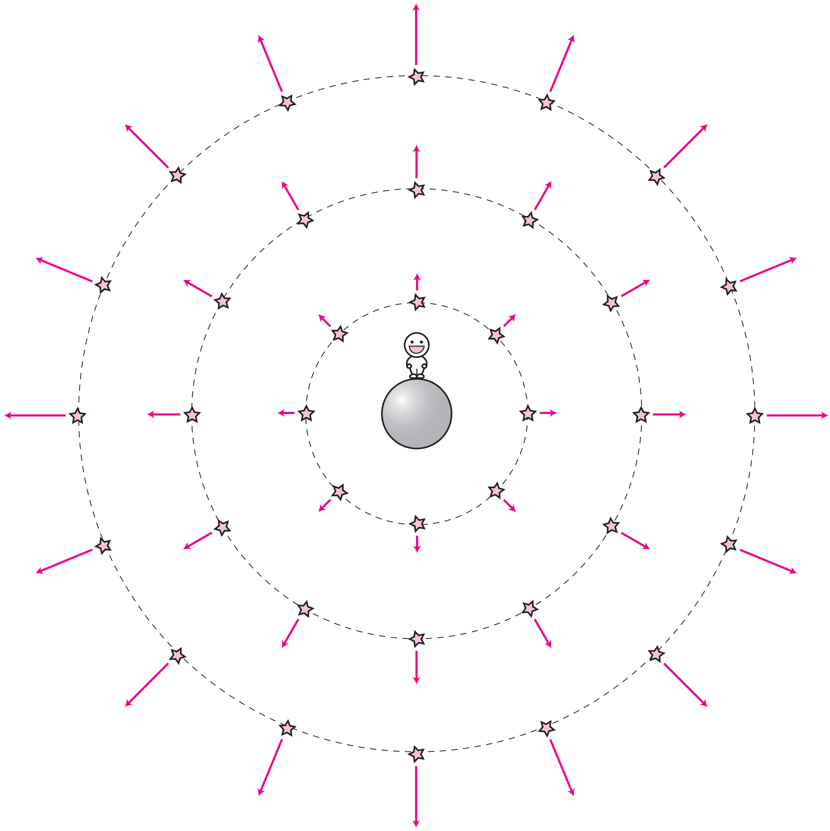
Notes

1 Named after Christian Andreas Doppler (1803–1853). He predicted the phenomenon in a paper in 1842.

2 The frequency f and the wavelength λ are related by the relation

$$f\lambda = c\,.$$

So if λ gets bigger, f becomes smaller.



7

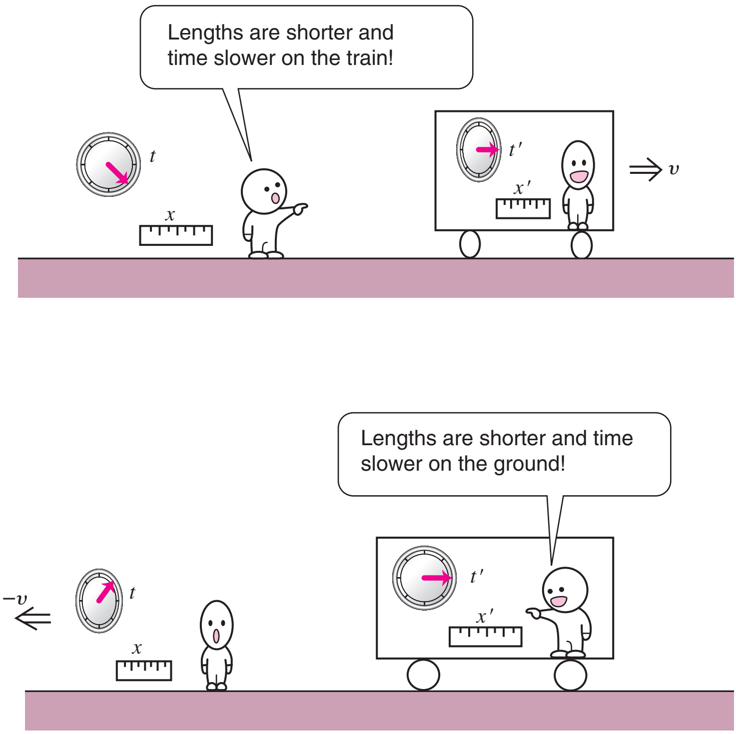
Summary of Part I

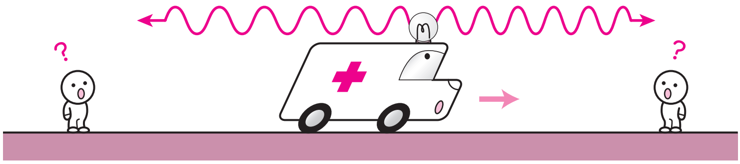
This concludes Part I of this book. I hope you have been able to grasp an outline of what relativity is all about. Let us summarize what we have learned:

- The 'Special Theory of Relativity' was constructed by Einstein to resolve the mystery of the speed of light. Einstein's solution was that the concept of *simultaneity* depended on the frame of reference. And the rule that relates the observations from different frames was given by the Lorentz transformation.

The predictions of Special Relativity such as time dilation and Lorentz contraction are as infamous as they are famous. The reason for the notoriety is due to the apparent paradoxical nature of the prediction: say we have two frames, A and B, moving relative to each other. According to Special Relativity, the observer in frame A will observe the clock in frame B to run slower than the clock in frame A, and the ruler in frame B to be shorter than the ruler in frame A. The observer in frame B will observe the exact opposite. Now how can both points of view be true *at the same time*?

Of course, the two points of view are NOT true *at the same time*. They are both true because they are NOT *at the same time*. Time dilation and Lorentz contraction were both consequences of the fact that different observers do not agree on what it is meant to be *at the same time*. Let us not forget this since otherwise we can be misled to all sorts of paradoxes which have nothing to do with the predictions of relativity.


Lengths are shorter and
time slower on the train!
t
x
t′
x′
v
Lengths are shorter and time
slower on the ground!
−v
t
x
t′
x′



PART II

Problems

8
Qualitative problems

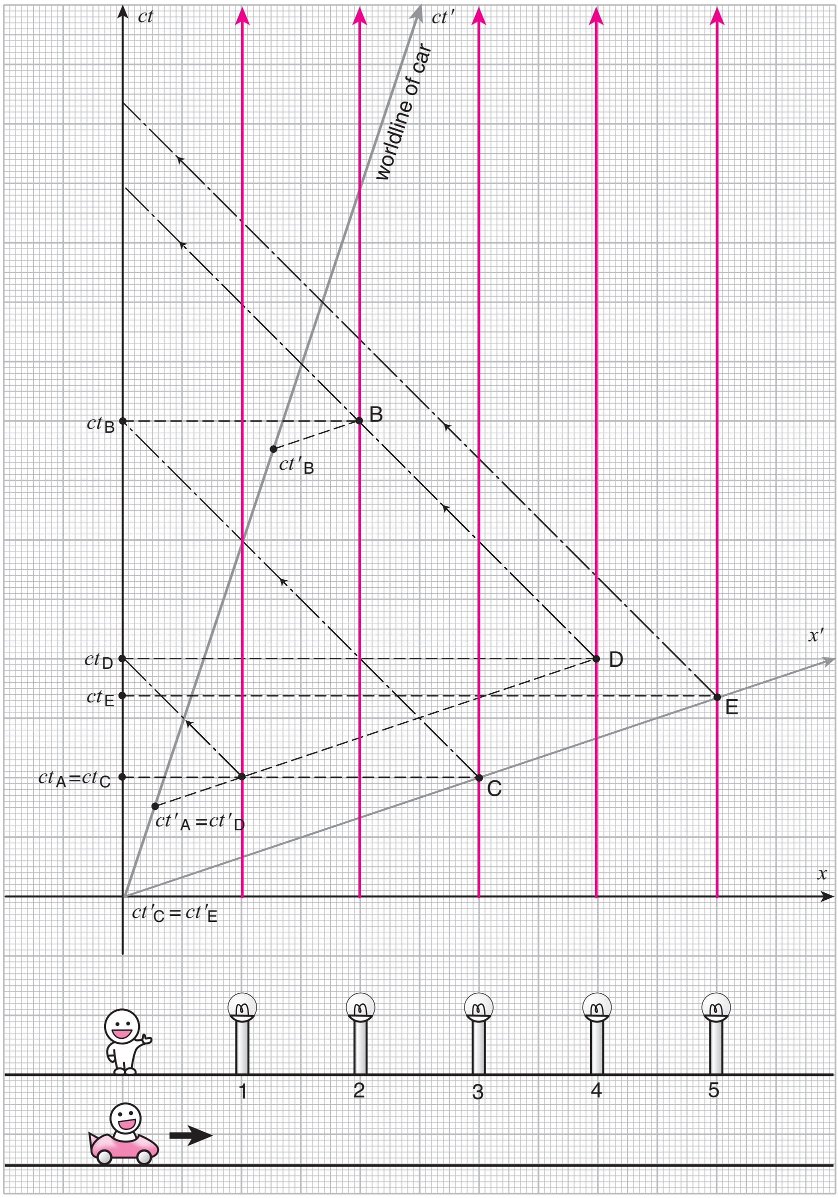
All the problems in this chapter are qualitative and you will be able to solve them if you can read spacetime diagrams. Try them out to test your understanding of Special Relativity.[1]

8.1 Reading the spacetime diagram

8.1.1 Street lamps

Five street lamps, numbered 1 through 5, are located on a straight line along the x-axis equal distance apart as shown in the figure. They turn on at points A, B, C, D, and E, respectively, on the spacetime diagram.

1. In what order do the lamps turn on in the ground-frame?
2. In what order does the light from the lamps reach the observer at the origin $x = 0$?
3. A car is moving to the right at constant speed relative to the ground. At $t' = t = 0$, it is at $x' = x = 0$. The space- and time-axes in the moving frame of the car are tilted with respect to those of the rest frame as shown in the spacetime diagram. In what order do the lamps turn on in the car-frame?
4. In what order does the light from the lamps reach the observer riding the car?
5. Where is the car when the light from street lamp 4 reaches it?



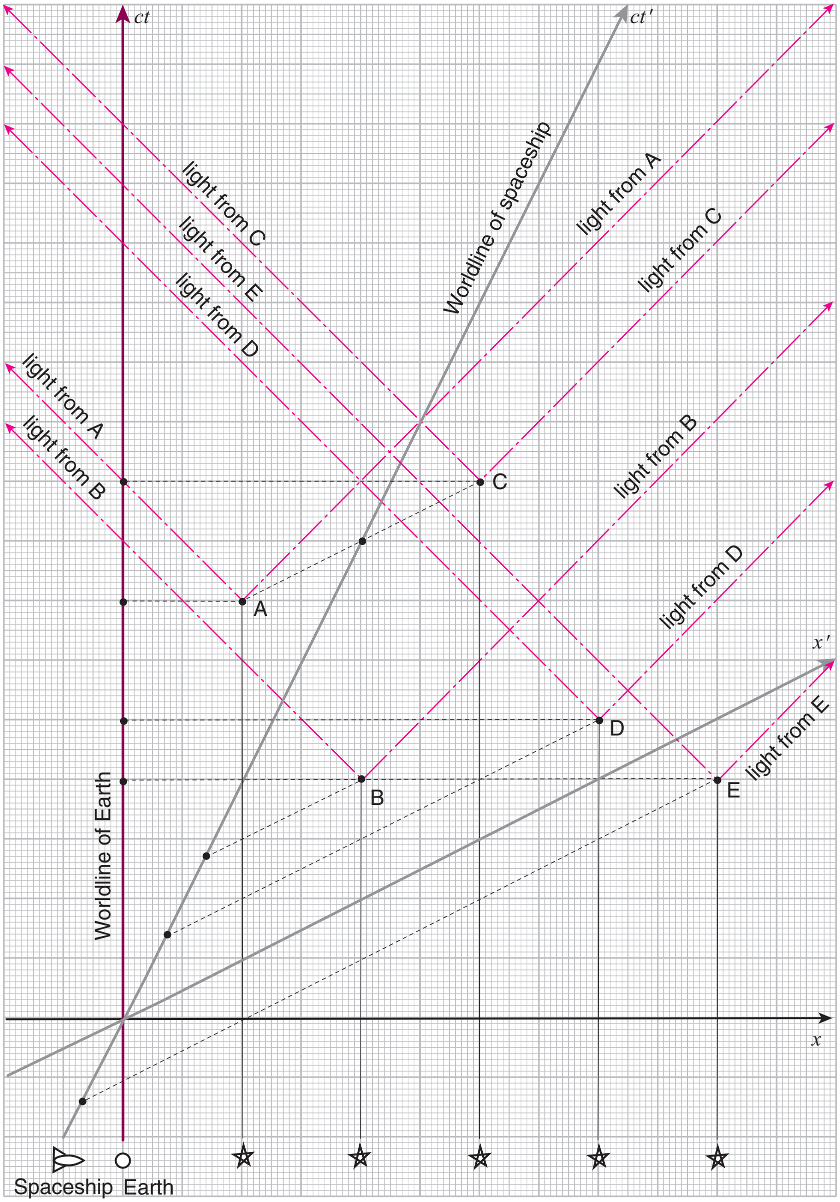
ct
ct′
worldline of car
B
ct_B
ct'_B
D
ct_D
ct_E
E
x′
$ct_A = ct_C$
C
$ct'_A = ct'_D$
x
$ct'_C = ct'_E$
1
2
3
4
5

8.1.2 Supernovae

The spacetime diagram shows five stars which go supernova (that is, explode) at spacetime points A, B, C, D, and E. These supernovae are observed by astronomers on the Earth, and also by scientists aboard a fast moving spaceship, the worldlines of which are also shown on the spacetime diagram.

Answer the following questions:

1. In which chronological order do the five supernovae occur in the Earth-frame of reference?
2. In which chronological order do the five supernovae occur in the spaceship-frame of reference?
3. In which chronological order do the astronomers on Earth see the supernovae?
4. In which chronological order do the scientists on the spaceship see the supernovae?
5. Is the chronological order in which supernovae A and B occur the same in all frames of reference? Explain.


ct
ct′
x′
x
light from C
light from E
light from D
light from A
light from B
Worldline of spaceship
light from A
light from C
light from B
light from D
light from E
A
B
C
D
E
Worldline of Earth
Spaceship
Earth

8.2 Questions on before and after

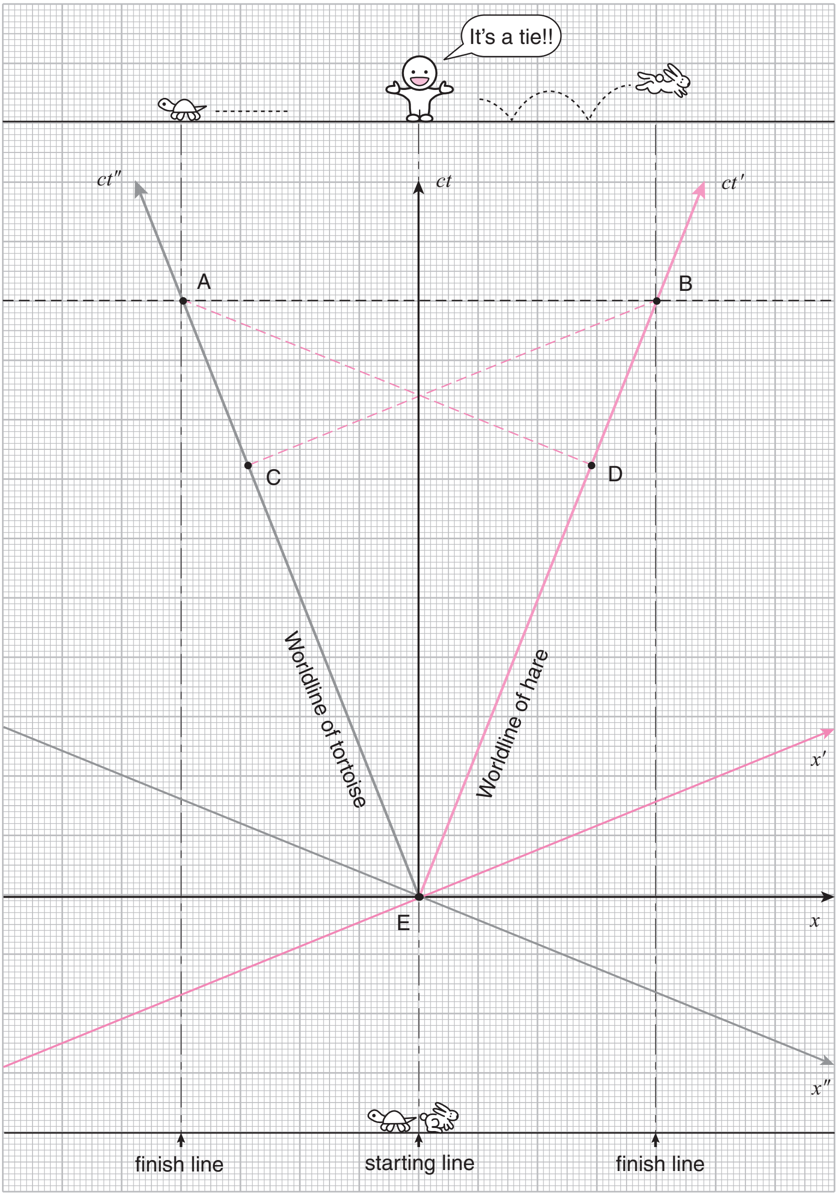
8.2.1 The hare and the tortoise 1

The hare and the tortoise decide to have another race. But instead of running in the same direction toward the finish line, they decide to run in opposite directions toward finish lines located at equal distances from the starting line, as shown in the figure.

The race takes place and, in the frame fixed to the ground, both animals cross their respective finish lines at the same time and the referee declares the race a tie. Assuming that the two animals move at a constant velocity from start to finish, what is the result of the race as seen from the frames of the two moving animals?

Choose from one of the following and explain the reason for your choice. Use the spacetime diagram to facilitate your explanation.

1. Both the hare and the tortoise think that they won.
2. Both the hare and the tortoise think that they lost.
3. The hare thinks it won and the tortoise thinks it lost.
4. The hare thinks it lost and the tortoise thinks it won.
5. None of the above.

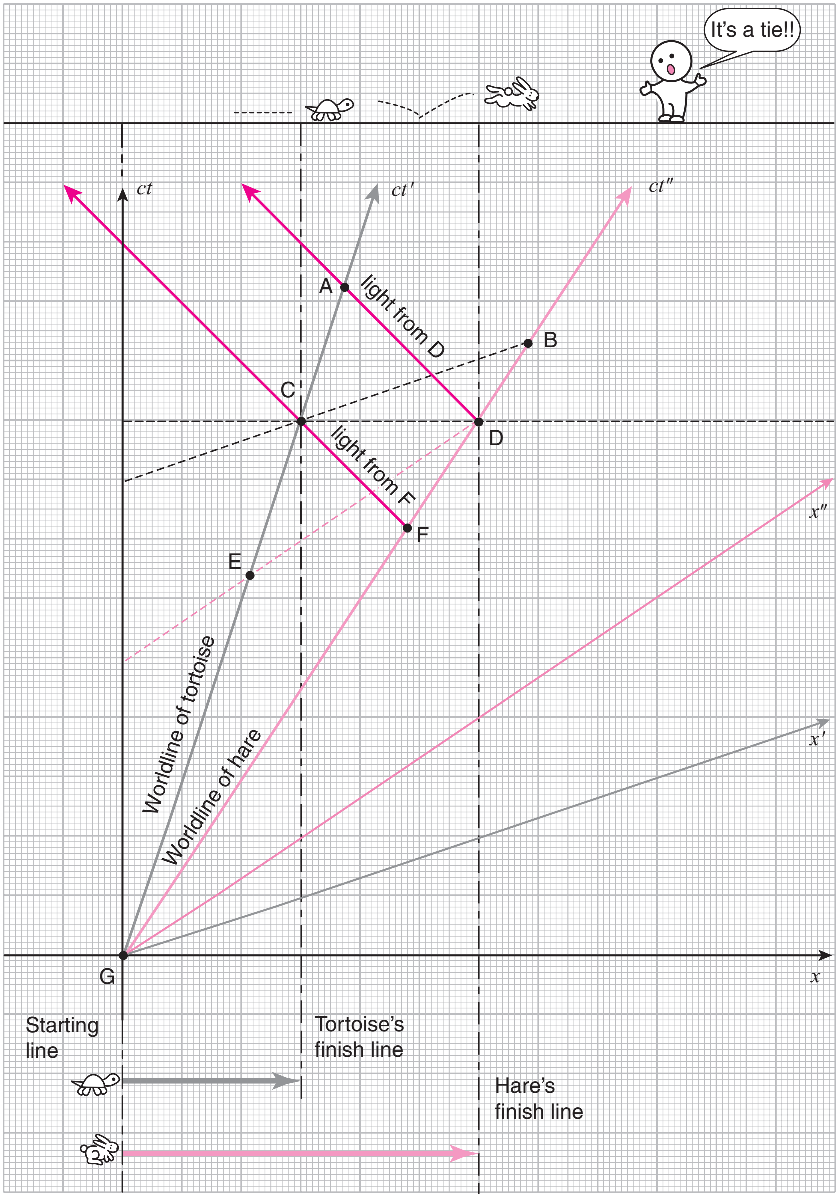

It's a tie!!
ct''
ct
ct'
A
B
C
D
Worldline of tortoise
Worldline of hare
x'
x
E
x''
finish line
starting line
finish line

8.2.2 The hare and the tortoise 2

The hare and the tortoise decide to have another race. They start from the same point and race in the same direction, but the hare decides to allow the tortoise a handicap by letting him race only half the distance, as shown in the figure. The race takes place and, in the frame fixed to the ground, both animals cross their respective finish lines at the same time and the referee declares the race a tie. The worldlines of the hare and the tortoise are shown on the spacetime diagram.

Answer the following questions:

1. At which of the points labeled A through G on the spacetime diagram does the hare cross its finish line?
2. At which spacetime point does the tortoise cross its finish line?
3. At which spacetime point is the tortoise when the hare crosses its finish line in the hare-frame?
4. At which spacetime point is the hare when the tortoise crosses its finish line in the tortoise-frame?
5. At which spacetime point does the tortoise actually *see* the hare cross its finish line?
6. What is the result of the race in the respective frames of the hare and the tortoise? Do they agree or disagree?
7. If the tortoise bases its conclusion on what it actually sees, will it think that it won or that it lost?

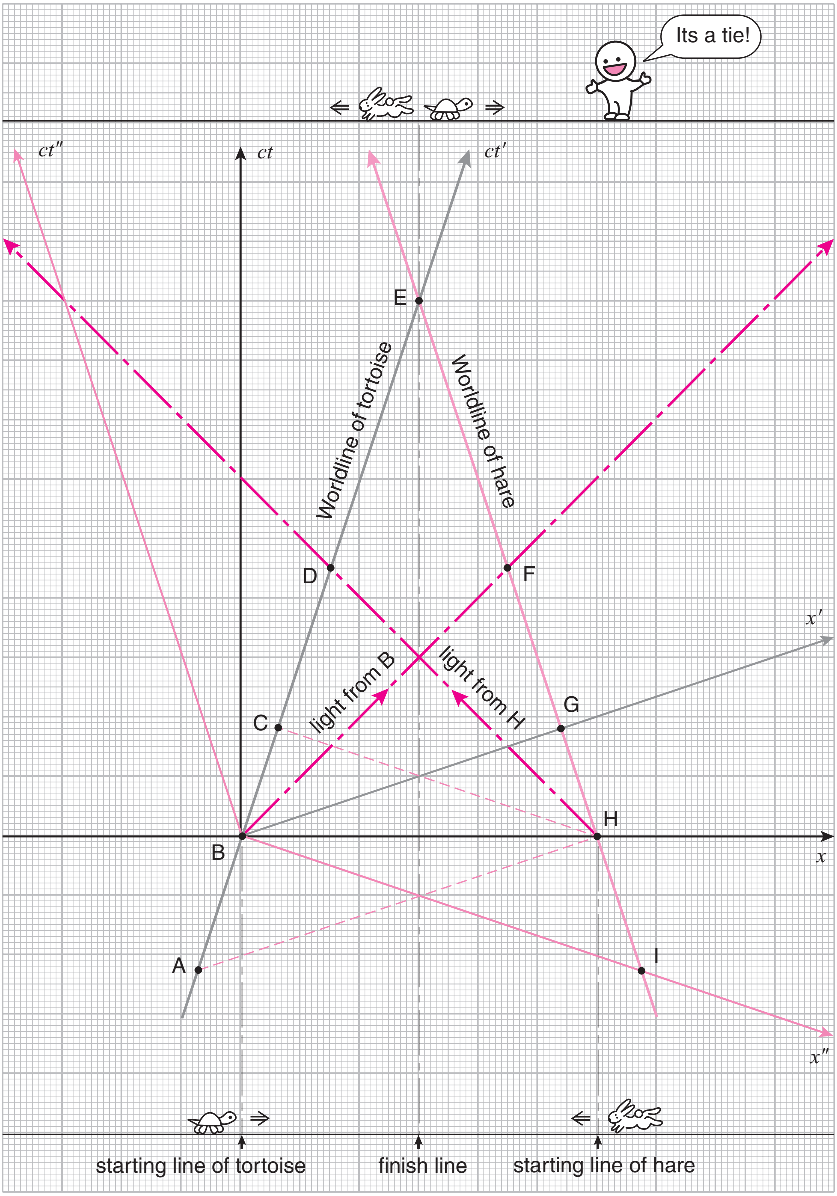

It's a tie!!
ct
ct′
ct″
A
light from D
B
C
D
light from F
F
E
x″
x′
Worldline of tortoise
Worldline of hare
G
x
Starting
line
Tortoise's
finish line
Hare's
finish line

8.2.3 The hare and the tortoise 3

The hare and the tortoise decide to have another race. This time, they start out from opposite directions the same distance away from the finish line, as shown in the figure, and race toward each other. In the frame of the referee fixed to the ground, the two animals cross their respective starting lines at the same time and then cross the finish line at the same time from opposite directions. The referee declares the race a tie. Assume that both animals were moving at constant velocities before, during, and after the race.

