Moses Fayngold

Special Relativity and Motions Faster than Light

WILEY-VCH

Author:
Moses Fayngold
Department of Physics,
New Jersey Institute of Technology, Newark.
e-mail: fayngold@ADM.NJIT.EDU

Illustrations:
Roland Wengenmayr, Frankfurt, Germany

Cover Picture:
Albert Fayngold, New York, NY

1st edition

Library of Congress Card No. applied for.

British Library Cataloguing-in-Publication Data:
A catalogue record for this book is available from the British Library.

Die Deutsche Bibliothek – CIP-Cataloguing-in-Publication Data:
A catalogue record for this publication is available from Die Deutsche Bibliothek.

Printed in the Federal Republic of Germany

Printed on acid-free paper

Typesetting ProSatz Unger, Weinheim
Printing betz-druck gmbh, Darmstadt
Bookbinding Großbuchbinderei
J. Schäffer GmbH & Co. KG, Grünstadt

ISBN 3-527-40344-2

Table of Contents

Special Relativity and Motions Faster than Light. Moses Fayngold

ISBN: 3-527-40344-2

Preface

This is a book about Special Relativity. The potential reader may ask why yet another book needs to be written on this subject when so many have already covered this ground, including some classical early popularizations. There are four answers to this question.

First, this book is intended to supplement the ordinary physics texts on Special Relativity. The author's goal was to write a book that would satisfy the demands of different categories of reader, such as college students on the one hand and college professors teaching physics on the other. To this end, many sections are written on two levels. The lower level uses an intuitive approach that will help undergraduates to grasp qualitatively, fundamental aspects of relativity theory. The higher level contains a rigorous analytical treatment of the same problems, providing graduate students and professional physicists with a good deal of novel material analyzed in depth. The readers may benefit from this approach. There are not many books having the described two-level structure (a rare and outstanding example is the monograph *Gravitation* by C. W. Misner, K. S. Thorne, and J. A. Wheeler [1]).

Second, the book explores some phenomena and delves into some intriguing areas that fall outside the scope of the standard treatments. For instance, in the current book market on relativity one can spot a "hole" – an apparent lack of information (but for just one or two books [2]) about faster-than-light phenomena. One of the purposes of this book is to fill in the hole. The corresponding chapters (Chap. 6–8) aim to elucidate areas related to faster-than-light motions, which at first seem to contradict relativity, but upon examination reveal the consistency, subtlety, and depth of the theory.

Third, there have appeared recently a good deal of new theoretical studies and corresponding experiments demonstrating superluminal propagation of light pulses, which, on the face of it, could appear to imply possible violation of causality. (A similar approach has been used to slow the light pulses dramatically and, finally, to "stop" light by encoding information it carried, into the physical state of the medium.) These experiments have been described in the most prestigious journals (see, for instance, Refs. [3–6]), and have attracted much attention in the physics and optics communities. This book describes the new results at a level accessible to an audience with a minimal background in physics (Chap. 7). It contains an analysis of a simpler version of this type of experiment [7–11], including a purely qualitative description, which can be understood by any interested person with practically no math.

Special Relativity and Motions Faster than Light. Moses Fayngold

ISBN: 3-527-40344-2

Fourth, there exists another "gap" in a vast pool of books (and textbooks especially) on the special theory of relativity: the significant lack of coverage of accelerated motions. This has produced another long-standing and widespread misconception (even among professional physicists!) that the theory is restricted to inertial (uniform) motions of particles that are not subject to external forces. I was surprised to find even in recently published books statements that the special theory of relativity is incomplete because it cannot describe accelerated motions of any kind.

Nothing can be farther from the truth than such statements. How could the particle accelerators that are routinely used in high-energy physics have been designed and work properly without the special theory of relativity? One of the goals of this book is to dispel the myth that accelerated motions cannot be treated in the framework of the Special Relativity. The reader will find a standard treatment of accelerated motion in Chapter 4, which is devoted especially to the *relativistic dynamics* of a point mass. In Chapter 5 we describe subtle phenomena associated with accelerated motion of extended bodies (Sects. 5.4 and 5.5), and motions in rotating reference frames, including famous experiments with the atomic clocks flown around the Earth (see references in Chap. 5, Sects. 5.7 and 5.8). In Chapter 6 the reader will find a description of the rotational motion of a rod and motion of charged particles in a magnetic field (Sects. 6.3 and 6.4), and in Chapter 8 accelerated superluminal motion is considered (Sects. 8.10 and 8.12).

Rather than being a textbook or a monograph, the book is a self-consistent collection of selected topics in Special Relativity and adjacent areas, which are all arranged in a logical sequence. They have been selected and are discussed in such a way as to provide the above-mentioned categories of readers with interesting material for study or future thought. The book provides numerous examples of some of the most paradoxical-seeming aspects of the theory. What can contribute more to the real understanding of a theory than resolving its paradoxes? Paraphrasing Martin Gardner [3], "you have to know where and why opponents of Einstein go wrong, to know something about relativity theory."

The first three chapters cover traditional topics such as the Michelson–Morley experiment, Lorentz transformations, etc.

A few chapters deal with the strange world of superluminal velocities and tachyons, and other topics hardly to be found elsewhere. Their investigation takes us to the boundaries of the permissible in relativity theory, exploring the remote domains of superluminal phenomena, while at the same time serving as the foundation of a deeper understanding of Einstein's unique contribution to scientific thought.

Initially the appearance of the theory of relativity, with its absolute insistence that no signal carrying information can travel faster than light in a vacuum, created the opinion among many that no superluminal motion of any kind was possible. In this book a great many phenomena are described in which superluminal motion seems to appear or does appear. Such phenomena may occur in some astrophysical processes, in physical laboratories, and even in everyday life. However incredible some of them might seem, they are all shown to be in accordance with Special Relativity, since in an almost mysterious accord with the overriding dictates of the theory, subtle details always conspire to insure that none of these phenomena can be used

for signal transmission faster than light in a vacuum. And Special Relativity is just the kind of theory for describing adequately this kind of motion.

A couple of decades ago there was a great controversy in the scientific literature about hypothetical superluminal particles – tachyons. After extensive discussion it was decided by the overwhelming majority of physicists that tachyons cannot exist since their existence would bring about violations in causality, plunging the Universe into unresolvable paradoxes, by changing the past. There are numerous papers which argue that the kind of tachyon hypothesized in the early discussions cannot exist (see the references in Chap. 8). Yet the reader of this book will find a description of real tachyon-like objects that can be "manufactured" in the laboratory. They possess a kind of duality, which allows one to represent a tachyon-like object as either a superluminal or subliminal object, depending on what physical quantities are chosen for its description.

Many of these topics are hardly to be found elsewhere, and some of them have so far only been published in a few highly specialized professional journals. In this respect this book should be a unique source of information for broad categories of readers.

As already mentioned above, the book is intended to satisfy also the demands of those readers with a minimal background in math. They will find in many descriptions an easy part showing the inner core of a phenomenon, its physical picture. These readers can stop at this point – they have grasped the main idea.

For the better prepared, after they have been made capable of seeing the rather complicated features involved, there follows a quantitative description with the equations and other details. Many of the examples discussed are unusual and thought provoking; they often start as unsolvable paradoxes, to be, after a few unexpected turns, finally resolved. One can find an example of such an approach in Chapter 5, Section 5.4.

Another example of this approach can be found in the discussion of phase and group velocities (Chap. 6, Sect. 6.12). They are discussed on three different levels. The first – intuitive – gives a pictorial representation of the phenomenon using a simple model. This will help the beginner *with no math at all* to grasp the relationship between the two velocities. Then the same relationship is obtained graphically. Finally, it is obtained by analyzing the superposition of two wave functions. The last two levels are appropriate for everybody familiar with college math. The first one may be good for two extreme categories of reader: the least prepared at the one pole, and the most sophisticated (e. g. college professors) at the other. The former may find it good to learn, while the latter may find it good to teach.

In summary, the book can be used as supplementary reading for college students taking courses in physics. High school and college teachers can use it as a pool of examples for class discussion. Further, because it contains much new material beyond standard college programs, it may be of interest for all those curious about the workings of Nature. A mathematical background on the undergraduate level will be helpful in understanding quantitative details. More advanced readers can find in the book much thought-provoking material, and professional physicists, while skipping the topics that are familiar to them, or written on the elementary level, may well find some new insights there or see a problem in a fresh light.

Acknowledgements

I am grateful to Boris Bolotowsky, Julian Ivanchenko, and Gregory Matloff, who encouraged me to keep on working on the book on its earlier stages. Stephen Rosen and Leo Silber helped me with their comments and good advise. Slawomir Piatek spent much of his time discussing with me a few sample chapters, and I used his insightful remarks in the revised version of the text. Yury Abramian in faraway Armenia helped me in my searches for a few references in Russian scientific literature. My elder son Albert made the front cover of the book. Roland Wengenmayr, in an extensive collaboration, which I found very rewarding, turned Albert's and mine initial crude sketches into line drawings, and then created in his illustrations a series of characters, which, in my opinion, perfectly match the text.

My special thanks to my younger son Vadim for his vicious, but constructive criticism of the first drafts of the manuscript and for his invaluable technical help; and also to David Green for his time and angelic patience in translating my version of the English language into English (any remaining linguistic and other errors that might have survived and slipped into the final text are to be blamed entirely on me).

I wish to thank the consulting editor Edmund Immergut for his professional guidance in finding the most appropriate publisher for this book.

I enjoyed working with Vera Palmer, the publishing editor at Wilew-VCH, and, on the latest stages, with Melanie Rohn and Peter Biel in the intensive copy-editing process.

I am deeply grateful to my wife Sophie who did all in her power to save me more time for writing.

Special Relativity and Motions Faster than Light. Moses Fayngold

ISBN: 3-527-40344-2

1
Introduction

1.1
Relativity? What is it about?

One of the cornerstones of the Special Theory of Relativity is the Principle of Relativity. A good starting point for discussing it may be a battlefield. So imagine a battlefield with deadly bullets whistling around and let me ask a question: could you catch such a bullet with your bare hands?

The likely answer is: "Not I. You'd better try to do it yourself!"

Which implies: that's impossible.

I remember that, as a schoolboy, I had given precisely the same answer to this question. But then I read a story about a pilot in World War I who had in one of his flight missions noticed a strange object moving alongside the plane, right near the cockpit. The cockpits could easily be opened in those times, so the pilot just stretched out his arm and grabbed the object. He saw that what he had caught was ... a bullet. It had been fired at his plane and was at the final stage of its flight when it caught up with the plane and was caught itself.

The story shows that you really can catch a flying bullet. Nowadays, having spaceships, one can, in principle, catch a ballistic missile. Assuming unlimited technological development, we do not see anything that would prevent us from "catching" any object by catching up with it – be it a solid, a liquid, or a jet of plasma – no matter how fast it is moving. If a natural object had been accelerated to a certain speed, then a human being, who is also a natural object, can (although, perhaps, at a slower rate) be accelerated up to the same speed.

We see that the velocity of an object is a sort of "flexible" characteristic. The bullet that is perceived by a ground-based observer to be moving appears to be at rest to the pilot. We will call such quantities observer-dependent, or relative.

Not all of the physical quantities are relative. Some of them are observer-independent, or absolute. For example, the pilot may have noticed that the bullet he had caught was made of lead and coated with steel, and the mass ratio of lead and steel in it is 24 : 1. This property of the bullet is absolute because it is true for anyone independently of one's state of motion. The gunner who had fired the bullet will agree with the pilot on the ratio 24 : 1 characterizing its composition. But he will disagree on its velocity. He will hold that the bullet moves with high speed whereas it is obviously at rest for the pilot.

Special Relativity and Motions Faster than Light. Moses Fayngold

ISBN: 3-527-40344-2

Another example: if a car with three passengers has a velocity 45 miles per hour, then the fact of its having this velocity is of a quite different category to the fact of its having three passengers inside. The latter is absolute because it is true for anyone regardless of one's state of motion. The former is relative because it is only true for those standing on the ground. But it is false, say, for a driver in another car moving along the same straight road. The driver will agree with you on the number of passengers in the first car but disagree on its velocity. He may hold that the first car has zero velocity because it has always been at the same distance from him.

Who is right – you or the second driver? Both are. And there is no contradiction here, because each observer relates what he sees to his own "reference frame".

Moreover, even one and the same observer can measure different velocities of the same object. A policeman in a car using radar for measuring speeds of moving objects will register two different values for the velocity of a vehicle, if he measures the velocity the first time when his own car just stands on the road, and the second time when his car is moving. We emphasize that nothing happens to the observed vehicle, it remains in the same state of motion with constant speed. And yet the value of this speed as registered by the radar is different for the two cases.

We thus see that the value of a speed does not by itself tell us anything. It only becomes meaningful if you specify *relative to what* this speed is measured. This is what we mean by saying that speed (more generally, velocity) is a relative physical quantity.

Understanding the relative nature of some physical quantities (and the absolute nature of some others) is the first step to acquiring the main ideas of Special Relativity. We will in this book outline its most characteristic features with all the contradictions between the old and new concepts.

Let us start first with an account of the theory of relativity widespread among general public:

"Einstein has proved that everything is relative. Even time is relative."

One of these statements is true and profoundly deep; the other one is totally misleading.

The true statement is that *time is relative*. Realization of the relative nature of time was a revolutionary breakthrough in our understanding of the world.

The wrong statement in the above "popular" account of relativity is that *everything* is relative. We already know that, for instance, the number of passengers in a car (or the chemical composition of a bullet) is not relative. One of the most important principles in relativity is that *certain physical quantities are absolute (invariant)*. One such invariable quantity is the speed of light in a vacuum. Also, certain combinations of time and distance turn out to be invariant. We will discuss these absolute characteristics in the next chapter. They are so important that we might as well call the theory of relativity the theory of absoluteness. It all depends on which aspect of the theory we want to emphasize.

We will now discuss the relativity aspect. Let us first recall the classical principle of relativity in mechanics. Suppose you are inside a train car that moves uniformly along a straight track. If the motion is smooth enough then, unless you look out of the window, you cannot tell whether the train is moving or is at rest on the track. For

instance, if you drop a book, it will fall straight down with the same acceleration, as it would do on the stationary platform. It will hit the floor near your feet, as it would do on the platform. If you play billiards, the balls will move, and collide, and bounce off in precisely the same manner as they do on the platform. And all other experiments will be indistinguishable from those on the platform. There is no way to tell whether you are moving or not by performing mechanical tests. This means that the states of rest or uniform motion are equivalent for mechanical phenomena. There is no intrinsic, fundamental difference between them. This general statement was formulated by Galileo, and it came to be known as his principle of relativity. According to this principle, the statement: "My train is moving" has no absolute meaning. Of course, you can find out that it is moving the moment you look out of the window. But the moment you do it, you start referring all your observations to the platform. You then can say: "My car is moving relative to the platform." Platform constitutes your reference frame in this case. But you may as well refer all your data to the car you are in. Then the car itself will be your reference frame, and you may say: "My car is at rest, while the platform is moving relative to it." Now, pit the last two quoted statements against each other. They seem to be in contradiction, but they are not, because they refer to different reference frames. Each statement is meaningful and correct, once you specify the corresponding frame of reference.

We see that the concept of reference frame plays a very important role in our description of natural phenomena. We can even reformulate the principle of relativity in terms of reference frames. To broaden the pool of examples (and make the further discussion more rigorous!) we will now switch from jittering trains, and from the spinning Earth with its gravity, far into deep space. A better (and more modern) realization of a suitable reference frame would be a non-rotating spaceship with its engines off, coasting far away from Earth or other lumps of matter. Suppose that initially the ship just hangs in space, motionless with respect to distant stars. You may find this an ideal place to check the basic laws of mechanics. You perform corresponding experiments and find all of them confirmed to even higher precision than those on Earth.

If you release a book, it will not go down; there is no such thing as "up" or "down" in your spaceship, because there is no gravity in it. The book will just hang in the air close by you. If you give it an instantaneous push, it will start to move in the direction of the push. Inasmuch as you can neglect air resistance, the book will keep on moving in a straight line with constant speed, until it collides with another object. This is a manifestation of Newton's first law of motion – the famous law of inertia. Then you experiment with different objects, applying to them various forces or combinations of forces. You measure the forces, the objects' masses, and their response to the forces. In all cases the results invariably confirm Newton's second law – the net force accelerates an object in the direction of the force, and the magnitude of the acceleration is such that its product by the mass of the object equals the force. This explains why the released book does not go down – in the absence of gravity it does not know where "down" is. With no gravity, and possible other forces balanced, the net force on the book, and thereby its acceleration, are zero. Then you push against the wall of your compartment and immediately find yourself being pushed back by

the wall and flying away from it. This is a manifestation of Newton's third law: forces always come in pairs; to every action there is always an equal and opposite reaction.

Let us now stop for a while and make a proper definition. Call a system where the law of inertia holds an inertial system or inertial reference frame. Then you can say that your ship represents an inertial system. So does the background of distant stars relative to which the ship is resting.

Suppose now that you fall asleep, and during your sleep the engines are turned on. The spaceship is propelled up to a certain velocity, after which the engines are turned off again. You are still asleep, but the ship is now in a totally different state of motion. It has acquired a velocity relative to the background of stars, and it keeps on coasting with this velocity due to inertia. The magnitude of this velocity may be arbitrary. But even if it is nearly as large as that of light, it will not by itself affect in any way the course of events in the ship. After you have woken up and checked if everything is functioning properly, you don't find anything unusual. All your tests give the same results as before. The law of inertia and other laws hold as they had done before. Your ship therefore represents an inertial reference frame as it had been before. Unless you look outside and measure the spectra of different stars, you won't know that your ship is now in a *different state of motion* than it had originally been. The reference frame associated with the ship is therefore also different from the previous one. But, according to our definition, it remains inertial.

What conclusions can we draw from this? First: any system moving uniformly relative to an inertial reference frame is also an inertial reference frame. Second: all the inertial reference frames are equivalent with respect to all laws of mechanics. The laws are the same in all of them. The last statement is the classical (Galilean) principle of relativity expressed in terms of the inertial reference frames.

The classical principle of relativity is very deep. It seems to run against our intuition. In the era of computers and space exploration, I may still happen to come across a student in my undergraduate physics class who would argue that if a passenger in a uniformly moving train car dropped an apple, the apple would not fall straight down, but rather would go somewhat backwards. He or she reasons that while the apple is falling down, the car is being pulled forward from under it, which causes the apple to hit the floor closer to the rear of the car. This argument (which overtly invokes the platform as a fundamental reference frame) overlooks one crucial detail: before being dropped, the apple in the passenger's hand had moved forward together with the car. This pre-existing component of motion persists in the falling apple due to inertia and exactly cancels the effect described by the student, so that the apple as seen by an observer in the car will go down strictly along its vertical path (Fig. 1.1). This conclusion is confirmed by innumerous observations of falling objects in moving cars. It is a remarkable psychological phenomenon that sometimes not even such strong evidence as direct observation can overrule the influence of a more ancient tradition of thought. About a century and a half ago, when the first railways and trains appeared, some people were afraid to ride in them because of their great speed. The same story repeated with the emergence of aviation. Many people were afraid to board a plane not only because of the altitude of flight, but also because of its great speed. Apart from the fear of a *collision* at high speed, it might have been

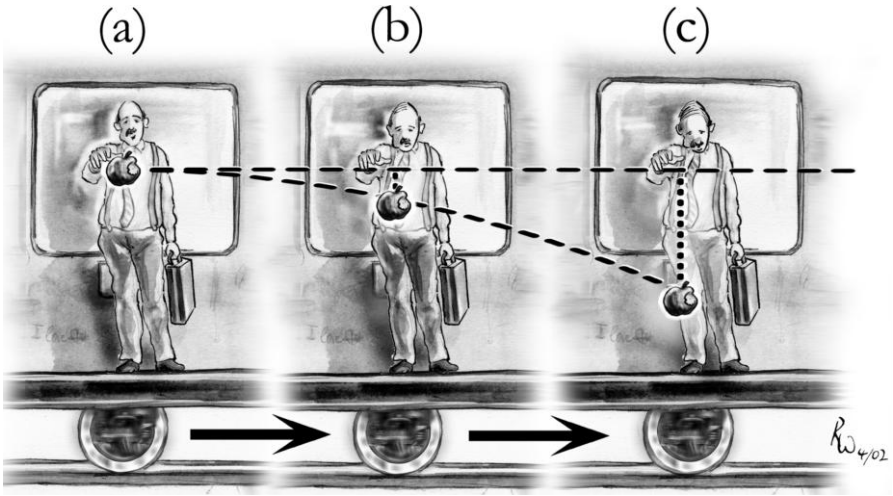


Fig. 1.1 The fall of an apple in a moving car as observed from the platform. (a), (b), and (c) are the three consecutive snapshots of the process. The passenger sees the apple fall vertically, while it traces out a parabola relative to the platform. The shape of a trajectory turns out to be a relative property of motion.

the fear of the speed itself. Many believed that something terrible would happen to them at such a speed. It took a great deal of time and new experience to realize that speed itself, no matter how great, does not cause any disturbance in the regular patterns of natural events so long as velocity remains constant. It is the *change* of velocity (deceleration, acceleration) during braking, collision, or turning that can be felt and manifest itself inside a moving system. If you are in a car that is slowing, you can immediately tell this by experiencing a force that pushes you forward. Likewise, if the car accelerates, everything inside experiences a force in the backward direction. It is precisely because of these forces that I wanted you to fall asleep during the acceleration of the spaceship, otherwise you would immediately have noticed the appearance of a new force and known that your ship was changing its state of motion, which I did not want you to do.

A remarkable thing about this new force is that it does not fit into the classical definition of a real force. It appears to be real because you can observe and measure it; you have to apply a real force to balance it; when unbalanced, it causes acceleration, as does any real force; it is equal to the product of a body's mass and acceleration, as is any real unbalanced force. In this respect, it obeys Newton's second law. Yet it appears to be fictitious if you ask the questions: Who exerts this force? Where does it come from? Then you realize that it, unlike all other forces in Nature, does not have a physical source. It does not obey Newton's third law, because *it is not a part* of an action–reaction pair. You cannot find and single out a material object producing this force, not even if you search out the whole Universe. Unless, of course, you prefer to consider the whole universe becoming its source when the universe is accelerated past your frame of reference.

The new force has been called the inertial force – and for a good reason. First, it is always proportional to the mass of a body to which it is applied – and mass is the measure of the body's inertia. In this respect, it is similar to the force of gravity. Second, its origin can be easily traced to a manifestation of inertia. Imagine two students, Tom and Alice. They both observe the same phenomenon from two different reference frames. Tom is inside a car of a train that has just started to accelerate, while Alice is on the platform. Alice's reference frame is, to a very good approximation, inertial, whereas Tom's is not. Tom looks at a chandelier suspended from the car's ceiling. He notices that the chandelier deflects backwards during acceleration. He attributes it to a fictitious force associated with the accelerating universe. Alice sees the chandelier from the platform through the car's window (Fig. 1.2), but she interprets what she sees quite differently. "Well," she says, "this is just what should be expected from the Newton's laws of motion. The unbalanced forces are exerted on the car by the rails and, maybe, by the adjacent cars, causing the car to accelerate. However the chandelier, which hangs from a chain, does not immediately experience these new forces. Therefore it retains its original state of motion, according to the law of inertia, which holds in my reference frame. At the start the chandelier accelerates back relative to the car only because the car accelerates forward relative to the platform. This transitional process lasts until the deflected chain exerts sufficient horizontal force on the chandelier."

"Finally," Alice concludes, "this force will accelerate the chandelier relative to the platform at the rate of the car, and there will be no relative acceleration between the car and chandelier." All the forces are accounted for in Alice's reference frame. In Tom's reference frame, the force of inertia that keeps the chandelier with the chain

Fig. 1.2 A chandelier in an accelerated car. To Alice, the tension force in the deflected chain acquires a horizontal component causing the chandelier to accelerate at the same rate as the car. Tom explains the deflection of the chandelier as the result of the inertial force. This force balances the horizontal component of the chain's tension.

off the vertical is felt everywhere throughout the car, but cannot be accounted for. This state of affairs tells Tom that his car is accelerating.

Tom has also brought along an aquarium with fish in it. When the train starts to accelerate, both Tom and Alice see the water in the aquarium bulge at the rear edge and subside at the front edge, so that its surface forms an incline (Fig. 1.3). Alice interprets this by noticing that the rear wall of the aquarium drives the adjacent layers of water against the front layers, which tend to retain their initial velocity. This causes the rear layers to rise. In contrast, the front layers sink because the front wall of the fish tank accelerates away from them, so the water surface tilts.

Tom does not see any accelerated motions within his car, but he feels the horizontal force pushing him towards the rear. "Aha," Tom says, "this force seems to be everywhere indeed. It pushes me and the chandelier back, and now I see it doing the same to water. It is similar to the gravity force, but it is horizontal and seems to have no source. Its combination with the Earth-caused gravity gives the net force tilted with respect to the vertical line." Being as good a student as Alice, Tom knows that the water surface always tends to adjust itself so as to be perpendicular to the net force acting on it. Since the latter is now tilted towards the vertical, the water surface in the aquarium becomes tilted to the horizontal by the same angle. The only trouble is that there is no physical body responsible for the horizontal component of the net force. "This indicates," Tom concludes, "that the horizontal component is a fictitious inertial force caused by acceleration of my car."