Answer the following questions:

1. At which spacetime point, labeled A through I, does the tortoise cross its starting line?
2. In the frame of the tortoise, at which spacetime point is the hare when the tortoise crosses its starting line?
3. At which spacetime point does the hare *see* the tortoise start?
4. Do both animals agree with the referee that they started at the same time? If not, explain in what chronological order the animals start in each animal's frame. Refer to the spacetime diagram in your explanation.
5. Do both animals agree with the referee that they finished at the same time? If not, explain in what chronological order the animals finish in each animal's frame. Refer to the spacetime diagram in your explanation.


Its a tie!
ct″
ct
ct′
E
Worldline of tortoise
Worldline of hare
D
F
x′
light from B
light from H
C
G
H
B
x
A
I
x″
starting line of tortoise
finish line
starting line of hare

8.2.4 The starship and the supernova

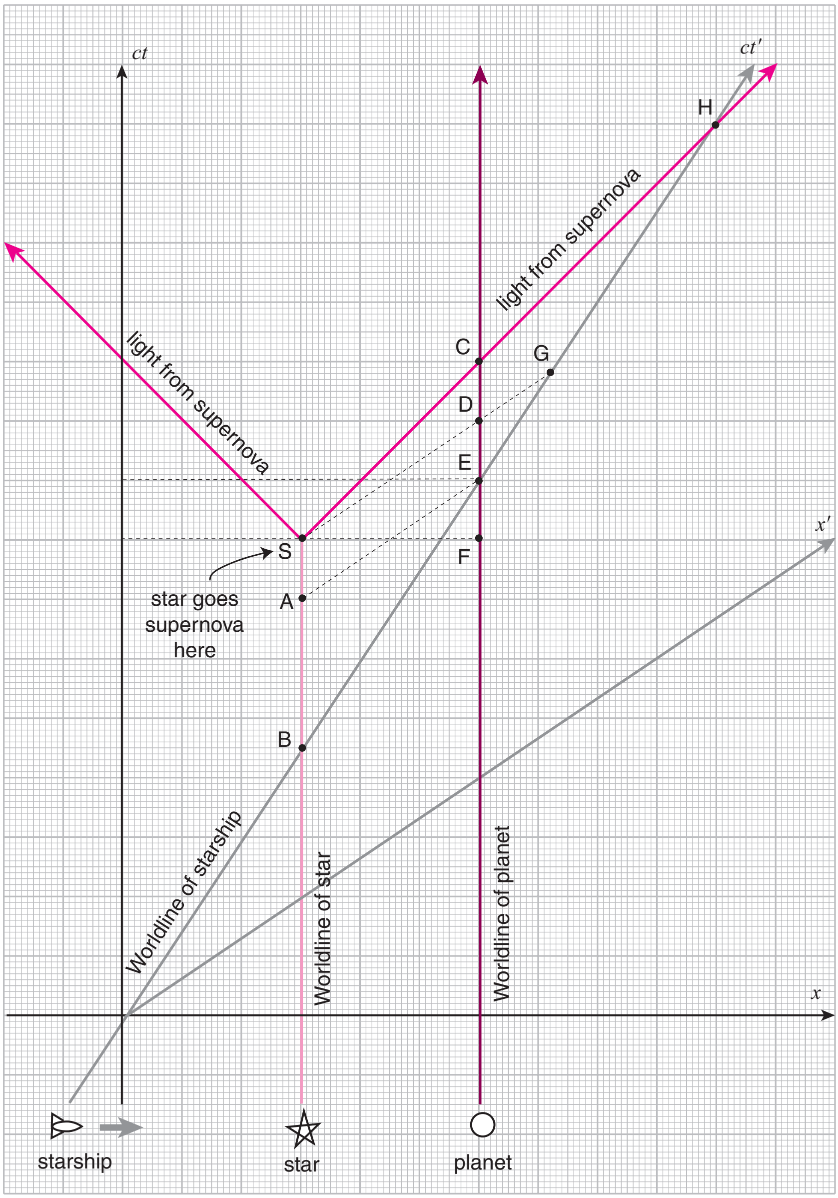
A star is about to go supernova and a planet orbiting it must be evacuated. The starship *Einstein* is sent to the planet to pick up some biology students on a field trip to observe the local flora and fauna.

The worldlines of the *Einstein*, the star, and the planet are shown on the spacetime diagram. Assume that the planet is not moving relative to the star. The *Einstein* will fly by at a constant velocity past the planet and beam up the students without stopping.

The star goes supernova at spacetime point S. The light from the supernova spreads out in both directions along the worldlines shown.

Answer the following questions:

1. At which spacetime point, labeled A through H, does the *Einstein* arrive at the planet?
2. In the frame moving with the *Einstein*, at which spacetime point is the star when the *Einstein* arrives at the planet?
3. In the *Einstein* frame, does the *Einstein* arrive at the planet before the star goes supernova, or after the star goes supernova? Explain your answer referring to the spacetime diagram.
4. In the frame fixed to the planet, does the *Einstein* arrive at the planet before the star goes supernova, or after the star goes supernova? Explain your answer referring to the spacetime diagram.
5. Do the students marooned on the planet *see* the supernova explosion before the *Einstein* arrives, or after the *Einstein* arrives? Explain your answer referring to the spacetime diagram.


ct
ct′
H
light from supernova
light from supernova
C
G
D
E
x′
S
F
star goes
supernova
here
A
B
Worldline of starship
Worldline of star
Worldline of planet
x
starship
star
planet

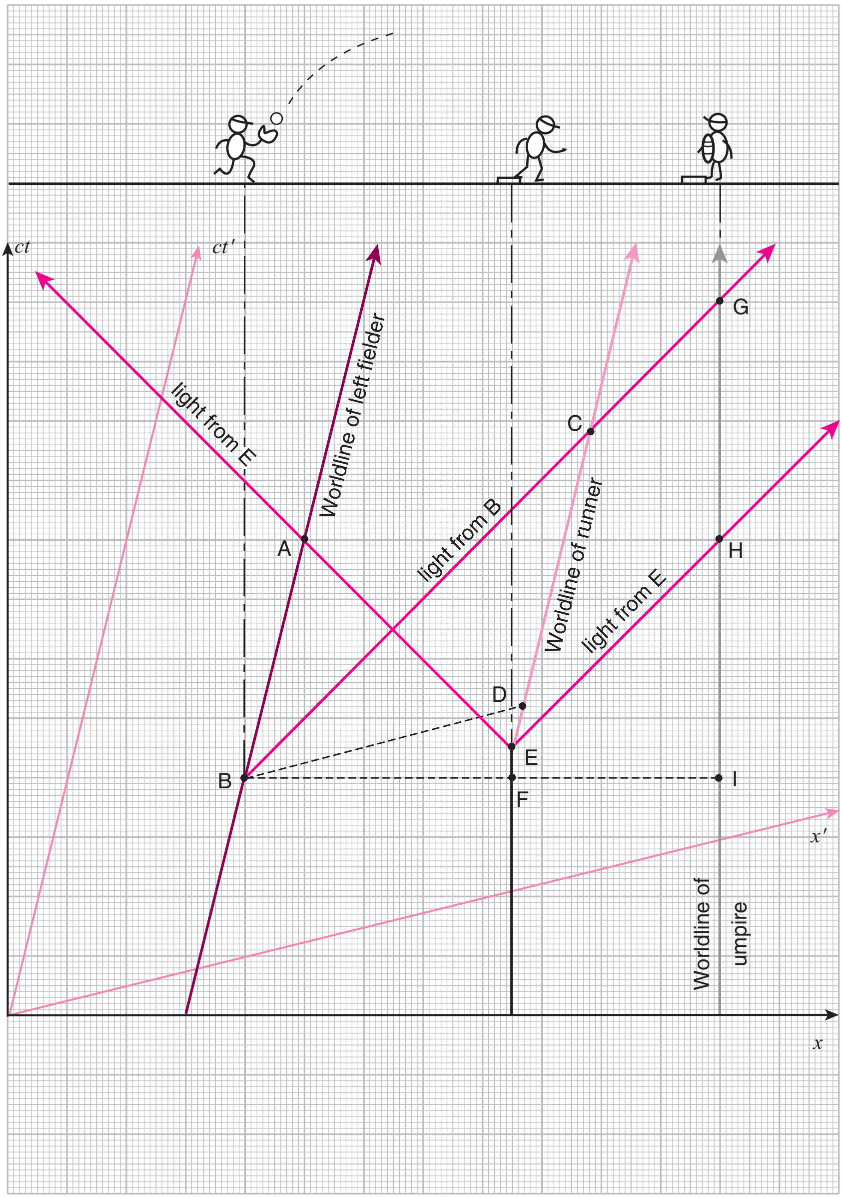
8.3 Relativistic sports

8.3.1 Tagging up in baseball 1

During a baseball game a fly ball[2] is hit toward left field along the third base line. The left fielder runs forward toward third base and catches the ball at point B on the spacetime diagram. The runner on third base tags up and starts off toward home base at point E on the spacetime diagram.

Answer the following questions:

1. At which spacetime point, labeled A through I, is the third base runner when the left fielder catches the ball at B in the umpire's frame of reference?
2. At which spacetime point, labeled A through I, is the third base runner when the left fielder catches the ball at B in the left fielder's frame of reference?
3. In the umpire's frame of reference, has the third base runner committed a foul? Explain, referring to the spacetime diagram.
4. In the left fielder's frame of reference, has the third base runner committed a foul? Explain, referring to the spacetime diagram.
5. If the umpire bases his judgment solely on the chronological order in which he actually sees the ball being caught and the runner leaving third base, will he call a foul? Explain, referring to the spacetime diagram.



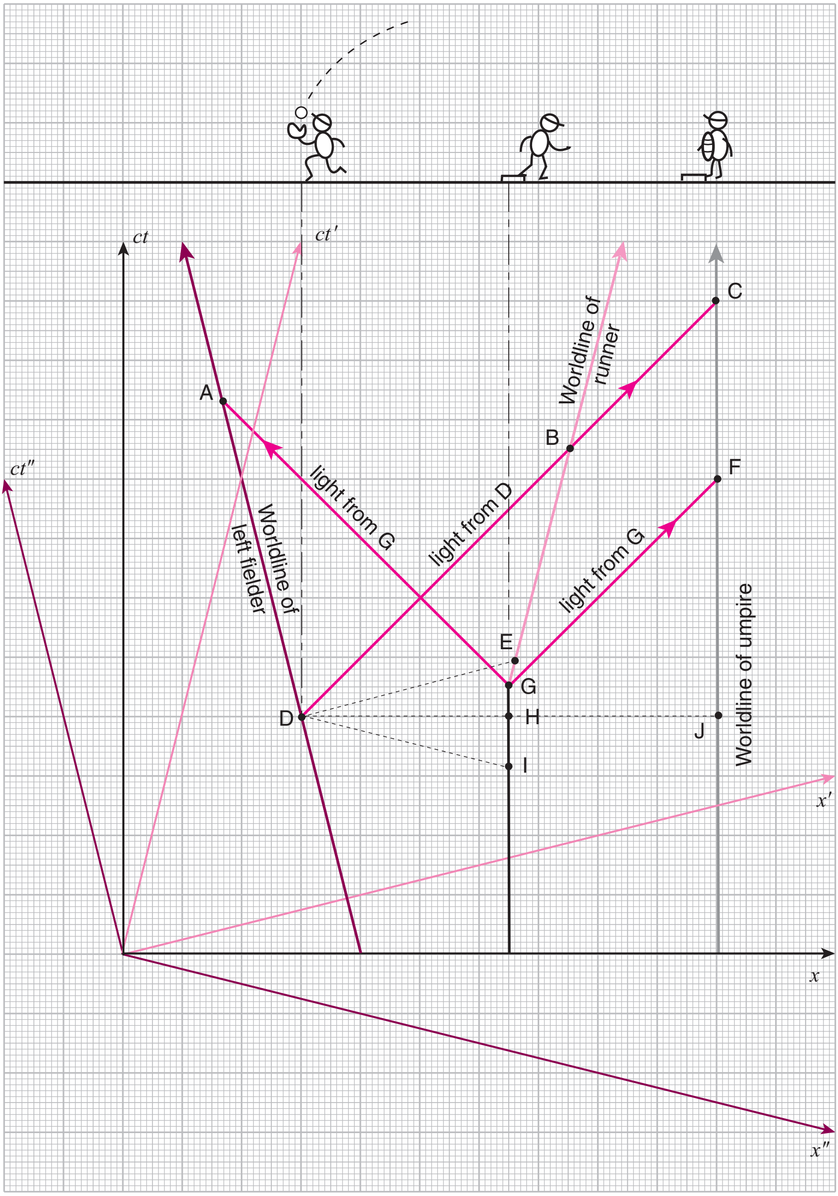
ct
ct′
Worldline of left fielder
light from E
light from B
Worldline of runner
light from E
A
B
C
D
E
F
G
H
I
x′
x
Worldline of
umpire

8.3.2 Tagging up in baseball 2

During a baseball game a fly ball is hit toward left field along the third base line. The left fielder runs backward away from third base and catches the ball at point D on the spacetime diagram. The runner on third base tags up and starts off toward home base at point G on the spacetime diagram.

Answer the following questions:

1. At which spacetime point, labeled A through J, is the third base runner when the left fielder catches the ball at D in the umpire's frame of reference?
2. At which spacetime point, labeled A through J, is the third base runner when the left fielder catches the ball at D in the left fielder's frame of reference?
3. At which spacetime point, labeled A through J, is the third base runner when the left fielder catches the ball at D in the runner's frame of reference after he left third base?
4. In which of the three frames discussed above, namely (1) the umpire's frame of reference, (2) the left fielder's frame of reference, and (3) the runner's frame of reference after he has left third base and is running toward home base, has the runner committed a foul? Explain, referring to the spacetime diagram.
5. If the three observers base their judgments solely on what they see, which of them, if any, would conclude that the runner has committed a foul? Explain, referring to the spacetime diagram.

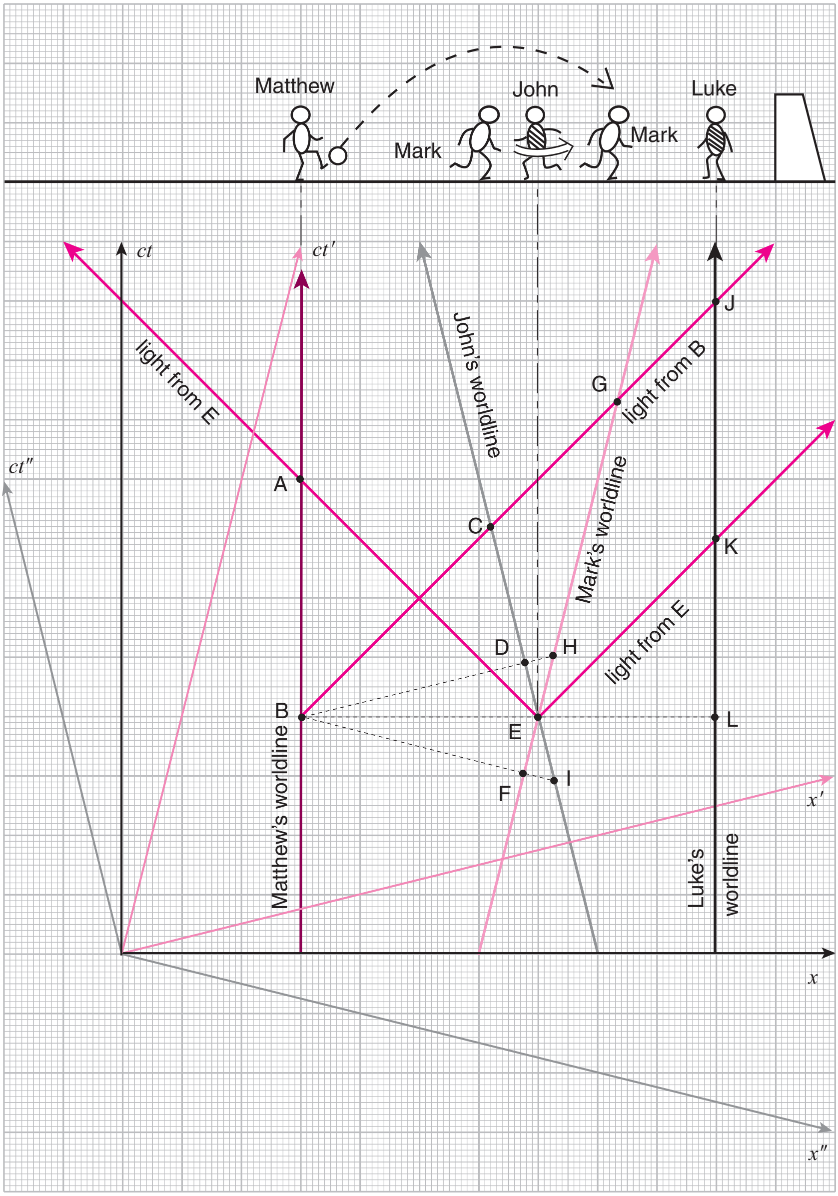

ct
ct′
ct″
x
x′
x″
A
B
C
D
E
F
G
H
I
J
Worldline of runner
light from G
light from D
light from G
Worldline of left fielder
Worldline of umpire

8.3.3 The offside rule in soccer

During a game of soccer, Matthew is initially in possession of the ball while his teammate Mark runs toward the goal being kept by Luke of the opposing team. Matthew kicks the ball toward Mark at point B on the spacetime diagram. Mark passes John of the opposing team and thereby enters the offside[3] position at E.

Answer the following questions:

1. At which spacetime point, labeled A through L, is Mark when the ball is kicked at B in Mark's frame of reference?
2. At which spacetime point, labeled A through L, is Mark when the ball is kicked at B in John's frame of reference?
3. In Mark's frame of reference, did Mark enter the offside position before the ball was kicked or after the ball was kicked? Explain, referring to the spacetime diagram.
4. In John's frame of reference, did Mark enter the offside position before the ball was kicked or after the ball was kicked? Explain, referring to the spacetime diagram.
5. What is the chronological order in which Luke, the goal keeper, sees the two events: the ball being kicked by Matthew, and Mark entering the offside position?



Matthew
Mark
John
Mark
Luke
ct
ct′
ct″
light from E
John's worldline
light from B
G
J
A
C
K
Mark's worldline
D
H
light from E
B
E
L
F
I
Matthew's worldline
Luke's
worldline
x′
x
x″

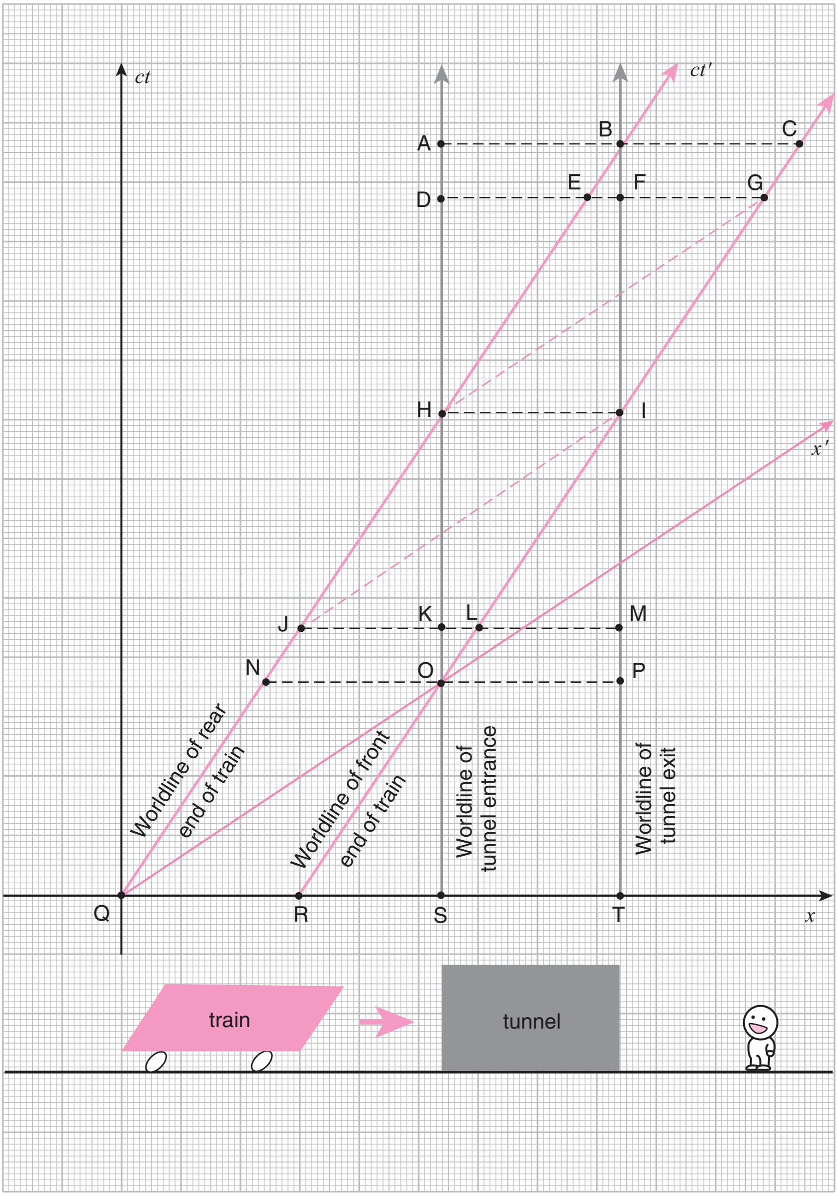
8.4 Lorentz contraction

8.4.1 Train and tunnel

A high speed train speeds through a tunnel at constant velocity. The worldlines of both ends of the train and both ends of the tunnel are shown in the spacetime diagram.

Answer the following questions:

1. At which of the points labeled A through T on the spacetime diagram does the front end of the train emerge from the tunnel?
2. At which point does the rear end of the train enter the tunnel?
3. At which point is the rear end of the train when the front end emerges from the tunnel in the train-frame?
4. At which point is the front end of the train when the rear end of the train enters the tunnel in the train-frame?
5. Does the train fit inside the tunnel in the tunnel-frame? How about in the train-frame?


ct
ct′
x′
x
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
Worldline of rear end of train
Worldline of front end of train
Worldline of tunnel entrance
Worldline of tunnel exit
train
tunnel

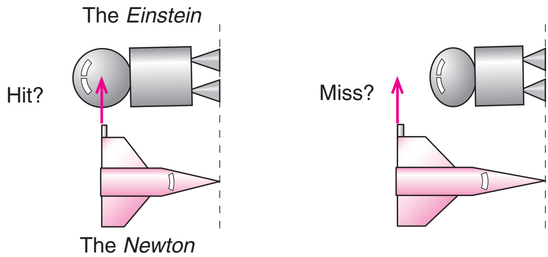
8.4.2 The starship and the enemy space cruiser 1

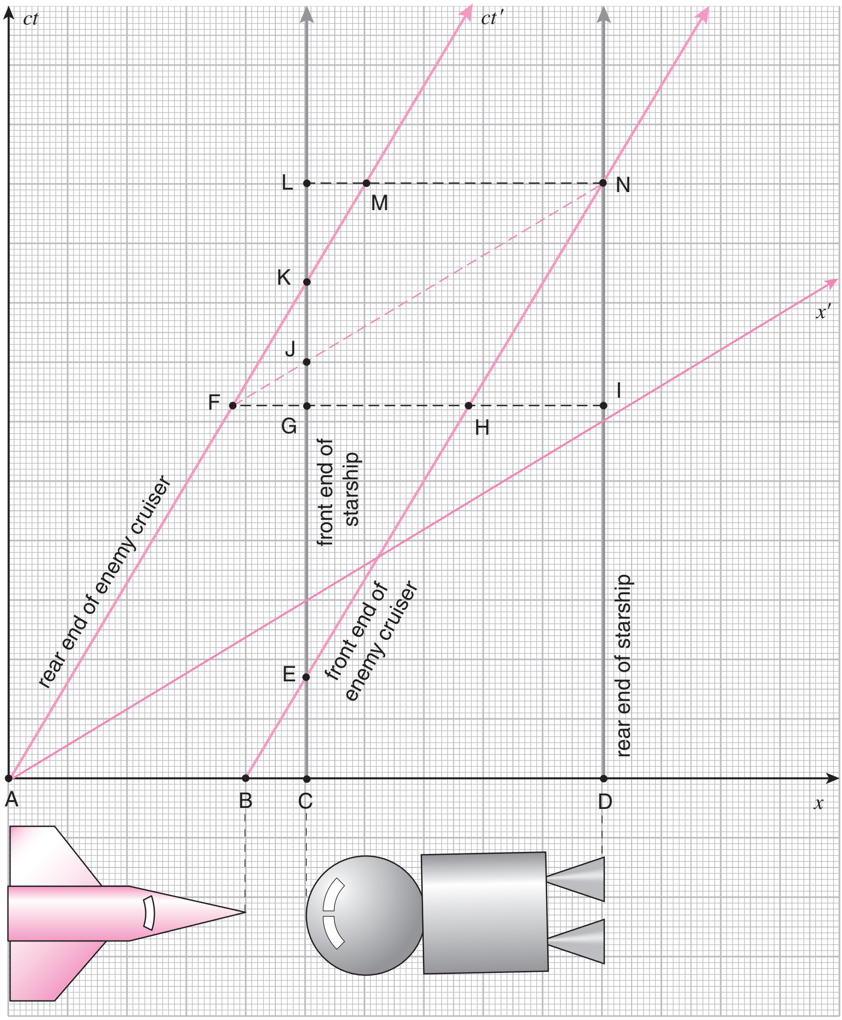
The Imperial space cruiser *Sir Isaac Newton* is equipped with a disruptor cannon mounted on its rear end which can only fire perpendicularly to its direction of motion. A standard Imperial Navy tactic is to fly by an enemy and fire this cannon as the front end of the space cruiser passes the enemy ship's rear end. In the following discussion, assume that the space cruiser flies by so closely that you can neglect the time it takes for the disruptor beam to travel from the cannon to the target.

The starship *Einstein*, which is the same length as the *Newton* when both are at rest, is about to be attacked. The ship's doctor is worried that, due to Lorentz contraction, the *Newton* will be shorter than the *Einstein* and therefore cannot miss. The science officer, on the other hand, insists that since the *Einstein* will be shorter than the *Newton* in the *Newton*'s frame, there is nothing to worry about. You, the ship's captain, must decide which of them is correct and take appropriate defensive measures.

Answer the following questions:

1. At which spacetime point labeled A through N, does the front end of the *Newton* pass by the rear end of the *Einstein*?
2. As the front end of the *Newton* passes by the rear end of the *Einstein*, at which spacetime point is the rear end of the *Newton* in the *Einstein's* frame?
3. As the front end of the *Newton* passes by the rear end of the *Einstein*, at which spacetime point is the front end of the *Einstein* in the *Newton's* frame?
4. At which spacetime point does the disruptor cannon fire?
5. When the disruptor cannon fires, at which spacetime point is the front end of the *Newton* in the *Einstein's* frame?
6. Explain to your ship's doctor and science officer which of them is correct. Refer to the spacetime diagram in your explanation.


The Einstein
Hit?
Miss?
The Newton

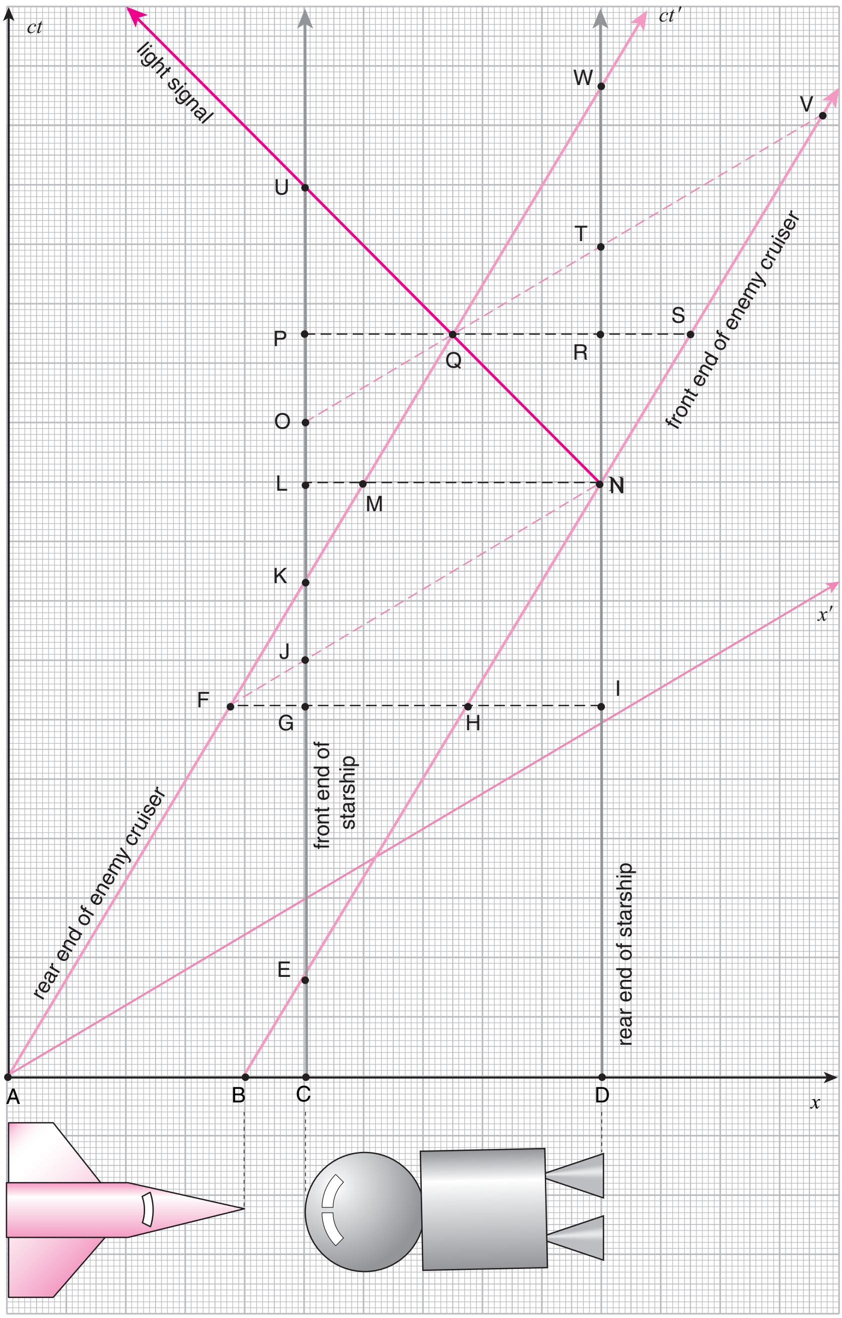

ct
ct′
x′
x
L
M
N
K
J
F
G
H
I
E
A
B
C
D
rear end of enemy cruiser
front end of starship
front end of enemy cruiser
rear end of starship

8.4.3 The starship and the enemy space cruiser 2

The Imperial Navy tactic assumed in the previous problem is actually unrealistic since there is no way for the gunner at the rear end of the *Newton* to know when the *Newton*'s front end has reached the rear end of the *Einstein*. So instead, assume that a light signal is sent from the front end of the *Newton* to its rear when a sensor detects the rear end of the *Einstein* pass by its front end at N, and the cannon fires when it receives this signal.

Answer the following questions:

1. At which spacetime point, labeled A through W, does the cannon fire in this case?
2. Is the *Einstein* hit this time? Explain, referring to the spacetime diagram.


ct
light signal
ct′
W
V
U
T
S
front end of enemy cruiser
P
Q
R
O
L
M
N
K
J
F
G
H
I
x′
front end of starship
rear end of enemy cruiser
E
rear end of starship
A
B
C
D
x

8.4.4 The duel of the space cruisers

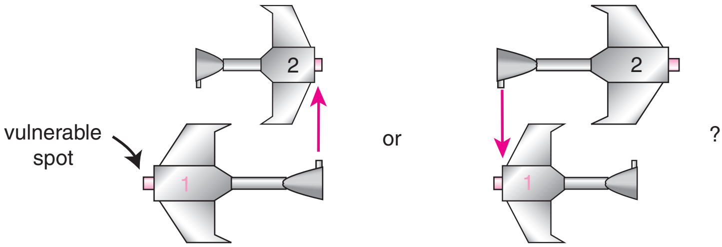
An Imperial space cruiser is equipped with a disruptor cannon mounted on its nose which can only fire in the direction perpendicular to its direction of motion. The cruiser is well protected by shields except for one vulnerable spot on its rear end.

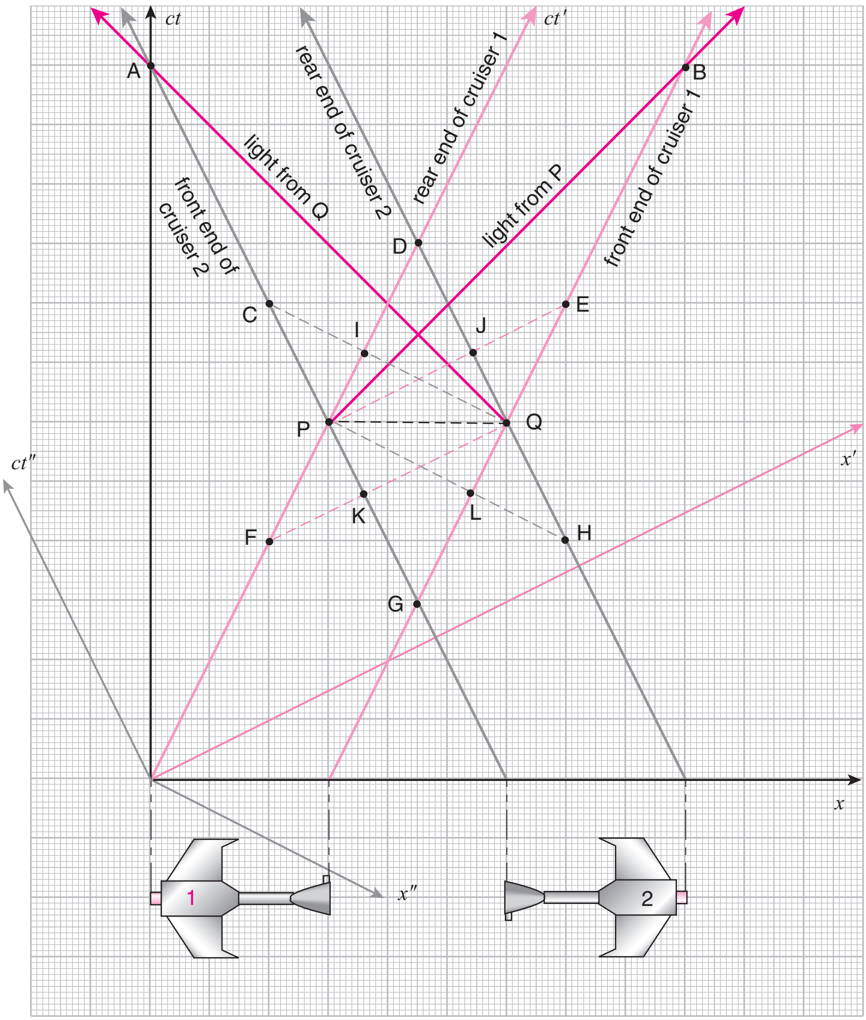
Two of these space cruisers challenge each other to a duel. They fly straight toward each other and fire their cannons as they pass by, trying to hit the vulnerable spot of the opponent. The commander of cruiser 1 thinks that she will win since, due to Lorentz contraction, her opponent's ship will be shorter than hers which will allow her cannon to fire at the opponent's vulnerable spot before the opponent gets a chance to fire his cannon at her vulnerable spot. The commander of cruiser 2 thinks that he will win due to the exact same reason.

The duel takes place and, from a frame fixed to a nearby planet, it is observed that both cruisers are the same length; they fire their respective cannons at the same time, and they are both hit. The worldlines of both cruisers are shown on the spacetime diagram. The cannon of cruiser 1 (coming in from the left) is fired at point Q, and the cannon of cruiser 2 (coming in from the right) is fired at point P.

Answer the following questions:

1. In the frame moving with cruiser 1 (coming in from the left), at which spacetime point, labeled A through L, is the nose of cruiser 2 (coming in from the right) when the cannon of cruiser 1 is fired at Q?
2. In the frame moving with cruiser 1, what is the chronological order in which the two cannons are fired?
3. When cruiser 2 is hit by cruiser 1's cannon at Q, a light signal is sent toward the nose of the cruiser to inform the commander of the damage. At which spacetime point does the signal reach the nose of cruiser 2?
4. What was wrong with the commanders' reasoning that they will win? Explain, referring to the spacetime diagram.


vulnerable spot
2
1
or
2
1
?


ct
ct′
ct″
x
x′
x″
A
B
C
D
E
F
G
H
I
J
K
L
P
Q
light from Q
light from P
rear end of cruiser 2
rear end of cruiser 1
front end of cruiser 2
front end of cruiser 1
1
2

8.4.5 Trains in a tunnel

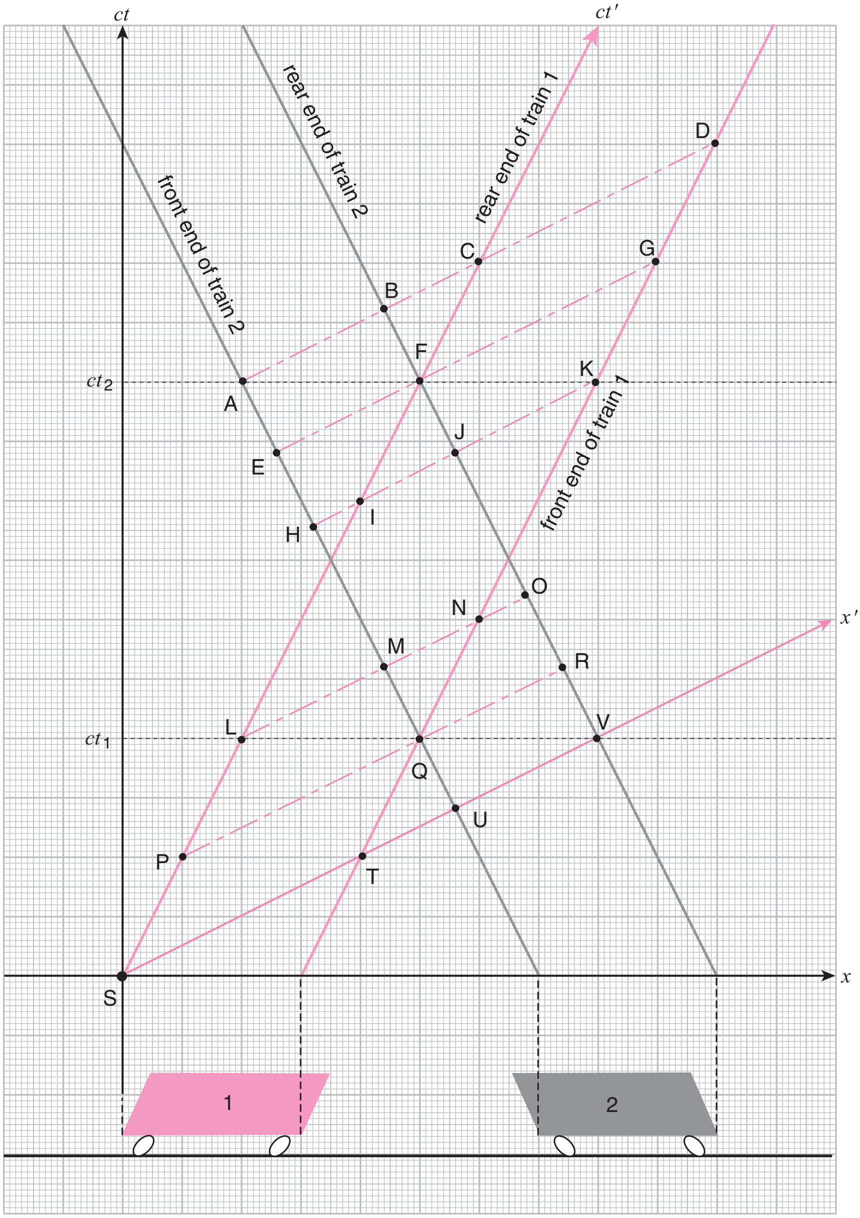
Two trains, 1 and 2, of equal lengths are traveling at the same speed in opposite directions inside a long, dark, and straight tunnel. At time t_1 in the ground-frame of reference, all the ceiling lights in the tunnel turn on simultaneously. Then, at time t_2 in the ground-frame of reference, they all turn off simultaneously.

In the ground-frame of reference, both trains are observed to be the same length (since they are both moving at the same speed) and both are immersed in light for the same amount of time. Therefore, an observer on the ground concludes that both trains are hit by the same number of photons.

However, an observer riding train 1 reasons that since train 1 is longer than train 2 in the train 1 frame of reference (since train 2 is shorter due to Lorentz contraction) train 1 must be hit by more photons than train 2. This seems to contradict the conclusion of the observer on the ground.[4]

Answer the following questions:

1. At which spacetime points, labeled A through V, are the front and rear ends of train 1 when its front end enters the field of light in the train 1 frame of reference?
2. At which spacetime points, labeled A through V, are the front and rear ends of train 1 when its rear end exits the field of light in the train 1 frame of reference?
3. At which spacetime points, labeled A through V, are the front and rear ends of train 2 when its rear end enters the field of light in the train 1 frame of reference?
4. At which spacetime points, labeled A through V, are the front and rear ends of train 2 when its front end exits the field of light in the train 1 frame of reference?
5. What was wrong with the reasoning of the observer on train 1? Explain, referring to the spacetime diagram.


ct
ct′
x′
x
front end of train 2
rear end of train 2
rear end of train 1
front end of train 1
ct_2
ct_1
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
1
2

Solutions to Chapter 8 problems

8.1 Reading the spacetime diagram

8.1.1 Street lamps – solution

1. 1 and 3 simultaneously, then 5, then 4, and then 2.
2. 1, then 3, then 2 and 4 simultaneously, and then 5.
3. 3 and 5 simultaneously, then 1 and 4 simultaneously, and then 2.
4. 1, then 3, then 2 and 4 simultaneously, and then 5.
5. Midway between lamps 1 and 2.

8.1.2 Supernovae – solution

1. B and E simultaneously, then D, then A, then C.
2. E, then D, then B, then A and C simultaneously.
3. B, then A, then D, then E, then C.
4. B, then D, then E, then A and C simultaneously.
5. The chronological order of A and B is the same in all frames of reference. The spacetime points A and B can be connected by an object or signal which travels slower than the speed of light. (In technical terms, A is in the future light-cone of B.) Therefore, B must happen before A in all frames.

8.2 Questions on before and after

8.2.1 The hare and the tortoise 1 – solution

Both the hare and the tortoise think that they won. The tortoise crosses its finish line at A, while the hare crosses its finish line at B. Though A and B are simultaneous in the referee's frame, they are not in the animals' frames. In the tortoise's frame, A is simultaneous with D on the hare's worldline, which is only about two thirds of the way to the hare's finish line. So in the tortoise's frame of reference, it won the race. In the hare's frame, B is simultaneous with C on the tortoise's worldline, which is only about two thirds of the way to the tortoise's finish line. So in the hare's frame of reference, it also won the race.

8.2.2 The hare and the tortoise 2 – solution

1. D 2. C 3. E 4. B 5. A
6. When the hare crosses its finish line at D, in the hare-frame the tortoise is still at E, and when the tortoise crosses its finish line at C, in the tortoise-frame the hare is already at B. So in both animals' frames the hare won.
7. The tortoise does not *see* the hare cross it's finish line until A, which is after C. So if the tortoise bases its conclusion on what it sees, and does not take into account the finite amount of time it takes for light to reach it from the hare's finish line, then it will think that it won.

8.2.3 The hare and the tortoise 3 – solution

1. B 2. G 3. F
4. No, they do not agree. The tortoise starts at B while the hare starts at H. In the tortoise-frame, B is simultaneous with G, which is chronologically later than H. In the hare-frame, H is simultaneous with C which is chronologically later than B. So both animals will think that the other animal started earlier than they did. (The tortoise will not see the hare starting until it is at D, and the hare will not see the tortoise starting until it is at F. So if the animals do not take into account the finite time it takes for light to reach them from their opponent's starting points, they may reach the opposite conclusion.)
5. Yes, they will agree. Both animals cross the finish line at E.

8.2.4 The starship and the supernova – solution

1. E 2. A
3. In the *Einstein*-frame, the supernova explosion S is simultaneous with D and G, which come after E. Therefore, the *Einstein* arrives at the planet *before* the star goes supernova.
4. In the planet frame, the supernova explosion S is simultaneous with F, which comes before E. Therefore, the *Einstein* arrives at the planet *after* the star goes supernova.
5. The light from the supernova explosion S reaches the planet at C, which is after E. So by that time the students would have been picked up by the *Einstein* and well on their way to safety. The light catches up with the *Einstein* at H, which is when the students will finally see what they were escaping from.

8.3 Relativistic sports

8.3.1 Tagging up in baseball 1 – solution

1. F 2. D
3. The third base runner leaves the base at E, which is after F, so he has not committed a foul in the umpire-frame.
4. The third base runner leaves the base at E, which is before D, so he has committed a foul in the left fielder-frame.
5. The umpire sees the third base runner leaving the base H, while he sees the ball being caught at G. H is before G so he will call a foul.

8.3.2 Tagging up in baseball 2 – solution

1. H 2. I 3. E
4. The runner leaves third base at G, which is after H and I but before E. So the runner has NOT committed a foul in the umpire- and left fielder-frames, but HAS committed a foul in the frame he is in while he is running.
5. The umpire will see the runner leave third base at F and see the left fielder catch the ball at C. Since he sees the runner leaving the base before the ball is caught, he will call a foul. The left fielder, on the other hand, sees the runner leave third base at A, well after he caught the ball at D. So the left fielder will not think the runner committed a foul. Finally, the runner sees the ball being caught at B, well after he has left third base at G, so he himself will think that he has committed a foul.

8.3.3 The offside rule in soccer – solution

1. H 2. F
3. In Mark's frame of reference, B is simultaneous with H. Since Mark entered the offside position at E, which is before H, it was before the ball was kicked.
4. In John's frame of reference, B is simultaneous with F. Since Mark entered the offside position at E, which is after F, it was after the ball was kicked.
5. Light from B reaches Luke at J. Light from E reaches Luke at K. Since K happens before I, Luke sees Mark enter the offside position first, and then the ball being kicked by Matthew.

8.4 Lorentz contraction

8.4.1 Train and tunnel – solution

1. I   2. H   3. J   4. G
5. In the tunnel-frame points H and I are simultaneous so the train fits inside the tunnel exactly. But in the train-frame I is simultaneous with J and H is simultaneous with G. So when the front end of the train emerges from the tunnel at I, the rear end is still at J, and when the rear end finally enters the tunnel at H, the front end is already at G, so the train is much longer than the tunnel in the train-frame.