In a similar way, one can detect a rotational motion, because the parts of a rotating body accelerate towards its center. We call this centripetal acceleration. For instance, we could tell that the Earth is rotating even if the sky was always cloudy so that we

Fig. 1.3 The water in an accelerated fish tank. The rear wall of the tank rushes upon the water, raising its adjacent surface, while the front wall accelerates away from the water, giving it extra room in front, which causes the water there to sink. To Tom, tilt of the water surface is caused by inertial force. The tilted chain of the chandelier makes the right angle with the tilted water surface.

would be unable to see the Sun, Moon, or stars. That is, we could not "look out of the window." But we do not have to. Many mechanical phenomena on Earth betray its rotation. The earth is slightly bulged along the equator and flattened at the poles. A freely falling body does not fall precisely along the vertical line (unless you experiment at one of the geographical poles). A pendulum does not swing all the time in one plane. Many rivers tend to turn their flow. Thus, in the northern hemisphere, rivers are more likely to have their right banks steep and precipitous and the left ones shallow. In one-way railways, the right rails wear out faster than the left ones because the rims of the trains' wheels are pressed mostly against the right rail. In the southern hemisphere the situation is the opposite. It is easier to launch a satellite in the east direction than in the north, south or west direction. All these phenomena are manifestations of the inertial forces.

We will illustrate the origin of these forces with a simplified model of a train moving radially on a rotating disk. Suppose that the train is moving down a radial track towards the center of the disk, and you observe this motion from an inertial stationary platform (Fig. 1.4). At any moment the instantaneous velocity of the train relative to the platform has two components: radial towards the center and transverse, which is due to the local rotational velocity of the disk. The peripheral parts of the disk have higher rotational velocity than the central ones. As the train moves toward the center, it tends, following the law of inertia that holds on the platform, to retain the larger rotational velocity "inherited" from the peripheral parts of the disk. This would immediately cause derailment on to the right side of the track, had it not been for the

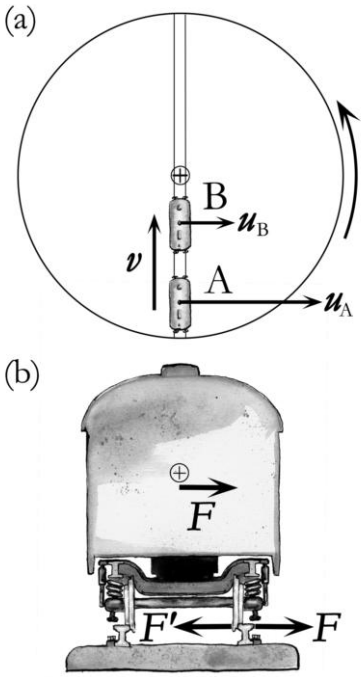


Fig. 1.4 Schematic of the inertial forces acting on a moving car in a rotating reference frame. (a) View from above. The train moves from A to B with speed v. Owing to inertia, the train tends to transport its original rotational velocity $\boldsymbol{u}_A$ from A to B. Since $\boldsymbol{u}_A$ is greater than $\boldsymbol{u}_B$, the train experiences transverse inertial force $\boldsymbol{F}$. (b) View from behind. The force $\boldsymbol{F}$ is balanced by force $\boldsymbol{F}'$.

wheels' rims that hold the train on the rails. The same effect causes the overall asymmetry between the left and right banks of rivers. We thus see that these phenomena are, in fact, manifestations of the inertia. Their common feature is that they permeate all the space throughout an accelerated system, and cannot be attributed to an action of a specific physical body. Because of them, the Earth can be considered as an inertial system only to a certain approximation. Careful observation reveals the Earth's rotation without anyone ever having to look up into the sky.

All these examples show that inertial systems in classical physics form a very special class of moving systems. The world when looked upon from such a system looks simpler because there are no inertial forces. You can consider any inertial system as stationary by choosing it to be your reference frame without bringing along any inertial forces. There is no intrinsic physical difference between the states of rest and uniform motion. All other types of motion are absolute in a sense that nature provides us with the criterion that distinguishes one such motion from all the others. We can also relate all observational data to an accelerated system and consider it motionless. However, there are intrinsic physical phenomena (inertial forces) that reveal its motion relative to an inertial reference frame. Not only can we detect this motion without "looking out of the window," we also can determine precisely all its characteristics, including the magnitude and direction of acceleration, the rate of rotation, and the direction of the rotational axis.

We thus arrive at the conclusion that Nature distinguishes between inertial and accelerated motions. It does not mean at all that the theory cannot describe accelerated motions. It can, and we will see examples of such a description later on in the book. The special theory of relativity can even be formulated in arbitrary accelerated and therefore non-inertial reference frames [13]. However, the description of motion in such systems is far less straightforward, to a large extent because of the appearance of the inertial forces. The General Theory of Relativity reveals deep connections between inertial forces in an accelerated system and gravitation. We will in this book be concerned with Special Relativity.

1.2 Weirdness of Light

The special theory of relativity has emerged from studies of the motion of light.

Let us extend our discussion of motions of physical bodies to situations involving light. Previously we had come to the conclusion that one can catch up with any object. Does this statement include light? This question was torturing a high school student, Albert Einstein, about a century ago and eventually brought him to Special Relativity. What we have just learned about velocity prompts immediately a positive answer to the question. Velocity is a relative quantity, it depends on a reference frame. It can be changed by merely changing the reference frame. For instance, if an object is moving relative to Earth with a speed v, we can change this speed by boarding a vehicle moving in the same direction with a speed V. Then the speed of the object relative to us will be

$$v' = v - V \tag{1}$$

We can change v by "playing" with the vehicle – accelerating or decelerating it. For instance, reversing the speed of the vehicle would result in changing the sign of V in the above equation, and, accordingly, would greatly increase the relative speed of the object without touching it. If we want to catch up with the object, we need to bring its relative velocity down to zero. We can do this by accelerating the vehicle to the speed $V = v$.

Because this works for objects such as bullets, planes, or baseballs, people naturally believed that is should also work for light. It is true that we never saw light at rest before. However, as an old Arabic saying has it, "if the mountain does not go to Mohammed, then Mohammed must go to the mountain." If we cannot stop the light on Earth, then we have to board a spaceship capable of moving relative to Earth as fast as light does, and use this "vehicle" to transport us in the direction of light. Let c be the speed of light relative to Earth, and V be the speed of a spaceship also relative to Earth. If Equation (1) is universal, then we can apply it to this situation and expect that the speed c' of light relative to the spaceship will decrease by the amount V:

$$c' = c - V \tag{2}$$

Suppose that the rocket boosters accelerate the spaceship; its velocity V increases, and c' decreases. When V becomes equal to c, the speed c' becomes zero. In other words, light stops relative to us, that is, we have caught up with light. The same principle that helped the pilot catch a bullet works here to help us catch the light. The law (1) of addition of velocities says that it is possible.

However, there immediately follows an interesting conclusion. We know that the Earth can to a good approximation be considered as an inertial reference frame, and all inertial reference frames, according to mechanics, are equivalent. Einstein thought that this principle could be extended beyond mechanics to include all natural phenomena. If this is true, then whatever we can observe in one inertial system can also be observed in any other inertial system. If light can be stopped relative to at least one spaceship, then it can be brought to rest relative to any other inertial system, including Earth. In physics, if Mohammed can come to the mountain, the mountain can come to Mohammed. To stop light relative to the spaceship, we need to accelerate the ship up to the speed of light. To stop light relative to earth, we may, for example, put a laser gun on this ship, and fire it backwards. Then the laser pulse, while leaving the ship with velocity c relative to it, will have zero velocity with respect to Earth. We will then witness a miraculous phenomenon of stopped light.

I can imagine an abstract from a science fiction story exploiting such a possibility, something running like this:

"Mary stretched her arm cautiously and took the light into her hand. She felt its quivering wave-like texture, which was constantly changing in shape, brightness, and color. Its warm gleam has gradually penetrated her skin and permeated all her body, filling it with an ecstatic thrill. She suddenly felt a divine joy, as though a new glorious life was being conceived in her."

But, alas! Beautiful and tempting as it may seem, our conclusion that freely traveling light can be stopped relative to Earth, or whatever else, is not confirmed by observation. It stands in flat contradiction with all known experiments involving light. As had already been established before Einstein's birth, light is electromagnetic waves. The theory of electromagnetic phenomena, developed by J. C. Maxwell, shows a remarkable agreement with experiments. And both theory and experiments show quite counterintuitive and mysterious behavior of light: not only is it impossible to catch up with light; it is impossible even to change its speed in a vacuum by a slightest degree, no matter what spaceship we board or in what direction or how fast it moves.

We have arrived at a deep puzzle. Light does not obey the law of addition of velocities expressed by Equation (1). The equation appears to be as fundamental as it is simple. And yet there must be something fundamentally wrong about it.

"Wait a minute," the reader may say, "Equation (1) is based on a vast number of precise experiments. It is therefore absolutely reliable, and it says that ..."

"What it says is true for planes, bullets, planets, and all the objects moving much slower than light. But it is not true for light," I answer.

"Well, look here: the speed of light as measured in experiments on Earth is about 300 000 km s^{-1}. Suppose a spaceship passes by me with the velocity 200 000 km s^{-1}, and I fire the laser pulse at the same moment in the same direction. Then 1 s later the laser pulse will be 300 000 km away from me, whereas the spaceship will be 200 000 km away. Is it correct?"

"Absolutely."

"Well, then, it must be equally true that the distance between the spaceship and the pulse will be 100 000 km, which means that the laser pulse makes 100 000 km in 1 s relative to the spaceship. It is quite obvious!"

"Apparently obvious, but not true."

"How can that be?"

"This is a good question. The answer to it gives one the basic idea of what relativity is about. You will find the detailed explanations in the next chapter. It starts with the analysis of one of the best known experiments that have demonstrated the mysterious behavior of light mentioned above. But in order to understand it better, let us first recall a simple problem from an Introductory Course of College Physics."

1.3
A steamer in the stream

The following is a textbook problem in non-relativistic mechanics; however, its solution may be essential for understanding one of the experimental foundations of Special Relativity.

So, let us begin!

A steamer has a speed of u km h^{-1} relative to water. How long will it take to swim the distance L km back and forth in a lake? The answer is

$$t_0 = 2\frac{L}{u} \tag{3}$$

"Is that all?," the reader may ask. No. It is just a preliminary exercise. The problem is this: the same steamer starts at point A on the bank of the river with the stream velocity v km h^{-1}. It moves downstream to the point B on the same bank at a distance L from A, immediately turns back and moves upstream. How long will it take to make round trip from A to B and back to A?

This is just a bit more complicated but still simple enough. Our reasoning may run like this: if the steamer makes u km h^{-1} relative to water, and the stream makes v km h^{-1} relative to the bank, then the steamer's velocity relative to the bank is $(u + v)$ km h^{-1} when downstream and $(u - v)$ km h^{-1} when upstream. We are interested in the resulting time, which is determined by the ratios of the distance to velocities. We must therefore use the speed averaged *over time*. The total time consists of two parts: one (t_{AB}), which is needed to move from A to B, and the other (t_{BA}) to move back from B to A. The time t_{BA} is always greater than t_{AB}, since the net velocity of the steamer is less during this time. Thus, the net velocity of the steamer is greater than u during the shorter time, and less than u by the same amount during the longer time. Therefore, its average over the whole time is less than u. As a result, the total time itself must be greater than t_0. It must become ever greater as v gets closer to u. This result becomes self-evident when $v = u$. Then the steamer after turning back is carried down by the stream at the same rate as it makes in the up direction. So it will just remain at rest relative to the bank at B, and will never return to A. This is the same as to say that it will return to A in the infinite future, that is, the total time is infinite.

What if v becomes greater than u, that is, the stream is faster than the steamer? Then the steamer after the turn is even unable to remain at B; it will be dragged down by the stream, getting ever further away from its destination. We can formally describe this situation by ascribing a negative sign to the total time t.

Let us now solve the problem quantitatively. The times it takes to go from A to B and then from B to A are, respectively,

$$t_{AB} = \frac{L}{u+v}\,, \quad t_{BA} = \frac{L}{u-v} \tag{4}$$

So the total time

$$t_{\downarrow\uparrow} = t_{AB} + t_{BA} = \frac{L}{u+v} + \frac{L}{u-v} = \frac{t_0}{1-\dfrac{v^2}{u^2}} \tag{5}$$

where t_0 is the would be time in the still water, given by Equation (3).

If we plot the dependence in Equation (5) of time against the stream velocity, we obtain the graph shown in Figure 1.5.

Equation (5) describes symbolically in one line all that was written over the whole page and, moreover, it provides us with the exact numerical answer for each possible situation. The graph in Figure 1.5 describes all possible situations visually. You see

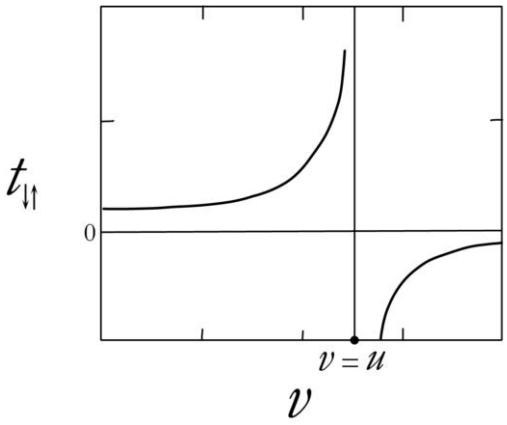


Fig. 1.5 The dependence of round-trip time $t_{\downarrow\uparrow}$ on speed v.

that for all $v < u$ the time t is greater than t_0, it becomes infinite at $v = u$, and negative at all $v > u$. When v is very small relative to u, Equation (5) gives $t_{\downarrow\uparrow} \approx t_0$. This is natural, since for small v the impact of the stream is negligible, and we recover the result in Equation (3) obtained for the lake.

Now, consider another case. The river is L km wide. The same steamer has to cross it from A to B right opposite A on another bank, and then come back, so the total distance to swim relative to the banks is again $2L$. How long will it take to do this?

The only thing we have to know to get the answer is the speed of the steamer u' in the direction AB right across the river. The steamer must head all the time a bit upstream relative to this direction to compensate for the drift caused by the stream. If during the crossing time the steamer has drifted l km downstream, then in order to get to B, it must head to a point B′ l km upstream of B. Thus, its velocity relative to water is u and directed along AB′, the velocity of the stream is v and directed along B′B, and the resulting sought for velocity of the steamer relative to the banks is directed along AB. These three velocities form a right triangle (Fig. 1.6), and therefore

$$u' = \sqrt{u^2 - v^2} = u\sqrt{1 - \frac{v^2}{u^2}} \tag{6}$$

Hence our final answer for the total time back and forth between A and B is

$$t_\perp = \frac{2L}{u'} = \frac{2L}{u\sqrt{1 - \frac{v^2}{u^2}}} = \frac{t_0}{\sqrt{1 - \frac{v^2}{u^2}}} \tag{7}$$

Note that Equations (6) and (7) give a meaningful result only when $v < u$ (a side of the right triangle is shorter than the hypotenuse). Then, according to Equation (7), time $t_\perp$ is also greater than t_0, but it is less than $t_{\downarrow\uparrow}$. Hence one can write

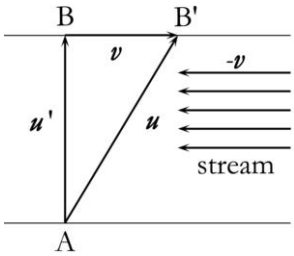


Fig. 1.6

$$t_0 < t_\perp < t_{\downarrow\uparrow} \tag{8}$$

If $v > u$, the triangle in Figure 1.6 cannot be formed. The steamer's drift per unit time exceeds its velocity u, and the steamer will not be able to reach the point B, let alone return to A. This circumstance is reflected in the mathematical structure of Equations (6) and (7): these equations yield imaginary numbers when $v > u$. They say that there is in this case no physical solution that would satisfy the conditions of the problem.

Now, what is the link between this problem and the experiment with light mentioned above? Take the running waves on the water surface instead of the steamer and you turn the mechanical problem into a hydrodynamic one. Then take the sound waves in air during the wind instead of the steamer on the water stream, and you get the same problem in fluid dynamics. And as the last step, consider the light that propagates in a moving transparent medium in transverse and longitudinal directions, and here you are with the optical problem that is identical with the initial mechanical one.

This is why I started the book with this Introductory Physics problem. On the one hand, its mathematical description is exactly the same as that of the problem ahead. On the other hand, its solution is psychologically easier just because a mechanical problem is more familiar to a great majority of people. I believe that even the less advanced students will feel more comfortable with this book if it starts with a familiar problem.

However, I want to stress here again that the treatment of the problem is based on unspoken assumptions about addition of velocities, which were shown later to be incorrect. The corresponding errors in the results obtained are negligible for a steamer or for sound in air, so we can use them for these cases; but they may become large in the case of light. What the physical nature of these misconceptions is, and how they are related to the nature of light, are discussed in the next chapter.

2
Light and Relativity

2.1
The Michelson experiment

In the history of the study of the world, one can trace a tendency to explain the greatest possible number of phenomena using the smallest number of basic principles. In the eighteenth and nineteenth centuries it seemed that the solution of this task was not far off. This period witnessed a spectacular flourish of Newtonian mechanics. Using its basic concepts, scientists made astonishing progress in astronomy, navigation, technology, earth studies, etc. Later the advance of the molecular–kinetic theory allowed the huge field of thermodynamic phenomena to be described in the language of mechanics.

This engendered a hypothesis that *all* natural phenomena can be reduced to mechanics, that is, one could construct an entirely mechanical picture of the world – a picture based on the laws of Newton and on the corresponding concepts of absolute time and space. Consequently, physicists sought to integrate electromagnetic phenomena and particularly the propagation of light into mechanical theory.

By that time it had been proved that light propagation is a wave process for which the phenomena of interference and diffraction, common for all waves, could be observed. And since all waves known in mechanics could propagate only in some medium with elastic properties, it seemed reasonable to assume that light waves are also mechanical oscillations of some elastic medium which penetrates all physical objects and fills all space in the Universe. This hypothetical medium was called the ether.

The ether hypothesis leads to a number of inferences, the examination of which may confirm or refute the hypothesis itself. In this section we will consider one of such inferences, the analysis of which has played an important role in the history of science.

Let us assume that the space is filled with ether. Then, since the Earth is traveling through the ether, an earthly observer may expect to discover an "ether wind." The speed of light in the ether as measured by the earthly observer may in this case depend on direction. If the wind has a speed v relative to the Earth, the observer would expect to measure for the speed of light $c_{\uparrow} = c + v$ in the direction of the wind and $c_{\downarrow} = c - v$ in the opposite direction. And what is the speed of light in the transverse

Special Relativity and Motions Faster than Light. Moses Fayngold

ISBN: 3-527-40344-2

direction? In order for light to move perpendicularly to the wind it is necessary to compensate for the lateral "drift," which means that the light's velocity relative to the ether must have a longitudinal component against the wind, equal to v. However, the total velocity of light relative to the ether is equal to c. Therefore, according to our results in the previous section, the transverse component must be equal to $c_\perp = \sqrt{c^2 - v^2}$ (Fig. 1.6 with $u = c$ and $u' = c_\perp$). If our reasoning is correct, the speed of light relative to the Earth must be anisotropic (that is, dependent upon the direction) owing to the Earth's motion in the ether. Conversely, an observation of such anisotropy would enable us to detect this motion and to find its speed. In other words, optical phenomena would reveal a fundamental difference between a moving reference frame and a "privileged" frame attached to the ether. This would mean that the relativity principle formulated by Galileo for mechanical phenomena is invalid for optical phenomena, and so we would be able to distinguish the state of uniform motion in a straight line from the state of "absolute rest."

The prominent physicist–experimenter Michelson, later accompanied by Morley, had actually tried to discover this effect in a series of experiments. The idea of these experiments was very simple and based on the interference of light waves. Consider two rays with the same oscillation frequency f, which have been obtained by splitting a beam from a small light source. The splitting of the beam occurs in a glass plate P which partially transmits and partially reflects light. At a certain position of the beam-splitter, the reflected and transmitted parts of the light wave propagate in two mutually perpendicular directions, and then come back, after reflection in the mirrors A and B (Fig. 2.1 a). Because the split beams have taken different routes, they may accordingly have spent different times traveling along their respective paths. As a result, their oscillations will have a certain phase shift with respect to one another when they recombine. The phase shift can be determined as a ratio of the relative time lag to the oscillation period T, multiplied by 2π. If the two waves of the same frequency and the same individual light intensity I_0 meet having a phase difference $\Delta\phi$ at a certain point, the net intensity at this point will be

$$I = 2\, I_0\, (1 + \cos \Delta\phi) \tag{1}$$

For waves oscillating in synchrony we have $\Delta\phi = 0$, and the waves reinforce each other, producing the net intensity equal to four individual intensities (constructive intereference). When the wave oscillations are totally out of phase ($\Delta\phi = 180^\circ$), the waves cancel each other out, giving zero net intensity at corresponding point. In this case light combined with light produces darkness (destructive intereference).

Generally, the phase shift $\Delta\phi$ is different for different points on the screen. Consider, for instance, an interferometer with its mirrors not ideally perpendicular to each other. Interference in this case is similar to that on a wedge-shaped layer of air between two interfaces. Imagine your eye placed at the screen (Fig. 2.1 b). Then you will see simultaneously the mirror B and the image A′ of the mirror A. If the mirrors are not ideally perpendicular, then the image A′ is not parallel to B, and the interference is equivalent to that on an air wedge BOA′. It is clearly seen from Figure 2.1 b that the further from the edge, the greater is the path difference between the interfer-

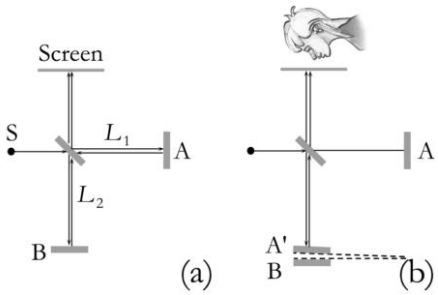


Fig. 2.1 (a) Schematic of the Michelson interferometer. S – source of light, A and B – mirrors. (b) An equivalent air-wedge A'OB produced by an angular misalignment of mirrors A and B.

ing beams, and accordingly the phase shift $\Delta\phi$. Hence the phase shift is a function of a distance y between the observation point and the image of the edge on the screen: $\Delta\phi = \Delta\phi\ (y)$. As you sweep across the screen, you will pass places with different phase shifts between the combining waves and accordingly different light intensity. The screen will display a pattern of bright and dark fringes (that is, alternating regions of high and low intensity). Such a pattern will be observed even when the "arms" of the interferometer (the distances between the center of the beam-splitter and the centers of the mirrors) are the same: $L_1 = L_2 = L$.

Let us consider this case and calculate an *additional* phase difference caused by a possible time lag due to hypothetical ether wind. Suppose that the wind "blows" along one of the arms of the interferometer. We can treat this problem in *total analogy with our treatment of the "Steamer in the stream"* in the previous section. The light here will play the role of the steamer, and the ether wind will be the stream. Then, by the same reasoning as before, the time required for the light to travel there and back along the "longitudinal" arm should be equal:

$$t_{\downarrow\uparrow} = \frac{L}{c+v} + \frac{L}{c-v} = 2\,\frac{Lc}{c^2+v^2} \approx 2\,\frac{L}{c}\,(1+\beta^2) \tag{2}$$

where β is the ratio v/c (which is much smaller than 1).

The round-trip time in the transverse direction is determined by the above-mentioned "transverse" speed $c_\perp$ and equals

$$t_\perp = \frac{2\,L}{c_\perp} = \frac{2\,L}{\sqrt{c^2-v^2}} = \frac{2\,L}{c\sqrt{1-\beta^2}} \approx \frac{2\,L}{c}\left(1+\frac{1}{2}\,\beta^2\right) \tag{3}$$

In the last two equations, we also wrote the approximations to the exact expressions to the accuracy of the second order of β. Thus, the time lag between these two waves will be

$$\Delta t = t_{\downarrow\uparrow} - t_\perp = \frac{L}{c}\,\beta^2 \tag{4}$$

The corresponding phase shift, according to the above definition, is

$$\Delta\phi_e = 2\pi \frac{L\beta^2}{cT} = 2\pi \frac{L\beta^2}{\lambda} \tag{5}$$

where $\lambda = cT$ is the wavelength of light (the distance traveled in one period).

As we see from Equation (5), the contribution from the ether wind depends only on the wavelength, the arm length, and the speed of the Earth relative to the ether. Therefore, it must be to a high accuracy the same for all points on the screen. Thus, the possible influence of the ether wind can be described as a constant [Eq. (5)], added to the phase $\Delta\phi$ (γ) in Equation (1). If a constant is added to a phase in the sine or cosine function, the graph of this function will just shift along the γ-axis. Therefore, with the ether wind, the observed interference pattern on the screen would be shifted relative to its position in the absence of the wind.

Suppose now that we have turned the whole device by 90°, so that the beam which was parallel to the "wind" now travels in the transverse direction, and vice versa. Then the wave that had previously arrived at a given point with delay will now arrive earlier; in other words, the time lag will change sign. This must result in the shift of the observed interference pattern corresponding to the change in phase difference by $2\Delta\phi_e$. Therefore, if there is no ether wind, the turning of the device will not affect the interference pattern. If the wind exists and affects the speed of light, the interference pattern will shift with the turning of the device. It was this shift that Michelson and Morley wanted to observe in their experiments.

In order to observe the effect, the pattern on the screen must shift a distance comparable to the fringe spacing, that is, the additional phase shift $\Delta\phi_e$ due to the expected "ether wind" must be comparable with 2π. According to Equation (5), this requires an experimental setup in which the distance L is of the order of λ/β^2. For the wavelengths of visible light and the speed of ether wind comparable to the speed of Earth's motion around the Sun, the length of the travel path of light in the device must be not less than 100 m. Therefore, the light in the Michelson interferometer was made to travel many times back and forth along either of the two paths before recombining to make the interference pattern on the screen [14]. The whole setup was state of the art by the time (1881–87) the experiments were carried out.

The experiments conducted based on this scheme and repeated many times thereafter with ever increasing accuracy did not produce the expected result. The ether wind, and thereby motion of Earth, could not be detected. This can be considered as evidence that motion of a reference frame does not affect the speed of light.

A plethora of studies have been devoted to the analysis of the Michelson experiment. In some of them the authors tried to retain the concept of ether. In order to account for the negative results of the Michelson experiment, they had to assume that the ether wind is precluded from being observed by some countereffect. For instance, the change of direction of the ether wind relative to the device could deform the interferometer's arms in such a way as to compensate the change of the interference pattern. As a result, no effect would be observed. Precisely such an explanation was proposed by the physicists H. A. Lorentz and G. F. FitzGerald.

Lorentz and FitzGerald had assumed that any system moving at a speed v relative to the ether contracts in the direction of motion by the amount $(1 - v^2/c^2)^{1/2}$. Such a

contraction explains the negative result of the Michelson experiment. Indeed, if we multiply the longitudinal size in Equation (2) by the above factor, the time $t_{\downarrow\uparrow}$ will become equal to $t_{\perp}$, which means that the light's traveling time for both rays and, correspondingly, the interference pattern, will no longer depend on the interferometer's orientation. Such an explanation is logically consistent, but it is unduly complicated. It implies the necessity of a few independent postulates:

1 The ether does exist (and in addition, it must possess a number of very special and hardly compatible properties, and each of them must also be postulated).
2 The motion of any system through the ether contracts the system in the longitudinal direction.
3 This contraction is such as to compensate all observable manifestations of the ether wind.

In addition to its complexity, the described scheme is faulty in two respects. First, its primary substance (ether), whose existence it postulates, does not reveal itself in the observed phenomena (the scheme itself has been designed to account for this fact). Second, it leads to a number of subsequent difficulties and complications. Therefore, it could not have become a foundation for a physical theory.

All these difficulties were eliminated in Einstein's special theory of relativity. This theory does not in any way mention ether. At the basis of the theory lies Einstein's principle of relativity, according to which all natural laws and thus all physical phenomena (and not only mechanical ones) look similar in all inertial reference frames. In other words, all inertial systems are absolutely equivalent.

This principle easily explains why no indications of the Earth's motion were detected in the Michelson experiment. Since the Earth's orbital motion is inertial with a high degree of accuracy on any small segment of its orbit, it cannot affect the outcome of any laboratory experiment.

Thus Einstein's principle of relativity makes the negative result of the Michelson experiment obvious from the very beginning. An interesting historical fact is that Einstein himself was probably unaware of the Michelson experiment when he published his first famous article on the theory of relativity. This does not mean, however, that such an experiment was unnecessary. Regardless of whether it was known to Einstein or not at the time, the Michelson experiment is one of the cornerstones of the experimental basis of the theory of relativity. Its result greatly facilitated the acceptance of this theory and helped to comprehend quickly its striking revelations about the basic properties of time and space. This is what comprises the historical role of the Michelson experiment.

2.2
The speed of light and the principle of relativity

Let us now try to interpret the results of the above-mentioned experiments with light. These results contradict our intuition based on observing motions much slower than

light. Our experience expressed in Equation (1) in Section 1.2 shows that the velocities of such motions just add together. In particular, this equation accurately describes a well known fact that if a surfer reaches the same speed as a running ocean wave by just riding it, then the speed of the wave relative to the surfer is zero.

However, what can be done with an oceanic (or sound) wave cannot be done with light. The experiments did not support the viewpoint that light waves are just perturbations in a specific medium (ether) permeating the whole space. And with no scientific evidence, it makes no sense to speak about such a medium. Therefore, we accept the viewpoint that space does not contain any light-carrying substance (ether), in which light could spread like the sound in air or waves in the ocean. A light wave can exist "all by itself" in a free space, and only in motion. A notion of "still" or even "slow" light waves in an empty space contradicts both electromagnetic theory and the experiment. Light always moves with the same universal speed. We cannot tell whether we are on a stationary platform, or in a uniformly moving car, or in a rushing spaceship with engines off, by measuring the speed of light: in either case the result is the same. Nor can we tell uniform motion from rest by observing any other electromagnetic phenomena. These phenomena, as well as the mechanical ones, are "insensitive" to a state of uniform motion of the observer. Einstein accepted this statement as part of a universal principle that he had formulated (Einstein's principle of relativity) – that *all* natural phenomena (rather than only mechanical ones) look the same in all inertial reference frames. In other words, Nature possesses a deep symmetry which is manifest in the equivalence of all inertial systems. All observed phenomena confirm this conclusion.

From my teaching experience, I can foresee a typical objection by a skeptical reader:

"Excuse me, but this conclusion seems ridiculous. I can understand that the invariance of the speed of light, difficult as it is to grasp, indicates that all inertial reference frames are equivalent. However, the speed of an object such as a stone or bullet is *not* invariant, and yet you say that this is also a manifestation of the same principle of relativity. How can that be?"

The answer to this is that the speed of a stone may vary even in *one* reference frame, depending on the initial conditions or on the applied forces. Therefore, any difference in such speed measured by different observers reflects only the difference in the initial conditions, not the difference in the laws of Nature. For instance, the falling item in Figure 1.1 in the Introduction has no initial velocity in the horizontal direction as seen from the train, and has an initial horizontal velocity equal to that of the train as seen from the ground. Therefore, it falls straight down relative to the train and traces out a parabola relative to the platform. But it might as well start moving without an initial horizontal component if dropped by the person on the platform, in which case it would fall straight down relative to the platform. Or, it might start moving in the train car with an initial horizontal component if pushed horizontally by the passenger, in which case it would move there in a parabola, as it does on the platform under similar conditions. Therefore, if we have two identical systems in two different inertial reference frames K and K′, and both systems start from identical initial conditions, they perform identical motions. Also, in either frame the speed of corresponding mass can vary within the same range – from zero to a speed ap-

proaching that of light. This is a more rigorous formulation of the principle of relativity for systems such as stones or planes.

Light, on the other hand, can move only with one fixed speed in one reference frame. The principle of relativity in this case requires that this fixed speed remains the same in any other inertial reference frame, regardless of the initial conditions.

But here the same thoughtful reader may ask another question:

"OK, this explanation is logically consistent, *if* we accept that the speed of light, unlike the speeds of most other objects, is the fixed quantity. But how can it be that light, which moves in the same space and time as do objects such as cars, bullets, and planets, does not obey the law of addition of velocities [Eq. (1) in Sect. 1.2] that applies to these objects?"

This question, as I noted in the Introduction, is crucial for understanding relativity.

Let us trace the origin of the law of addition of velocities. Consider two inertial reference frames K and K′. Let K′ move relative to K in the x-direction with a speed V, and the origins of both systems coincide at the moment $t = 0$. Consider an object M at a later (non-zero) moment t. By this moment the origin of system K′ will have traveled a distance Vt (Fig. 2.2). Therefore, the x-coordinate of the object in K at this moment will differ from its x'-coordinate in K′ by this distance:

$$\left.\begin{aligned} x &= x' + V t \\ t &= t' \end{aligned}\right\} \tag{6}$$

The second of Equations (6) expresses the obvious fact that time is the same in both systems. The relations (6) between the space and time coordinates of an event observed in two different reference frames is known as galilean transformations. The law of addition of velocities follows directly from these transformations. The speed of the object in K is $v = \mathrm{d}x/\mathrm{d}t$. Its speed in K′ is $v' = \mathrm{d}x'/\mathrm{d}t'$. Since $t = t'$, we have

$$v = \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}x'}{\mathrm{d}t} + V = \frac{\mathrm{d}x'}{\mathrm{d}t'} + V = v' + V \tag{7}$$

This is the addition rule expressed by Equation (1) in Section 1.2. For $v = c$ we recover Equation (2) in Section 1.2 as a special case.

However, since Equation (7) does not hold for light, it must be generally wrong, even though it describes accurately the slow motions. But how can it be wrong if it follows directly from most fundamental properties [Eq. (6)] of space and time? There can be

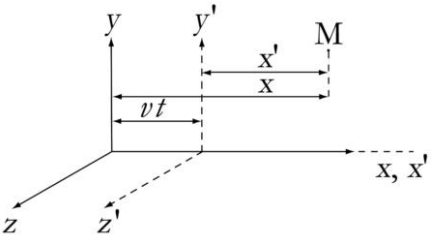


Fig. 2.2

only one answer: the "fundamental" properties [Eq. (6)] that we had considered as self-evident must themselves be generally wrong and need critical revision. That was Einstein's brilliant idea, that became a starting point for his theory of relativity.

2.3 "Obvious"" does not always mean "true"!

When we enter the area of speeds comparable to the speed of light, we must generalize the law of velocities addition in such a way that one equation would describe both the simple addition of low-speed motions and the "weird" behavior of light. To do this, we will analyze in more detail here the initial premises on which the law of velocities addition is based.

Consider the following situation: a spaceship (system K′) moves at a speed V relative to an inertial system K, assumed to be stationary. An object moves inside the spaceship from its rear to its front (i. e. parallel to the spaceship's velocity) at a speed v'. The speed v of the object relative to K is then given by the "obvious" Equation (7).

Let us now scrutinize the definition of speed used in the previous section. The object's speed v' relative to the spaceship is $v' = \dfrac{\Delta x'}{\Delta t'}$, where $\Delta x'$ is the length of the spaceship and $\Delta t'$ is the time it takes for the object to travel this length. Thus, Equation (7) means that

$$v = V + \frac{\Delta x'}{\Delta t'} \tag{8}$$

Scrutinize the meaning of all the terms in this equation. The first two terms (the speeds v of the object and V of the spaceship relative to K) are measured using rulers and clocks, which belong to the system K and *do not participate in the spaceship's motion*. The last term (the speed of the object relative to the spaceship) is measured by the spaceship's crew using *the rulers and the clocks they find on the spaceship*. Of course, the rulers and clocks in K and K′ are identical in the sense that they have been constructed in the same way (the possibility of their structures being identical is guaranteed by the identity of all the laws of nature in both systems, i. e. by Einstein's principle of relativity). However, the two systems of rulers and clocks are *moving* relative to each other, and we do not know beforehand how this will affect the result of their *direct comparison with each other*. That is why it is utterly wrong to measure both items on the right of Equation (8), which contribute to the net speed v, in the units belonging to *different* reference systems. The rulers and clocks of system K may be affected by its motion relative to K.

The correct equation, corresponding precisely to the definition of velocity of an object in K , is

$$v = V + \frac{\Delta x}{\Delta t} \tag{9}$$

where Δx is the length of the spaceship measured *in units of system K* and Δt is the corresponding time (i. e. the time it takes for the object to move from the rear to the front of the spaceship) measured *using the clocks of the system K*.

The correct Equation (9) can be reduced to Equation (7) only if we make two *additional assumptions*:

1. The distance $\Delta x'$ in K′ (in our case the length of the spaceship measured by its own rulers) is transferred without any change to the system K (that is, $\Delta x' = \Delta x$).
2. The duration $\Delta t'$ of a process (in our case the time that the object spends in motion) in system K′ is the same as its duration in system K (that is, $\Delta t' = \Delta t$).

In other words, objects' sizes (or distances between objects) and durations of processes (or time intervals between events) had been assumed to be absolute regardless of the state of motion of the system to which we attach our clocks and scales. The absoluteness of distances and the invariance of time in all reference systems must result in simple addition [Eq. (7)] of velocities. However, since the "simple addition" law, when applied to light, clashes with experiment, it must be generally wrong. Therefore, the assumption that space and time are absolute must also be wrong. We have already emphasized that the belief in absoluteness of space and time was "born" in the world of low speeds. However, the speed of light is not low! It follows that the concepts of absolute time and space upon which Equation (7) was based must be changed in such a way as to obtain a description of the world that would hold for *any* motions, slow or fast.

2.4 Light determines simultaneity

It is natural that light, whose "weird" behavior has prompted us to revise the concepts of time and space, is itself suggesting the direction of such a revision. In fact, not only does it suggest it, but rather it points unambiguously to the only possible solution.

Light propagates in the same physical space where other objects are moving. However, while the speeds of most objects can change (in particular, after transition into another reference frame), the magnitude of the velocity of light remains constant. The properties of time and space must be reconciled with this fundamental fact.

The invariance of the speed of light suggests, as the above analysis shows, that the time interval $\Delta t'$ between two events at different points of the system K′ is generally different from the time interval Δt between the same events in the system K, that is, $\Delta t \neq \Delta t'$.

In particular, this means that if $\Delta t = 0$ (if both events occur simultaneously in K), then $\Delta t'$ may be different from zero, and these same events will not be simultaneous in system K′. It is at this point where the most fundamental break with Newtonian concepts lies.

The classical notion of absolute simultaneity is based upon the intuitive idea that time is something universal and is the same at any moment for all points in space

and in any reference frame. Space itself is perceived as the locus of all points (or, more precisely, "events"), "snapped" at some moment of time.

But what does it mean – one and the same moment of time for two points A and B a way apart?

Let two clocks with identical structure be placed at the points of interest. We call two events occurring at these points simultaneous if the clocks A and B show the same time readings at the corresponding moments. But this definition is based on an unspoken assumption that both clocks had been started at the same time. It follows that the simultaneity of the two given events at A and B depends upon the definition of simultaneity of another pair of events (the starts of clocks A and B). Since a concept cannot be defined in terms of itself, it is necessary to find some other definition.

The concept of simultaneity for spatially separated events (and thereby the mere idea of space "at a given moment") can only have a clear physical meaning (that is, be based on a realizable experimental procedure) if there exists a universal means to overcome the disconnection of events at different places. Light provides us with such a means! The process of propagation of light (or, more generally, electromagnetic interactions) is precisely what makes it possible to connect the time "there" with the time "here." Being a universal "messenger" between different regions of space, light makes it possible to judge the simultaneity of spatially separated events. The experimental fact that the speed of light is independent of the reference frame allows to define an electromagnetic procedure for clocks' synchronization, which is *uniform for all inertial systems*. The clocks A and B that are at rest at adistance x from one another *in a given reference frame* are synchronized if the light signal emitted from A at the moment t_A arrives at B at the moment $t_B = t_A + x/c$. It follows from this definition that the two light signals from a flash at a moment t_C at the point C just in the middle of the segment AB reach the ends A and B simultaneously:

$$t_A = t_B = t_C + \frac{1}{2}\frac{x}{c} \tag{10}$$

If we reverse this procedure, we will come to Einstein's definition of simultaneity: two events at points A and B are simultaneous if the light signals from these events meet exactly in the middle between A and B.

Because of the invariance of the speed of light, the conclusion about *relativity* of simultaneity follows immediately from Einstein's definition. Let us consider again the spaceship from the previous section, assuming that its walls are transparent and that a detector of light signals is positioned in the middle of the spaceship. This detector does not respond to a signal coming from only one direction or to signals arriving from the opposite directions at different moments. If, however, the detector is lit from both directions *simultaneously*, a wiring device switches on, and the detector explodes. A similar detector is put at the point C of the "stationary" system K (Fig. 2.3 a). Suppose that precisely at the moment when both detectors were coincident ($t_C = t'_C = 0$) we marked the instantaneous positions of the end points A and B of the spaceship in system K. The phrase "precisely at the moment" now has a clear physical meaning due to the definition of simultaneity: it means that if two flashes of light occur at points A and B at this moment,

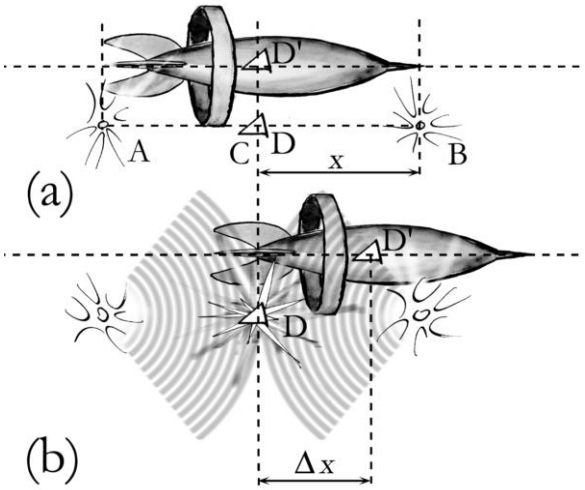


Fig. 2.3 A thought experiment illustrating the relativity of simultaneity. (a) The initial moment: two flashes at the end points of a moving spaceship are simultaneous in K. (b) The final moment: the photons from the flashes explode detector D, but not detector D′.

then the emitted signals will meet exactly in the middle of the segment AB, i.e. at the point C, where our detector is located, and the latter, being lit simultaneously from the opposite directions, will explode. Since the spaceship is transparent, we can also observe the course of events in the spaceship while remaining outside (Fig. 2.3 b). We will watch the spaceship's detector moving toward one signal and running away from the other while it goes past point C. By the moment when both signals meet at C, exploding our detector, the detector on the spaceship will reach some other point C′ and remain intact because it will be lit by only one (oncoming) signal. Thus, in the spaceship's system, the signals will meet not at its center but at some other point and so the detector will not explode. On the other hand, relative to the spaceship, both signals move with the same speed c as they do in system K! So how is it possible that they do not meet at the middle of the spaceship? Or, more precisely, why does the signal from the front travel a longer distance than that from the rear? There is only one plausible answer: because it was emitted earlier! To put it another way, in the spaceship's system the flashes were *not* simultaneous. The flash in the front occurred earlier than the flash in the rear.

Let us calculate how much earlier. Let $\Delta x'$ be the distance between the center of the spaceship and the point C′ where the signals meet, *measured relative to the spaceship*. We will call it the proper distance. To play it safe, we will avoid the statement that $\Delta x'$ is equal to $\Delta x = \mathrm{CC}'$ with CC′ being the distance measured in system K (later we shall see that such a precaution is justified). In contrast, since the segment $\Delta x'$ is moving together with the spaceship at a speed V relative to K, its length $\Delta x = \mathrm{CC}'$, measured in system K, might differ from its length $\Delta x'$ measured in system K′ by a factor $\gamma\,(V)$, which depends on V:

$$\Delta x = \gamma^{-1}(V)\,\Delta x' \tag{11}$$

On the other hand, in the system K the distance Δx is $\Delta x = Vt$, where t is the time interval between the flashes at A and B, and the detector's explosion at C. If we denote

the distance AC = CB (i. e. half of the spaceship's length in the system K) as x, then $t = x/c$, and therefore

$$\Delta x = \frac{V}{c}\, x \tag{12}$$

Exluding Δx from Equations (11) and (12), we find

$$\Delta x' = \gamma(V)\, \frac{V}{c}\, x \tag{13}$$

Thus, we have found that the signal coming from the front of the spaceship has traveled a distance longer by $2\Delta x' = 2\gamma(V)\, \frac{V}{c}\, x$ than the signal coming from the rear. This means that it was emitted earlier by the time interval $2\Delta t' = 2\Delta x'/c$. Suppose now that a clock has been attached to each of the two detectors and that both clocks read zero time ($t_C = t'_C = 0$) at the moment when they were coincident. We will then obtain the following result for the moments of two flashes at A and B:

$$\left.\begin{aligned} &\text{In system K:} \quad t_A = t_B = 0 \\ &\text{In system K}': \quad t'_A = \frac{\Delta x'}{c} = \gamma(V)\, \frac{V}{c^2}\, x\,; \qquad t'_B = -t'_A \end{aligned}\right\} \tag{14}$$

We can put it this way: when the flash occurred at A, the spaceship's clock located at that point reads the time $t'_A = \Delta x'/c$, and when the flash occurred at B, the corresponding clock reads $t'_B = -\Delta x'/c$, *if both clocks on the the spaceship had previously been synchronized in their reference frame* according to Einstein's definition of simultaneity.

We want to emphasize the importance of this conclusion. We are discussing natural phenomena. A pair of spatially separated events is being considered. And it turns out that these events are simultaneous in one reference frame but non-simultaneous in another. This means that *simultaneity is relative*. Its relativity is due to the fact that the speed of light is invariable. If light obeyed the simple law of velocities addition, the light signal would travel faster from the front to the rear than from the rear to the front in the reference frame of the spaceship. This would account for the fact that the two signals do not meet in the center of the spaceship. The flashes would remain simultaneous. In that case, however, the laws for electromagnetic phenomena (e. g. the speed of light propagation!) and the corresponding procedures used to define simultaneity would not be the same for all inertial systems. There would be only one "privileged" system of reference, where the speed of light would be the same in all directions. The clocks of all other systems would be set according to the clocks of this "absolutely still" system, which would bring us back to Newtonian concept of absolute time.

However, light does not leave us such a possibility, because it moves with one fixed speed in all inertial systems, rendering them all equivalent. Thus, Einstein's principle of relativity, together with invariance of the speed of light, implies the relativity of time.

2.5 Light, times, and distances

The relativity of time causes the relativity of distances and time intervals: these quantities are different in different reference systems.

Let us consider a vertical cylinder of length $\Delta l'$ with mirrored butt-ends; a light signal is traveling back and forth periodically inside the cylinder (Fig. 2.4 a). In a system K′ attached to the cylinder, the time interval between two successive arrivals of the signal to a chosen end is equal to

$$\Delta t' = 2\,\frac{\Delta l'}{c} \tag{15}$$

The interval $\Delta t'$ can be called an eigen (proper) period of the signal's motion. Now we can analyze the whole process in system K, in which the cylinder moves horizontally at speed V. What is the time of this process in system K? Denote this time as Δt. In system K, light participates simultaneously in two motions: in the vertical direction (along the cylinder's axis) and in the horizontal direction (together with the cylinder). As a result, during the period Δt of one "oscillation" up and down, the signal will travel the distance $V\Delta t$ in the horizontal direction, and so its trajectory will become a broken line AB′A″ (Fig. 2.4 b).

The length $l_{AB'A''}$ of the element AB′A″ is equal to

$$l_{AB'A''} = 2\sqrt{\Delta l'^2 + \left(\frac{1}{2}V\Delta t\right)^2} = \sqrt{c^2\,\Delta t'^2 + V^2\,\Delta t^2} \tag{16}$$

It is greater than $2\,\Delta l'$. At the same time, the *speed* of light along the broken line in the system K must remain equal to c. To travel a greater distance at the same speed takes a longer time. Indeed, putting $l_{AB'A''} = c\Delta t$ for the element's length in Equation (16), we will obtain the following relationship between Δt and $\Delta t'$:

$$c\Delta t = \sqrt{c^2\,\Delta t'^2 + V^2\,\Delta t^2} \tag{17}$$

It follows that

$$\Delta t' = \Delta t\sqrt{1 - \frac{V^2}{c^2}} \tag{18}$$

As we can see from Equation (18), the period of the same process – one complete oscillation of the light signal inside the cylinder – is different in different systems and is smallest in a system where the cylinder is at rest.

There is another way to see it. In system K light moves along AB′ with speed c, while moving horizontally with speed V. We can see from Figure 2.4 that the vertical component of its motion must be

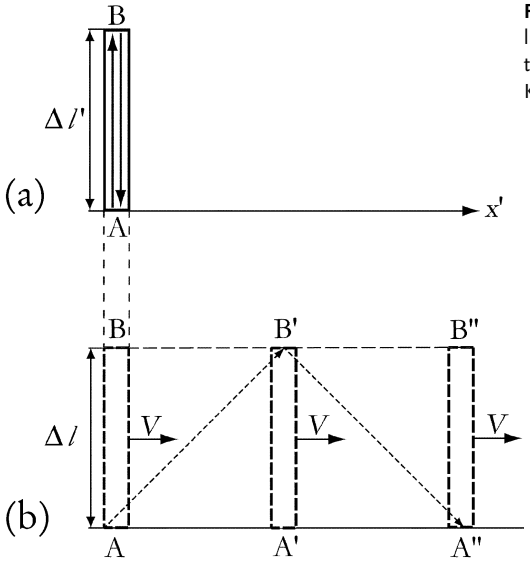


Fig. 2.4 The light pulse in a vertical cylinder that is moving horizontally. (a) In the rest frame of the cylinder (system K′). (b) In system K.

$$v = \sqrt{c^2 - V^2} = c\sqrt{1 - \frac{V^2}{c^2}} \tag{19}$$

The motion of light along the cylinder is slower than c when the cylinder is moving, and its period Δt is accordingly greater than $\Delta t'$ by the same factor, which is the essence of Equation (18).