8.4.2 The starship and the enemy space cruiser 1 space – solution

1. N, where the worldlines of the rear end of the *Einstein* and the front end of the *Newton* meet.
2. M   3. J
4. F, since the *Newton* will fire in its own frame of reference.
5. H
6. The science officer is correct and the ship's doctor is wrong. The front end of the *Newton* will pass by the rear end of the *Einstein* at point N. In the *Newton*-frame, N is simultaneous with F and J, so the *Einstein* is extended from J to N while the *Newton* is extended from F to N. So when the *Newton* fires its cannon at F, it will miss. In the *Einstein*-frame, F is simultaneous with G, H, and I, so when the *Newton's* cannon fires at F, the *Newton* is extended from F to H, while the *Einstein* is extended from G to I. Clearly the *Newton* is shorter than the *Einstein* due to Lorentz contraction, but the cannon will miss since the front end of the *Newton*, at H, has not reached the rear end of the *Einstein*, at I, yet.

8.4.3 The starship and the enemy space cruiser 2 – solution

1. Q
2. Yes, it is hit. In the *Newton*-frame Q is simultaneous with O, T, and V, so the *Newton* is extended from Q to V while the *Einstein* is extended from O to T. On the other hand, in the *Einstein*-frame Q is simultaneous with P, R, and S, so the *Newton* is extended from Q

to S while the *Einstein* is extended from P to R. No matter how you look at it, point Q is smack in the middle of the *Einstein*.

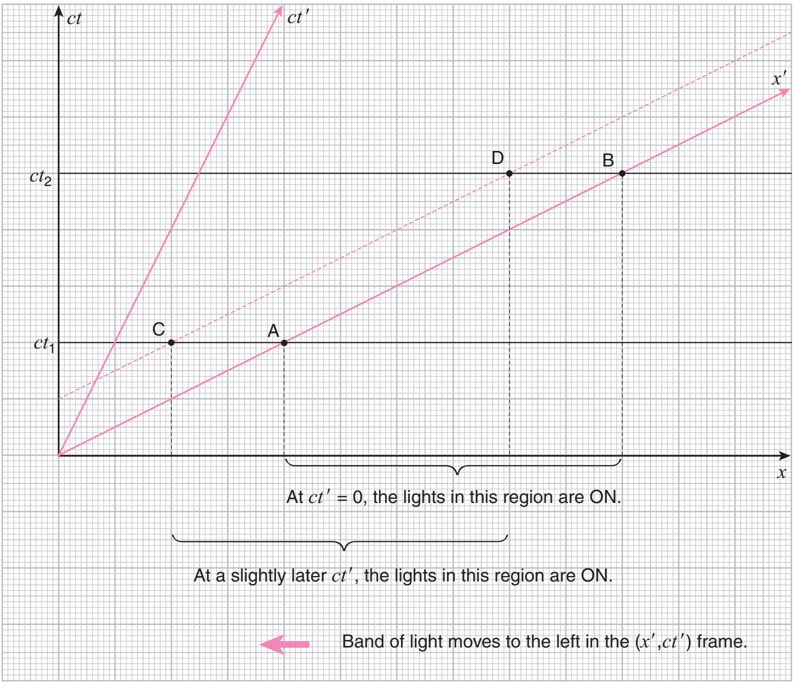
8.4.4 The duel of the space cruisers–solution

1. K
2. In the cruiser 1-frame, Q is simultaneous with K which comes before P. So cruiser 1 fires its cannon first, and cruiser 2 second.
3. A
4. When cruiser 2 is hit at Q, the damage can only propagate forward at or slower than the speed of light. Therefore, the nose of cruiser 2 will not feel the damage until A, giving it plenty of time to fire its cannon back at P. (In technical terms, the spacetime points P and Q are space-like separated so they cannot be causally connected.)

8.4.5 Trains in a tunnel–solution

1. The front end of train 1 enters the light field at Q. Q is simultaneous with P and R in the train 1-frame. Therefore, the front end of train 1 is at Q, and the rear end is at P.
2. The rear end of train 1 exits the light field at F. F is simultaneous with E and G in the train 1-frame. Therefore, the front end of train 1 is at G, and the rear end is at F.
3. The rear end of train 2 enters the light field at V. V is simultaneous with S, T, and U in the train 1-frame. Therefore, the front end of train 2 is at U, and the rear end is at V.
4. The front end of train 2 exits the light field at A. A is simultaneous with B, C, and D in the train 1-frame. Therefore, the front end of train 2 is at A, and the rear end is at B.
5. Even though train 2 is shorter than train 1 in the train 1 frame of reference, as can be seen by comparing the lengths of AB and CD, it spends more time in the light field. The rear end of train 2 starts entering the light field at V, and is completely immersed in light (QR) by the time the front end of train 1 starts entering the light field at Q. The rear end of train 1 exits the light field at F, but at that time train 2 is still completely immersed in light (EF). So the effect of Lorentz contraction is completely cancelled out by the fact that train 2 spends a longer time in the light than train 1.

The important thing to notice is that the lights inside the tunnel do not turn on and off simultaneously in either of the trains' frames. In the train 1-frame, what you observe is a band of light moving from the right to the left. See the spacetime diagram shown below. Since train 1 is moving in the opposite direction as the band of light while train 2 is moving in the same direction, train 1 spends less time in the light than train 2. By the way, the band of light travels at a speed faster than the speed of light, but that does not violate causality since nothing is really moving, nor is any information being transmitted.



Notes

1 To instructors: please feel free to use these problems, or variations of them, in your courses. If you think of any new problems, please email them to the author. I plan to create a separate booklet of relativity problems.

2 A "fly ball" is a ball that is hit high into the air, as opposed to a "ground ball," which is hit downwards toward the ground. In baseball, the runner cannot run while the fly ball is in the air and must wait until it is either caught (in which case the batter is out) or it bounces on the ground. Otherwise, an obvious tactic would be for the batter to hit as high a fly ball as possible and have the runner run while the opposing team can do absolutely nothing about it.

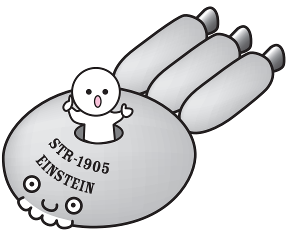
3 In soccer, one cannot pass a ball to a teammate unless there is at least one other member of the opposing team between your teammate and the opposing team's goal keeper. When there is none, your teammate is said to be in the "offside" position. The point of this rule is to give the goal keeper a fair chance to defend the goal. It is, however, permissible to send out a pass into the offside zone if your teammate only enters the offside zone "after" the pass is kicked.

4 I thank David Seppala for informing me of this problem.

9
Quantitative problems

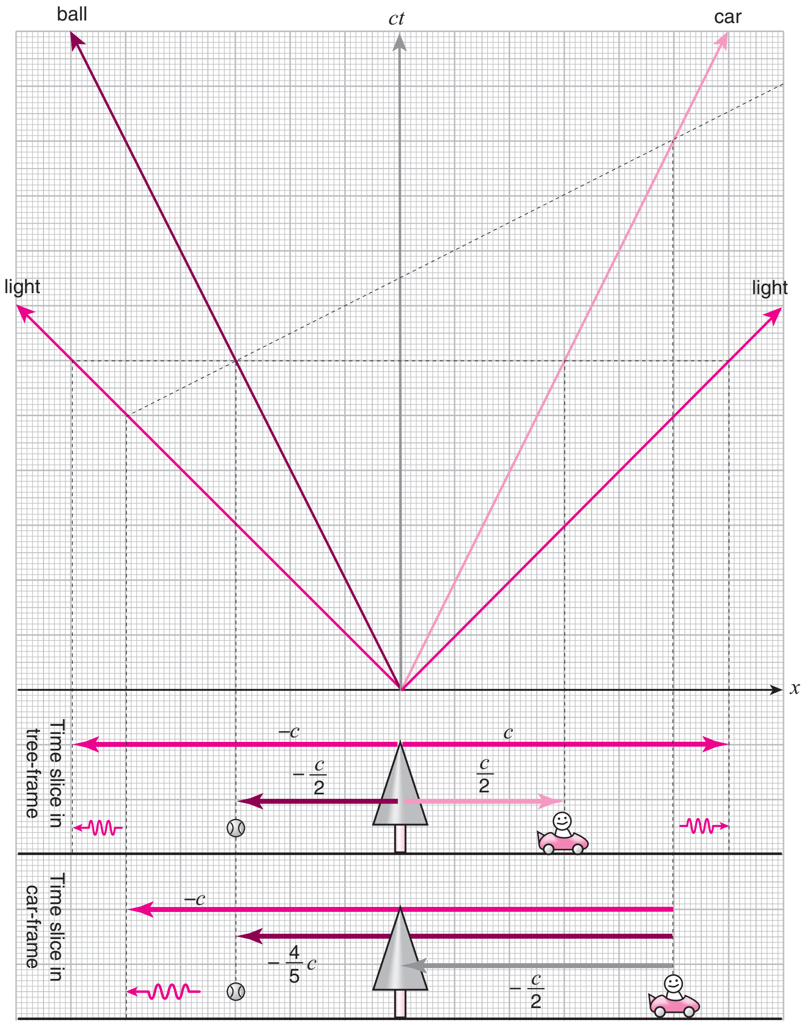
9.1 Addition of velocities

Solve the following problems pictorially using spacetime diagrams. (Do not resort to the equation provided in the endnotes of Chapter 4.)



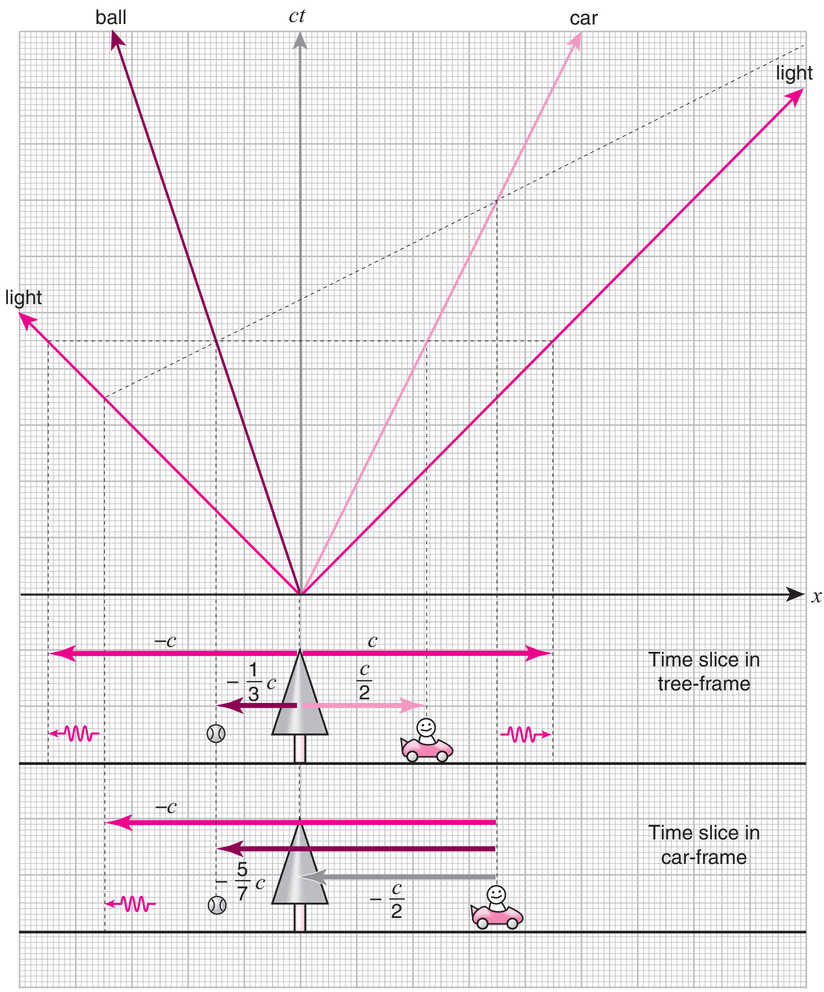
1. A tree is at rest on the ground, and a car is traveling to the right at speed $\frac{1}{2}c$. If a ball is traveling to the left at speed $\frac{1}{2}c$ in the tree-frame, what is its speed in the car-frame?

   Solution: See figure.



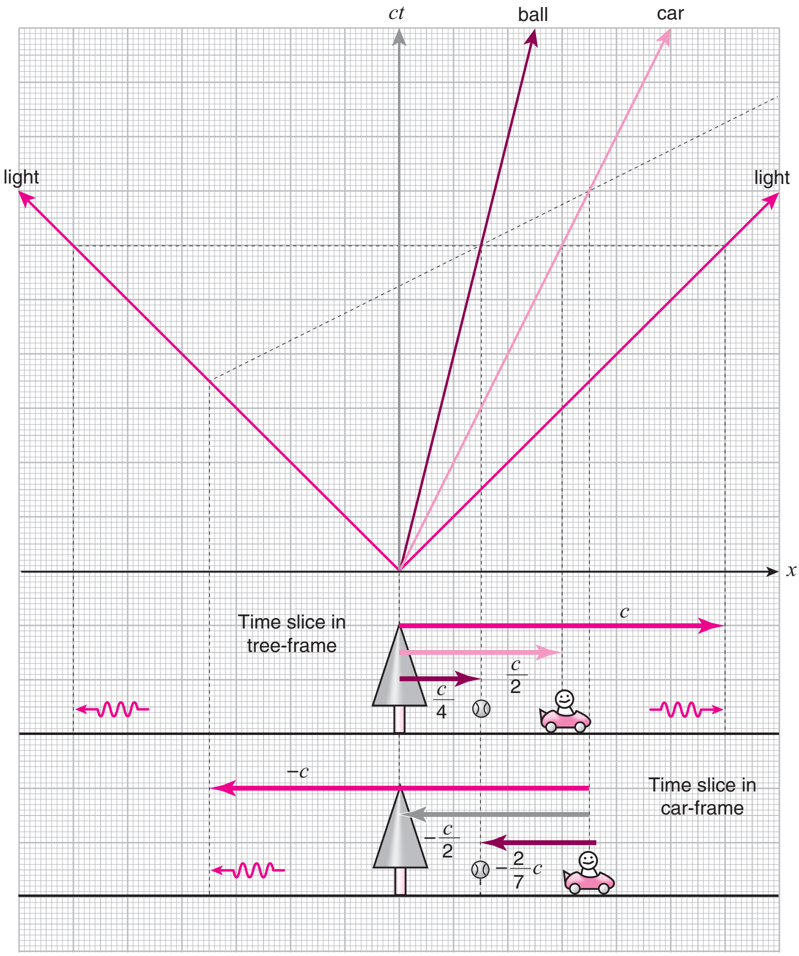
2. A tree is at rest on the ground, and a car is traveling to the right at speed $\frac{1}{2}c$. If a ball is traveling to the left at speed $\frac{1}{3}c$ in the tree-frame, what is its speed in the car-frame?

   Solution: See figure.



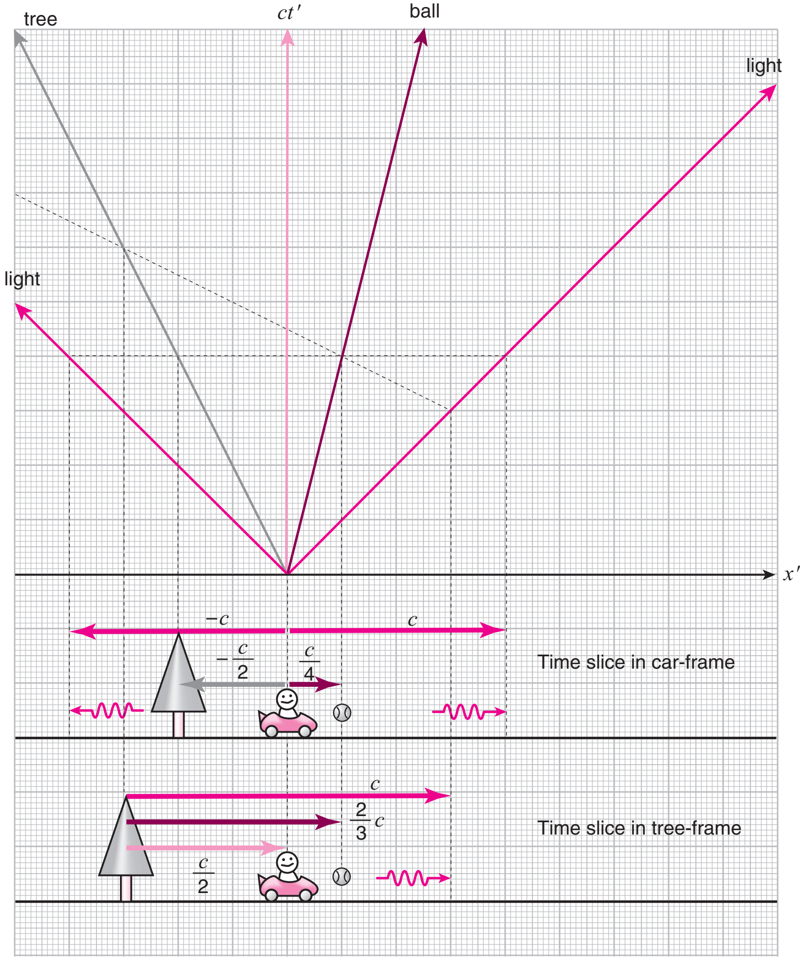
3. A tree is at rest on the ground, and a car is traveling to the right at speed $\frac{1}{2}c$. If a ball is traveling to the right at speed $\frac{1}{4}c$ in the tree-frame, what is its speed in the car-frame?

Solution: See figure.



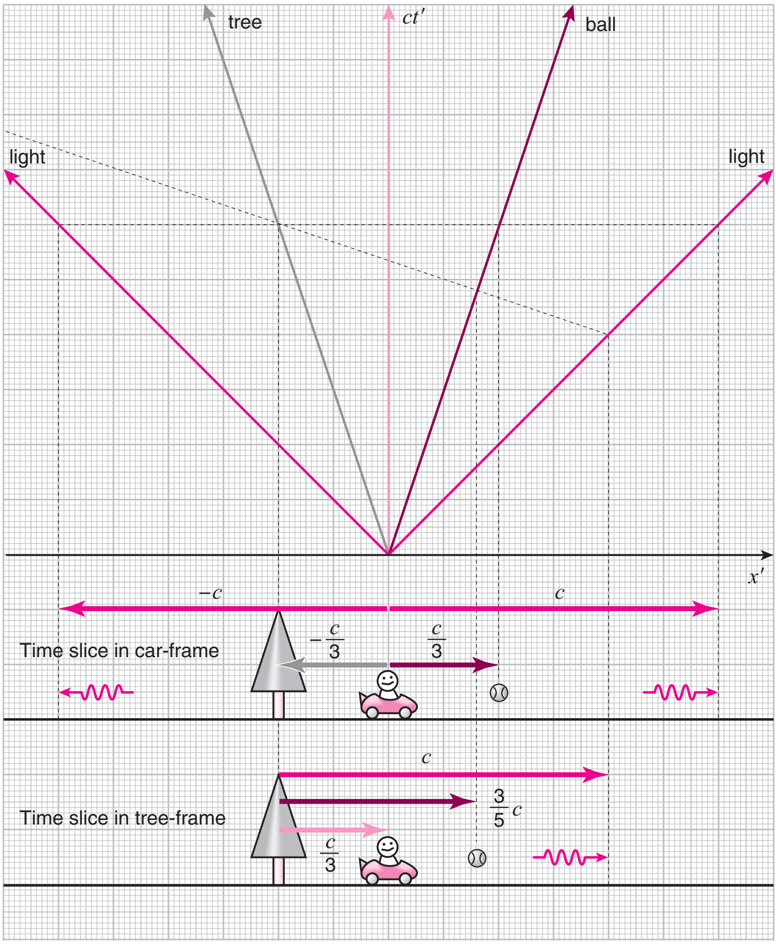
4. A tree is at rest on the ground, and a car is traveling to the right at speed $\frac{1}{2}c$. If a ball is traveling to the right at speed $\frac{1}{4}c$ in the car-frame, what is its speed in the tree-frame?

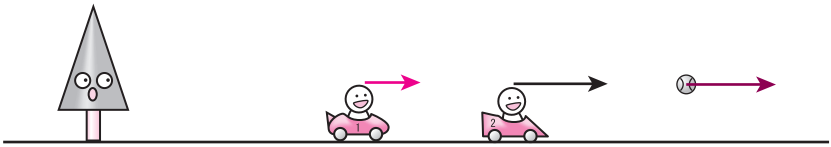
   Solution: See figure.



5. A tree is at rest on the ground, and a car is traveling to the right at speed $\frac{1}{3}c$. If a ball is traveling to the right at speed $\frac{1}{3}c$ in the car-frame, what is its speed in the tree-frame?

   Solution: See figure.




1
2

PART III

Dynamics: Relativity with a few equations

10

The world's most famous equation

In Part III, we will discuss the world's most famous equation[1]

$$E = mc^2$$

The symbol E in this equation represents the *energy* of an object at rest, m represents the object's *mass*, and c is the speed of light in vacuum as before. Obviously, in order to understand what this equation means, we must first understand what *energy* and *mass* are. And for that, we must first understand Newtonian dynamics. So as in Part I, I will first discuss the Galilei–Newton theory before going on to the Einsteinian theory. In the following, I cannot avoid using equations altogether, since $E = mc^2$ itself is an equation, but I will continue to use drawings as much as possible.

Notes

1 This equation did not appear in Einstein's first paper on relativity, which we referred to at the beginning of Part I, namely the one titled "On the electrodynamics of moving bodies." Instead, it appeared in a short paper submitted to the journal *Annalen der Physik* several months later titled "Does the inertia of a body depend upon on its energy-content?" ("Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?" in the original German). English translations for both papers are included in [4].

11
The problem

Recall the questions about motion that I listed as the "more advanced" ones in section 2.1:

Q4+Q5: Is the velocity changing with time?
Q6: If the velocity is changing, what is causing it?

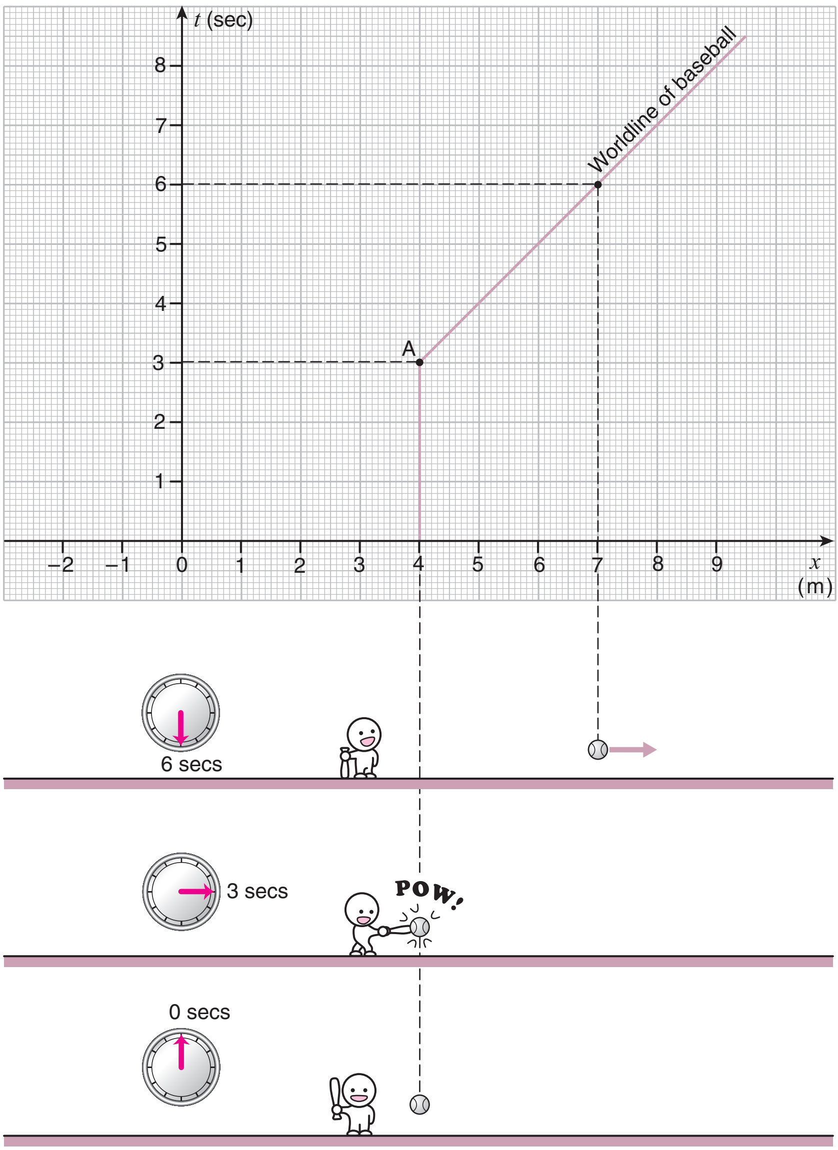
Let us ask these questions of the motion of a baseball whose worldline is shown here on the spacetime diagram. The worldline has a kink at point A, the time-coordinate of which is $t = 3$ seconds, so we can tell that the velocity of the baseball changed at $t = 3$ seconds. The baseball has been hit by a bat at this point. The worldline before A is vertical, so the baseball was at rest before being hit. From the slope of the worldline after A, we can tell that the velocity of the baseball after being hit was +1 meters per second. So the answers to the above questions in this case are:

A4+A5: The velocity of the baseball changed instantaneously at $t = 3$ seconds from 0 meters per second to +1 meters per second.
A6: The velocity of the baseball changed because it was hit by a bat at $t = 3$ seconds.