Further, we have suggested [Eq. (11)] that the "longitudinal" size of an object, that is, its length in the direction of the velocity of its relative motion, may be relative, too. To find the law for the length transformation, we will modify our experiment slightly by directing the axis of the cylinder *along* its relative velocity (Fig. 2.5).

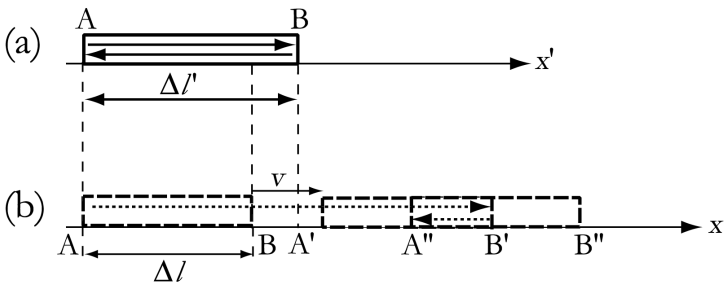


Fig. 2.5 The same as in Figure 2.4, but now the cylinder is horizontal. (a) In the rest frame of the cylinder (system K′). (b) In system K. AB is the initial position of the cylinder (the pulse starts at A); A′B′ is its intermediate position (the pulse reaches the front at B′); A″B″ is its final position (the reflected pulse returns to the rear at A″).

Obviously, in the system K′, this operation will not affect the cylinder's length $\Delta l'$ (the size of an object in its rest system does not change when the object is turned!) Correspondingly, the period of motion [Eq. (15)] of the light signal will remain the same. However, if the time interval $\Delta t'$ between two events (emission and return of the signal) at one point (the point A of the cylinder) in system K′ does not depend on orientation of the cylinder, then the corresponding time interval Δt between those same events considered from system K also will not change. Therefore, Equation (18) must also hold for the cylinder in the horizontal position. Using the relationship in Equation (18) between Δt and $\Delta t'$, we obtain

$$\Delta t = \frac{\Delta t'}{\sqrt{1 - \frac{V^2}{c^2}}} = \frac{2\,\Delta l'}{c}\left(1 - \frac{V^2}{c^2}\right)^{-1/2} \tag{20}$$

Now let us express the time interval Δt in terms of the "longitudinal" length Δl of the cylinder in system K. In this system the light now travels in a moving horizontal "corridor" of length Δl, catching up with the mirror B which runs away from it at a speed V. How long does it take light to catch up with the front end B? Denote this time interval as Δt_1. The distance traveled by the light pulse from point A to point B′ where it catches up with the front of the cylinder is $c\Delta t_1$. The same distance can be expressed in terms of the length of the cylinder as $V\Delta t_1 + \Delta l$ (Fig. 2.5). Thus we have $c\Delta t_1 = V\Delta t_1 + \Delta l$, so that $\Delta t_1 = \Delta l/(c - V)$. After the reflection from mirror B, the light returns to point A (rear of the cylinder), which moves *towards* it at speed V. The time Δt_2 it takes the reflected signal to return to this point can be found in a similar way and is equal to

$$\Delta t_2 = \Delta l/(c + V) \tag{21}$$

The total time Δt the signal spends between its departure and return to the same end of the cylinder is

$$\Delta t = \Delta t_1 + \Delta t_2 = \frac{\Delta l}{c - V} + \frac{\Delta l}{c + V} = \frac{2\,\Delta l}{c}\left(1 - \frac{V^2}{c^2}\right)^{-1} \tag{22}$$

It turns out to have been calculated in the same way as the corresponding time for the "horizontal" beam in the Michelson experiment. However, we must keep in mind that here the quantities $c - V$ and $c + V$ are the rates of change of the distance between the signal and the butt-ends of the cylinder *in system K*, and they by no means represent the speeds of the signal *relative to the cylinder*, i. e. its speed in system K′. Similarly, the distance Δl is the length of the cylinder in *system K*, and, as we will now show, it is indeed different from its length measured in its rest frame.

Comparing Equations (20) and (22), we obtain

$$2\,\frac{\Delta l}{c}\left(1 - \frac{V^2}{c^2}\right)^{-1} = 2\,\frac{\Delta l'}{c}\left(1 - \frac{V^2}{c^2}\right)^{-1/2} \tag{23}$$

that is

$$\Delta l = \Delta l' \sqrt{1 - \frac{V^2}{c^2}} \tag{24}$$

According to Equation (24), the length of a moving segment is smaller than its length in its rest frame by a factor of $(1 - V^2/c^2)^{-1/2}$. In other words, the sizes of moving objects are contracted in the direction of motion. This effect is called Lorentz contraction. However, it has a completely different meaning from the contraction introduced by Lorentz in connection with the Michelson experiment. Lorentz assumed that the longitudinal contraction appears *only* when an object is moving relative to the ether which serves as a universal system of reference. As a consequence, a segment that is stationary relative to the ether possesses the greatest length. In reality, the length contraction is observed for any object moving relative to *any* inertial system. And the segment has the greatest length in its own system of rest, which may be moving relative to a given inertial system at an arbitrary speed V. The ratio

$$\frac{\Delta l}{\Delta l'} = \gamma^{-1}(V) \equiv \sqrt{1 - \frac{V^2}{c^2}} \tag{25}$$

is precisely the same proportionality coefficient as in Equation (11) between the length of a moving segment and its "rest" length that was introduced in the previous section. Now, analyzing our thought experiments with light, we have found the exact value of this coefficient. Because we will come across this coefficient fairly often (indeed, we will see later that not a single relativistic effect can be described without it), we shall give it a special name – the Lorentz factor, and stick to our symbol $\gamma(V)$, remembering that the explicit form [Equation (25)] of the function $\gamma(V)$ is known to us. Using this symbol, we can rewrite the essential Equations (18) and (24) for time and length transformation:

$$\Delta t = \gamma(V)\,\Delta t' \tag{26}$$

$$\Delta l = \gamma^{-1}(V)\,\Delta l' \tag{27}$$

It is essential for the correct use of these equations that we understand clearly the physical meaning of the related quantities: $\Delta t'$ is the time between two different events taking place at *one and the same point* of the system K′; Δt is the time between *the same* events in K, where these events are being observed at different points of space. Similarly, $\Delta l'$ is the length of a segment in its rest system K′; Δl is its length in K that slides along the segment with a speed V. Since in the system K the segment is moving, the coordinates of its end points must be recorded at *one and the same moment in K*; the quantity Δl represents the distance in K between these two instantaneous positions. As a consequence of the relative nature of simultaneity, the recordings of instantaneous positions of the end points, performed simultaneously in the system K, are not simultaneous in the system K′.

2.6 The Lorentz transformations

The examples considered in the previous sections demonstrate how the invariance of the speed of light leads to the relativity of time and space. Now we shall proceed with the deduction of the equations which provide a complete description of the fundamental properties of time and space. These equations are called the Lorentz transformations. They show how the coordinates of any arbitrary event (the Cartesian coordinates of a point where the event has occurred, and the corresponding moment of time) become transformed after transition from one inertial system to another.

Let axes x, y, z of a system K be parallel to the axes x', y', z' of system K′, moving in the direction x at a speed V relative to K (Figure 2.2). Let the origins O and O′ of both systems coincide at the moment $t = t' = 0$ on their clocks there. This can always be achieved by a proper choice of the initial moments of time in both systems.

Let at this moment a flash of light at the origin produce a diverging spherical wave. *Because of the invariance of the speed of light, this wave will be spherical in both reference frames*. Let us express this fact in mathematical terms.

By the moment t in system K the wave front will form a spherical surface of radius $r = ct$ centered at the origin; this surface is described by the equation

$$x^2 + y^2 + z^2 - c^2 t^2 = 0 \tag{28}$$

Similarly, we conclude that the space and time coordinates of the expanding wave front in system K′ must also satisfy the equation of the spherical surface centered at the origin of K′ and having radius $r' = ct'$:

$$x'^2 + y'^2 + z'^2 - c^2 t'^2 = 0 \tag{29}$$

Equating the left parts of these two equations, we obtain

$$c^2 t^2 - x^2 - y^2 - z^2 = c^2 t'^2 - x'^2 - y'^2 - z'^2 \tag{30}$$

The expression

$$\sqrt{c^2 t^2 - x^2 - y^2 - z^2} \tag{31}$$

is called the space–time interval. It generalizes the spatial distance between two events[1]. We can see from Equation (30) that the value of this interval is invariant. If we use identical synchronized clocks, as well as Cartesian coordinates with identical length units in all systems of reference, then the mathematical form of this expression (the combination of squares of all four coordinates, each square taken with a definite sign) will be maintained. This property of the space–time interval is called covariance.

1) The "spatial" part of the interval (31) is $-r^2$, where r is the length of the vector with coordinates (x, y, z).

Now we must find the relations between the coordinates (t, x, y, z) and (t', x', y', z'), which will satisfy the requirement that the interval be covariant. First of all, these relations, that is, functions $t\,(t', x', y', z')$, $x\,(t', x', y', z')$, $y\,(t', x', y', z')$, and $z\,(t', x', y', z')$, have to be linear:

$$\left.\begin{aligned} ct &= a_{00}\,ct' + a_{01}\,x' + a_{02}\,y' + a_{03}\,z' \\ x &= a_{10}\,ct' + a_{11}\,x' + a_{12}\,y' + a_{13}\,z' \\ y &= a_{20}\,ct' + a_{21}\,x' + a_{22}\,y' + a_{23}\,z' \\ z &= a_{30}\,ct' + a_{31}\,x' + a_{32}\,y' + a_{33}\,z' \end{aligned}\right\} \tag{32a}$$

For those familiar with the formalism of linear algebra, this can be represented as a matrix equation:

$$\begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ a_{10} & a_{11} & a_{12} & a_{13} \\ a_{20} & a_{21} & a_{22} & a_{23} \\ a_{30} & a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} \tag{32b}$$

Equations (32) follow from the fact that only linear relationships between coordinates can ensure the covariance of the expression (31) for the interval. If instead of Equations (32) we put into expression (31) any other function of the primed coordinates, we will not obtain the sum of squares of these coordinates. Further, the linear function is the only one whose inverse function is also linear. No other mathematical relationship possesses this property, which must hold here because the equivalence of systems K and K′ implies that the inverse relationship between (t', x', y', z') and (t, x, y, z) has the same mathematical form as the direct one.

These mathematical conditions have a simple physical meaning: the linearity of the relationship $(ct, x, y, z) \Leftrightarrow (ct', x', y', z')$ expresses both the equivalence of all points of space and moments of time, and the equivalence of inertial systems K and K′.

Thus, the physics of the considered phenomena dictates the linearity of the transformation between the coordinates of an event observed in different inertial systems. In our case, when the velocity $\mathbf{V}$ is parallel to the x axis, the transverse coordinates must have the same values in both systems:

$$y = y'; \quad z = z' \tag{33}$$

The reason is that the speed of relative motion of the systems K and K′ is zero in the directions y and z. Given Equations (33), the matrix Equation (32b) simplifies to

$$\begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a_{00} & a_{01} & 0 & 0 \\ a_{10} & a_{11} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} \tag{34a}$$

that is,

$$\left.\begin{aligned} c\,t &= a_{00}\,ct' + a_{01}\,x' \\ x &= a_{10}\,ct' + a_{11}\,x' \\ y &= y' \\ z &= z' \end{aligned}\right\} \tag{34 b}$$

However, the point O′ moves at a speed V relative to K, so by the moment t its position in K is given by the coordinate $x = Vt$. Assuming $x' = 0$ and putting into Equation (34 b) $x = Vt$, we have

$$\left.\begin{aligned} c\,t &= a_{00}\,c\,t' \\ V\,t &= a_{10}\,c\,t' \end{aligned}\right\} \tag{35}$$

from which there follows

$$\frac{a_{10}}{a_{00}} = \frac{V}{c} \tag{36}$$

Thus we have the following matrix equation for the transformation of x and t:

$$\begin{pmatrix} c\,t \\ x \end{pmatrix} = \begin{pmatrix} a_{00} & a_{01} \\ \dfrac{V}{c}\,a_{00} & a_{11} \end{pmatrix} \begin{pmatrix} c\,t' \\ x' \end{pmatrix} \tag{37}$$

To find the remaining three unknowns, we substitute the linear Equation (37) into the equation

$$c^2\,t^2 - x^2 = c^2\,t'^2 - x'^2 \tag{38}$$

obtaining

$$(a_{00}\,ct' + a_{01}\,x')^2 - (a_{00}\,Vt' - a_{11}\,x')^2 = (c\,t')^2 - x'^2 \tag{39}$$

Now we demand that the expression on the left might be reduced to the form $c^2\,t'^2 - x'^2$. This is only possible if

$$\left.\begin{aligned} &a_{00}^2\left(1 - \frac{V^2}{c^2}\right) \equiv a_{00}^2\,\gamma^{-2}\,(V) = 1 \\ &a_{01} - \frac{V}{c}\,a_{11} = 0 \\ &a_{01}^2 - a_{11}^2 = -1 \end{aligned}\right\} \tag{40}$$

We have a system of three equations for three unknowns. The solution is simple and gives

$$a_{00} = a_{11} = \gamma(V); \quad a_{01} = \frac{V}{c}\gamma(V) \tag{41}$$

Taking into account Equation (36), we finally obtain

$$\begin{pmatrix} ct \\ x \end{pmatrix} = \gamma(V) \begin{pmatrix} 1 & \frac{V}{c} \\ \frac{V}{c} & 1 \end{pmatrix} \begin{pmatrix} ct' \\ x' \end{pmatrix} \tag{42}$$

or

$$\left.\begin{aligned} ct &= \gamma(V)\left(ct' + \frac{V}{c}x'\right) \\ x &= \gamma(V)(x' + Vt') \end{aligned}\right\} \tag{43}$$

This is the Lorentz transformation. Changing the sign of V, we can obtain the inverse transformation:

$$\left.\begin{aligned} ct' &= \gamma(V)\left(ct - \frac{V}{c}x\right) \\ x' &= \gamma(V)(x - Vt) \end{aligned}\right\} \tag{44}$$

If we substitute these transformations into Equation (30), it will become an identity, which means that the expression (31) for the interval is covariant with respect to the Lorentz transformation. In a special case when the speed V is much smaller than c, the ratio V/c is negligible, the Lorentz factor approaches 1, and Equation (44) reduces to Galilean transformations in Equations (6).

2.7 The relativity of simultaneity

Using the Lorentz transformation in its general and explicit form, we can arrive at the results that are already familiar to us, and in particular the conclusion about relativity of time. This relativity can not only be observed, but also be described quantitatively, with the help of the following imaginary experiment. Let us arrange equally spaced identical clocks along the x-axis in the system K, assuming the clocks to be synchronized in K according to the previously described Einstein's definition of simultaneity. Let us place identical clocks at equal distances from each other along the x' axis of K′. As previously, the expression "identical" means that the clocks of the system K′ are constructed in this system from the materials and under conditions identical with the materials and conditions in system K; the possibility of such identical conditions is guaranteed by Einstein's relativity principle (independence of all laws of nature from a state of inertial motion.) All clocks of the system K′ are also

synchronized in this system using the light signals according to Einstein's procedure. In the final result, we have two linear sets of identical clocks synchronized in their respective systems and moving relative to one another at a speed V.

Let us define the initial moment so that when the origins of both systems coincide ($x = x' = 0$), the corresponding local clocks there would read the same time $t = t' = 0$. Consider other points of space *in system K* at this moment. The phrase "at this moment" means that all the K-clocks placed at these points read the same time $t = 0$. All events that occur at this moment form a set that is simultaneous in K. (What we call "space" is just an infinite and continuous set of different but simultaneous events!) In system K′, however, these same events *do not* form a simultaneous set, and thus the system's clocks have different readings at different points. To ascertain this fact, we use the Lorentz transformation. Substituting $t = 0$ into Equation (43) or (44), we obtain

$$t' = -\gamma(V)\,\frac{V}{c^2}\,x \tag{45}$$

i. e. the already familiar result, Equation (14).

We see that the spatially separated events occuring all at one moment $t = 0$ in K occur at different moments in K′, depending on location of a K′-clock. Namely, a K′-clock to the right of the origin of K′ (at $x' > 0$) reads earlier time than does the K′-clock currently passing by the origin. A clock to the left of the origin (at $x' < 0$) reads later time than the K′-clock at the origin. According to Equation (45), the discrepancy between readings of "moving" and "stationary" clocks is proportional to their distance from the origin (Fig. 2.6).

In should be emphasized that this conclusion is quite general in its character, since it does not depend on any assumptions about the nature or internal structure of the clocks. A clock can be based on some periodic process such as mechanical or electromagnetic oscillations, or radioactive decay of some nuclei. It does not matter what physical phenomenon provides the basis for functioning of a given clock. In any case the result will be the same. This allows us to ignore the properties of specific

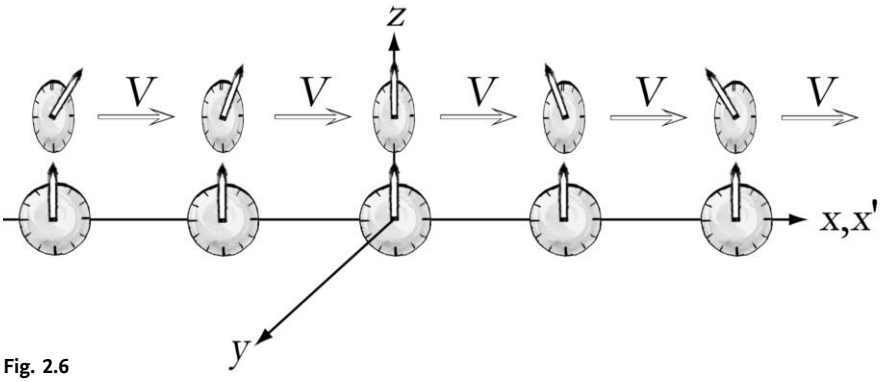


Fig. 2.6

clocks or physical processes and focus on the properties of time itself. Let us repeat the essential point of our previous discussion: spatially separated events that are simultaneous in one system of reference are generally not simultaneous in another one. An exception occurs when the given events are taking place in a plane perpendicular to the direction of the systems' relative motion.

All relativistic phenomena are in fact the consequences of this fundamental property of time, which itself is due to the existence of the universal speed c. We shall find out shortly how relativity of time affects the appearance of a process or of an object in different reference frames.

2.8 A proper length and a proper time

Consider a point (x', y', z') in a system K′. Suppose the clock records the moments t'_1 and t'_2 of two successive events at this point. The difference $\Delta t' = t'_2 - t'_1$ is the time interval between these events in the system K′. The time interval between the two events occuring at the same point is called the proper time. What time between these two events will be measured in another reference system K?

Using the Lorentz transformation (43), we have for a fixed value of $x'_1 = x'_2 = x'$:

$$\left.\begin{aligned} ct_1 &= \gamma(V)\left(ct'_1 + \frac{V}{c}x'\right) \\ ct_2 &= \gamma(V)\left(ct'_2 + \frac{V}{c}x'\right) \end{aligned}\right\} \tag{46}$$

and hence it follows that

$$\Delta t = t_2 - t_1 = \gamma(V)(t'_2 - t'_1) = \gamma(V)\,\Delta t' \tag{47}$$

Since $\gamma(V)$ is always greater than 1, we find that $\Delta t > \Delta t'$, that is, in system K, where the considered events are observed at *different* points of space, they are separated by a longer time interval than in system K′. A process localized at one point in some reference frame lasts the shortest when it is observed from this reference frame. Any change of an object's location due to its motion is accompanied by slowing down of its evolution. A moving clock "ticks" slower than a stationary clock by a factor $\gamma(V)$.

Let t'_1 and t'_2 be the moments of birth and decay of a particle, for instance, a μ-meson, in its proper system. Then the time interval $\Delta t' = t'_2 - t'_1$ is its proper lifetime. For the μ-meson this time is about $2 \cdot 10^{-6}$ sec. But in a system K, in which the μ-meson is moving at a speed V, its birth and decay take place at different points, i.e. because of the particle's motion, its life is "spread" in space. Accordingly, its life span in the new system is $\gamma(V)$ times longer than its "proper" lifetime. If V is large enough (e.g. $v \to c$), then Δt will be much greater than $\Delta t'$, and the particle will cover a vast distance before its decay. Precisely this phenomenon had been discovered for μ-mesons created in the upper atmosphere (at altitudes of about 100 km) under the influence

of cosmic rays. The greatest distance a particle can travel in $2 \cdot 10^{-6}$ s is $\Delta r = c\Delta t' \approx 2 \cdot 10^{-6}\,\mathrm{s} \times 3 \cdot 10^{8}\,\mathrm{m\,s^{-1}} = 600$ m. Therefore, if time were absolute, μ-mesons would decay not far from the point of their creation – practically at the same altitude of about 100 km above the Earth's surface. In fact, however, these particles can be detected at sea level, which means that they manage to travel trough the whole thickness of the Earth's atmosphere during their lifetime. In this fact the relativity of time is revealed in its full sway: since in the Earth's reference system the μ-meson moves at a subluminal speed, its lifetime in this system greatly exceeds its "proper" lifetime. This allows it to travel such a long distance.

Now let us consider the consequences of the Lorentz transformation for space.

A rod of length

$$\Delta x' = x'_2 - x'_1 \tag{48}$$

is positioned parallel to the x' axis in system K′. What is the rod's length in K? As mentioned previously, in order to measure the length, we have to mark the instantaneous coordinates x_1 and x_2 of the edges of the rod at the same moment t in *system K*. The clearing between the marks gives us the wanted length:

$$\Delta x = x_2 - x_1 \tag{49}$$

According to the Lorentz transformations, we have for the moment t

$$x'_1 = \gamma(V)(x_1 - Vt), \quad x'_2 = \gamma(V)(x_2 - Vt) \tag{50}$$

so that

$$\Delta x' = x'_2 - x'_1 = \gamma(V)(x_2 - x_1) = \gamma(V)\Delta x \tag{51a}$$

or

$$\Delta x = \gamma^{-1}(V)\Delta x' \tag{51b}$$

in full agreement with Equations (11) and (27). We have obtained an already familiar result: the length of the rod is the greatest in a system where the rod is at rest (the proper length). In a system where the rod is moving while being parallel to its velocity, its length diminishes by a factor of $\gamma(V)$.

The described effect of the length contraction allows one to explain the result of the experiment with atmospheric μ-mesons from the viewpoint of a fictitious observer traveling together with these particles. Relative to this observer, the μ-mesons are motionless while the Earth is rushing toward them at a speed V. The observer measures the lifetime of μ-mesons in his system of reference and, naturally, finds it equal to their *proper* lifetime $2 \cdot 10^{-6}$ s. In such a small time interval the Earth, even at a speed close to c, will travel less than 600 m toward him. How can it be then that the μ-mesons born in the upper layers of atmosphere, i. e. about 100 km from the Earth, can be de-

tected near the Earth's surface? The explanation lies in the effect of Lorentz contraction of the Earth in the direction of relative motion. The 100 km length that we have mentioned is the *proper* thickness of the atmosphere. In the system of μ-mesons moving towards the Earth, both the Earth and its atmosphere are flattened owing to the length contraction. The atmosphere's thickness of 100 km is contracted to about 600 m – just enough to pass by the observer in as small a time as 10^{-6} s!

Thus, the entire picture of the process becomes intrinsically consistent. Time and space transform in such a way that even though different inertial observers will measure different values for these quantities, they all register the same result: a μ-meson is created in the upper layers of the atmosphere and decays at the sea level or even in the depths of the ocean. The relativity of time and space is consistent with the covariance of dynamic laws.

The described experiment with μ-mesons demonstrates the unity of Nature. Our statement about relativity of time and space was deduced from studies of electromagnetic interactions. Meanwhile, interactions that lead to decay of μ-mesons are not electromagnetic. They have a different nature and are called weak interactions. But the time dilation of μ-meson decay follows the same Lorentz transformation rules that we have obtained analyzing the properties of light. Different kinds of interaction turn out to have common properties.

This suggests that at a deeper level all kinds of interaction will prove to be manifestations of yet unknown universal interaction just like the electric and magnetic fields, which initially appeared fundamentally different, have proven to be merely special cases of the electromagnetic field. Now we have a developed theory that unifies electromagnetic and weak interactions. According to this theory, both interactions, despite all the differences between them (we have just pointed out that they are different!), are manifestations of a more fundamental *electro-weak* interaction. The theory had predicted hitherto unknown phenomena, which have now received an experimental confirmation!

In the last years of his life, Einstein tried unsuccessfully to find a theory that would unify all observable physical phenomena ("the theory of everything"). It is only now, at a more developed level of our knowledge, that the approaches to the long coveted "Great Unification" of all known forces appear to be emerging.

Thus Einstein's principle of relativity that had been postulated on the basis of mechanics and electromagnetism about a century ago still remains a powerful guide in our search for an understanding of the workings of Nature.

2.9 Minkowski's world

Our study of the relationship between space and time can be made geometrically clear by using a remarkable construction – so called space–time diagrams. They were introduced by an outstanding mathematician, Germann Minkowski, who had been the young Einstein's instructor in the Zurich Polytechnicum. After Einstein had published his theory of relativity, Minkowski allegedly said: "To tell the truth,

I did not expect this from Einstein." Einstein, on his part, also allegedly, remarked after the publication of Minkowski's work: "After the mathematicians had taken care of my theory, I no longer can understand a bean in it." He joked, of course. The geometrical approach introduced by Minkowski was an important contribution to the theory. It was later used by Einstein himself to extend his theory to include gravitation. Now this approach is universally used in both the special and general theories of relativity.

A space–time diagram is in its essence just a graph of a displacement versus time dependence for a given motion. The basic element in this construction is a concept of an *event*. An event is any phenomenon so fleeting and occupying so small a region that it can be considered as instantaneous and point-like. It therefore can be characterized by one moment of time t and three spatial coordinates (x, y, z). Hence to any possible event there corresponds a set of four numbers (t, x, y, z), and vice versa; each set of four numbers (t, x, y, z) specifies an event in space and time. For instance, a set of numbers (43 s, 2 m, –54.6 m, 0.33 m) labels an event that happened at 43 s a.m. today, at a point with coordinates 2 m, –54.6 m, and 0.33 m along the x, y, and z directions respectively. All possible events that have ever happened, are happening, and are going to happen, form a four-dimensional set which is a combination of three spatial and one temporal dimensions. This set was called Minkowski's world [15, 16].

In order for all four coordinates of Minkowski's world to have common dimensions, it is convenient to use ct rather than t as a time coordinate. This means that instead of measuring time in seconds, we measure it in equivalent units of length – so-called light seconds: 1 light second is the distance traveled by light in 1 second. Accordingly, we might express spatial distances in these new units of length. We thus would obtain one common unit for all four dimensions of our world.

As a first illustration of a space–time diagram, consider a light signal moving in the positive x direction: $x = ct$ (Fig. 2.7). This motion is represented graphically by a straight line OS which is a bisector of the angle between the axes x and ct (it is convenient to direct the ct axis upwards). The light signal moving in the opposite direction $(x = -ct)$ will be represented graphically by a straight line OS′. We can consider either of the two lines as a trajectory of the light pulse in a plane (x, ct). We call such a trajectory *the world line*. Thus, OS is a world line of a light signal moving in the positive x direction, and OS′ is the world line of a signal moving in the negative x direction. Now, what is it good for?

Consider two events: the first is the emission of a light pulse from a laser gun at a point (x_1, y_1, z_1) at a moment t_1, and the second is the absorption of this signal by a detector at point x_2, y_2, z_2 at a moment t_2. The distance between the gun and the detector is

$$r_{12} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} = c\,(t_2 - t_1) \tag{52}$$

We can write the relation (52) between the distance r_{12} and time $t_{12} \equiv t_2 - t_1$ in the form

$$c^2\,(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2 = 0 \tag{53}$$

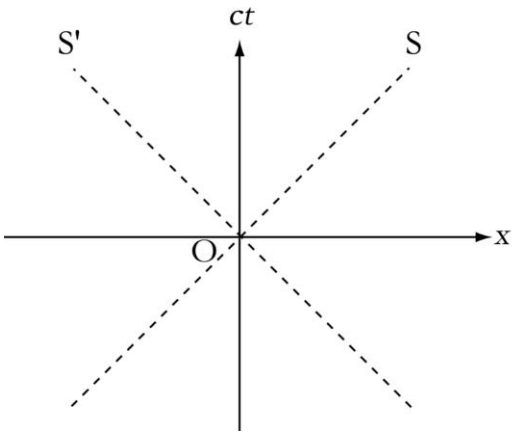


Fig. 2.7

Denote the expression in the left-hand side of this equation as s_{12}^2. Then the equation for the light pulse can be written as

$$s_{12}^2 = 0 \tag{54}$$

The quantity s_{12} is called the four-dimensional interval (or just the interval) between events 1 and 2 [recall Eq. (31)]. We see that the interval for a light signal in vacuum is equal to zero.

Suppose that the same process is being observed from another inertial reference frame K′. The space and time coordinates for the same two events (emission and absorption of the signal) in system K′ will be $(ct'_1, x'_1, y'_1, z'_1)$ and $(ct'_2, x'_2, y'_2, z'_2)$, respectively. The distance r'_{12} between the laser gun and the detector and the time interval $t'_{12} \equiv t'_2 - t'_1$ between the emission and absorption of the signal will also be different. However, because of the invariance of the speed of light there must be $r'_{12}/t'_{12} = c$, or $c^2 (t'_{12})^2 - (r'_{12})^2 = 0$. Therefore, if an observer in system K′ also considers the quantity $(s'_{12})^2 \equiv c^2 (t'_{12})^2 - (r'_{12})^2$, he can write

$$(s'_{12})^2 = s_{12}^2 = 0 \tag{55}$$

The zero value for a light signal's interval turns out to be a universal property independent of reference frame.

Thus far, the invariance of the four-dimensional interval under four-dimensional rotations (Lorentz transformations) was established in Equations (54) and (55) only for the light intervals as a direct consequence of the invariance of the speed of light in a vacuum. This result, however, turns out to be much more general. Consider two *arbitrary* events (t_1, x_1, y_1, z_1) and (t_2, x_2, y_2, z_2), that are not necessarily connected by the light signal (do not lie on the light's world line), and introduce the corresponding interval

$$s_{12}^2 = c^2 (t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2 \tag{56}$$

Applying the Lorentz transformations (43), the reader can prove that this interval also turns out to be independent of the choice of the reference frame:

$$s_{12} = s'_{12} \tag{57}$$

Thus, our geometrical interpretation is not limited to the processes involving the propagation of light, but holds for *all* possible processes. It reflects fundamental physical properties of time and space as a whole.

Now, let us look at this result from a geometrical viewpoint. In a three-dimensional space, the distance between two points (the length of the segment connecting them) does not depend on the orientation of axes (x, y, z). When we switch to another system of axes (x', y', z'), each point is assigned a new set of coordinates. The new coordinates of a point are linear combinations of the old ones with the coefficients depending on the angles between the old and new axes; however, the length of a segment does not change.

The behavior of the four-dimensional interval displays a striking similarity to this property of the segment: as we switch to another reference system K′, the new coordinates of an event are expressed in terms of the old ones by linear equations (Lorentz transformations), whose coefficients depend on the relative velocity between the two systems; however, the interval itself does not change. We can therefore consider the interval s_{12} as the "distance" between corresponding points in four-dimensional space; the Lorentz transformations from this viewpoint can be thought of as a rotation from one four-dimensional system of axes to another.

However, the analogy found with ordinary three-dimensional space is not complete. Unlike the distances in ordinary space, the squares of the temporal and spatial coordinates in the interval (56) enter with the *opposite* signs. This fact is a manifestation of the fundamental physical difference between time and space. In order to emphasize this difference, their combination is sometimes called (3 + 1)- rather than four-dimensional space. It is a four-dimensional space where one dimension is physically different from the other three. This leads to a peculiar property that the value of s_{12}^2, unlike the square of the ordinary distance, can be negative or equal to zero even if the interval connects two *different* events. As a result, for any chosen point (that is an event t_1, $\boldsymbol{r}_1$) in Minkovsky's world, the set of all other events can be divided into three different regions:

1. The region where

$$s_{12}^2 > 0 \tag{58}$$

that is $c^2 (t_2 - t_1)^2 > \boldsymbol{r}_{12}^2$ (the temporal component of the interval is greater than the spatial component). We call such intervals the time-like intervals. In particular, when $\boldsymbol{r}_1 = \boldsymbol{r}_2$ that is $\boldsymbol{r}_{12} = 0$ (both events occur in the same place), the four-dimensional interval reduces to just the time interval between the events (a proper time!), multiplied by c.

2. The region where

$$s_{12}^2 < 0\,, \tag{59}$$

that is $c^2 (t_2 - t_1)^2 < r_{12}^2$ (the spatial component of the interval is greater than the temporal component). We call such intervals space-like. In particular, when $t_1 = t_2$ (the two events are simultaneous), the interval is determined only by the distance between the events. We see from Equations (59) and (56) that the spatial "contribution" to s_{12}^2 is the negative square of the distance r_{12}.

3. The region where

$$s_{12}^2 = 0 \tag{60}$$

(the temporal and spatial components in the interval have equal weights). We call such intervals the zero- or else isotropic intervals. The term "zero" is self-evident from the definition in Equation (60). The term "isotropic" stems from ordinary three-dimensional geometry, where a vector $\boldsymbol{r}_{12} = 0$ reduces to a single point and therefore cannot be characterized by any specific direction. In this respect it is isotropic. The reader must be careful about this part of the analogy, because a *four*-dimensional zero-interval does not necessarily reduce to a single point! It can connect two *different* points. These points represent the events that can be linked to one another by a light signal. The zero length of this four-dimensional interval reflects the fact that the squares of its temporal and spatial parts, being equal in magnitude, enter the interval with the opposite signs.

Because of the invariance of the interval, its belonging to one of the three different types has an absolute character, that is, it does not depend on the choice of the reference frame.

Juxtaposition of physical events and points of the four-dimensional space–time allows us to visualize the relationships between different events by using space–time diagrams. It is true that we cannot adequately image all four dimensions of Minkowski's world on a sheet of paper having only two dimensions. We will therefore plot on the graph only one spatial dimension x and the time dimension ct. Physically, this means that we consider only the set of events that happen on one straight line (the x-axis). Looking at it slightly differently, we may say that we consider the history of this axis.

Pick up one of the events in this history (call it event O) as a reference event. It will serve as an origin for measuring time ct and the spatial coordinate x. Draw the ct-axis through this point up perpendicularly to the x-axis (Fig. 2.7). Let us take the directions indicated by arrows in Figure 2.7 to be positive. Now, all the events that happen at the "origin" of the line x after the event O, are represented by the points on the ct-axis above point O; all events preceding the event O belong to the part of the ct-axis below this point. We say that the ct-axis forms *the world line* of the point $x = 0$ (that is, a set of all events successively happening at this point). On the other hand, all the points of the x-axis at the moment $t = 0$ are represented by the x-axis itself. We may call it a world line of all events on x that are simultaneous with the event O. Alternatively, we may call it a world line of a hypothetical superluminal particle that traces out all the line x at one moment $ct = 0$ (it must therefore move with an infinite speed).

If we attempt to take into the picture the y-axis (which is perpendicular to both ct and x), then the world lines of all the photons moving in the xy plane and passing through the origin form generatrices of a conical surface with the open angle 90° (Fig. 2.8). This surface had been called the *light cone*. Because there is also the z-axis, the (ct, x, y) light cone forms a sort of "subsurface" of the four-dimensional Minkowski's world. If one wants to include the z direction to make the complete description, one has to consider all the photons moving in three-dimensional space (x, y, z) and passing through the origin at the zero moment. Their world lines will form a three-dimensional conical surface in four-dimensional space–time. Mathematicians call this type of surface a "hyper-surface". We can try somehow to imagine it, but we cannot model or depict it in real space.

From the geometrical viewpoint, any world line in space–time is just a graph of motion of some particle or process. It forms a trajectory in space–time. We should not confuse it with a purely spatial trajectory of the given motion. For example, the *space–time* trajectory (the world line) of a stationary particle at point $x = 0$ is the ct-axis. The purely *spatial* trajectory of the same point is the point $x = 0$ itself. The spatial trajectory is in this case reduced to a point in space (because the particle is stationary!) The spatial trajectories of the photons moving through the origin in the plane xy all lie in this plane, whereas their world lines all lie on the surface of the light cone in Figure 2.8.

Let us now plot on our diagram all possible types of intervals connecting the event O with other events. We will then see that the three types of intervals indicated above fill out three different regions of Minkovski's space-time.

Indeed, all time-like intervals [of the type in Eq. (58)] fall inside the light cone. The space-like intervals [of the type in Eq. (59)] turn out to lie outside the light cone. And, finally, all zero intervals [of the type in Eq. (60)] lie along generatrices of the light cone itself. Thus, three different types of intervals analytically distinguished by the criteria (58)–(60) correspond to their locations in geometrically different domains. This geometrical difference, in turn, corresponds to fundamental physical

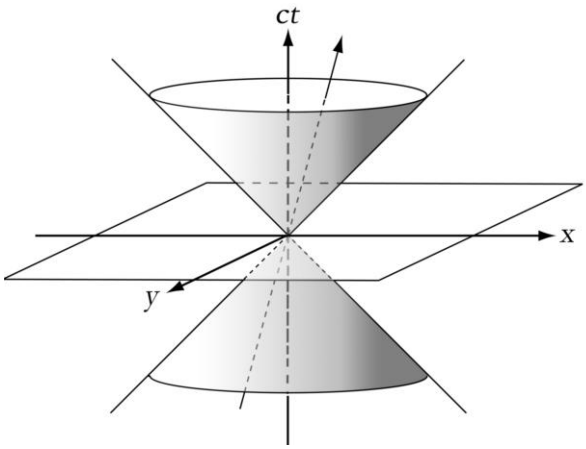


Fig. 2.8

difference between them. We can see it with full clarity if we write down Equation (56) for an interval in terms of the speed of corresponding object or process:

$$v = \frac{|\boldsymbol{r}_2 - \boldsymbol{r}_1|}{t_2 - t_1} = \frac{r_{12}}{t_2 - t_1} \tag{61}$$

We will have

$$s_{12}^2 = c^2 (t_2 - t_1)^2 - r_{12}^2 = c^2 (t_2 - t_1)^2 \left(1 - \frac{v^2}{c^2}\right) = c^2 (t_2 - t_1)^2 \gamma^{-2}(v) \tag{62}$$

It follows that for the time-like intervals $\gamma^2(v) > 0$ and the speed $v < c$; for the zero intervals $\gamma(v) = \infty$ and $v = c$; and finally, for space-like intervals, $\gamma^2(v) < 0$, which corresponds to $v > c$. The intervals of the first two types lie inside or on the surface of the light cone in Figure 2.8. They connect the events which can be either the cause or the effect of the event O. In contrast, the intervals of the third type lie outside the light cone; they may connect events only by a process characterized by superluminal speed $v > c$. As we will show later, this type of process cannot be used for signaling; therefore, there is no causal connection between corresponding events. The same event outside the light cone with the vertex at O can be observed as occurring earlier or later than the event O, depending on a chosen reference frame. In particular, we can find a system of reference in which the event occurs simultaneously with O. The concept "earlier–later–simultaneous" is relative for such pairs of events.

We can illustrate this relativity figuratively using Minkovski's diagrams, if we depict in one graph the coordinate axes for two different inertial systems K and K′. We have already noticed that the time axis of an inertial system is coincident with its origin's world line. Let now a real particle move uniformly along x with a speed v and pass the point $x = 0$ at the moment $t = 0$, so that

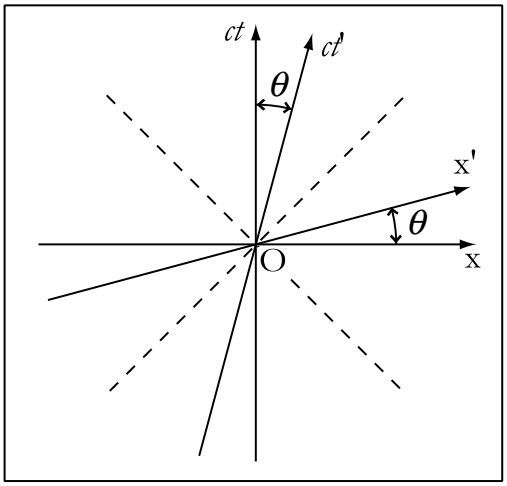
$$x = vt = \frac{v}{c} \cdot ct \tag{63}$$

The particle is stationary in a reference frame K′ co-moving together with it. If the particle is at the origin of this reference frame, Equation (63) describes the world line of the origin (Fig. 2.9). On the other hand, we have found that the world line of an origin is the time axis of the associated reference frame. Thus, the time axis ct' of system K′ is just the world line of its origin traced out in system K. As is seen from Figure 2.9, the axis ct' makes with ct an angle

$$\theta = \arctan \frac{v}{c} \tag{64}$$

For positive v this line is in the first and third quadrants of the (ct, x) plane. For a particle moving in the negative x direction we have $v < 0$, and the corresponding world line passes in the second and fourth quadrants of the plane (ct, x). If this particle is

Fig. 2.9



a photon, then $v = c$, and $\theta = \pm 45^\circ$. We see again that the world lines of photons moving in the positive or negative x direction form mutually perpendicular bisectors of the angles between the coordinate axes.

Since the world line of the origin O′ represents the time axis of the corresponding system K′, all the events happening at O′ after the moment $t' = 0$ are represented by the points on ct' above the origin; all the events that occur there before this moment are on the lower part of line ct'.

Let us now turn to the x'-axis. How is it represented in our diagram? Earlier, you remember, we have defined a spatial coordinate axis as the set of all events on this axis that are simultaneous with the event O. Because now we are interested in the x'-axis of moving system K′, we must consider the events on this axis that are simultaneous with the event O *by clocks of system K′*. All these events are characterized by one moment $t' = 0$ in readings of all K′-clocks. Using Equations (44), we find that in K all these events occur at *different* moments depending on their coordinate x:

$$ct = \frac{v}{c}\,x \tag{65}$$

However, this is just the equation for the line passing through the origin and making with the x-axis the same angle as in Equation (64). This line represents the x'-axis of system K′ from the viewpoint of system K.

We see that the axes ct' and x' turn out to have been rotated through the same angle relative to ct and x respectively (Fig. 2.9). However, in contrast with the usual geometry of purely spatial rotations, the rotations here are in the opposite senses (this is yet another manifestation of the physical difference between space and time). As a result, the coordinate system (ct', x') seems skewed (deformed) from the viewpoint of system K [in exactly the same way, the system (x, ct) would look distorted from the viewpoint of system K′]. Precisely because of such a "deformation", the photon's

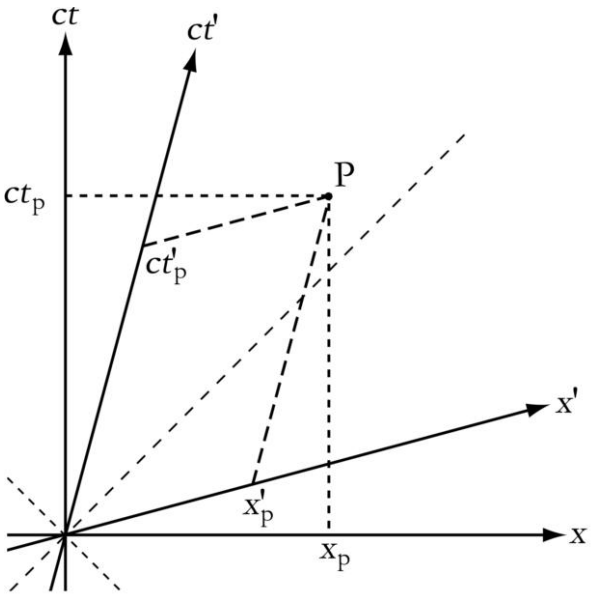


Fig. 2.10

world line remains, as it should, the bisector of the angle between the ct'- and x'-axes, that is, the speed of light in K′ is also equal to c. But now the photon's world line in Figure 2.9 makes the angle $\alpha = 45^\circ - \theta < 45^\circ$ with the axes ct', x'. This is due to the fact that these axes are plotted in the system K, which is "alien" for them. At $v \to c$ the angle θ approaches 45°, and α approaches zero, that is, the axes ct' and x' approach the photon's world line; however, the ct'-axis always remains inside, whereas the x'-axis remains outside the light cone.

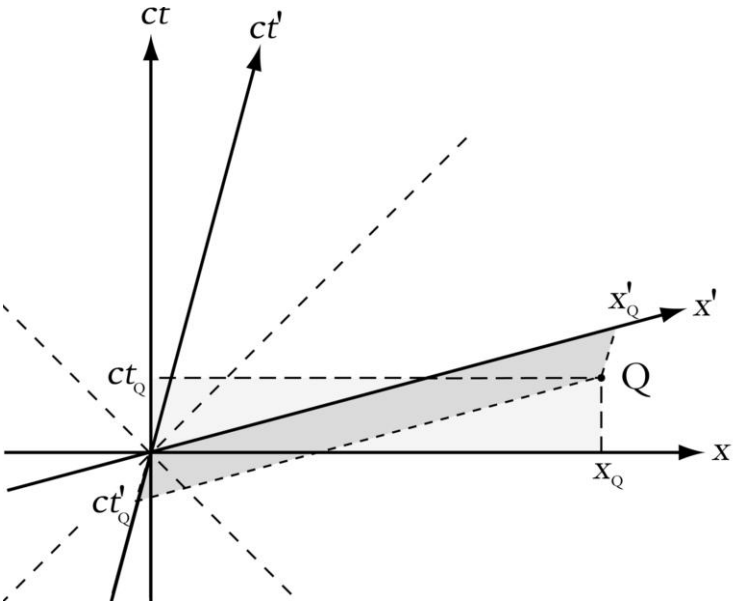
If we now consider an event P, then its coordinates (ct_P, x_P) in K are the normal projections of point P on the axes ct, x, and the coordinates (ct'_P, x'_P) are the oblique projections of this point on the axes ct', x' (Fig. 2.10).

We can now consider the relation "earlier–later" between different events graphically, from the viewpoint of the systems K and K′. Let, for example, the event P lie inside the light cone with the vertex at O. This means that the events O and P admit a causal connection. If $ct_P > 0$ (P lies in the upper fold of the light cone), then event P can result as an effect (the consequence) of event O. Projecting point P onto the ct'-axis, we see that at any possible tilt of this axis, projection ct'_P lies in the upper semiplane, that is, $ct'_P > 0$. The event P occurs after event O for all possible observers. We say that it is in *absolute future* with respect to O.

If the point P lies in the lower fold of the light cone, then event P can be the cause of the event O and accordingly it will be observed *before* O in any reference frame. In this case it is in *absolute past* with respect to O.

Consider now another point Q lying *outside* the light cone (Fig. 2.11). Let this point lie above the x-axis, that is, its projection $ct_Q > 0$. Physically, it means that Q occurs later than O in the inertial system K. But the projection ct'_Q of the very same point onto the ct' axis can be negative (this will be the case for any system with the x'-

Fig. 2.11



axis passing above Q, as is seen from Fig. 2.11). Physically it means that in corresponding inertial system K′ the event Q occurs earlier than O. The relation earlier–later is not invariant for pairs of events like O and Q. For such events one can always find a reference frame for which their time ordering is switched to the opposite. Also there exists such a system for which both events of the pair are simultaneous ($ct_Q = 0$). In this system the interval between the events is determined entirely by their spatial separation r'_{OQ}. Therefore, all the events outside the light cone can be called *absolutely remote* with respect to the event O. They are connected with O by space-like intervals.

It is easy to see that any communications between the event O and an event outside the light cone would involve signals moving faster than light in a vacuum. We will discuss later the relation between such signals and a fundamental physical principle of causality.

We conclude this section by demonstrating how the concept of interval can be used to derive the equation for proper time. Consider two successive events A and B on a stationary body K. The proper time between the events is the difference $\Delta t = t_B - t_A$ between the moments t_A and t_B of these events read by a clock K. Since both events occur at the same place, we have $x_{AB} = x_B - x_A = 0$, and the interval between the events in this reference frame is reduced to $s^2_{AB} = c^2(\Delta t)^2$. In another reference frame K′ moving with speed V along the x-axis, the body K moves in the opposite direction with the same speed V. The clocks of system K′ read the moments t'_A and t'_B for the events A and B, respectively, so that the time interval between the events is $\Delta t' = t'_B - t'_A$. In system K′ these events are spatially separated by the distance $x' = V\Delta t'$. The four-dimensional interval between the events expressed in terms of

the coordinates of system K′ is $(s'_{AB})^2 = c^2(\Delta t')^2 - V^2(\Delta t')^2$. Because of the invariance of the interval we have $c^2(\Delta t)^2 = (c^2 - V^2)(\Delta t')^2$ or

$$\Delta t = \gamma^{-1}(V)\,\Delta t' \tag{66}$$

This is the result in Equation (18) obtained in different way in Section 2.5.

2.10 What is horizontal?

Mister O'Bryen is a very rich man. When back on Earth from his space missions, he likes to travel in his own jet ship with a swimming pool in it. Each time when the pool is being filled with fresh sea water, Mr. O'Bryen likes to watch the water rising. Knowing some physics, he never misses an opportunity to note that the surface of the rising water remains horizontal in the uniformly moving ship as it does when the ship is anchored in a harbor. "What a sound manifestation of the principle of relativity," Mr. O'Bryen murmurs. "A ship's resident cannot tell whether the ship is moving or anchored by observing the rising water in the ship's tank."