These answers further bring to mind the following questions:

Q7: If the baseball is given the "same impact" as it received from the bat at $t = 3$ seconds again, what will its velocity be?
Q8: If the "same impact" is given to a different object at rest, say something lighter like a ping-pong ball, or something heavier like a bowling ball, what will the object's velocity be?

Let us first look at the answers provided by Newtonian dynamics.


t (sec)
Worldline of baseball
A
8
7
6
5
4
3
2
1
−2
−1
0
1
2
3
4
5
6
7
8
9
x
(m)
6 secs
3 secs
POW!
0 secs

12

Newtonian dynamics

12.1 The mass–momentum vector

Situations that can be addressed within Newtonian dynamics are cases in which the velocities of objects are much slower than the speed of light c, and the Galilei transformation suffices as the transformation from one inertial frame to another.

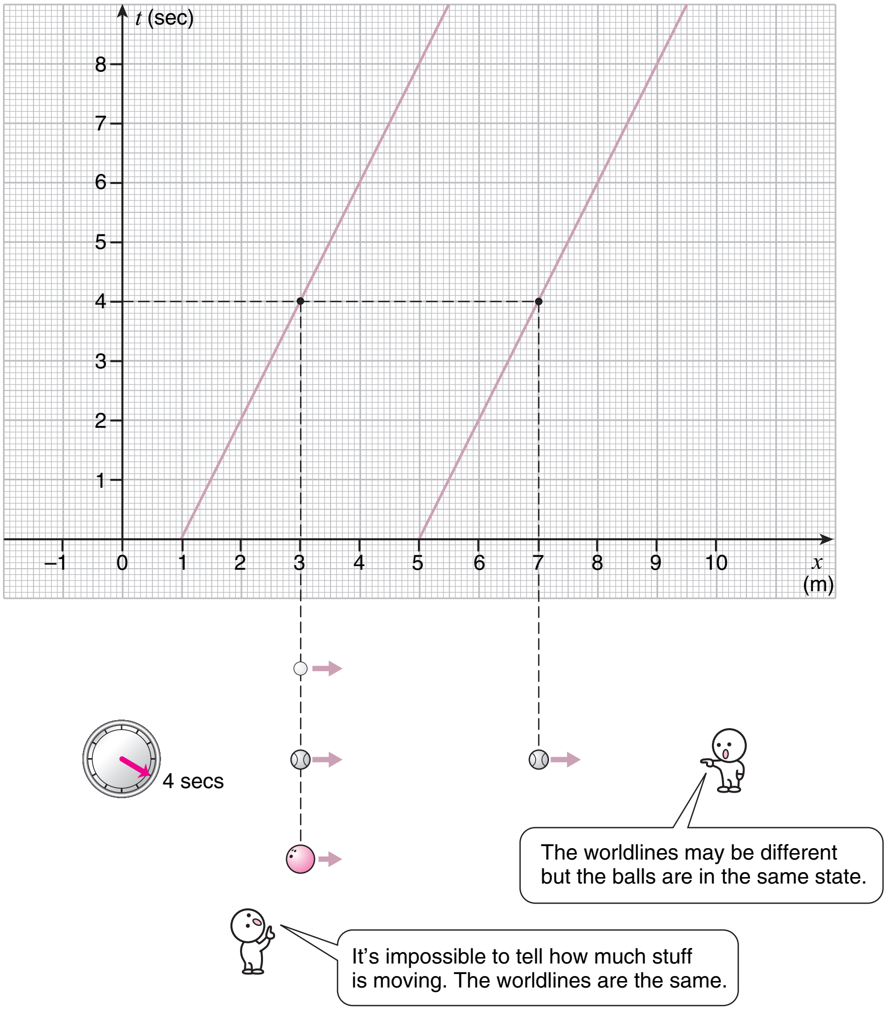
Now, we would like to answer the questions posed in the previous section using diagrams. But for that we must be able to represent pictorially what we mean by the term "same impact." In the current case, the "impact" we are talking about is that which accelerates the baseball from "a state in which it is at rest" to "a state in which it is traveling at $+1$ meters per second." But this in turn means that we must first be able to represent pictorially the "state of the baseball moving at velocity v" for generic velocities v.

So how can we do this? Of course, the motion of any object is described by its worldline on the spacetime diagram, and the object's velocity is encoded in the slope of the worldline. But there are two reasons why the worldline is not an appropriate representation of the "state of motion" of an object:

1. Depending on whether the object is a baseball, a ping-pong ball, or a bowling ball, the amount of "stuff" that is moving is different, but the worldline does not give you that information.
2. No matter where the object is spatially, the velocity of the object will be the same as long as the slope of its worldline is the same. However, worldlines distinguish the spatial location of the object also.

In other words, to specify the "state of motion" of an object, we need to specify not only its velocity but also how much "stuff" is contained in it (this is called the *mass* of the object), while where the object is located spatially is redundant information.

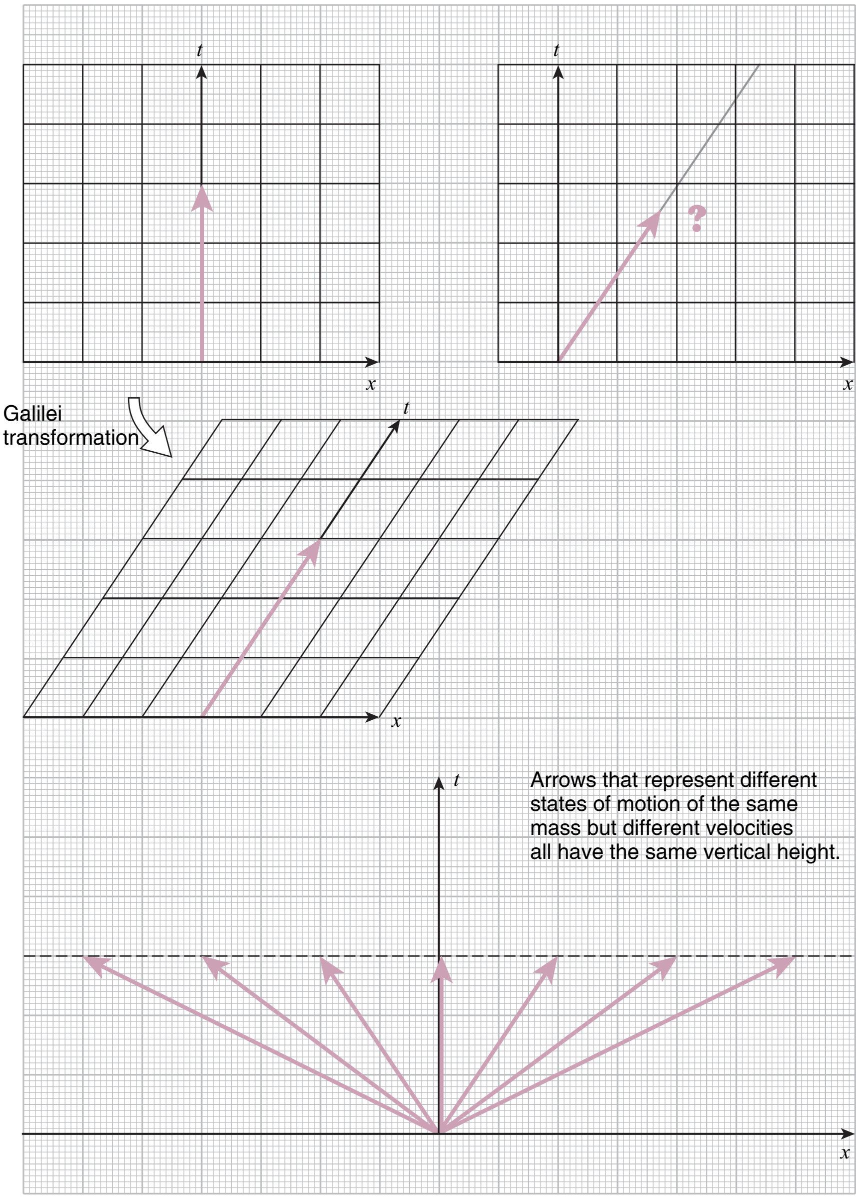
The reason why we need to specify the *mass* of the object here is because the *inertia* of the object will depend on what its *mass* is, and the change in its velocity when the "same impact" is applied will be different. Just imagine hitting a baseball, a ping-pong ball, and a bowling ball with a bat.



Let us begin by thinking about how to represent pictorially the "state in which the object is at rest." Since it does not matter where the object is at rest as long as it is at rest, consider it to be at rest at the spatial origin. Then, the worldline of the object will overlap with the time-axis of the spacetime diagram. Now, let's draw an arrow along this worldline starting from the spacetime origin, and let the length of the arrow be proportional to the *mass* of the object. If the mass of the object is doubled, we also double the length of this arrow. If the mass of the object is halved, we also halve the length of the arrow. This will allow us to encode not just the fact that the object is at rest, but also how much *mass* is at rest. Of course, we have to specify the correspondence between the length of the arrow and the object's mass, but before we do that, let's think about representing the "state of motion" in which the object is moving.

When the object is moving at velocity v, its worldline is tilted. As in the case where it is at rest, it does not matter where the object is spatially, so let this worldline go through the spacetime origin. We would like to represent the "state of motion" of this object with an arrow along this worldline starting at the spacetime origin and with a length proportional to the mass of the object. (See the figure on the opposite page, upper-right.) Now, if the arrow which represents the same object at rest is given by that shown in the upper-left figure, what should the length of this arrow which represents it in motion be?

You may think that the length of the arrow should stay the same as when the object was at rest, but this is incorrect. This is because the "state in which the object is moving with velocity v" must be the same as the "state in which the object is at rest" observed from a frame moving at velocity $-v$ relative to the first. If we Galilei transform the spacetime diagram with the arrow representing the object at rest to such a frame, we obtain the figure shown on the opposite page, middle-left. As you can see, the vertical height of the arrow, and not its length along the object's worldline, must stay the same as when the object was at rest. This tells us that the "states of motion" of objects with the same *mass* but different velocities should be represented by arrows with the same vertical height as shown in the bottom figure. We will call the vertical height of the arrow from the spacetime origin its *time-component*.


t
x
t
?
x
Galilei
transformation
t
x
t
Arrows that represent different
states of motion of the same
mass but different velocities
all have the same vertical height.
x

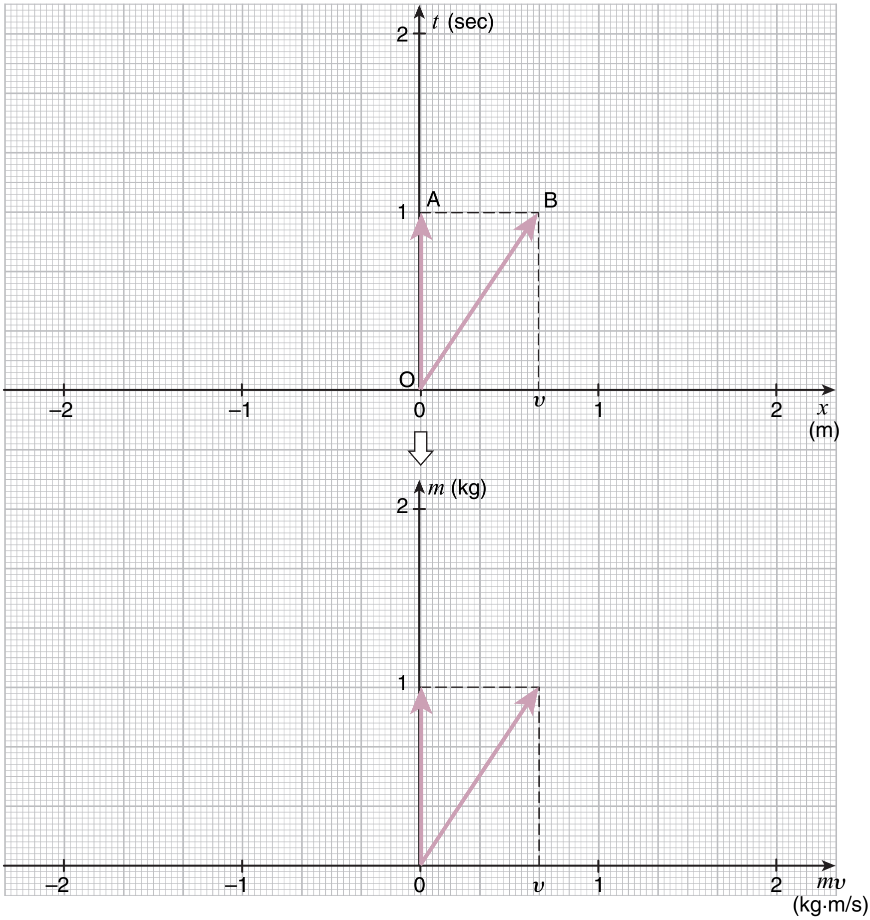
We must now specify how to assign a length to the arrow for a given *mass*. Now, the SI unit for *mass* is *kilogram* (kg).[1] If you live in a country other than the USA, kilograms may seem like a unit of *weight* to you, but *weight* and *mass* are completely different things.[2] *Weight* refers to the strength of gravity acting on the object, so it will be different depending on whether it is measured on the Earth, or on the Moon, or on Mars. On the other hand, *mass* refers to the "amount of stuff" that the object contains, so it will not change depending on the locale. However, as long as we measure the *weight* of an object on the surface of one particular astronomical object, say the Earth, it will be proportional to the object's *mass*,[3] and this allows us to use the same unit for both. That is, if an object's *weight on the surface of the Earth* is 1 kilogram, then its *mass* is also 1 kilogram.[4]

The length of the arrow must be proportional to the *mass* of the object. So let's just assign a time-component of 1 second to the arrow representing the "state of motion" of an object with a mass of 1 kg. Then, the "state of motion" of the 1 kg object at rest will be represented by the arrow OA on the upper diagram. The "state of motion" of the same object moving at velocitiy v meters per second will be represented by the arrow OB. The time-component of OB is 1 second, as was the case with OA, but its space-component, that is, its horizonal extent, is equal to v meters. This is because the object will move by v meters in 1 second.

If the *mass* of the object is m kg,[5] then the length of the arrow should be m times that of the 1 kg case. So its time-component will be m seconds, and its space-component will be mv meters. Now, since the original units of m is kg, and that of mv is $\mathrm{kg} \cdot \mathrm{m/s}$ (read *kilogram meters per second*), converting the units and drawing the corresponding arrow into the spacetime diagram can be a source of confusion and error. So let's just lift the arrows off of the spacetime diagram and place them on a separate graph with kg notched onto the vertical axis, and $\mathrm{kg} \cdot \mathrm{m/s}$ notched onto the horizontal axis, as shown on the lower figure. This graph represents the "space" of all possible "states of motion."

To summarize, the "state of motion" of an object of mass m (kg) and velocity v (m/s) will be represented by an arrow whose time-component is m (kg) and its space-component is mv ($\mathrm{kg} \cdot \mathrm{m/s}$). The quantity mv, that is, the product of the mass m and the velocity v, is known as the *momentum* of the object. As the name implies, the *momentum* quantifies what might be called the "tenacity of the motion" in the spatial direction. The larger it is, the more difficult it will be to bring the object to a complete stop. On the other hand, as will see later, the mass m

quantifies the *inertia* of the object and can be though of as the "tenacity of the motion" in the time direction. We will call this arrow which represents pictorially the "state of motion" of the object the *mass–momentum vector*.[6]



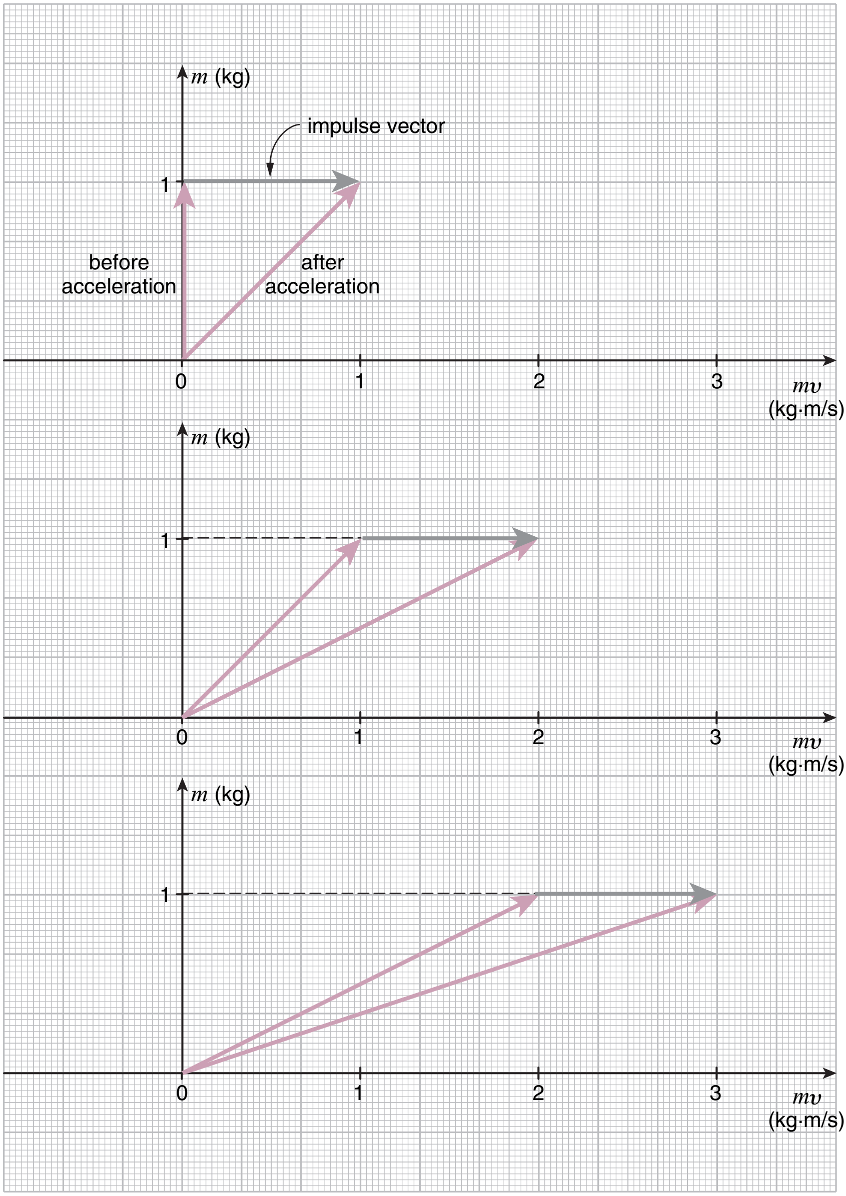
12.2 The impulse vector

Returning to the baseball problem, let's assume for the sake of simplicity that the mass of the baseball is 1 kg. (A bit heavy for a baseball, perhaps.) This baseball was accelerated from the state in which it was at rest to the state in which it is moving at $+1\,\mathrm{m/s}$, so the *mass–momentum vectors* that represent its states of motion **before** and **after** the bat-impact are as shown in the top figure. Now, draw an arrow connecting the tip of the *mass–momentum vector* **before** the impact to the tip of the *mass–momentum vector* **after** the impact, as shown, and let's use it to represent the effect of the impact of the bat. We will call this arrow the *impulse vector*.

Define the *sum* of two vectors to be the vector you get if you attach the starting-point of the second vector onto the tip of the first vector, and then connect the starting-point of the first vector and the tip of the second vector. Using this definition, we can say that the *mass–momentum vector* **after** the impact is the *sum* of the *mass–momentum vector* **before** the impact and the *impulse vector*. In other words, *adding* the *impulse vector* onto the *mass–momentum vector* **before** the impact will give us the *mass–momentum vector* **after** the impact.

If the change in the velocity of the baseball due to the bat-impact is described by this addition of the *impulse vector* onto the *mass–momentum vector*, then answering the questions Q7 and Q8 posed in Chapter 11 is an easy matter. First Q7, namely, the question of what the velocity of the baseball will be if the "same impact" is applied to it a second time: if we add the *impulse vector* which represents the impact in question (in the current case, it is an *impulse vector* with space-component $+1\,\mathrm{kg}\cdot\mathrm{m/s}$) to the *mass–momentum vector* which represents the state of the baseball traveling at $+1\,\mathrm{m/s}$, the sum will be the *mass–momentum vector* shown in the middle figure. And from the slope of this new vector, we can read off the velocity of the baseball after the second impact to be $+2\,\mathrm{m/s}$. This is as expected since the process should be the same as observing the baseball at rest being accelerated to $+1\,\mathrm{m/s}$ from a frame moving at velocity $-1\,\mathrm{m/s}$ relative to the first.

If we continue to apply the "same impact" to the baseball repeatedly, the effect will be described by the repeated addition of the *impulse vector* onto the baseball's *mass–momentum vector*, and each addition will increase the velocity of the baseball by $+1\,\mathrm{m/s}$. See the bottom figure.


m (kg)
impulse vector
1
before
acceleration
after
acceleration
0
1
2
3
mv
(kg·m/s)
m (kg)
1
0
1
2
3
mv
(kg·m/s)
m (kg)
1
0
1
2
3
mv
(kg·m/s)

12.3 Inertial mass

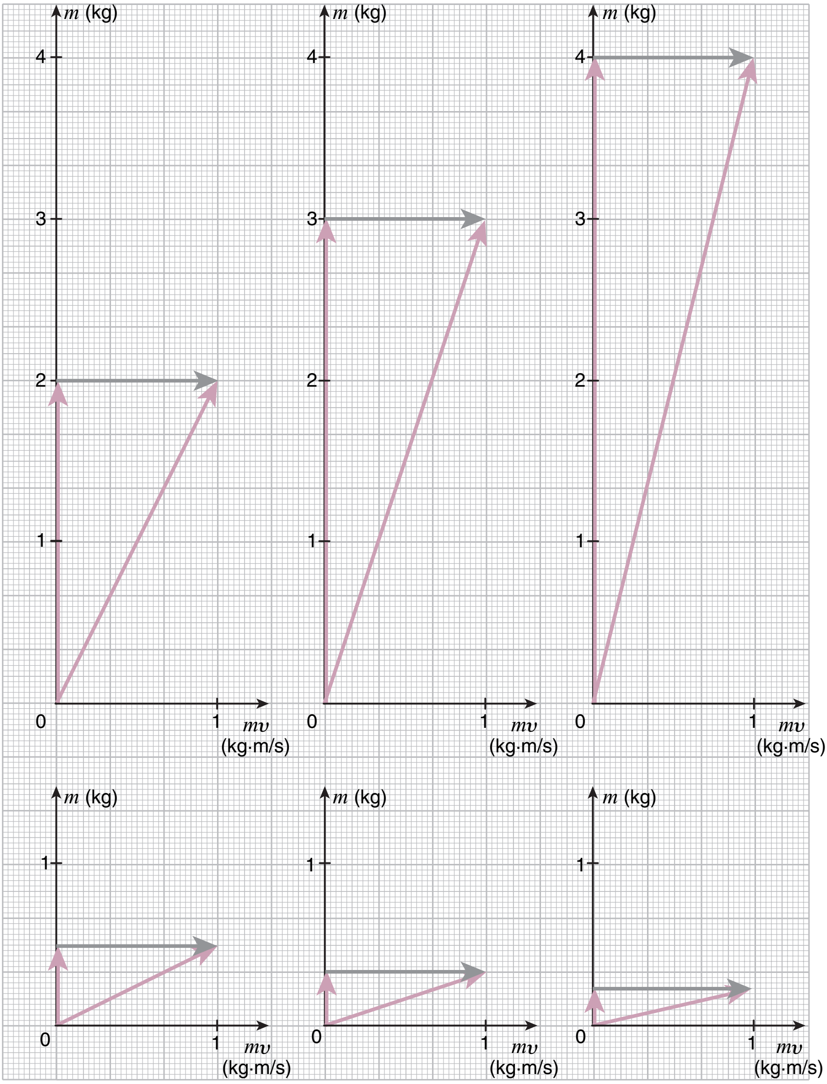
Next Q8, namely, the question of what would happen if the "same impact" is applied to a different object at rest with a different mass. Let's consider a bowling ball with mass 2 kg. (A bit light for a bowling ball, perhaps.) Its initial state of motion at rest will be described by a vertical mass–momentum vector, **twice as long** as that for the baseball, as shown in the upper-left figure. If we add the *impulse vector* with space-component equal to $+1\,\mathrm{kg}\cdot\mathrm{m/s}$ to it, the resulting *mass–momentum vector* will be that also shown in the upper-left figure, and we can read off the velocity of the bowling ball after impact to be $+0.5\,\mathrm{m/s}$. Because the mass of the bowling ball was **double** that of the baseball, the "same impact" only accelerated the bowling ball by **half** as much as the baseball.

Similarly, if the mass had been **three times** that of the baseball, the change in velocity would have been **one-third**, and if the mass had been **four times** that of the baseball, the change in velocity would have been **one-fourth**. In general, if the mass were n times that of the baseball, the change in velocity due to the "same impact" would have been $1/n$. (See figures above-center and above-right.)

Next, let's consider an object like a ping-pong ball which is lighter than the baseball, say with mass equal to 0.5 kg. (A bit heavy for a ping-pong ball, perhaps.) This time, the *mass–momentum vector* which represents its initial state is **half as long** as that for the baseball as shown in the lower-left figure. Adding the *impulse vector* with space-component equal to $+1\,\mathrm{kg}\cdot\mathrm{m/s}$ to it, the resulting *mass–momentum vector* will be that also shown in the lower-left figure, and we can read off the velocity of the ping-pong ball after impact to be $+2\,\mathrm{m/s}$. Because the mass of the ping-pong ball was **half** that of the baseball, the "same impact" has accelerated the ping-pong ball **twice** as much as the baseball.

Similarly, if the mass had been **one-third** that of the baseball, the change in velocity would have been **three times** that of the baseball, and if the mass had been **one-fourth** that of the baseball, the change in velocity would have been **four times** that of the baseball. In general, if the mass were $1/n$ times that of the baseball, the change in velocity due to the "same impact" would be n times that of the baseball. (See figures bottom-center and bottom-right.)

As you can see, the larger the mass of the object is, the longer the time-component of its *mass-momentum vector* will be. And the longer the time-component, the smaller the amount of deflection from the vertical


m (kg)
4
3
2
1
0
1
mv
(kg·m/s)
m (kg)
4
3
2
1
0
1
mv
(kg·m/s)
m (kg)
4
3
2
1
0
1
mv
(kg·m/s)
m (kg)
1
0
1
mv
(kg·m/s)
m (kg)
1
0
1
mv
(kg·m/s)
m (kg)
1
0
1
mv
(kg·m/s)

that the *mass–momentum vector* undergoes when its space-component is increased by the addition of the same *impulse vector*. Therefore, the time-component of the *mass–momentum vector*, namely, the *mass* of the object, quantifies how much *inertia* the object has. The larger the *mass*, the more difficult it is to change the object's velocity. Another way to think about this is to consider the *mass* as representing the "tenacity of the motion" in the time-direction, just like the *momentum* represents the "tenacity of the motion" in the space-direction. Both quantify the tendency of the motion to continue in their respective directions in spacetime.

12.4 Newton's Second Law

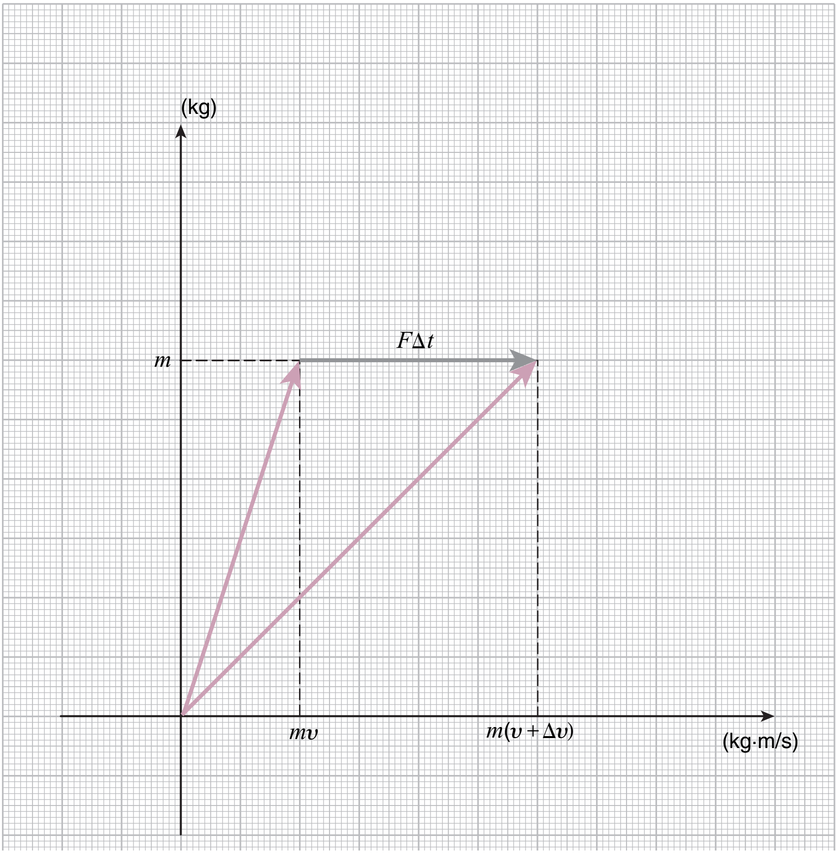
At this point, let me express in equations what we have discussed so far using drawings. We have seen that the change in an object's *mass–momentum vector* due to an impact with another object is given by the *impulse vector*. But since the *impulse vector* only has a space-component, only the space-component of the *mass–momentum vector*, namely the *momentum* will change. And since the *momentum* is equal to the product of the *mass* and the *velocity*, and since the *mass* of the object does not change, the change in the *momentum* is equal to the product of the *mass* and the change in the *velocity*. So the equation which expresses the same thing as our drawings is

$$\text{impulse} \;=\; \text{momentum change} \;=\; \text{mass} \times \text{velocity change}\ .$$

Using the symbols commonly used in physics textbooks, this becomes

$$F\Delta t \;=\; \Delta p \;=\; m\,\Delta v\ .$$

Here, m is the *mass* of the object, Δt is the infinitesimal time interval during which the impulse acts on the object, Δp is the change in the *momentum*, and Δv is change in the *velocity*. The symbol F represents the "*force*," which is the amount of *impulse* applied to the object per unit time, so it is equal to the *impulse* divided by the time Δt. Consequently, the *impulse* is given by the product $F\Delta t$. The reason for writing the *impulse* this way is because no one ever invented a separate symbol for it.[7] The law expressed by the above equation is known as *Newton's Second Law*.[8]

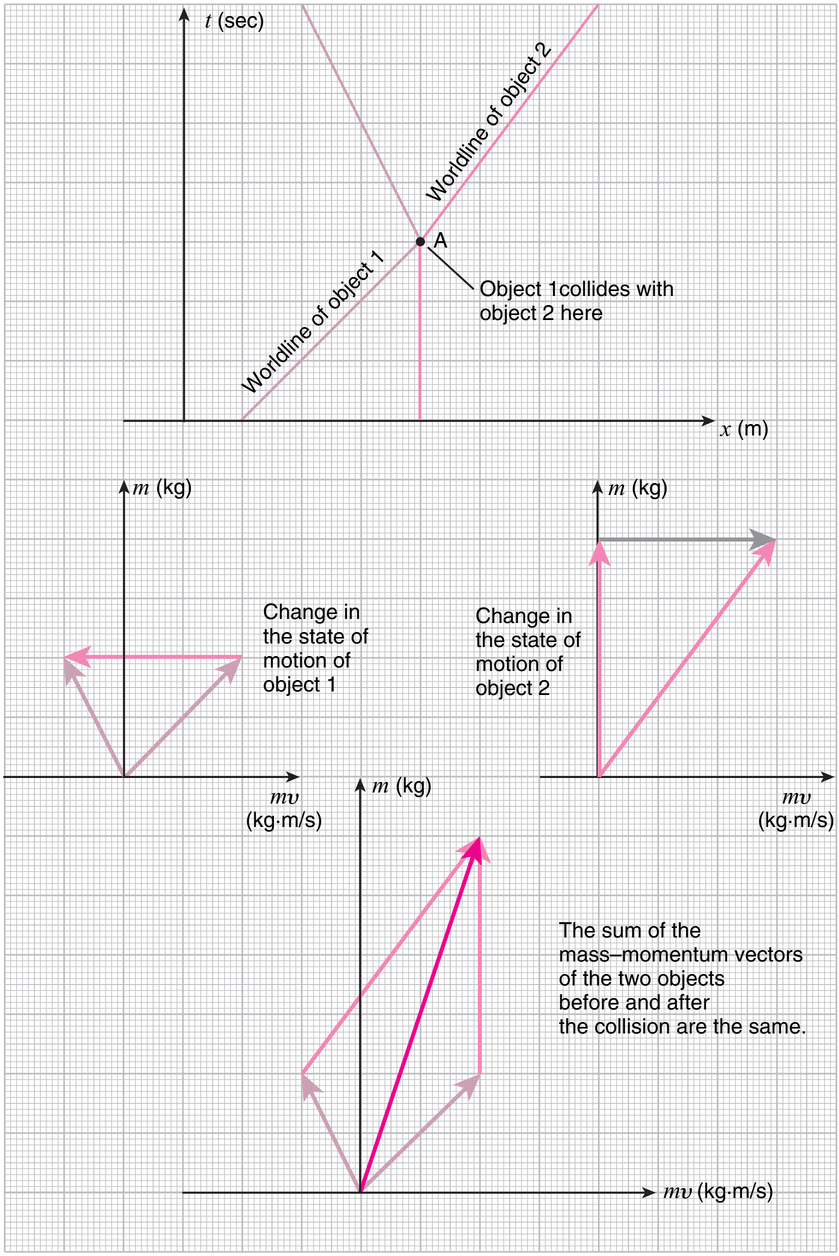

(kg)
$F\Delta t$
m
mv
$m(v + \Delta v)$
(kg·m/s)

12.5 Newton's Third Law and the conservation of mass–momentum

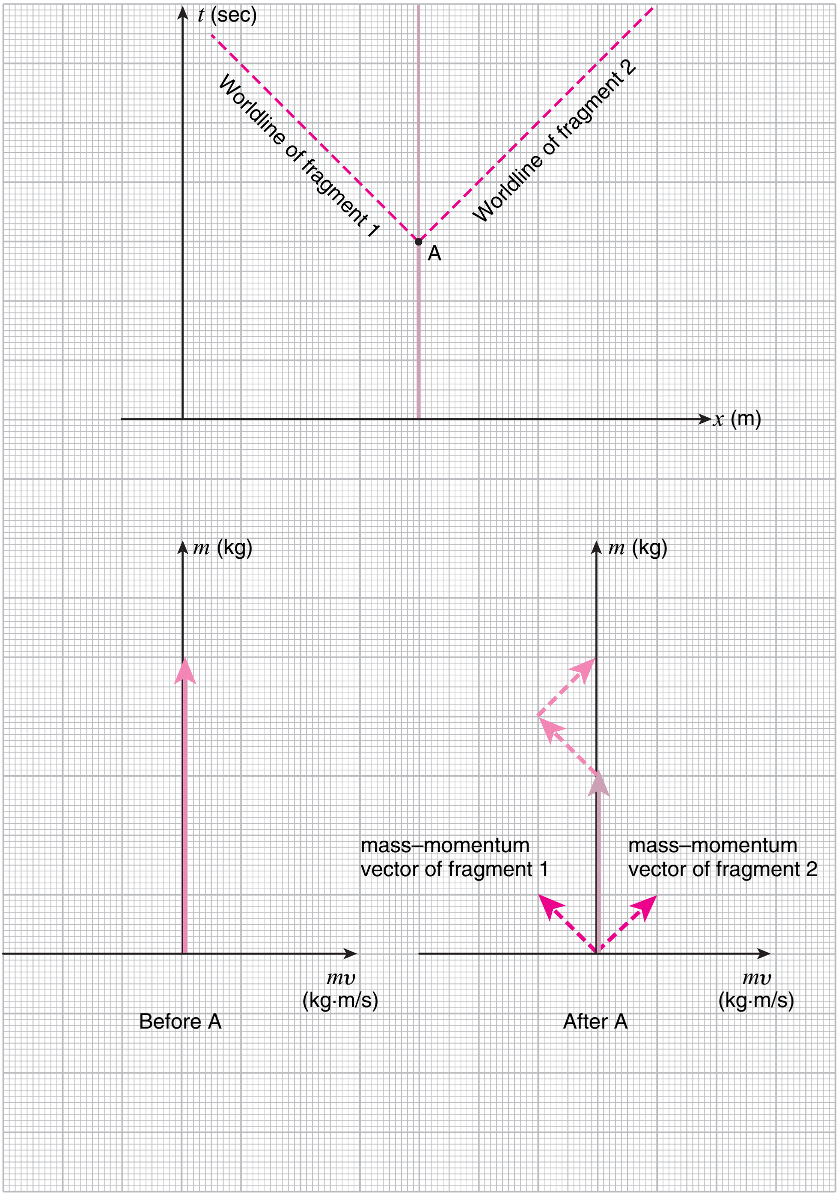
Since I have told you about Newton's Second Law, let me also tell you about Newton's Third Law for the sake of completeness.

Newton's Third Law[9] is the law which governs how two objects interact and exchange *momenta*[10] with each other. As an example, consider the process depicted in the spacetime diagram shown top-right: at point A on the spacetime diagram, a moving object 1 collides with another object 2 at rest. Newton's Third Law states that during this collision, the *impulse vector* applied to object 2 by object 1, and the *impulse vector* applied to object 1 by object 2, are of the same magnitude but pointing in opposite directions. Since the *impulse* was equal to the change in the *momentum*, this is the same thing as saying that the change in the momentum of object 1, and the change in the momentum of object 2, are of the same magnitude but pointing in opposite directions. In other words, the amount of *momentum* gained by one object is the same as the amount of *momentum* lost by the other: that is, the *total momentum* carried by the two objects before and after the collision is the same. The total mass is, of course, unchanged, so we can also say that the *total mass–momentum vector* is unchanged.

The law can easily be extended to more generic situations with n interacting objects. Just think of the interaction of the n objects as a series of interactions involving one pair of objects at a time. Since momentum is conserved in each of these interactions, the total momentum of the n objects will be conserved.


t (sec)
Worldline of object 2
A
Worldline of object 1
Object 1collides with
object 2 here
x (m)
m (kg)
Change in
the state of
motion of
object 1
mv
(kg·m/s)
m (kg)
Change in
the state of
motion of
object 2
mv
(kg·m/s)
m (kg)
The sum of the
mass–momentum vectors
of the two objects
before and after
the collision are the same.
mv (kg·m/s)

As an example of what this law can tell us, consider the following situation: at point A on the spacetime diagram shown here, an object at rest releases two projectiles of the same mass in opposite directions with the same speed. The law states that the *mass–momentum vector* of the object before point A, and the sum of the *mass–momentum vectors* of the object and the two projectiles after point A must be the same. By drawing a simple picture, as shown in the bottom figure, we can conclude that the mass–momentum vector of the object after its release of the two projectiles must stay vertical, that is, the object stays at rest, while its length is diminished by the masses carried away by the two projectiles. (This may seem like an esoteric example, but we will use an analogue of this later.)


t (sec)
Worldline of fragment 1
Worldline of fragment 2
A
x (m)
m (kg)
mv
(kg·m/s)
Before A
m (kg)
mass–momentum
vector of fragment 1
mass–momentum
vector of fragment 2
mv
(kg·m/s)
After A

Notes

1 A *kilogram* was originally defined as the mass of one liter of water at 4 °C. (Water is densest at this temperature.) But since this definition was inconvenient as a standard of mass, a chunk of metal of the same mass was made and its mass was defined as a *kilogram*. This, too, is not that convenient, so proposals have been made recently, to redefine the *kilogram* from the mass of a single atom using Avogadro's number, or from the energy of a photon using Planck's constant.

2 In terms of symbols used in physics, *mass* is m while *weight* is mg.

3 This means that the *inertial mass* and the *gravitational mass* are equal.

4 Strictly speaking, the *weight* acting on 1 *kilogram* of *mass* on the surface of the Earth is called 1 *kilogram-weight* and distinguished from kilograms.

5 The *mass* of an object is often represented by the symbol m. This should not be confused with the symbol for meters.

6 You can think of the term "*vector*" as just a nerdy way of saying *arrow*.

7 *Impulse* is the same thing as Δp after all. The *force* $F = \Delta p/\Delta t$ is basically the *rate of momentum transfer* to the object. Note that this definition of the term "*force*" is quite different from its use in everyday English. Not understanding this has lead to a lot of confusion among beginning physics students.

8 In most physics textbooks, both sides of the equation are divided by Δt and written as

$$F = m\,\frac{\Delta v}{\Delta t}\,,$$

or

$$F = ma\,,$$

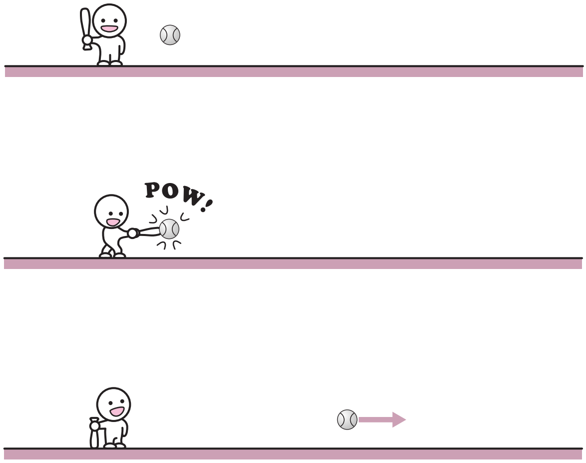
where

$$a = \frac{\Delta v}{\Delta t}$$

is the rate of change of the velocity, and is called the *acceleration*.

9 Newton's Third Law is also known as the *action–reaction law*. When two objects interact with each other, momentum is exchanged between the two. *Action* refers to the rate of momentum transfer from object 1 to object 2, while *reaction* refers to the rate of momentum transfer from object 2 to object 1. Newton's Third Law states that *action* and *reaction* are equal in magnitude, but opposite in direction, which is just a mathematical way of saying that what is lost by one must be gained by the other. Note that the definitions of the terms "*action*" and "*reaction*" are completely different from their use in everyday English. This unfortunate selection of terminology by Newton has led to endless confusion, about Newton's Third Law, when all it is saying is that *total momentum is conserved*.

10 The plural of *momentum*.


POW!

13
Relativistic dynamics

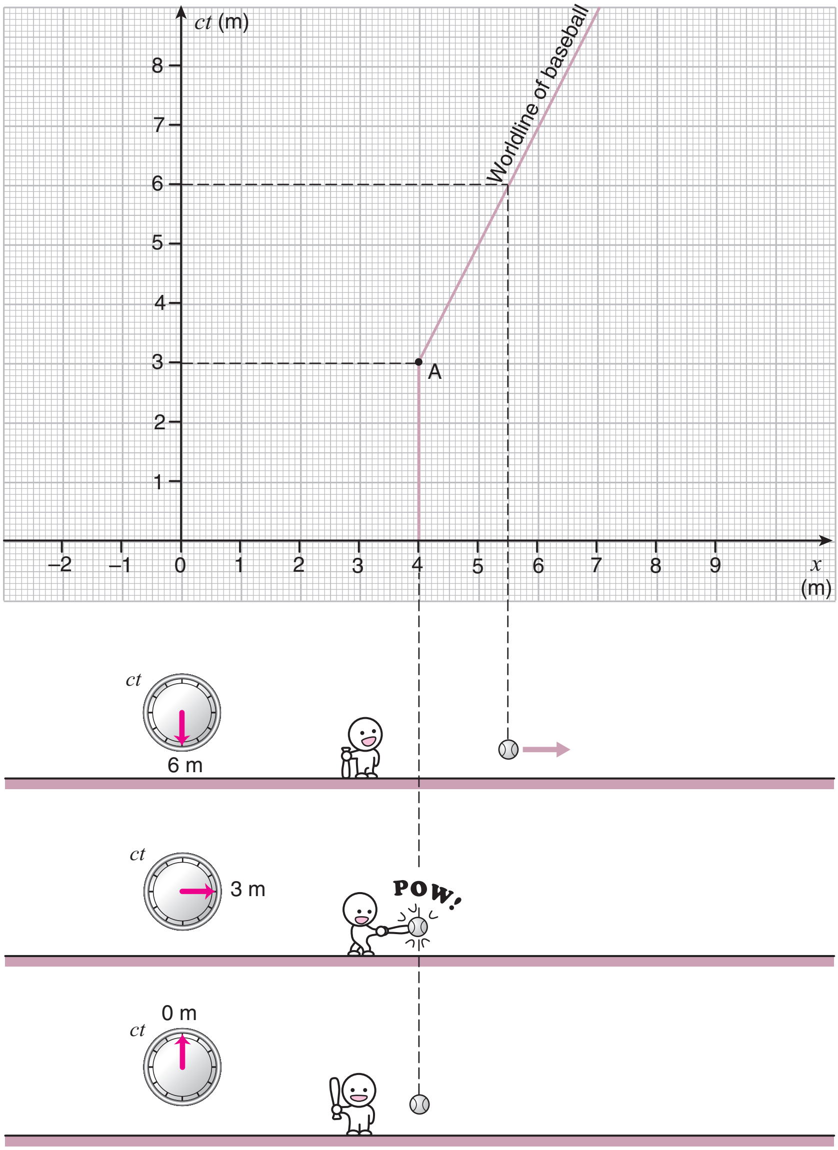
13.1 The energy–momentum vector

Next, let us consider the case when the velocities of objects are close to the speed of light c, and the Lorentz transformation must be used for relating observations from different inertial frames.