At this very time an observer on the sea shore also watches the same process, and notices an unusual phenomenon: the surface of the rising water in the moving ship is tilted to the horizon, so that the water at the rear edge of the pool is higher than that at the front edge (Fig 2.12). The faster the ship's motion, the steeper is the tilt. It

Fig. 2.12 The rising water in a moving tank as observed from the shore. The tilt occurs because the events A and B that are simultaneous in the shore's reference frame are not simultaneous in the ship. Event B occurs later in the ship's history than event A. Accordingly, the water at B is higher than it is at A. The line AB as observed from the shore *now* is tilted. The tilt is highly exaggerated.

even occurs to the observer (who also happens to be a good engineer) that this phenomenon could be used for measuring the ship's speed. "What a sound manifestation of the relativity of time," thinks the engineer. "Two simultaneous events separated by a distance x along the ship's motion are not simultaneous in the moving ship. They are separated by a time $t' = \gamma(u)\, ux/c^2$, if u is the ship's speed. The points A and B that I mark *now* at the front and the rear edges of the water surface are marked with the different moments of time by the ship's synchronized clocks. *Now* the ship's clock at A lags behind the ship's clock at B by the amount t'. If the water in the ship's tank rises at a rate $v' = \mathrm{d}y'/\mathrm{d}t'$ by the ship's clock, its level at B must be higher than at A by the amount $y = y' = v' t'$, since it had more time to rise. Using the above expression for t', one would obtain for y

$$y = \gamma(u)\, \frac{uv'}{c^2}\, x = \gamma^2(u)\, \frac{uv}{c^2}\, x \tag{67}$$

where v is the rate of the water rise in the shore's reference frame. Thus the water surface observed by me must be inclined by the angle α:

$$\tan\alpha = \frac{y}{x} = \gamma^2(u)\, \frac{uv}{c^2} = \gamma(u)\, \frac{uv'}{c^2} \tag{68}$$

If I know v', the engineer concludes, then measuring α (provided that my instruments are sensitive enough) will give me a quantitative measure of the ship's velocity u through Equation (68).

Now, let us read this equation. It says that there is no tilt when $u = 0$. This makes sense, doesn't it? When the ship is anchored, the rising water surface in its tank is horizontal at any moment for either Mr. O'Bryen or the engineer. If $v' = 0$ (the water is still) there is again no tilt, which also makes sense. The surface of *still* water is horizontal for both observers no matter how fast the ship moves so long as the motion remains uniform. It is only when the ship and the water in its tank are both moving that the effect is being observed by the engineer. When $v' < 0$ (the water in the tank is sinking), the sign of $\tan\alpha$ is negative, which means that the water level as observed by the engineer is higher at the front edge than at the rear (Fig. 2.12). Thus the surface of the same pool of water is horizontal in one inertial reference frame (the ship), and it may be inclined in another one (e.g. the sea shore). The property of being horizontal turns out to be relative even for the two observers in the same locality, because of relativity of time.

Some readers may be tempted to think that the phenomenon is of the same nature as one observed by Alice and Tom (recall the experiment with the aquarium in the Introduction). However, the water tilt in that experiment was caused exclusively by the *acceleration*, not the velocity. And this makes all the difference. The water tilt in the "aquarium effect" is determined by the ratio a/g where a is the acceleration of Tom's reference frame, and g is the acceleration due to gravity. The effect is easy to observe since the acceleration a comparable to g is easy to achieve. And it was observed for even a still water ($v' = 0$!) by *both* Alice and Tom. It is therefore of the ef-

fects that we would call absolute. It reveals the acceleration of a system to an insider. In contrast, the "pool effect" is observed only by the engineer when the water rises or sinks ($v' \neq 0$!) and the ship moves *without* acceleration at a speed u. We have no g as a scaling constant here. Instead, a new constant, c, enters the picture. The water surface tilt is in this case determined by *two* ratios, u/c and v'/c. In the real conditions both of them are small, and accordingly the surface tilt α would be ridiculously small. For example, if Mr. O'Bryen's ship rushes at a speed of 360 km h^{-1} (like a small aeroplane!) and the water in his tank rises at a rate of 1 m s^{-1}, Equation (68) gives $\alpha \approx 1.1 \cdot 10^{-13}$! We see that we have badly overestimated the precision of the engineer's measurements in our thought experiment. However, the equations that we use are correct. For other conditions they can give, as we shall see further, noticeable values of α. Generally, they describe a new relativistic effect that is seen and registered differently in different inertial frames.

Now, suppose that our inertial observers, Mr. O'Bryen and the engineer, have both focused on the motion of the same element of the water surface in the pool. As the water in the pool rises, the surface element moves vertically as seen by Mr. O'Bryen, and sideways as seen by the engineer (Fig. 2.13). The engineer denotes the velocity of the surface element as $\boldsymbol{V}$; he measures u for its horizontal component and $v = v'/\gamma(u)$ for the vertical component. So the velocity vector makes an angle θ with the vertical, where

$$\tan\theta = \frac{u}{v} = \frac{u}{v'}\,\gamma(u) \tag{69}$$

Comparison with Equation (68) shows that $\theta \neq \alpha$, that is, the direction of $\boldsymbol{V}$ is not generally perpendicular to the surface element. Well, we have already become psychologically prepared for various manifestations of relativity, and here is just another one: the water surface moves perpendicularly to itself as seen by Mr. O'Bryen, and not perpendicularly to itself as observed by the engineer.

Is there at least a small island of absoluteness in this ocean of relativity? Does there exist such a speed of water rise in the pool at which the water surface would move perpendicularly to itself in all inertial reference frames? And if it does, what is its value?

Let us introduce a unit vector $\boldsymbol{n}$ perpendicular to the water surface in the engineer's reference frame. By definition, vector $\boldsymbol{n}$ makes an angle α with the vertical. So the

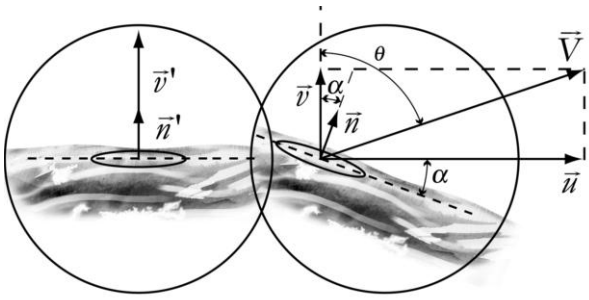


Fig. 2.13 Relative orientation of the surface element and its velocity. (a) In the ship's reference frame. (b) In the shore's reference frame.

mathematical condition for our question and the sought for value of v is $\boldsymbol{V} \| \boldsymbol{n}$, that is, $\tan\alpha = \tan\theta$, or

$$\frac{u}{v} = \gamma^2(u)\,\frac{uv}{c^2} \tag{70}$$

Since $v' = \gamma(u)\,v$, the last equation is satisfied for $v' = c$. The water in Mr. O'Bryen's tank should rise or sink with the speed of light!

Let us for a while set aside the question of whether the water can do such a feat; let us first focus on the result itself. We know already that the value of c is an invariant of Lorentz transformations – it is absolute. We therefore suspect that in this case the magnitude of $\boldsymbol{V}$, the surface element's speed as measured by the engineer, must also remain equal to c. Let us check this. For $v' = c$ the transverse component of $\boldsymbol{V}$ measured by the engineer will be $v = v'\gamma^{-1}(u) = c\gamma^{-1}(u)$. The magnitude of V is therefore

$$V = \sqrt{v^2 + u^2} = \sqrt{c^2\gamma^{-2} + u^2} = c \tag{71}$$

A remarkable result! *If* the water in the tank could rise (or sink) with the speed of light in a vacuum, then any shore-based observer would also see it rising or sinking with the same speed in the direction perpendicular to its surface. The surface's property of moving perpendicular to itself becomes absolute when it acquires the absolute speed – that of light in a vacuum! But since the surface is now tilted, the *direction* of its motion must also be tilted by the same angle with respect to the vertical direction.

Now, if the reader compares the above conclusion with that shown in Fig. 2.4 from a quite different viewpoint, he will find them identical. No wonder! Once we have admitted the possibility of the water surface moving with the speed of light, we have naturally arrived at the result shown in Fig. 2.4 for a light wave.

The water cannot, of course, move with the speed c, but it can, in principle, move with a speed arbitrarily close to c. That would definitely be an utterly exotic situation. But the equations we used describe the same effect for any moving surface. We can therefore, first, substitute the water surface with a wave front of any nature. It may be the rising shock wave produced by hitting the bottom of the tank, or the waves from a pebble thrown on to a still water surface. It may be a light wave in any transparent medium. It may also be a shock wave propagating in an ultra-dense medium, say the interior of a neutron star. Second, we may consider a reference frame moving much faster than Mr. O'Bryen's jet ship. In the last two cases the values of α and θ may be measurable. Consider, for example, the light wave rising from a bulb in the water tank of a spaceship from a "Star Trek" serial. The spaceship moves horizontally with speed 300 km s^{-1} (which is real for an advanced technology, and very slow by a science fiction standard). The speed of light in water is about 1.33 times less than c. For these conditions, Equations (68) and (69) give $\alpha = 7.5 \cdot 10^{-4}$ and $\theta = 1.33 \cdot 10^{-3}$, respectively. The difference in values for α and θ shows that vector $\boldsymbol{V}$ does not make a right-angle with the wave surface. This result illustrates our previous conclusion: the surface element or the element of the wave front moving perpendicular to itself in one reference frame does not generally move perpendicularly to itself in another in-

ertial reference frame but for one exception: when an element moves with the speed of light in a vacuum.

We can summarize these results in the following way. Generally, the wave front in a moving medium is not necessarily perpendicular to the wave velocity ***V***. Particularly, waves from a pebble thrown into Mr. O'Bryen's pool are circular as seen by Mr. O'Bryen, and non-circular for the engineer. The deviation from concentricity will be very small for these waves. However, for the case of the light waves diverging from a bulb in a water tank inside a spaceship the difference would be noticeable for the two different inertial observers. Again, the waves would be spherical for Mr. O'Bryen and non-spherical for the engineer, and this non-sphericity could be measured. Mr. O'Bryen interprets what he sees as yet another manifestation of the principle of relativity: he cannot tell whether his ship is at rest or on the move by observing waves in his pool. The engineer interpreted his observations as a kinematic effect: he sees the very same waves diverging non-spherically because the medium supporting the waves (water in the pool) is itself moving with the ship. Naturally, this motion singles out a special direction among all the others – the direction of the ship's velocity, that of ***u***. As a result, some physical properties of such moving medium (e. g. its ability to transmit waves) depend on the angle between ***u*** and the wave vector ***k***. We call such a medium non-isotropic (or anisotropic). Our conclusion is that any isotropic medium, when moving, is equivalent to a fictitious stationary anisotropic medium. In this respect, it can resemble a certain type of crystals, whose physical properties are different for different directions, or else it resembles a fluid at rest in an external electric or magnetic field.

However, if a disturbance in a fluid can propagate, as light in a vacuum, with the speed c, this propagation, according to our equations, will not be affected by translational motion of the fluid. Whether this fluid moves or not becomes immaterial, since the motion of this medium will not be revealed in observations. Why, then, talk about the medium at all? According to all experiments, a disturbance propagating with $v' = c$ can in all respects be considered as propagating just in a vacuum. Stop talking of the medium! It was just what happened to ether about 100 years ago. Since 1905 people have almost never mentioned ether. They refer to vacuum when talking about light propagation in the intergalactic space or in a laboratory container with all the air pumped out. It is true that, as was found out later, vacuum is not an absolute emptiness, and all the particles, photons included, can be described as excited states of vacuum. But this is a different story. And no matter what the revealed complexity of vacuum may be, when it comes to light's motion, its description does not require any carrier necessary to support the light waves. In this respect (as we can see things today), light exists and propagates in its own right!

Well, what happens if we apply our equations to light in a vacuum? Consider Mr. O'Bryen during one of his space missions trying to signal with a laser pulse from his spaceship to the engineer. The pulse is fired in the direction perpendicular to the ship's motion (Fig. 2.14). But the engineer sees the pulse propagating in a slightly different direction that makes an angle θ with the perpendicular line. Accordingly, he observes the wave front of the pulse tilted through the same angle (since $\alpha = \theta$ in this case). He finds the angle by putting in the equations $v' = c$:

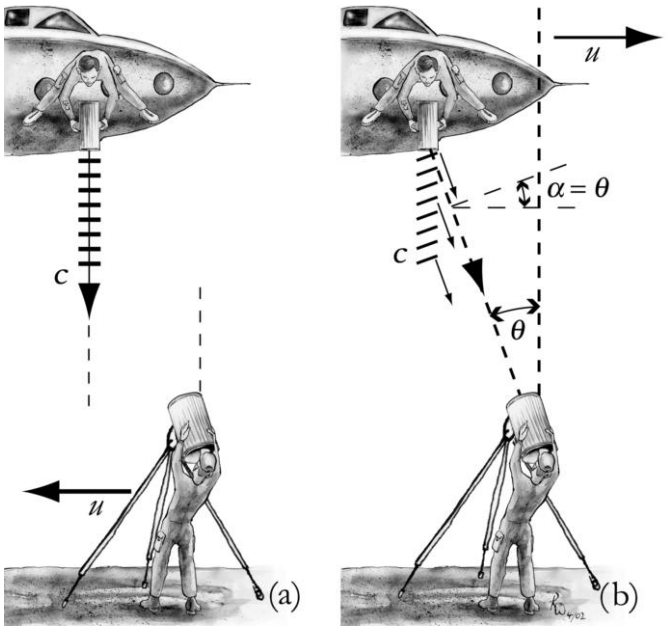


Fig. 2.14 Laser signaling from the spaceship to the engineer. (a) Observed by Mr. O'Bryen. (b) Observed by the engineer.

$$\tan\alpha = \tan\theta = \frac{u}{c}\,\gamma\,(u) \tag{72}$$

He accordingly has to tilt his detector to achieve better acceptance of the pulse.

Now, substitute the spaceship with a distant star moving in the direction transverse to the line connecting it with the engineer (or an astronomer, for that matter). Then we will get the same result for the light from that star! The astronomer will have to tilt his telescope to achieve a better image of the star (or the star's image will be shifted slightly from the center of the vision field). There is a subtle point here, however: in the case of light the astronomer cannot tell *relative to what* does he have to tilt his telescope, since the actual line of sight to the star is his only reference direction. The effect can only be observed if the astronomer's reference frame changes the direction of its motion. With the position of the telescope rigidly fixed in this reference frame, the change in its motion will cause the star's image to shift slightly with time. This shift can be found by comparing the star's photographs taken at different times. But precisely such a situation is realized on Earth, which changes the direction of its translational motion due to its orbiting around the Sun. Such motion must be manifest in periodic (with a period of 1 year) circular or elliptic motion of a star's image in the photographic films obtained with a fixed telescope. This phenomenon was noticed long ago by astronomers and is known as the *aberration of light* [17]. But what is less well known is this: while the

deviation of the light's velocity vector from the perpendicular line can be easily explained in terms of non-relativistic addition of velocities, the corresponding tilt of the wave front, which remains perpendicular to the velocity vector, is a *purely relativistic effect*. Actually, it is the *manifestation of relativity of time*, discussed by the engineer at the beginning of this section.

3
The Velocities' Play

3.1
The addition of collinear velocities

We now have sufficient background to establish a universal law of velocity transformation. It would relate the velocities of an object measured in two different reference frames.

The most general expression for the result of superposition of two motions (along the same line) with velocities V and v has the form

$$v = \eta(V, v')(V + v') \tag{1}$$

The function $\eta(V, v')$ must be such that for $V \ll c$, $v' \ll c$, the value of $\eta(V, v')$ approaches 1. On the other hand, the speed of light added with the speed V must again give c in the result. In other words, for $v' = c$ there must be

$$v = \eta(V, c)(V + c) = c \tag{2}$$

It follows that

$$\eta(V, c) = \frac{c}{V + c} = \frac{1}{1 + \dfrac{V}{c}} \tag{3a}$$

The symmetry with regard to the interchange between V and v' demands that

$$\eta(c, v') = \frac{1}{1 + \dfrac{v'}{c}} \tag{3b}$$

To satisfy both requirements, $\eta(V, v')$ must have the form

$$\eta(V, v') = \frac{1}{1 + \dfrac{Vv'}{c^2}} \tag{4}$$

Special Relativity and Motions Faster than Light. Moses Fayngold

ISBN: 3-527-40344-2

Thus, the result of addition of two collinear velocities, which complies with the postulate that c is constant, is given by

$$v = \frac{V + v'}{1 + \dfrac{V v'}{c^2}} \tag{5}$$

Note that in deriving Equation (5) we did not use the Lorentz transformation explicitly. Our reasoning was based only upon the general requirement that the expression has to be symmetrical with regard to the velocities being added and that if one of the velocities equals c, the result of their addition must also be equal to c. However, as we have seen in Chapter 1, this latter requirement leads immediately to the Lorentz transformation for spatial and time coordinates. That is why the use of this condition in the derivation of Equation (5) is equivalent to an implicit use of the transformation itself.

Let us derive Equation (5) in another way, where the Lorentz transformation is used explicitly. For that purpose, we will take the correct Equation (9) in Section 2.5 and express the length Δx in the system K in terms of the proper length:

$$\Delta x = \gamma^{-1}(V)\,\Delta x' \tag{6}$$

and the time of the object's motion, in terms of the corresponding quantities $\Delta x'$ and $\Delta t'$, according to Equation (43) in Section 2.6:

$$\Delta t = \gamma(V)\left(\Delta t' + \frac{V}{c^2}\,\Delta x'\right) \tag{7}$$

The reader should remember that Δx here is the length in K of a moving segment whose ends' coordinates are being measured simultaneously, and Δt is the time in K corresponding to the motion of some point along this segment. The beginning and the end of this motion are being observed both in K an K′ at different points in space, namely at the beginning and the end of the segment Δx. This is why both $\Delta t'$ and $\Delta x'$ are present in Equation (7) for Δt. The time $\Delta t'$ in Equation (7) is not a proper time of some process in K′, since the point in question is moving in both reference frames.

Finally, we compose the ratio

$$\frac{\Delta x}{\Delta t} = \gamma^{-2}(V)\,\frac{\Delta x'}{\Delta t' + \dfrac{V}{c^2}\,\Delta x'} \tag{8}$$

and add the resulting expression to V:

$$v = V + \frac{\Delta x}{\Delta t} = V + \left(1 - \frac{V^2}{c^2}\right)\frac{v'}{1 + \dfrac{V v'}{c^2}} = \frac{V + v'}{1 + \dfrac{V v'}{c^2}} \tag{9}$$

The result we obtained is in agreement with Equation (5).

3.2 The addition of arbitrarily directed velocities

The Lorentz transformation also allows us to obtain the general law of addition of velocities with arbitrary directions.

Suppose again that the system K′ is moving relative to K at a speed V along the x-axis. Consider an object in a state of relative motion at a velocity v' with respect to K′. The nature of this object has no importance whatsoever: it can be a massive particle, a photon, or just a mathematical point with a given law of motion. The components of velocity v' are

$$v'_x = \frac{\mathrm{d}x'}{\mathrm{d}t'}\,, \qquad v'_y = \frac{\mathrm{d}y'}{\mathrm{d}t'}\,, \qquad v'_z = \frac{\mathrm{d}z'}{\mathrm{d}t'} \tag{10}$$

Now, what is the velocity of the same object as observed from the system K? The answer follows almost automatically from the Lorentz transformation. According to the definition of velocity, one has in system K

$$v_x = \frac{\mathrm{d}x}{\mathrm{d}t}\,, \qquad v_y = \frac{\mathrm{d}y}{\mathrm{d}t}\,, \qquad v_z = \frac{\mathrm{d}z}{\mathrm{d}t} \tag{11}$$

Here the infinitely small time interval $\mathrm{d}t$ and corresponding increments of coordinates $\mathrm{d}x$, $\mathrm{d}y$, $\mathrm{d}z$, separate two events: the passing of the object through two near points. For the observer in K′ the same events are separated by increments $\mathrm{d}t'$, $\mathrm{d}x'$, $\mathrm{d}y'$, $\mathrm{d}z'$. Now all we have to do is to express the increments in K in terms of the corresponding increments in K′. Taking differential of the Lorentz transformation in Equation (43) in Section 2.6, we have

$$\mathrm{d}x = \gamma(V)\,(\mathrm{d}x' + V\,\mathrm{d}t')\,;\;\; \mathrm{d}y = \mathrm{d}y'\,;\;\; \mathrm{d}z = \mathrm{d}z'\,;\;\; \mathrm{d}t = \gamma(V)\left(\mathrm{d}t' + \frac{V}{c^2}\,\mathrm{d}x'\right) \tag{12}$$

It follows immediately that the sought-for components of velocity are

$$\left.\begin{aligned} v_x &= \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}x' + V\,\mathrm{d}t'}{\mathrm{d}t' + \dfrac{V}{c^2}\,\mathrm{d}x'} = \frac{V + v'_x}{1 + \dfrac{V\,v'_x}{c^2}} \\ v_y &= \frac{v'_y}{\gamma(V)\left(1 + \dfrac{V\,v'_x}{c^2}\right)}\,; \quad v_z = \frac{v'_z}{\gamma(V)\left(1 + \dfrac{V\,v'_x}{c^2}\right)} \end{aligned}\right\} \tag{13}$$

We have expressed the velocity components in the stationary system in terms of the velocity components in the moving system.

If we regard K′ as stationary and K as moving, we can obtain the reciprocal transformation by just swapping primed and unprimed components and changing the sign of the relative velocity:

$$v'_x = \frac{V - v_x}{1 - \dfrac{V v_x}{c^2}}; \quad v'_y = \frac{v_y}{\gamma(V)\left(1 - \dfrac{V v_x}{c^2}\right)}; \quad v'_z = \frac{v_z}{\gamma(V)\left(1 - \dfrac{V v_x}{c^2}\right)} \tag{14}$$

The reader can obtain the same result by solving Equation (13) for the primed components of the velocity $\boldsymbol{v}'$.

We have thus arrived at the sought-for general equations for velocity transformation, which describe both the simple addition of slow (relative to light) motions and the "weird" behavior of light.

3.3 The velocities' play

Let us now "play" a little with the derived equations in order to see how they work in different situations.

First, we want to make sure that for the motions slow with respect to light our equations reduce to simple addition of velocities. Setting in Equation (14) $V \ll c$, $v'_x \ll c$, we obtain

$$v_x \approx V + v'_x; \quad v_y \approx v'_y; \quad v_z \approx v'_z \tag{15}$$

The first of these relationships is equivalent to Equation (7) in Section 2.2. The theory of relativity has not overturned the previous theory, it has only revealed its approximate character and clearly charted the domain of its applicability. One only has to keep in mind that the borders of this domain are themselves relative – they depend on the accuracy of our measurements. On increasing this accuracy, we can in principle notice the approximate character of the relations (15) even at very small velocities. We will discuss one such situation in detail in Section 5.7.

The new and unexpected (from the viewpoint of Newtonian mechanics) phenomena become significant when the velocities involved are comparable to c.

Consider an object moving in system K′ with velocity $\boldsymbol{v}' \perp \boldsymbol{V}$, that is, $v'_x = 0$. Then the exact Equations (13) yield

$$v_x = V; \quad v_y = \gamma^{-1}(V)\, v'_y; \quad v_z = \gamma^{-1}(V)\, v'_z \tag{16}$$

We see that the transverse components v_y, v_z are diminished in system K by a factor of $\gamma(V)$, in total accord with the effect of time dilation.

Let now $v'_y = c$, $v'_z = 0$; this corresponds to the experiment with the light pulse within a vertical cylinder, moving horizontally, which has been considered in Section 2.5. According to Equations (16), the speed of the pulse along the cylinder as measured in system K is equal to

$$v_y = \gamma^{-1}(V)\, c \tag{17}$$

which is less than c by a factor of $\gamma(V)$. However, it does not follow from here that the light pulse moves slower in system K. Indeed, Equation (17) is not yet the *whole* result of the transformations (13), but only one part of it, related to the transverse component of the velocity. Let us now take into account its second part – the emergence of the longitudinal component in system K:

$$v_x = V \tag{18}$$

which is caused by the motion of the cylinder itself. Owing to this component, the velocity vector turns out to have been rotated, so that in system K the velocity $\boldsymbol{v}$ is not perpendicular to $\boldsymbol{V}$.

In our case, when the object moving inside the cylinder is a photon, this longitudinal component gives the precisely right contribution to the full velocity to compensate for the slowdown of the transverse motion, so that the total speed is

$$v = \sqrt{V^2 + v_\gamma^2} = \sqrt{V^2 + c^2\left(1 - \frac{V^2}{c^2}\right)} = c \tag{19}$$

Thus, the light velocity vector is oriented differently in system K than in system K′, but its magnitude remains equal to c.

Now we are going to consider a few cases where the velocity of an object is parallel to the relative velocity of the system K and K′.

1. Let system K′ be a spaceship moving relative to K with a speed V close to the speed of light. There is a particle accelerator working inside the spaceship, which produces high-energy particles moving from the rear to the front with a speed v' relative to the spaceship. This speed is also close to that of light, so that we can write

   $$V = c - \Delta V\,, \quad v' = c - \Delta v' \tag{20}$$

   where

   $$\Delta V \ll c\,, \quad \Delta v' \ll c \tag{21}$$

   What is the particles' speed relative to system K?
   According to the "obvious" pre-relativistic Equation (1) in Section 1.2 we would have

   $$v = V + v' = 2\,c - \Delta V - \Delta v' \tag{22}$$

   that is, the particles' speed in K must be nearly twice the speed of light. In reality, however, the particles' speed in K is described by Equation (7) in Section 2.2, leading to Equation (5), so that we have

   $$v = \frac{V + v'}{1 + \dfrac{V\,v'}{c^2}} = \frac{2\,c - \Delta V - \Delta v'}{2 - \dfrac{\Delta V + \Delta v'}{c} + \dfrac{\Delta V + \Delta v'}{c^2}} = \frac{c}{1 + \dfrac{\Delta V\,\Delta v'}{c\,(2\,c - \Delta V - \Delta v')}} \tag{23}$$

Taking into account the condition (21), we can to a high accuracy approximate this exact equation by

$$v \cong c\left(1 - \frac{1}{2}\frac{\Delta V\,\Delta v'}{c^2}\right) \tag{24}$$

We see that v remains less than c.

2. Imagine now two relativistic spaceships moving away from each other. They represent two different inertial systems K_1 and K_2, and their speeds v_1 and v_2 are measured relative to one and the same system K. Because in this case we use the same time t when measuring both speeds, the simple addition equation

$$\frac{d(x_2 - x_1)}{dt} = v_1 + v_2 \tag{25}$$

retains a certain physical meaning. It shows the rate of change of the distance between the spaceships in system K. For instance, if the spaceships fly apart in system K at a speed $0.9\,c$ each, then the separation between them, as seen by an observer in K, increases at a rate $1.8\,c$. This does not in any way contradict the theory of relativity, because the $1.8\,c$ is *not* the speed of their relative motion. To determine their relative speed, one has to use the measuring devices of *one of the spaceships*, that is, to transfer to the rest frame of this spaceship, and then measure the speed of the other spaceship. In this case one will obtain

$$v = \frac{v_1 + v_2}{1 + \frac{v_1\,v_2}{c^2}} = \frac{1.8\,c}{1 + (0.9)^2} \cong 0.995\,c \tag{26}$$

that is, v remains less than c again.

Thus, the number $1.8\,c$ in this example emerges as an intermediate result of the algebraic calculation. It has a physical meaning of the rate of separation change between the two objects in the "alien" reference frame K, in which they are both moving. This number is *not* the experimental result of the direct measurement of the speed of one of the objects relative to the other.

One could ask: "I see two objects flying apart each with a speed $0.9\,c$ relative to me, don't I therefore see them flying apart with the relative speed $1.8\,c$?"

The answer to this would be "No. They do fly apart with the speed $1.8c$, but this is not the speed of their *relative* motion. The relative speed is observed in the rest frame of one of the objects."

This example shows the difference between the speeds $v_1 + v_2$ and the relative speed v.

3. The situation is again the same as in case 1, except that the objects in question are photons, because instead of particle accelerator, the crew of the spaceship is now using a laser. Then $v' = c$. Equation (1) in Section 1.2 would give for the photon speed in K the value $v = c + V$; in reality, however, we will have

$$v = \frac{V + c}{1 + \frac{Vc}{c^2}} = c \tag{27}$$

Here we do not even need any special assumptions about value of V (whether it is much smaller than c or close to c.) At *any* V the relativistic law of addition of velocities gives for the speed of a photon the same value c. The equation works as a simple and at the same time intricate machine: no matter what speed is entered into the machine together with the "c," "c" always comes out in the output. The same result follows if the speed "c" of a photon is determined first in a stationary system K and then one looks at its speed in the spaceship dashing after the photon with a speed V. Here we come back to the question from which Einstein had started his musings that lead him to the theory of relativity. And the theory gives an immediate answer: the spaceship's system K′ is equivalent to K, and in it one observes the same laws of nature and the same rule of addition of velocities. One only has to change V to $-V$, because relative to K′ the system K where the photon speed c has been measured is moving in the opposite direction. We obtain

$$v' = \frac{c - V}{1 - \frac{Vc}{c^2}} = c \tag{28}$$

No matter how hard is the spaceship accelerating, the photon will run away from it with the same speed c.

We have considered these examples, perhaps even in too minute detail, in order to emphasize once again the basic fact lying at the core of the theory of relativity: the invariance of the speed of light. We can express this invariance using a simple geometrical construction. Let us represent all theoretically possible speeds as points on a "speed axis." Then the speed c will be also represented by a point on this axis. It turns out then that this point is singled out from all the rest by its immobility. Indeed, any speed different from c changes under the Lorentz transformations along the direction of motion, and therefore the point representing such a speed changes its position on the axis. However, the point corresponding to the speed c remains in place. This can be considered as a geometrical representation of the physical fact of the invariance of the speed of light in a vacuum.

From this representation, one can deduce another important property of the speed of light. As we have found from the rule of addition of velocities, no speed less than c can be made equal to c. If at some moment of time an object is moving slower than light, then no matter how many Lorentz transformations we apply (no matter how hard and long we accelerate the object), the representing point can approach the point "c" infinitesimally closely, but will never merge with it (otherwise one could by a succession of reverse Lorentz transformations change the position of the point "c," which would contradict its basic property – immobility.) It follows that the value of "c" is not only a fundamental constant, but also the unattainable limit for all other

speeds. The speed of light turns out to be a barrier that not only cannot be crossed, but cannot even be attained.

Will it then be legitimate to ask what there is behind the barrier? Does it make sense to discuss superluminal velocities?

We will consider these questions in Chapters 6–8.

4
Relativistic Mechanics of a Point Mass

4.1
Relativistic kinematics

We can now develop the relativistic kinematics of a particle. The most natural way to do this would be to use the concept of Minkowski's world. By analogy with the four-dimensional vector of an event $(ct, \boldsymbol{r}) = (ct, x, y, z)$, we introduce a general concept of a four-dimensional vector as any quantity characterized by four components which transform as the four components (coordinates) of an event in Minkowski's world. As an example directly related to our topic, we will first consider four-dimensional velocity (or just 4-velocity) of a particle. Suppose we picked up two close points on the particle's world line: $A_1 = (ct_1, x_1, y_1, z_1)$ and $A_2 = (ct_2, x_2, y_2, z_2)$. They are connected by a small interval ds which can be specified by its projections $(c\mathrm{d}t, \mathrm{d}x, \mathrm{d}y, \mathrm{d}z)$ onto coordinate axes, where $\mathrm{d}t = t_2 - t_1$, $\mathrm{d}x = x_2 - x_1$, $\mathrm{d}y = y_2 - y_1$, $\mathrm{d}z = z_2 - z_1$. In the Minkowski world, we can consider these projections as the components of a four-dimensional displacement with the temporal part $c\mathrm{d}t$ and the spatial part $\mathrm{d}\boldsymbol{r} = (\mathrm{d}x, \mathrm{d}y, \mathrm{d}z)$.

In Newtonian mechanics the motion of a particle from A_1 to A_2 is characterized by velocity $\boldsymbol{v} = \mathrm{d}\boldsymbol{r}/\mathrm{d}t$. The velocity $\boldsymbol{v}$ is a three-dimensional vector in space. We obtain it by dividing 3-displacement d$\boldsymbol{r}$ by corresponding time interval dt, which is the invariant under *three-dimensional* rotations and Galilean transformations. If we want to obtain a *four-dimensional* analog of velocity in Minkowski's world, we must take the 4-displacement $(\mathrm{d}t, \mathrm{d}\boldsymbol{r})$ and divide it by a quantity that is related to the motion of a particle between the points A_1 and A_2 and at the same time is invariant under *four-dimensional* rotations, that is, under Lorentz transformations. Such a quantity is the interval ds! Therefore, we can define 4-velocity as

$$u = (u_t, u_x, u_y, u_z) = \left(\frac{c\mathrm{d}t}{\mathrm{d}s}, \frac{\mathrm{d}x}{\mathrm{d}s}, \frac{\mathrm{d}y}{\mathrm{d}s}, \frac{\mathrm{d}z}{\mathrm{d}s}\right) \tag{1}$$

The 4-velocity thus defined turns out to be dimensionless, as is, for instance, the ratio $\beta \equiv v/c$. It also follows from this definition that the square of any 4-velocity is equal to unity:

Special Relativity and Motions Faster than Light. Moses Fayngold

ISBN: 3-527-40344-2

$$u_t^2 - u_x^2 - u_y^2 - u_z^2 = \frac{c^2\,\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2}{\mathrm{d}s^2} = 1 \tag{2}$$

This means that in a four-dimensional space on whose axes are plotted the components of a 4-velocity vector, the tip of this vector always remains on the surface of the "sphere" of unit radius, centered about the origin. Mathematicians call such a sphere a pseudo-sphere.

How can such a vector of constant "length" characterize the real physical velocity of a particle, which can vary over continuous ranges of values and directions?

This becomes clear if we turn to *components* of the 4-velocity and express them in terms of ordinary velocity $\boldsymbol{v}$. To this end, we write

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 = c^2\,\mathrm{d}t^2 - \mathrm{d}\boldsymbol{r}^2 = c^2\,\mathrm{d}t^2\left(1 - \frac{v^2}{c^2}\right) = \gamma^{-2}(v)\,c^2\,\mathrm{d}t^2 \tag{3}$$

Therefore,

$$u_t = \gamma(v)\,; \quad u_x = \frac{v_x}{c}\,\gamma(v)\,; \quad u_y = \frac{v_y}{c}\,\gamma(v)\,; \quad u_z = \frac{v_z}{c}\,\gamma(v) \tag{4a}$$

or

$$u = \left[\gamma(v)\,, \frac{\boldsymbol{v}}{c}\,\gamma(v)\right] \tag{4b}$$

We see that even though the square of a 4-velocity is always equal to 1 for any possible motion, its components are variables depending on $\boldsymbol{v}$. For a photon, the components of its 4-velocity are infinite. They can combine into a vector of unit length only because the squares of its temporal and spatial components add with the opposite signs. The infinite values for the components of the 4-velocity of a photon emphasize the special role of the speed of light, and its unattainability for the particles with non-zero rest mass.

In Newtonian mechanics the product of the particle's velocity by its mass gives the momentum:

$$\boldsymbol{P} = m\boldsymbol{v} \tag{5}$$

In relativistic mechanics, the analogous product of the rest mass by 4-velocity and by c gives 4-momentum:

$$P_j = m_0\,u_j\,c \tag{6}$$

Here and hereafter subscript j can take on the values 0, 1, 2, 3, which stand for the axes ct, x, y, z, respectively. When considering only three spatial dimensions, we will often use the Greek letter α ranging through 1, 2, 3 for x, y, z, respectively.

All the components of the 4-momentum have a simple physical meaning. For the spatial components we have a regular three-dimensional momentum:

$$P_a = m_0\,\gamma(v)\,v_a = m\,v_a\,, \quad \text{or} \quad \boldsymbol{P} = m\boldsymbol{v} \tag{7}$$

where

$$m(v) = m_0\,\gamma(v) \tag{8}$$

and the subscript a, as just defined, stands only for the spatial components x, y, z.

The factor m multiplying $\boldsymbol{v}$ in relativistic momentum is called relativistic mass. We see that this mass depends on the speed of the particle. When the speed is small, m is practically indistinguishable from the particle's rest mass m_0. This is the situation that we have been used to in non-relativistic physics. However, at high enough speeds the mass increases. As the speed approaches c, mass m becomes infinitely large.

For the temporal component of the 4-momentum we have

$$P_0 = m_0\,c\,\gamma(v) = mc \tag{9}$$

where the subscript 0 on P stands for the "*ct*" component of momentum; do not confuse it with the similar subscript on m to indicate the *rest* mass. Combining Equation (9) with Einstein's famous equation relating total mass to the energy, $E = mc^2$, we can rewrite it as

$$P_0 = \frac{E}{c} \tag{10}$$

In other words, the temporal component of 4-momentum has the physical meaning of the energy of the particle.

These expressions can be written in the form showing the similarity between time–space on the one hand and energy–momentum on the other. Using Equations (7), (9), and (10), we obtain direct relationship between E and $\boldsymbol{P}$:

$$\frac{E^2}{c^2} - P_{\mathrm{x}}^2 - P_{\mathrm{y}}^2 - P_{\mathrm{z}}^2 = \frac{E^2}{c^2} - \boldsymbol{P}^2 = m_0^2\,c^2 \tag{11}$$

The algebraic structure of the left-hand side of this equation is precisely the same as the structure of the expression for the interval in Minkowski's world. The values of E/c and $\boldsymbol{P}$ are different for different observers (that is, in different reference frames), but, because the rest mass of a particle is constant, the difference of their squares is invariant. Therefore, the values E/c, P_x, P_y, and P_z can be considered as coordinates of a point in a ficticious four-dimensional space similar to Minkowsky's space–time (the momentum space). The coordinate E/c in this space is similar to the time coordinate ct, and P_x, P_y, and P_z are similar to the spatial coordinates x, y, and z. It fol-

lows that the quantities E/c and $\boldsymbol{P}$ must transform in the same way as do ct and $\boldsymbol{r}$ when one switches between two inertial frames moving with relative velocity V:

$$\left.\begin{aligned} &E' = \gamma(V)(E - V P_x)\,, \\ &P'_x = \gamma(V)\left(P_x - \frac{V}{c^2} E\right); \quad P'_y = P_y\,; \quad P'_z = P_z \end{aligned}\right\} \tag{12}$$

If we know the energy and momentum of a given particle, we can find its velocity. The general expression for the velocity of the particle in terms of its energy and momentum is the same as in Newtonian mechanics:

$$v = \frac{\mathrm{d}E}{\mathrm{d}P} \tag{13}$$

Differentiating Equation (11) with respect to P, we obtain

$$v = \frac{\mathrm{d}E}{\mathrm{d}P} = \frac{P}{E}\, c^2 = c\sqrt{1 - \frac{m_0^2\, c^4}{E^2}} \leq c \tag{14}$$

Because the total energy of the particle is always higher than its rest energy, the expression under the square root sign is positive and less than 1, and the particle's speed automatically comes out always less than c. For a photon, which does not have a rest mass because it never rests ($m_0 = 0$), Equation (14) yields $v = c$.

4.2 Relativistic dynamics

The change in our concepts of space and time causes similar changes in our concepts of motion and force. We have already seen that the velocity of a moving object behaves "paradoxically" when we consider it from different reference frames. In particular, it does not obey the "obvious" law of addition of velocities. We can rephrase this in the language of transformations by stating that the velocity behaves differently with respect to different transformations: so far as we are confined to one reference frame (for instance, we rotate the axes in our three-dimensional space, but remain within the same system K), time t behaves as a scalar quantity, and the velocity $\boldsymbol{v}$ transforms like a regular three-dimensional vector. However, when we perform rotations in four-dimensional space–time involving the time axis (which physically corresponds to a transition to another reference frame), the velocity does not behave as a four-dimensional vector. This role is taken up by the "representative" – the 4-velocity. We have defined the 4-velocity in Equation (1) as $u_i = \mathrm{d}x_i/\mathrm{d}s$, $i = 0, 1, 2, 3$, where index 0 corresponds to the time axis, and 1, 2, and 3 correspond to the three spatial axes, whereas the interval $\mathrm{d}s$ is a scalar under 4-rotations.

The same can be done to describe the dynamic characteristics – momentum, energy, and force. By analogy with the definition of 3-momentum as a vector with three com-

ponents, $p_a = m_0 v_a = m_0 \mathrm{d}x_a/\mathrm{d}t$, we have defined 4-momentum as a vector with four components:

$$P_i = m_0 u_i = m_0 \frac{\mathrm{d}x_i}{\mathrm{d}s} \tag{15}$$

How do these changes affect the concept of force and accelerated motion?

The basic definition of force is

$$\boldsymbol{f} = \frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t} \tag{16}$$

Formally it is similar to definition of velocity as $v = \mathrm{d}\boldsymbol{r}/\mathrm{d}t$. Like the velocity, force behaves as a three-dimensional vector under restricted Lorentz transformations not involving time (pure spatial rotations). The 3-momentum $\boldsymbol{P}$ is the product of mass and velocity:

$$\boldsymbol{P} = m\boldsymbol{v} = m_0 \gamma(v)\, \boldsymbol{v} \tag{17}$$

Because of the relativistic factor $\gamma(v)$, the relationship between the mass, acceleration, and force is more subtle than in non-relativistic mechanics. To show it, let us express the force directly in terms of acceleration. This will require a little algebra.

Putting Equation (17) for $\boldsymbol{P}$ into definition (16) and performing differentiation, we obtain

$$\boldsymbol{f} = m_0 \frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{v}\, \gamma(v) = m_0 \left[\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} \cdot \gamma(v) + \boldsymbol{v} \frac{\mathrm{d}}{\mathrm{d}t} \gamma(v) \right] \tag{18}$$

Now, since $v^2 = \boldsymbol{v}^2$, the Lorentz factor can be written as $\gamma(v) = (1 - \boldsymbol{v}^2/c^2)^{-1/2}$. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t} \gamma(v) = \frac{\mathrm{d}}{\mathrm{d}t} \left(1 - \frac{\boldsymbol{v}^2}{c^2}\right)^{-1/2} = -\frac{1}{2} \left(1 - \frac{v^2}{c^2}\right)^{-3/2} \cdot \left(-2 \frac{\boldsymbol{v}}{c^2} \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}\right) = \gamma^3(v) \frac{\boldsymbol{a}\boldsymbol{v}}{c^2} \tag{19}$$

Putting this into the previous equation yields

$$\boldsymbol{f} = m_0 \gamma(v) \left[\boldsymbol{a} + \gamma^2(v) \frac{\boldsymbol{a}\boldsymbol{v}}{c^2} \cdot \boldsymbol{v} \right] \tag{20}$$

Now, let us read this equation. We see that the relation between force and acceleration is not that simple as it is in non-relativistic mechanics. Its most interesting feature is that acceleration, even though it is caused by a force, does not generally point in the direction of the force!

We will illustrate the relation (20) by considering a few special cases.

First, consider slow motions. Then the second term in the brackets is negligible

compared with the first, and the Lorentz factor is close to 1. The equation then reduces, as it should, to the known non-relativistic limit $\boldsymbol{f} = m_0 \boldsymbol{a}$.

Let velocity now be arbitrary, and change only in direction. Geometrically this means that the moving mass traces out an arch of a circle. The acceleration is accordingly perpendicular to the velocity, and the scalar product $\boldsymbol{av}$ is zero. Then

$$\boldsymbol{f} = m_0 \gamma(v) \boldsymbol{a} \tag{21}$$

Next consider the case when velocity changes only in magnitude (the force is applied along the velocity). Then $\boldsymbol{av} = av$, the mass moves in a straight line, and we have

$$\boldsymbol{f} = m_0 \gamma(v) \left[\boldsymbol{a} + \gamma^2(v) \frac{v^2}{c^2} \cdot \boldsymbol{a} \right] = m_0 \gamma^3(v) \boldsymbol{a} \tag{22}$$

Now, compare the two results. Don't you notice something strange about them?

In non-relativistic mechanics, we actually have two definitions of mass: as a measure of the amount of matter in a body, and as a measure of a body's inertia. The first can be measured, for instance, by weighing the body, and the second can be determined as a ratio of the applied force to the resulting acceleration. Both definitions had gone peacefully together, hand in hand.

In relativistic mechanics, the situation is different, and the two definitions of mass turn out to represent two characteristics with different behavior. The first definition – mass as the amount of matter in a body – applies directly to the body at rest. It tells us what we get if we stop the body, and is represented by a *constant* factor m_0 – rest mass. For instance, if we try to stop the photon, we get nothing, and we accordingly say that the rest mass of a photon is zero. But we do not actually have to stop a moving object or to catch up with it to measure its rest mass – we can instead measure its energy E and momentum $\boldsymbol{P}$ and then find the rest mass as the invariant $(E^2 - P^2c^2)^{1/2}/c^2$. If we apply this procedure to a photon, we will get zero.

The second definition – mass as the amount of body's inertia (call it inertial mass) – has a direct and clear physical meaning only in situations when it can be measured as the ratio of the force to acceleration. Generally such interpretation is not applicable because, according to Equation (20), the force and acceleration are not parallel, but it can be applied in special cases such as Equations (21) and (22).

We then obtain two different results and accordingly introduce two different notations for corresponding inertial mass:

$$m_\perp = m_0 \gamma(v) \tag{23a}$$

and

$$m_{\updownarrow} = m_0 \gamma^3(v) \tag{23b}$$

The mass $m_{\updownarrow}$ turns out to be larger than the mass $m_\perp$. Physically this means that the body is more inert in the longitudinal direction than in the transverse direction.

When you push it in the direction of its trajectory, the body resists harder than when you push it the direction perpendicular to its trajectory. The inertial mass is anisotropic!

How can we understand this result? Actually, there is no need for any deeper understanding here, this is how Nature works, the equations we use give the adequate description of motion in excellent agreement with experiment and predict correctly the outcome of other experiments, and this is all one should expect from a theory.

But I would still offer the somewhat naive comment that might appeal to our drive to have one common explanation for many different "faces" of the world.

Relativity forbids any physical body with non-zero rest mass to reach the speed of light. To implement this ban, it has made the inertial mass a variable quantity *increasing* with speed, so that the closer the speed is to c, the more vigorous is the body's resistance to a further change of speed. It already resists harder than the rest mass would do when the applied force is perpendicular to the velocity, even though such force cannot change the speed. But when the force pushes *along* the velocity to increase the speed, the resistance is much harder still – as if to make the ban more efficient! It is true that the ban works both ways – according to Equation (23 b), the resistance is the same for the force tending to accelerate the body and for the force tending to decelerate the body. But the main thing is that the "longitudinal" inertia increases with speed faster than the "transverse" inertia – and this is precisely what Equations (23 a and b) tell us!

Equations (23 a and b) describe only two special cases. One might need an expression that would determine acceleration in the general case of arbitrary orientation between the velocity and the force. In non-relativistic mechanics the corresponding expression $\boldsymbol{a} = m^{-1}\boldsymbol{f}$ does not depend on orientation because the mass is a constant. Since this is not so in relativistic mechanics, the derivation of the wanted expression is more difficult. Unfortunately, we cannot factorize the acceleration from the right-hand side of Equation (20), because it is entangled with velocity in a scalar product $\boldsymbol{av}$. If we rewrite the equation as

$$\boldsymbol{a} = \frac{\boldsymbol{f}}{m_0\,\gamma\,(v)} - \gamma^2\,(v)\,\frac{\boldsymbol{av}}{c^2}\,\boldsymbol{v} \tag{24}$$

it would be of little help because $\boldsymbol{a}$ is also present in the right-hand side. But we can do the following trick. Multiply Equation (24) by $\boldsymbol{v}$ to form a scalar product on both sides. Since $\boldsymbol{vv} = v^2$ we have

$$\boldsymbol{av} = \frac{\boldsymbol{fv}}{m_0\,\gamma\,(v)} - \gamma^2\,(v)\,\frac{v^2}{c^2}\,\boldsymbol{av} \tag{25}$$

This can be rewritten as

$$\left[1 + \gamma^2\,(v)\,\frac{v^2}{c^2}\right]\boldsymbol{av} = \frac{\boldsymbol{fv}}{m_0\,\gamma\,(v)} \tag{26a}$$

or

$$\gamma^2(v)\,\boldsymbol{a}\boldsymbol{v} = \frac{\boldsymbol{f}\boldsymbol{v}}{m_0\,\gamma(v)} \tag{26b}$$

Put this back into Equation (20)! Immediately, we will obtain the desired expression with $\boldsymbol{a}$ on only one side:

$$\boldsymbol{a} = \frac{1}{m_0\,\gamma(v)}\left(\boldsymbol{f} - \frac{\boldsymbol{f}\boldsymbol{v}}{c^2}\,\boldsymbol{v}\right) \tag{27}$$

Read this equation. It says that acceleration is determined not only by the force, but also by the velocity of the body, and generally is not parallel to the force.

So far we had discussed the properties of force in one reference frame. Our next question is how the force would behave if we transfer to another frame of reference. How would the force measurements carried out by different observers relate to each other?

One way to obtain the answer would be using the analogy with the velocity vector. In the Minkowski space it is represented by 4-velocity $u_i = \mathrm{d}x_i/\mathrm{d}s$. Similarly, we can form a four-dimensional vector (4-force) representing the physical force, and use its transformation properties to find the transformation rules for the force $\boldsymbol{f}$.

We define the 4-force in Minkowski's space–time as the derivative of 4-momentum with respect to corresponding interval [16]. Recalling the definition in Equation (6) of 4-momentum, we have

$$G_i = \frac{\mathrm{d}P_i}{\mathrm{d}s} = m_0\,c\,\frac{\mathrm{d}u_i}{\mathrm{d}s} \tag{28}$$

It is easy to notice that 4-force as we have defined it is perpendicular to the 4-velocity:

$$Gu \equiv G_0\,u_0 - G_a\,u_a = m_0\,c\left(u_0\frac{\mathrm{d}u_0}{\mathrm{d}s} - u_a\frac{\mathrm{d}u_a}{\mathrm{d}s}\right) = \frac{1}{2}\,m_0\,c\frac{\mathrm{d}u^2}{\mathrm{d}s} = 0 \tag{29}$$

In this equation we applied Einstein's "summation rule" according to which we perform the summation over the ranging indexes appearing twice in an expression. Because the 4-velocity has a constant kinematic length, its derivative is zero. And if the dot product of the two vectors is zero, the vectors are perpendicular, just as is the case in three-dimensional space.

Let us now express 4-force G explicitly in terms of the 3-force. Writing $\mathrm{d}s = c\,\mathrm{d}t/\gamma(v)$ [recall Equation (3)!] and keeping in mind that $\mathrm{d}\boldsymbol{P}/\mathrm{d}t = \boldsymbol{f}$, we have

$$G_i = \frac{\gamma(v)}{c}\frac{\mathrm{d}P_i}{\mathrm{d}t} = \frac{\gamma(v)}{c}\left(\frac{\mathrm{d}P_0}{\mathrm{d}t}, \frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t}\right) = \frac{\gamma(v)}{c}\left(\frac{\mathrm{d}E}{c\,\mathrm{d}t}, \boldsymbol{f}\right) = \frac{\gamma(v)}{c}\left(\frac{\boldsymbol{f}\boldsymbol{v}}{c}, \boldsymbol{f}\right) \tag{30}$$

The time component of the 4-force turns out to be connected with the work of the force.