Assume that a baseball which was initially at rest is accelerated to half the speed of light by an impact with a bat at point A on the spacetime diagram. The questions we wanted answers for in Chapter 11 were:

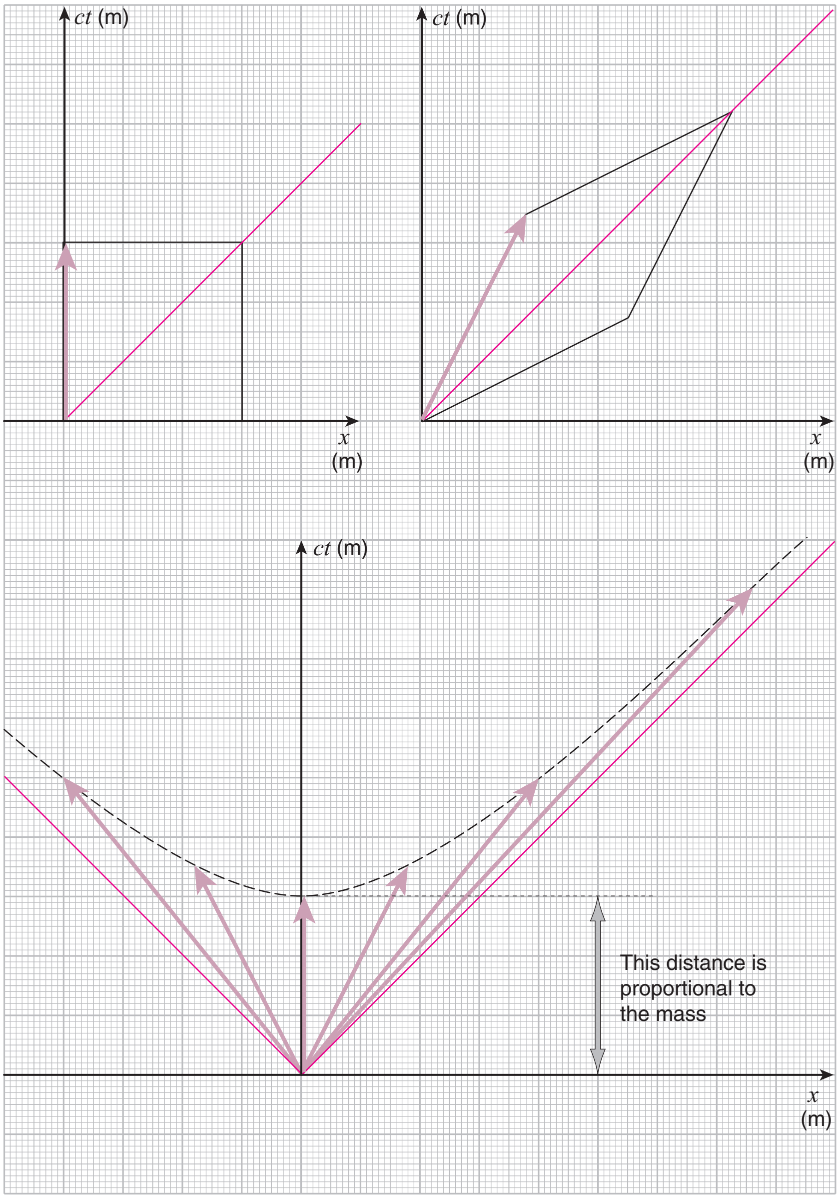
Q7: If the baseball is given the "same impact" again, what will its velocity be?

Q8: If the "same impact" is given to an object at rest with a different mass, what will the object's velocity be?


ct (m)
8
7
6
5
4
3
2
1
–2
–1
0
1
2
3
4
5
6
7
8
9
x (m)
Worldline of baseball
A
ct
6 m
ct
3 m
POW!
0 m
ct

As in the Newtonian case, let's represent the "state of motion" of an object with a vector on the spacetime diagram. First, to represent the "state in which the object is at rest," we follow the Newtonian case and use a vector pointing vertically up from the spacetime origin with length proportional to the object's mass. (See the figure on the opposite page, top-left.) Next, the "state in which the object is moving with velocity v" is the same as "the state in which the object is at rest" but observed from an inertial frame moving at velocity $-v$ relative to the first. So it must be represented by the vector you obtain by Lorentz transforming the vector which represents the "state in which the object is at rest" to that frame. For instance, if we want to find the vector which represents the "state in which the object is moving at velocity $v = \frac{1}{2}c$, we must Lorentz transform the "at rest" vector to a frame moving at velocity $-\frac{1}{2}c$. The result is shown in the figure opposite, top-right. Notice that, unlike the Galilei transformed case, not only the space-component but also the time-component of the vector has changed. What stays invariant is the area of the diamond with the vector as one of its sides, and the diagonals at 45° angles from the horizontal. (See figure.) When the object is at rest, this diamond becomes a square with the length of its sides proportional to the mass of the object, so its area is proportional to the mass squared.

In a similar fashion, we can determine the vectors that represent the motion of the object at all other velocites. When we do this, the tip of the vector will move along the curve shown in the bottom figure. This curve is of a kind known as a *hyperbola*, and its *asymptotes* are the light-cone of the spacetime origin. The distance between the spacetime origin and the hyperbola is proportional to the object's mass.


ct (m)
x (m)
ct (m)
x (m)
ct (m)
This distance is proportional to the mass
x (m)

Now, as in the Newtonian case, it is convenient to draw a graph for the "state of motion" vectors separately from the spacetime diagram. For the Newtonian case, we used a graph with *mass* notched on the vertical axis, and *momentum* = *mass* × *velocity* notched on the horizontal axis. In the relativistic case, however, because the vertical axis of the spacetime diagram is ct and not t, the corresponding "state of motion" graph will also have the vertical axis multiplied by c. So the quantity notched on the vertical axis will be *mass* × (*speed of light*), while that on the horizontal axis will be *mass* × *velocity* as before. The units for both axes will be $\text{kg} \cdot \text{m/s}$.

The vector representing the "state of the object of mass m at rest" on this graph is then one with time-component equal to mc, and space-component equal to zero. The vector representing the "state of the object of mass m moving at velocity v" will be the Lorentz transform of the "at rest" vector, as shown in the figure. As we noted above, the time-component is longer than what it was at rest. We express the factor by which the time-component has lengthened with the greek letter γ, that is:

$$(\text{time-component at velocity } v) \;=\; \gamma(\text{time-component at rest}) \;=\; \gamma mc\,.$$

Then, since the ratio of the time-component and the space-component is determined by the velocity v, we must have

$$(\text{space-component at velocity } v) \;=\; \gamma mv\,.$$

Note that the factor γ is a number that depends on v and changes with it. When $v = 0$, there is no lengthening of the time-component, so $\gamma = 1$. However, as v grows toward c, the diamond shown in the figure will get flatter and flatter and collapse onto the light cone, while still maintaining a constant area. For this to happen γ must diverge to infinity as v approaches c.[1]

Now, the time-component of our vector represents the "tenacity of the motion" in the time-direction, while its space-component represents the "tenacity of the motion" in the space-direction. The space-component γmv is called the *momentum*, just as in the Newtonian case (though its numerical value differs by the factor γ), and is represented by the symbol

$$p \;=\; \gamma mv\,.$$

The time-component γmc represents the *inertia* of the object, so you may be inclined to call it the *mass* (times c)[2]. However, we would like to reserve the term for m, which is the amount of "stuff" contained in the object, and independent of the frame. So we will call the time-component of our vector by another name: the *energy*. To be more precise, the *energy* E is defined as the time-component of the vector times c:

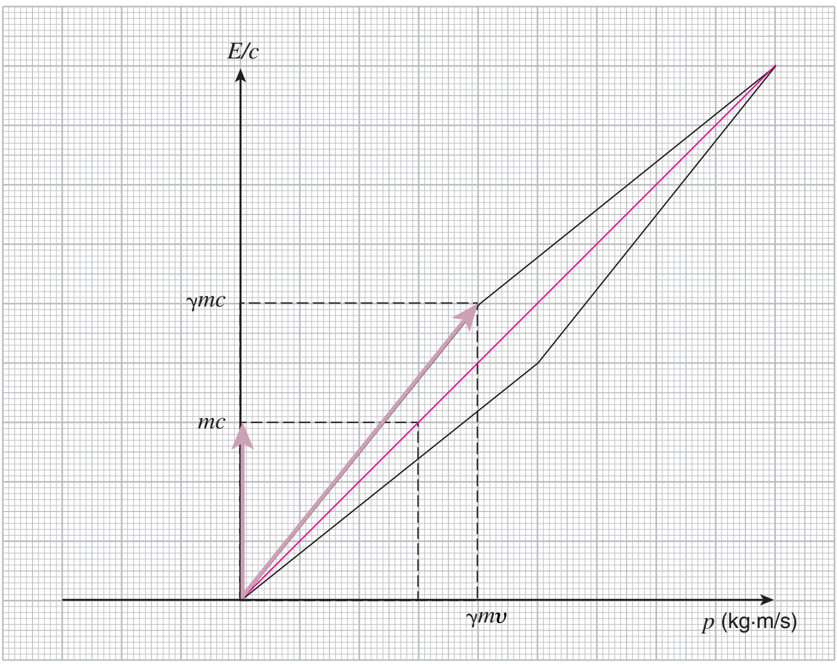
$$E = \gamma mc^2 .$$

Thus, the vector we have defined will be called the "energy–momentum vector."[3]

Notice that when the object is at rest, $v = 0$, the γ-factor is equal to one, and the above equation reduces to

$$E = mc^2 .$$

$E = mc^2$ is the energy that the object has when it is at rest, so it is called the *rest energy*. Now, this is Einstein's famous formula, but at this point, we have just taken the *mass*, multiplied it by c^2, and given it the alias *rest energy*, so it does not have any physical meaning yet. We will revisit this equation later.



13.2 The energy–momentum vector of a photon

The *photon* is the particle associated with light, so it travels at the speed of light c. What kind of *energy–momentum vector* will represent its motion?

If a particle's mass is m, the tip of its *energy–momentum vector* will move along a hyperbola as shown in the top figure.[4] The distance between the graph's origin and the hyperbola was equal to mc.[5] Now, imagine letting the particle mass m get smaller and smaller until it finally reaches zero. Then, the distance between the hyperbola and the origin will get smaller and smaller, until the hyperbola finally collapses onto the light-cone, as shown in the bottom figure. In this limit, the particle's *energy–momentum vector* will be at a 45° angle from the horizontal, which means that the particle is traveling at the speed of light. Therefore, particles like the *photon* that travel at the speed of light must have *mass* equal to zero. Conversely, a *massless* particle can only travel at the speed of light.

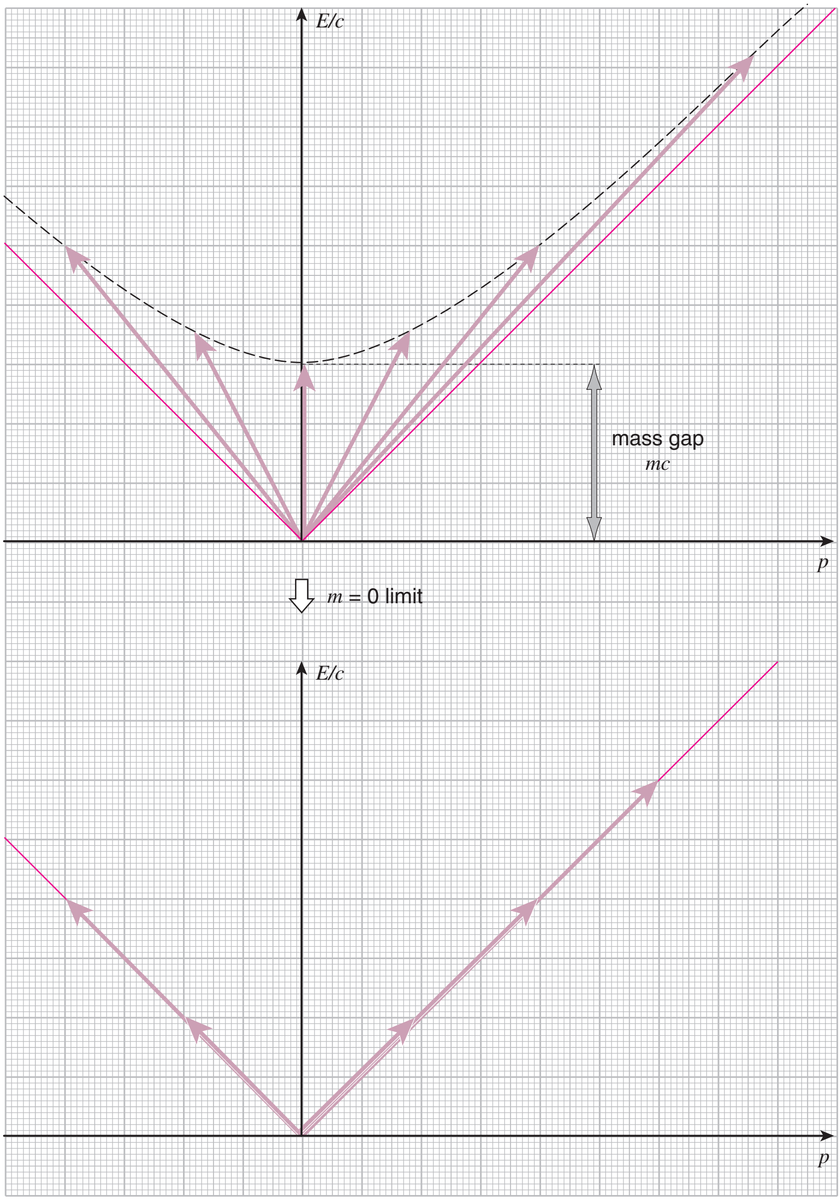
Because the *energy–momentum vector* is at a 45° angle, the *energy* E and *momentum* p of a photon must be related by

$$E = pc.$$

That is, the photon's *energy* and *momentum*[6] are not independent and determining one will determine the other. But then, what determines the *energy* of each photon?[7] Again, it was Einstein, who again in 1905 [10] showed that the energy of a photon E is proportional to the *frequency* of the corresponding light wave f, and is given by the relation

$$E = hf,$$

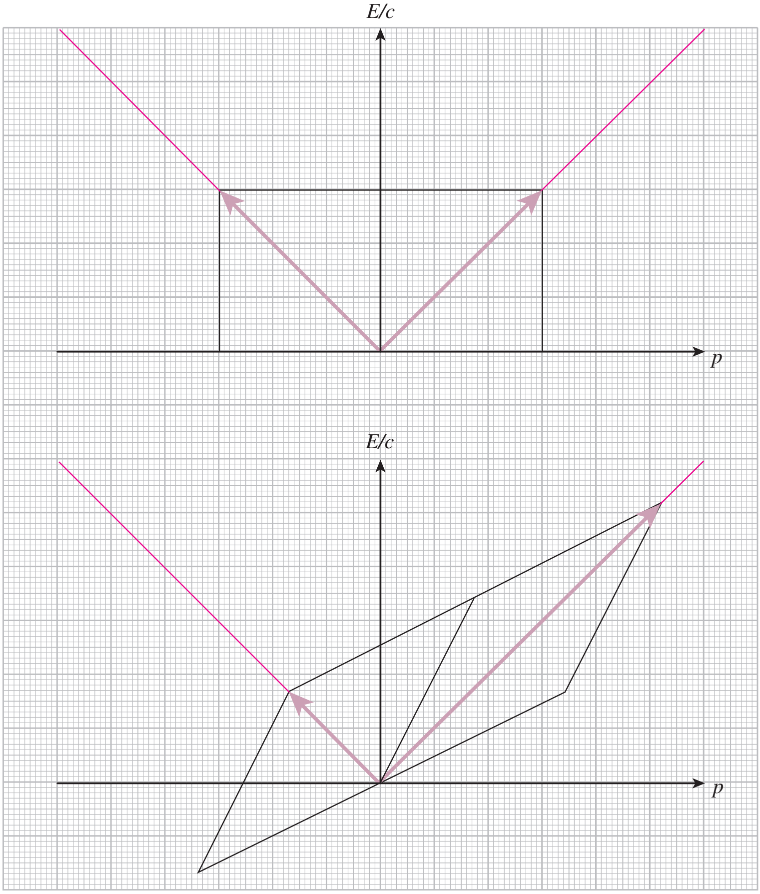
where h is a number known as *Planck's constant*. We are not going to discuss where this equation comes from in this book. However, note that for this equation to be correct in all frames of reference, both sides of the equation must transform in the same way under Lorentz transformations.



E/c
mass gap
mc
p
m = 0 limit
E/c
p

For instance, if we have two photons of the same frequency, one propagating toward the right and the other propagating toward the left, their *energy–mometum vectors* will have the same length as shown in the top figure. If the same two photons are observed from a frame moving with velocity $-\frac{1}{2}c$ relative to the first, the *energy–momentum vectors* of the two photons will be as shown in the bottom figure, which can be obtained from the top figure by a Lorentz transformation. As you can see, the *energy–mometum vector* of the photon propagating toward the right will become longer, while that for the photon propagating toward the left will become shorter. This means that the energy of the photon will become smaller if observed from a frame moving in the same direction as the photon, while it will become larger if observed from a frame moving in the opposite direction from the photon.

If there is a proportionality relation between the photon's energy and the frequency of light as given above, then the frequency of light must depend on the frame in exactly the same way as the energy. And indeed, as we discussed in Part I, the frequency of light does depend on the frame due to the Doppler effect! We did not calculate by how much the frequency will change in Part I, but it is possible to prove that the rate of change of the frequency is exactly the same as the rate of change of the energy, thereby justifying the relation $E = hf$.


E/c
p
E/c
p

13.3 The work–impulse vector

We are now ready to tackle questions Q7 and Q8 posed at the beginning of section 13.1. Let us begin with Q7. The *energy–momentum vectors* which represent the "state of the baseball at rest," and the "state of the baseball moving at $\frac{1}{2}c$" are given by OA and OB, respectively, in the figure. (I will not specify the mass m since the numerical value of mc will end up being huge due to the large value of c.)

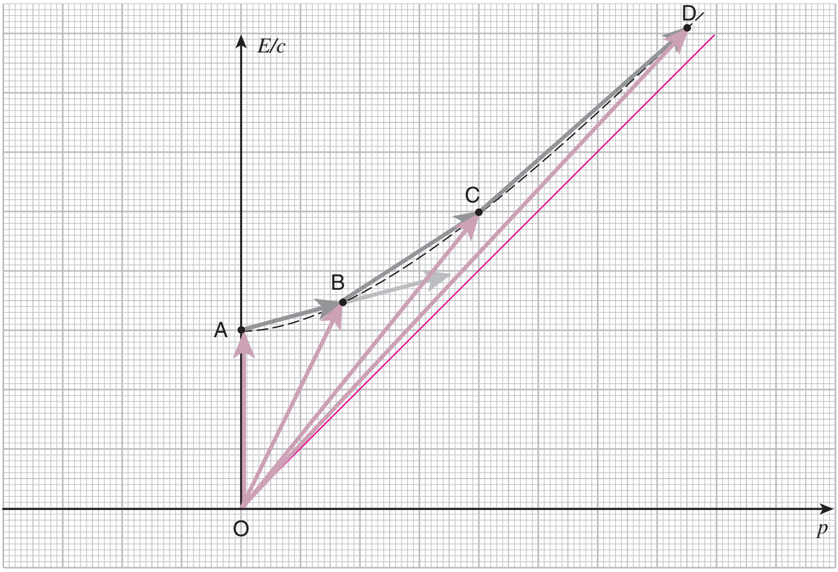
As in the Newtonian case, the vector AB that connects the tips of the two *energy–momentum vectors* represents the "impact" of the bat. This time, because the time-component of the *energy–momentum vector* has changed as well as the space-component, the vector that represents the "impact" has a non-zero time-component as well. This time-component is called *work*, and together with the space-component which was called the *impulse*, the vector AB is called the *work–impulse vector*.

Now, we would like to apply the "same impact" to the baseball traveling at $+\frac{1}{2}c$, but unlike the Newtonian case, we must be careful with what we mean by "same impact." For, as we discussed in section 4.10, if we observe the process of the baseball being accelerated from rest to $+\frac{1}{2}c$ from a frame moving at velocity $-\frac{1}{2}c$ relative to the first, it will be observed as a process of the baseball being accelerated from $+\frac{1}{2}c$ to $+\frac{4}{5}c$, and the *work–impulse vector* that connects these two states, represented by OB and OC on the diagram, will be given by the vector BC, which is clearly different from vector AB.

From the point of view of the accelerating baseball, however, AB and BC are the "same impact" since, in both cases, the baseball will observe the frame it was in before the impact to be moving at velocity $-\frac{1}{2}c$ relative to it after the impact. Furthermore, the *work–impulse vector* CD which accelerates the baseball from $+\frac{4}{5}c$ to $+\frac{13}{14}c$ is also the "same impact" from the point of view of the baseball (*cf.* section 4.10). Therefore, if by "same impact" we mean the "same impact from the point of view of the accelerating object," then the result of applying the "same impact" to the baseball traveling at $+\frac{1}{2}c$ will be the baseball traveling at $+\frac{4}{5}c$, and the "same impact" applied again will take the baseball to the state in which it is traveling at $+\frac{13}{14}c$, and so on.

However, when observed from the frame in which the baseball was initially at rest, AB, BC, and CD are all different *work–impulse vectors*. Neither the time-component (*work*), nor the space-component (*impulse*) are equal. So even though the "impacts" are "the same" from the point of view of the baseball, the *work–impulse vectors* that represent them are all different and will depend on the frame from which they are being observed.

So what if we define "same impact" to mean "the same *work–impulse vector* when observed from the initial frame"? That is, instead of adding BC to OB to obtain OC, we try to add AB to OB. If you try this, you will find that, in fact, you cannot add AB to OB: the resulting sum will be a vector that does not fall on the hyperbola that the tip of the *energy–momentum vector* must follow. That is, it is not an allowed state.

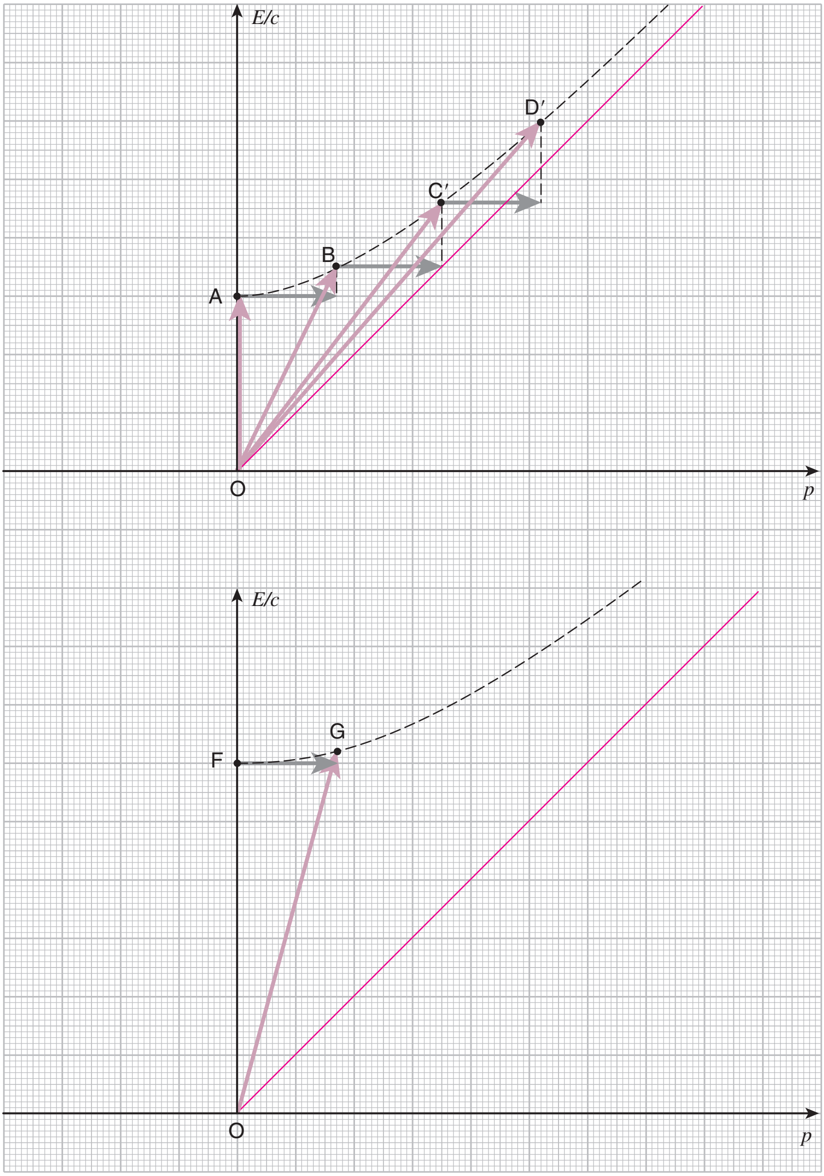


However, instead of specifying both the *work* and the *impulse* using the vector AB, we could ask what would happen if we add a *work–impulse vector* to OB with the same space-component as AB, but with its time-component adjusted appropriately so that the resulting vector would be an allowed state. The resulting *energy–momentum vector* is shown in the figure as OC′. If you compare this vector with OC on the diagram on the previous page, you can tell that the state represented by OC′ has velocity slightly smaller than that of OC which was $+\frac{4}{5}c$. Though it is a bit difficult to read off from the diagram, the velocity of OC′ is about $+0.76c$. Similarly, if we add to OC′ a *work–impulse vector* with the same space-component as AB but with its time-component appropriately adjusted, we obtain OD′, the velocity of which is about $+0.87c$.

In this way, the answer to Q7 will depend on what we mean by the term "same impact." However, in either case, the change in the velocity of the baseball due to the second impact is smaller than $+\frac{1}{2}c$, and that due to the third impact is smaller still. If we keep on iterating the process, the change in velocity per impact will continue to decrease, and no matter how many impacts are applied, the baseball is never accelerated beyond c. And the reason behind this is that, as the baseball is accelerated, its *energy–momentum vector* will get longer and longer in the time-direction, making its *inertia* larger and larger. In the limit that the velocity approaches the speed of light, the *inertia* will grow to infinity, making it impossible to accelerate the baseball further.

What about question Q8? If we consider a ball with twice the mass of the baseball, the vector representing its state at rest is given by OF, shown on the bottom figure, which is twice as long as OA. Adding the same *work–impulse vector* as AB to OF will not result in an allowed state, so again we consider adding a vector with the same space-component as AB, but with its time-component appropriately adjusted. The resulting vector is OG. In Newtonian dynamics, the change in velocity when the same impulse was applied was inverse proportional to the object's mass, so in the current case, it would predict the change in velocity to be half of $+\frac{1}{2}c$, namely $+\frac{1}{4}c$. However, if you look at OG carefully, you can tell that its velocity is slightly larger: it is about $+0.28c$. The reason behind this may sound a bit counter-intuitive, but when subjected to the same impulse, the extra mass prevents the velocity from increasing as quickly, which prevents the inertia from growing as quickly, which in turn results in a larger increase in velocity than the Newtonian case.[8]

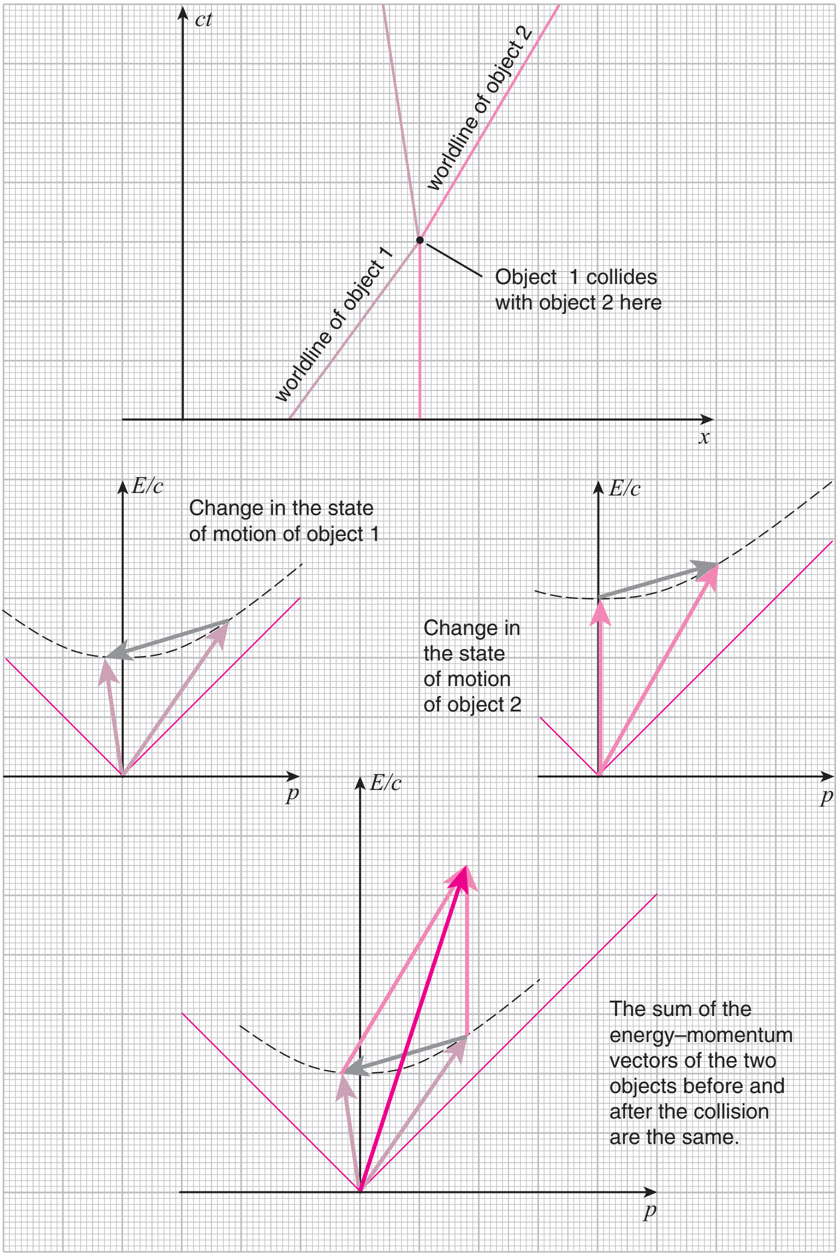
In this way, Einstein's relativistic dynamics is quite a bit more complicated than that of Galilei–Newton. Consequently, we cannot write a simple equation like $F\Delta t = m\Delta v$ to predict the outcome of a dynamic process.[9]



13.4 Conservation of energy–momentum

Even though we cannot express relativistic dynamics in a simple equation, the law that governs the exchange of *energy* and *momenta* between two objects that interact, say via a collision, is simple. The amount of *energy* and *momentum* lost by object 1 must be equal to the amount of *energy* and *momentum* gained by object 2, so that the *total energy* and *total momentum* of the two object system is conserved. In other words, the sum of the *energy–momentum vectors* before and after the interaction must stay the same. The law can be extended to a generic system with n objects: the total sum of all the *energy–mometum vectors* will be conserved.

The important point here is that it is the *energy* and not the *mass* that is conserved. The *total energy* before and after an interaction are always the same, but the *total mass* before and after an interaction are not necessarily so. To see how *energy* can be conserved while *mass* is not, we will consider a process that was discussed by Einstein himself in [9].


ct
worldline of object 2
worldline of object 1
Object 1 collides
with object 2 here
x
E/c
Change in the state
of motion of object 1
p
E/c
Change in
the state
of motion
of object 2
p
E/c
The sum of the
energy–momentum
vectors of the two
objects before and
after the collision
are the same.
p

13.5 $E = mc^2$

An object of mass m is at rest. At point A on the spacetime diagram, this object emits two photons of the same frequency back to back: one toward the right and the other toward the left. The energies of the two photons are the same since their frequencies are the same, as well as the magnitudes of their momenta (*cf.* section 13.2), the directions of which are opposite.

Since the object is initially at rest, the momentum of the object is initially zero. This means that the total momentum of the object and the two photons after A must also be zero. Since the momenta of the two photons are equal in magnitude but opposite in direction, they sum to zero. Therefore, the momentum of the object after emitting the photons must also be zero for the total momentum to be conserved. We can conclude that the object will stay at rest.

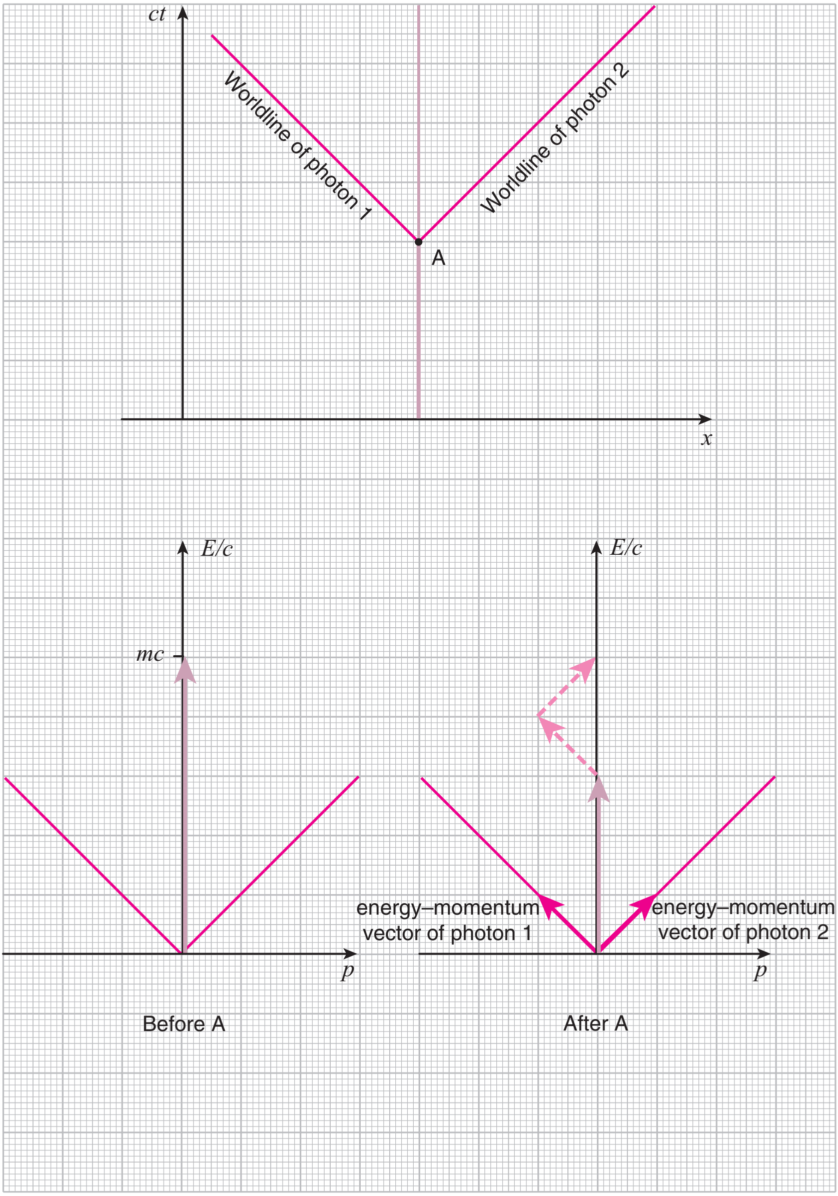
The total energy is also conserved, so the energy of the object must decrease by the amount that was taken away by the two photons. However, due to the relation

$$E = mc^2$$

between the *rest energy* and the *mass*, the *mass* of the object, which stays at rest, must also decrease by the corresponding amount. Note that since photons are *massless*, they do not carry away any *mass*. Therefore, even though the *total energy* is conserved in this process, clearly *total mass* is not.

Since photons can be considered chunks of pure *energy* without any *mass*, we can say that a portion of the object's *mass* has been converted into *energy* through this process. From this point of view, $E = mc^2$ is not just a relation which provides the *mass* with the alias "*rest energy*," but a physical relation which dictates the *conversion rate* from *mass* to *energy*.

Now, photons are emitted from an object when the object's *internal energy*, such as heat, is converted to *electromagnetic energy*. But objects can lose *internal energy* not just through photon emission, but, for instance, by losing heat to another object with a lower temperature. The exchange of heat does not involve any exchange of *mass*, but since the object's *internal energy* decreases just as in the photon emission process, its *mass* must also decrease. Conversely, if an object absorbs external energy, whether in the form of heat or photons, its *internal energy* will increase, so its *mass* must also increase.


ct
Worldline of photon 1
Worldline of photon 2
A
x
E/c
mc
p
Before A
E/c
energy–momentum
vector of photon 1
energy–momentum
vector of photon 2
p
After A

For instance, when water at 100 °C with an initial mass of 1 kg cools to 0 °C, it loses mass corresponding to the amount of heat that it has released. However, the change in mass is only about 0.000 000 005 grams. Because the speed of light c is so large, unless the released energy E is huge, the change in mass $m = E/c^2$ will not be large enough to be measurable. So in our everyday lives we do not have to worry about this effect, just like most other relativistic effects, and there is nothing wrong in thinking that mass is conserved.

13.6 Common misconception about $E = mc^2$

The only processes in which the release of energy is large enough so that the change in the mass is measurable are nuclear processes such as fission and fusion. Einstein himself suggests, at the end of his original paper [9], that $E = mc^2$ may be confirmed by looking at such processes.[10] For this reason, many people believe that $E = mc^2$ describes the operational principle behind weapons of mass destruction, namely the atomic (fission) and hydrogen (fusion) bombs. However, this is a misconception.

The release of energy is not *caused* by the decrease in mass. Rather, the decrease in mass is a consequence of the release of energy. $E = mc^2$ just tells us by how much the mass decreases, and does not dictate the amount of energy released. Therefore, the amount of energy released in a nuclear reaction has nothing to do with $E = mc^2$, just like the amount of energy released when water cools from 100 °C to 0 °C has nothing to do with $E = mc^2$ either. The amount of energy released in a nuclear reaction is actually determined by how strongly the nucleons (protons and neutrons) inside an atomic nucleus are bound together, and the amount of energy released when water cools is determined by how strongly the water molecules are bound together in liquid water. The only difference is that the amount of energy released in a nuclear reaction is large enough to see the mass difference, while that for water cooling is not.

Notes

1 The v dependence of γ can be derived as follows. When the object is traveling at velocity v, the time- and space-components of the vector representing its state of motion are related via

$$\frac{\text{space-component}}{\text{time-component}} = \frac{v}{c}\,.$$

On the other hand, the area of the diamond with the vector as one of its sides can be shown, using elementary geometry, to be equal to $(\text{time-component})^2 - (\text{space-component})^2$, and since this must be equal to the area of the square with mc as one of its sides, we have the relation

$$(\text{time-component})^2 - (\text{space-component})^2 = (mc)^2\,.$$

Solving these two equations for the time- and space-components, we find:

$$\text{time-component} = \frac{mc}{\sqrt{1-\frac{v^2}{c^2}}}\,, \qquad \text{space-component} = \frac{mv}{\sqrt{1-\frac{v^2}{c^2}}}\,.$$

Defining the γ-factor as

$$\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$$

allows us to write the above relations compactly as

$$\text{time-component} = \gamma mc\,, \qquad \text{space-component} = \gamma mv\,.$$

The time-component is enhanced by a factor of γ compared to when the object is at rest, and the space-component is also enhanced by the same factor compared to the Newtonian case.

2 In some old textbooks on relativity, γm is indeed called the *mass*, while m is called the *rest mass* to distinguish the two. However, this terminology is completely obsolete and no one uses it anymore. After all, once we decide to call $E = \gamma mc^2$ the *energy*, which is the same thing as γm since c^2 is just a constant, we do not need two names for the same thing.

3 The reason why the quantity $E = \gamma mc^2$ is called the *energy* is the following: When the velocity of the object v is small compared to the speed of light c, the γ-factor can be approximated by

$$\gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \approx 1 + \frac{v^2}{2c^2} + \cdots\,,$$

so E is approximately equal to

$$E = \gamma mc^2 \approx mc^2 + \frac{1}{2}mv^2 + \cdots\,.$$

The quantity $\frac{1}{2}mv^2$ is known as the *kinetic energy* in Newtonian Dynamics. Therefore, it makes sense to think of E as the *total energy* of the particle, mc^2 being the *rest energy*, and any extra portion being due to the motion of the particle.

4 This hyperbola on the *energy–momentum diagram* is known as the *mass shell*.

5 This separation between the *mass shell* and the origin of the *energy–momentum diagram* is known as the *mass gap*.

6 More precisely, the *magnitude of the momentum* is not independent of E. The direction of the momentum is arbitrary.

7 Clearly, the relation $E = \gamma mc^2$ is meaningless since in the limit $m \to 0$, the γ-factor diverges to infinity and the limit of E is indeterminate.

8 An alternative way to see why this would be the case is to consider what would happen if you decrease the mass. In the Newtonian case, the change in the velocity for the same impulse could be made as large as one liked by making the mass approach zero. In the relativistic case, however, the velocity of objects cannot exceed the speed of light. Consequenty, the increase in velocity cannot be doubled for the same impact by halving the mass.

9 The relativistic equation of motion can actually be written compactly as $f^\mu = dp^\mu/d\tau$. However, the simplicity is only superficial and this equation hides a lot of complications.

10 In [4], the relevant sentence is translated as: "It is not impossible that with bodies whose energy-content is variable to a high degree (e.g. with radium salts) the theory may be successfully put to the test."

14
Summary of Part III

This concludes Part III of this book. I hope you have been able to grasp the basic logic of where the equation $E = mc^2$ comes from. To summarize the important points:

- In Einstein's relativistic dynamics, the state of motion of an object is represented by an arrow called the *energy–momentum vector*. The vector's time-component (vertical component) is the *energy*, and the space-component (horizontal component) is the *momentum*, and they represent what might be called the "tenacity of the motion" or the "tendency of the motion to continue as is" in their respective directions in spacetime.
- The *energy–momentum vector* depends on the frame from which the observation is being made. However, the area of the diamond with the *energy–momentum vector* as one of its sides and the diagonals at 45° from the horizontal is invariant and equal to $(mc)^2$ where m is the object's mass.
- Changes in the motions of objects are represented by changes in their *energy–momentum vectors*. In a system of interacting objects, the *energy–momentum vector* of each individual object will change via interactions, but the *total energy–momentum vector* of the system will be conserved.
- Though the *total energy* of a system is conserved, the *total mass* is not. As a result, processes can occur which can be interpreted as the *mass* being converted to *energy*. In such cases, the equation $E = mc^2$ determines the conversion rate between *energy* and *mass*.

Afterword

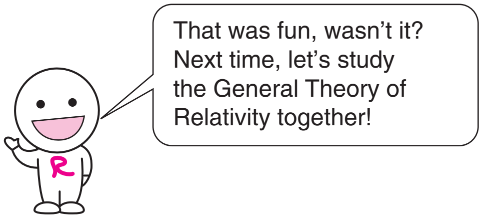
This brings us to the end of my exposition of Einstein's Special Theory of Relativity (SR). I have attempted to explain everything that is usually explained using equations using drawings only so that you can literally *see* what I am talking about. I hope you have found this approach more tractable, eye-opening, and fun.

Partly because of its name, Einstein's "Theory of Relativity" is often misunderstood to have discarded Newton's notions of *space* and *time* that were both "objective" and "absolute," and to have pronounced that both *space* and *time* were "relative," and even "subjective" concepts. In truth, Einstein was a firm believer in objective reality, and SR assumes the existence of an "objective" and "absolute" *spacetime*. All SR is claiming is that when the motion of objects in *spacetime* is observed from different inertial frames, things like *velocity* and *length* will be frame-dependent. And this dependence comes about because the way the *time-* and *space-axes* are introduced into the "absolute" *spacetime* differs from inertial frame to inertial frame. The frame-dependence of the *time-axis* already existed in Newton's theory, and as a consequence *velocity*, not surprisingly, was frame-dependent. In SR, however, in addition to a frame-dependent *time-axis*, the concept of *simultaneity* depends on the frame and results in a frame-dependent *space-axis* also. This leads to the frame-dependence of things like *length*, which we normally do not think of as a frame-dependent quantity.

In fact, it was not Einstein himself but Max Planck[1] (1858–1947) in 1906 who named Einstein's theory the "Theory of Relativity." Einstein did not necessarily like the name and is quoted as saying:

'Now to the term *relativity theory*. I admit that it is unfortunate, and has given occasion to philosophical misunderstandings.'[2]

He thought a better name would be the "theory of invariants" since SR is a theory which is concerned with what remains *invariant* under Lorentz transformations. As we mentioned in section 2.6, after his 1905 paper on SR, which was the "Theory of invariants for all inertial frames," Einstein spent the next 10 years working on extending it to the "Theory of invariants for ALL frames of reference." He completed this theory in 1915 as the "General Theory of Relativity" (GR) [4]. GR is a monumental theory, which not only extends the theory of motion to general frames of reference, but encompasses gravity within its framework as well. I hope to tell you all about it in my next book.



Notes

1 The discoverer of Planck's constant we encountered in section 13.2.
2 See [3], page 229.

References

[1] Takeuchi, T. (2005), *No Equations! Special Relativity That You Can See*, Japan: Iwanami Shoten (in Japanese).

[2] Takeuchi, T. (2007), *Theory of Relativity without Equations*, Taiwan: Shy Mau Publishing Company (in Chinese).

[3] Calaprice, A. (ed.) (2005), *The New Quotable Einstein*, enlarged commemorative edition, foreword by Freeman Dyson, Princeton, NJ: Princeton University Press.

[4] Hawking, S. (ed.) (2003), *On the Shoulders of Giants*, Philadelphia, PA: Running Press. (A collection of the great works by Copernicus, Galileo, Kepler, Newton, and Einstein.)

[5] An English translation of Newton's *Principia* can be found in [4]. Other translations and commentaries currently available are: Newton, I., *The Principia: Mathematical Principles of Natural Philosophy*, translated by I. Bernard Cohen and Anne Whitman, Berkeley, CA: University of California Press (new translation 1999); Newton, I., *The Principia*, translated by Andrew Motte in 1729, Amerhurst, NY: Prometheus Books (new edition 1995); Chandrasekhar, S. (2003), *Newton's Principia for the Common Reader*, New York: Oxford University Press.

[6] Einstein, A., *Relativity: The Special and the General Theory, The Masterpiece Science Edition*, New York: Pi Press (new hardback edition 2005); Thousand Oaks, CA: BN Publishing (paperback edition 2005).

[7] Bruce, C. (1998), *The Einstein Paradox: And Other Science Mysteries Solved by Sherlock Holmes*, New York: Basic Books.

[8] Weinberg, S. (2008), *Cosmology*, Oxford: Oxford University Press.

[9] Einstein, A. (1905), "Ist die Trägheit eines Korpers von seinem Energieinhalt abhängig?," (Does the inertia of a body depend upon its energy-content?), *Annalen der Physik*, **18**, 639–641. (An English translation can be found in [4].)

[10] Einstein, A. (1905), "Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt," (On a heuristic viewpoint concerning the production and transformation of light), *Annalen der Physik*, 17, 132–148.

Index

$E = mc^2$, 209, 246, 248
c, 2, 26, 59

acceleration, 56
addition of velocities
  in Galilei–Newton theory, 54
  in Lorentz–Einstein theory, 110
atomic bomb, 248

baseball, 178, 180
blue shift, 158

causal structure, 128
causality, 120
centrifugal force, 26
clock, 5
  location in frame of reference, 10, 130
clock synchronization, 130
Copernicus, Nicolaus (1473–1543), 24

Doppler effect, 154, 238
Doppler, Christian Andreas (1803–1853), 160
dynamics
  Newtonian, 212
  non-relativistic, 212
  relativistic, 230

Einstein, Albert (1879–1955), 2
energy, 235
  electromagnetic, 246
  internal, 246
energy-momentum, 230
  conservation of, 244
energy-momentum vector, 230, 235
expansion of the universe, 160

faster-than-light
  communication, 124
  travel, 126
fission, 248
fission bomb, 248
Foucault pendulum, 28
frame of reference, 5, 8
  Earth, 24
  inertial, 18, 24
  location of clock in, 10, 130
  non-inertial, 20, 22, 24
fusion, 248
fusion bomb, 248

Galilei transformation, 40, 108
Galileo Galilei (1564–1642), 2, 14, 59

hare and tortoise, 170, 172, 174
hydrogen bomb, 248

impulse, 218, 240
inertia, 16, 213, 217
  dependence on speed, 116
inertial force, 26
instantaneous communication, 124

kilogram, 216, 228

Law of Inertia, 14, 16, 18
light-cone, 128
light-day, 64
light-minute, 64
light-year, 64
Lorentz contraction, 140, 146, 184
Lorentz transformation, 94, 100
  low velocity limit of, 108
Lorentz, Hendrick Antoon (1853–1928), 118

mass, 212, 228
  gravitational, 58, 228
  inertial, 58, 220, 228
  unit of, 216
mass gap, 250
mass shell, 250
mass–momentum vector, 212, 217, 218
meter
  definition of, 118
momentum, 216, 234
  conservation of, 224
motion, 2

neutron, 248
Newton's First Law, 14, 16, 18
Newton's Second Law, 56, 222
Newton's Third Law, 224
Newton, Sir Issac (1642–1727), 2, 14
nucleon, 248

origin, 6

photon, 66, 236
  energy–momentum of, 236
Planck's constant, 228, 236, 253
Planck, Max (1858–1947), 253
Principia, 2
proton, 248

red shift, 156, 160
relativity
  in Galilei–Newton theory, 30
  in Lorentz–Einstein theory, 59
  of instance in time, 78
  of location in space, 6, 10, 12
  of simultaneity, 78, 81, 82, 84

SI units, 5
soccer, 182
space, 2, 4
space shuttle, 22, 26
space-component, 216
space-like separation, 128
spacetime, 34, 252
  conservation of spacetime volume, 88
spacetime diagram, 34
speed of light, 2, 26
  constancy of, 59
supernova, 168, 176

time, 2, 4
time dilation, 132, 138
time-component, 214
time-like separation, 128
twin paradox, 148

vector, 228
  sum of vectors, 218
velocity, 4

weight, 216, 228
  unit of, 228
work, 240
work–impulse vector, 240
worldline, 36