Once we have defined G as a 4-vector, we know its transformation properties. We thus can relate the components of the 4-force measured by two different inertial observers. On the other hand, we can express the components G' of this force in another reference frame in terms of the 3-force $\boldsymbol{f}'$. This will give us the equations connecting the components of the 3-force as measured in two different frames. The diligent reader can try to do this as an exercise.

However, we can obtain the same result more easily by using the definition in Equation (16) of the 3-force and applying the transformation rules to momentum and time. Consider two frames K and K′ with parallel axes. Suppose the frame K′ slides relative to K along the x-axis with a speed V. Then we have

$$\left.\begin{aligned} \mathrm{d}P_{x'} &= \gamma(V)\left(\mathrm{d}P_x - \frac{V}{c^2}\mathrm{d}E\right), \quad \mathrm{d}P_{y'} = \mathrm{d}P_y, \quad \mathrm{d}P_{z'} = \mathrm{d}P_z \\ \mathrm{d}t' &= \gamma(V)\left(\mathrm{d}t - \frac{V}{c^2}\mathrm{d}x\right) \end{aligned}\right\} \tag{31}$$

Suppose the motion takes place in the plane xy, so that we can consider only x and y components of force. Using the result in Equation (13), we find

$$\left.\begin{aligned} f_{x'} &= \frac{\mathrm{d}P_{x'}}{\mathrm{d}t'} = \frac{\mathrm{d}P_x - \frac{V}{c^2}\mathrm{d}E}{\mathrm{d}t - \frac{V}{c^2}\mathrm{d}x} = \frac{\mathrm{d}P_x}{\mathrm{d}t}\,\frac{1 - \frac{V}{c^2}\frac{\mathrm{d}E}{\mathrm{d}P_x}}{1 - \frac{V v_x}{c^2}} = \frac{\mathrm{d}P_x}{\mathrm{d}t} = f_x \\ f_{y'} &= \frac{\mathrm{d}P_{y'}}{\mathrm{d}t'} = \frac{\mathrm{d}P_y}{\gamma(V)\,\mathrm{d}t\left(1 - \frac{V v_x}{c^2}\right)} = \frac{f_y}{\gamma(V)\left(1 - \frac{V v_x}{c^2}\right)} \end{aligned}\right\} \tag{32}$$

The component of force *along* the direction of relative motion of the two systems is unchanged. The transformation law for the transverse component of force is more complicated – it depends on the velocity of a moving mass. In a special case when the mass moves together with the reference frame K′, we have $v_x = V$, and the second Equation (32) reduces to

$$f_{y'} = \gamma(V) f_y \tag{33}$$

Consider a possible physical situation illustrating this result. If we measure the gravity force acting on a car moving with a speed V on a horizontal track, then the driver of this car would measure a slightly greater force: since the direction of this force is perpendicular to the direction of motion, the force in the driver's reference frame is given by Equation (33). The difference is very small because the Lorentz factor is practically indistinguishable from unity in this case. However, in the case of relativistic velocities the difference may be very large. We will see possible dramatic consequences of such a difference in Section 5.4.

5
Imaginary Paradoxes

5.1
The three clocks paradox

The relativistic effects following from Einstein's relativity principle have caused a long-standing controversy in the scientific literature. Critics thought that these effects lead to paradoxes. In the first place this pertained to the conclusion that time is relative. This statement, the critics declared, resulted in the following contradiction. According to Einstein's theory, a moving clock A is always slow compared with a stationary clock B. However, according to the same theory, it makes no difference which one of the clocks we regard as moving and which one as stationary. We could as well describe clock A as stationary and clock B as moving. Then we must inevitably conclude that it is clock B that is slow compared with A. So, one method of analysis implies that A is slower than B and the other that B is slower than A. The question is which clock is actually slower.

An attempt to solve the paradox put in this way would be fallacious because we are asked here to answer the wrong question. Indeed, the alleged "contradiction" is based upon the conviction that "in reality" *only one* certain clock can be slower than another and that otherwise we would get a logical nonsense.

Meanwhile, in reality, there is no contradiction whatsoever. The situation is similar to an example with two antipodes, each of whom is convinced that he is standing "normally" while his antipode is turned upside-down. We know that they are both right, yet this does not cause a logical contradiction because the very idea of "down" is relative: it is a direction towards the center of the Earth, *which is different for each antipode*. In exactly the same way, the statement that any clock, A or B, can be slower than its counterpart is not self-contradictory because each possibility is linked to *its own observational procedure*.

When we assume the clock B_1 to be moving, it is not enough to have only one sample of stationary clocks (for instance, clock A_1) to compare their rates. Comparing the clocks' *rates* involves more than just comparing their instantaneous readings. It requires taking their readings at different moments of time. Suppose that when clock B_1 passes by clock A_1, their readings are the same. However, at a later moment the moving clock B_1 will pass by another place, so we will now have to compare its reading with the reading of *another* stationary clock A_2 at that place. That is how we

Special Relativity and Motions Faster than Light. Moses Fayngold

ISBN: 3-527-40344-2

discover that B_1 has a slower rate. If we now "switch" to a system K′, where B_1 is stationary, and start taking readings of *one* sample of K-clocks (say, clock A_1), then, as a consequence of this clock's motion relative to K′, we will have to use a pair of K′-clocks: first B_1 and then B_2. Now A_1 will be slower than B_1. But there is no paradox here, because two *different* pairs of events are being considered. In the first case we have clock B_1 passing by A_1 and A_2, and in the second case clock A_1 passing by B_1 and B_2. It is no surprise that two *different* cases give two different results. The process of clocks' juxtaposition is asymmetric. In this process, the clock that turns out to be slower is always the one that is compared with a *number* of clocks in the other reference frame – *no matter whether the given clock is considered as moving or stationary.* The last statement means that if both observers – one in K and the other in K′ – agreed to consider *one and the same* pair of events, then both would obtain *the same* result. However, the observers would offer different explanations of this result. The first observer, who is stationary relative to the chosen pair of clocks, would explain the result by the above-considered "dilation" of time. The second observer, from whose viewpoint this pair of clocks is moving, would point to a discrepancy between readings of the moving clocks.

This argument is important for a real understanding of the time dilation effect, and it is worthwhile to consider it in more detail. Suppose that two series of observations are being carried out by two observers, Alice and Tom. In the first series Tom in his system T compares *one* sample of Alice's clocks with a pair of his spatially separated but synchronized clocks. He finds that the rate of Alice's clock is $\gamma(v)$ times slower than the rate of his clocks. Similarly, Alice in her system A compares *one* sample of Tom's clocks with a pair of her clocks synchronized in her system A. She notices that the rate of the Tom's clock is $\gamma(v)$ times slower. After analyzing the procedure of measurements, both observers will come to the conclusion that each of them is right, and their results appear to contradict each other only because they were considering *different* pairs of events.

The second series consists of only one experiment. Both observers have agreed to consider *the same* pair of events, for example, the moments of coincidence of Alice's clock A with Tom's separated clocks B_1 and B_2. For Tom, who is stationary relative to his clocks, this experiment is an exact repetition of the first series of measurements, and naturally it yields the same result

$$t'_{B_2} - t'_{B_1} = \gamma(v)\,\Delta t_A \tag{1}$$

where t'_{B_1}, t'_{B_2} are the moments of clocks' coincidences in T as determined by the clocks B_1 and B_2, and Δt_A is the time interval between these coincidences by Alice's clock A.

But for Alice, this experiment is *different* from the ones she had performed in the first series. She had found then that every single clock B is slower as compared to her *pair* of clocks. Now, however, she compares *one sample* of her clock A with *two* of Tom's clocks B_1 and B_2 that pass one after onother by her chosen clock. Since the actual readings of any clock do not depend upon who is looking at it, Alice will record the same readings t'_{B_1}, t'_{B_2} and Δt_{A_1} of the respective clocks as Tom did, and therefore

will find that Equation (1) is correct. However, this equation states that interval Δt_{A_1} of her A-clock's proper time between its meetings with B_1 and B_2 is $\gamma(v)$ times *shorter* than the interval $t'_{B_2} - t'_{B_1}$. This result seems natural to Tom, but not to Alice. "How can it be," she asks, "that moving clocks B_1 and B_2, whose rate is *slowed down* by a factor of $\gamma(v)$, nevertheless show a $\gamma(v)$ times *greater* difference between their readings than my stationary clock A?"

The reason behind this "paradox" lies in the relativity of simultaneity. The clocks B_1 and B_2 have been synchronized in their rest frame T. However, they show *different* readings at any single moment in system A. Using Equation (45) in Sec. 2.7, which gives the dependence of these readings on the clocks' positions, we find that at the moment t'_{B_1} when the clocks B_1 and A are coincident, the clock B_2 reads the time

$$t'_{B_1} + \gamma(v)\,\frac{v}{c^2}\,x \tag{2}$$

where

$$x = v\,\Delta t \tag{3}$$

is the distance in A between the clocks B_1 and B_2. Putting this expression for x into Equation (2), we obtain for this reading

$$t'_{B_1} + \gamma(v)\,\frac{v^2}{c^2}\,\Delta t_{A_1} \tag{4}$$

Thus, the clocks B_1 and B_2 synchronized in their rest frame are not synchronized in system A. For Alice, at the moment when clocks A and B_1 are coincident, clock B_2 shows a greater time reading than B_1 by $\gamma(v)\,\frac{v^2}{c^2}\,\Delta t_{A_1}$. This effect of the initial "time gain" of the clock B_2 over clock B_1 (observed in A) is larger than the effect of "slowing down" of both clocks. It results in a greater difference between the readings of both clocks as they pass by A than the time Δt_{A_1} between these events measured by clock A itself. Clock B_2 passes by A Δt_{A_1} seconds after B_1. But according to the "time dilation" effect, the time interval Δt_{A_1} is linked to the interval $\Delta t'$ for the moving clocks: $\Delta t' = \gamma^{-1}(v)\,\Delta t_{A_1}$. Adding this time to Equation (4), we obtain

$$t'_{B_2} = t'_{B_1} + \left[\gamma(v)\,\frac{v^2}{c^2} + \gamma^{-1}(v)\right]\Delta t_{A_1} \tag{5}$$

It immediately follows from here that

$$t'_{B_2} - t'_{B_1} = \gamma(v)\,\Delta t_{A_1} \tag{6}$$

which is simply Equation (1).

Thus, Alice and Tom arrive at the same result. However, it is only for Tom that the difference between readings of B_1 and B_2 has a meaning of *time* between the events. This is why Tom regards the results of his measurements as an experimental manifestation of the time dilation effect. For Alice, however, the difference $t'_{B_2} - t'_{B_1}$ is not the duration of any process because the clocks B_1 and B_2 read *different* times at the same moment of her time in system A. Alice will explain the result in Equation (1) by this initial time discrepancy between the moving clocks. If she, too, wants to compare the flow of time in different reference systems, she must go back to the first series of experiments, i.e. consider the passing of Tom's *one clock* B by the pair of her clocks A_1 and A_2. Then she (and Tom!) will obtain the result opposite to Equation (1). But these opposite results do not in any way contradict each other, because they relate to different pairs of events.

5.2 The dialog of two atoms

We have seen that to compare the "flow of time" in two different inertial systems, it is necessary to have at least three samples of identical clocks. Two of them belong to one system (in which they are synchronized) and one to another system of reference. As a result, the comparison procedure is asymmetric with respect to the two systems. The asymmetry is in this case evident from the oddness of the minimal number of clocks that are needed for the experiment. Obviously, the clocks here cannot be distributed evenly between two reference frames.

Is it possible to find such a procedure of comparing two clocks that would be symmetrical? To do that, we must compare one clock in one reference frame with only one clock in another reference frame. It is true that, as was previously emphasized, the number of clocks compared cannot be less than three. This, however, pertains to situations when we use nothing else but clocks. But what if we employ, for instance, light signals? Couldn't we then get by without the third clock?

Consider the following thought experiment. Suppose Alice has an excited hydrogen atom A. The atom emits light of a certain frequency ω_0 uniformly in all directions. (We disregard here the quantum nature of radiation, according to which every single photon is found to have been emitted within a small solid angle around a certain direction. What is essential for the experiment considered here is the statistically averaged picture of this process, and this is equivalent to the continuous radiation of a diverging spherical wave.) Then Alice measures the wavelength of the emitted light:

$$\lambda_0 = 2\pi \frac{c}{\omega_0} \tag{7}$$

Suppose Tom is moving at a speed V along the x-axis in this system. Tom also has a hydrogen atom B in the same exited state as the atom A. Naturally, Tom's atom also emits spherical light waves with the same (proper!) frequency ω_0 and wavelength

$\lambda_0 = 2\pi \frac{c}{\omega_0}$ *in its rest frame B* (since there is now only one atom in each system, we can call each atom by the name of corresponding system).

Each atom represents a perfect sample of a clock; both clocks are absolutely identical and have the same *proper* period of one oscillation:

$$T_0 = \frac{2\pi}{\omega_0} \tag{8}$$

This period can be measured as the time interval between two successive passings of the crests and troughs of emitted waves through a point stationary relative to the given atom. Thus, the "oscillations of the atomic clock" are accompanied by radiation that carries information about the oscillation period. The electromagnetic "ticking" of the *stationary* atomic clock is reproduced with the same period at all other points of the given reference system. This eliminates the need to supply every point of space with its own clock. (This is precisely what actually happens in Nature: the information about time is being carried by radiation.) Hence the possibility arises of comparing the rates of two clocks moving relative to each other without any additional clocks: we will use instead the light from the considered clocks themselves. Wouldn't that be a symmetrical comparison procedure?

Our observers start counting the light wave oscillations at the moment when the atom B passes by A, that is, when the line AB connecting the atoms is perpendicular to the x-axis. This initial moment is taken as a zero moment in both systems (the simultaneity of the two events holds for both systems because at the zero moment both atoms lie on a line which is perpendicular to the direction of their relative motion).

We already know that in system A, the oscillation rate of the moving atom B is slower. How will this effect manifest itself when we compare the pattern of oscillations of the two atoms?

In system A, the atom B oscillates with a period

$$T = \gamma(v)\, T_0 \tag{9}$$

This period is greater than T_0. Since T_0 is also the proper period of the atom A, the latter will emit more than one wave during the time $T > T_0$. To put it another way, it will take T s for one wave emitted by the atom B to pass through A, during which time the atom A itself will emit a $\gamma(v)$ times greater number of waves (Fig. 5.1 a). Accordingly, the frequency of radiation from B, measured by Alice at the origin of her system A, is $\gamma(v)$ times smaller than its proper radiation frequency ω_0. Such a decrease in the wave frequency from a rapidly moving source, when the direction of oncoming radiation is perpendicular to the relative velocity $\boldsymbol{V}$, had been observed by Eves and Stilwell in their beautiful experiments with ion beams [18] and is known as the transverse Doppler effect. These experiments (and also the observations of atmospheric μ-mesons described in Section2.8), provide another demonstration of the relativity of time. Once again we observe a slowing of the rate of a moving clock compared with the identical stationary clock. But we know that any description of a sys-

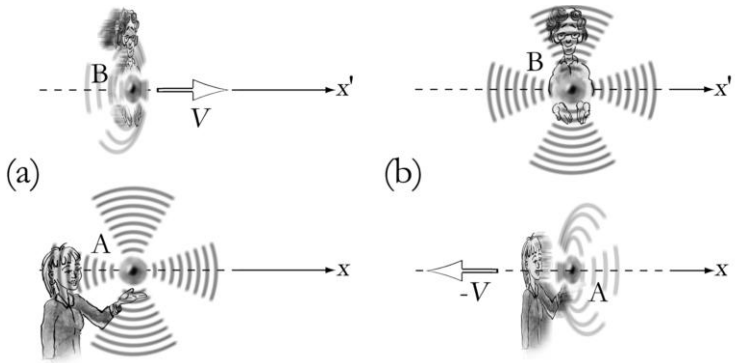


Fig. 5.1 The two radiating atoms passing by each other: (a) as observed by Alice; (b) as observed by Tom.

tem as moving or stationary is relative. And in the current example the effect of time dilation manifests itself in what appears to be a symmetrical procedure: we took only one clock from each of the systems! We immediately arrive at a logical contradiction: a direct comparison of the clocks' rates shows that clock B is slower than A and clock A is slower than B at the same time!

Of course, there can be no logical contradictions in Nature – they can only occur in our mind. In this case the paradox has emerged because we had assumed the comparison procedure to be symmetrical for the only reason that there were just two samples of clocks. Yet this fact alone is still not sufficient for the procedure to be symmetrical! Indeed, are the clocks in the procedure being compared directly with one another? Not at all! Let us take another glance at Figure 5.1. We see that the readings of clock A are matched not with those of B, but with the sequence of light waves from B arriving at A – and this makes quite a difference. In such a matching we consider two different pairs of events: Alice records the passing of two successive waves from B through her clock A (Fig. 5.1 a). But A is moving relative to B! Tom in B, using a similar procedure, records the passing of two wave crests from A through *his clock B* (Fig. 5.1 b). He will thereby observe a decrease in frequency of waves from A. However, the conclusions of both observers would not contradict each other, since they pertain to different pairs of events.

Thus, the elimination of the third clock did not eliminate the asymmetry in the comparison procedure. In order to compensate for one missing clock we are forced to introduce some new element, e. g. light. It then turns out that with the use of light the roles played by the two clocks in question are again not equivalent. In each reference frame one considers the passing of waves from the moving clock through a point that is *at rest* relative to that frame. It is here where the asymmetry lies: the moving clock is used as a source of signals, and the clock at rest as a measuring device. As a result, any source of periodic signals, while used as a detector counting a certain number of passing signals from an identical moving source, emits a greater number of its own signals. And once again, just as in the case of three clocks, the conclusion does not depend on whether we regard the given source as moving or stationary.

For the proof we assume again that two series of experiments are being conducted. In the first series each observer, having counted the number of waves from the other's clock, realizes that the frequency ω of passing radiation is less by a factor of $\gamma(v)$ than the proper frequency ω_0. By looking into the measuring procedure the observers conclude that their results do not contradict each other since they pertain to different sets of events.

Then the observers perform another series of experiments which are reduced to a single experiment: both observers agree to consider one *and the same* pair of events, say, the passing of two successive wave crests from A by clock B. For Tom, who is at rest relative to B, this experiment is just a repetition of the experiments from the previous series and gives the same result:

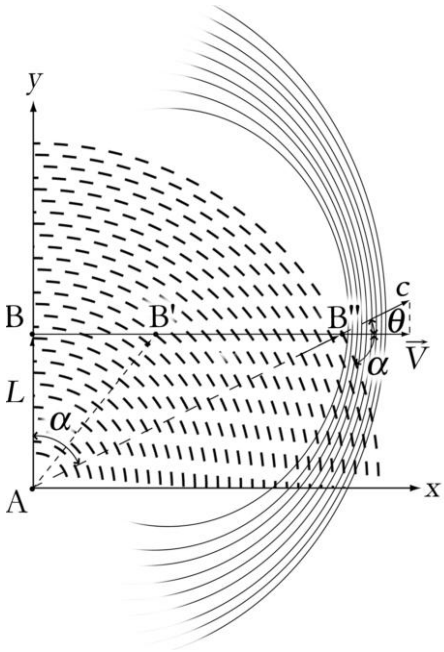
$$T' = t'_2 - t'_1 = \gamma(v)\, T_0 \tag{10}$$

where $T' = t'_2 - t'_1$ is the time interval by clock B between the passings through B of two successive wave crests from A, and T_0 is the proper period of B.

For Alice this experiment is different from those she performed in the first series. She then had counted the waves from B passing through her clock A. Now she counts the waves from A passing through clock B that is itself moving at a speed v. But since the number of waves between the two given events does not depend on who is counting them, Alice will obtain the same Equation (10) that Tom did. And it states that clock B will emit $\gamma(v)$ waves while only *one* wave from A passes by it.

If we look once again at Figure 5.1 a, which illustrates the picture of waves as seen by Alice, it may appear that the opposite is true. The crests of the waves diverging from A are closer in space than the crests of the waves from B; it seems that the waves from A must pass by B faster. How come, then, that clock B emits more of its own waves than it meets waves from A? In order to account for this "paradox," it suffices to note that by the moment when the first (counting from the initial zero moment) wave crest from B arrives at A, the atom B itself will reach the point B′ (Fig. 5.2), that is, the instantaneous direction toward the moving source of radiation is different from the direction of light arriving from the source. As to the first crest of the wave from A, we can see from Figure 5.2 that it reaches atom B at a still further point B″. At this point the velocity $\boldsymbol{v}$ of the atom B forms an angle α with the wave crest, so the atom's velocity will have a component perpendicular to the wave crest and directed along the radius drawn from A. In other words, B *runs away* from A's waves and therefore successive wave crests reach B with a certain delay. The waves from A pass by the receding B less frequently than by a *stationary* point of system A. This effect of delay, caused by the fact that B is running away from A, turns out to be stronger than the difference in density of the light waves' "ripples" along the line AB (Fig. 5.2). Moreover, Figure 5.2 explicitly shows that the direction AB now plays no role at all; in fact, its use may only cause more confusion. Rather, what we have to do is to compare frequencies of "ripples" of the waves from both sources along the instantaneous direction AB″. The motion of the source B causes its own "light ripples" to be different in different directions: the wave crests are packed more densely in front of the source and less densely behind

Fig. 5.2



it. It turns out that along AB″ the frequency of "light ripples" from B is greater than that of ripples from atom A. As a result, B actually emits more than one wave between the encounters with two successive crests from A. How much more? Let us calculate this effect quantitatively.

As we see from Fig. 5.2, the angle θ between the direction AB″ and OX is given by

$$\cos\theta = \frac{v\,t_1}{c\,t_1} = \frac{v}{c} \tag{11}$$

where t_1 is the time measured in system A it takes light from atom A to reach atom B. Therefore, the component of the velocity $\boldsymbol{v}$ along the line AB″ is

$$v_{\substack{\text{running}\\\text{away}}} = v\cos\theta = \frac{v^2}{c} \tag{12}$$

In the absence of this component of velocity the next wave crest would reach the atom B at the moment $t_1 + \dfrac{\lambda_0}{c} = t_1 + T_0$. However, since this component is present, the crest will reach B at a later moment:

$$t_2 = t_1 + \frac{\lambda_0}{c - v_{\substack{\text{running}\\\text{away}}}} = t_1 + \frac{\lambda_0}{c - \dfrac{v^2}{c}} = t_1 + \frac{c\,T_0}{c - \dfrac{v^2}{c}} = t_1 + \gamma^2(v)\,T_0 \tag{13}$$

Thus, in system A, two successive passings of wave crests from atom A through atom B are separated by the time interval[1)]

$$t_2 - t_1 = \gamma^2(v)\, T_0 \tag{14}$$

This interval is a factor $\gamma^2(v)$ greater than T_0! Comparing it with Equation (9), we see that it is a factor of $\gamma(v)$ greater than the time T (as measured in A) between two successive emissions of waves by atom B. Thus, even though the period of light waves from B, observed in A, is $\gamma(v)$ times greater than T_0, the period $t_2 - t_1$ turns out to be $\gamma^2(v)$ times greater. As a result, dividing Equation (14) by Equation (9), we find that during the time interval $t_2 - t_1$ the atom B emits $\gamma(v)$ waves of its own. This is precisely the resulting Equation (14) obtained by Tom, who states quite correctly that "his" atom B, considered as a clock, emits light waves $\gamma(v)$ times faster than "Alice's" one.

We thus see that if both observers register *one and the same process* – the arrival of waves at B – then both of them (in different ways and independently of each other) come to *the same conclusion* that the atom B emits $\gamma(v)$ times more waves of its own than it encounters waves from A. However, the explanations of this fact offered by the two observers will be absolutely different. Tom will point to the slower rate of oscillations of atom A, which moves relative to him. Alice will explain the same effect by the fact that Tom's atom B is moving away from the wave emitted by her atom A; as a consequence, the number of waves emitted by B is by a factor of $\gamma(v)$ greater than the number of waves coming to B even though B emits radiation $\gamma(v)$ times slower than A does.

If, however, Alice considers the arrival at A of light signals from B, while Tom considers the arrival at B of signals from A, then the two observers record different sets of events and, naturally, they come up with opposite results. Still, they do not contradict each other. Each result is correct – but only in relation to corresponding reference system.

As the last touch to complete the dialog of the two atoms, consider the following situation. Suppose that instead of maintaining both atoms in a state of continuous radiation of monochromatic light, we allow them to return from their initial excited state to their "normal" (non-excited) state after emitting only one photon each. Consider these two photons from the two atoms as signals. Suppose each atom emits its respective signal simultaneously at the zero moment of time in both systems (at this moment the line connecting the atoms – line AB – is perpendicular to their relative velocity $\boldsymbol{v}$.) The atoms at this moment are at the distance of the closest approach to each other. Call this distance L. Each atom emits its photon in the direction where it can hit the other atom. In Alice's reference frame these two directions are BA (for the photon from atom B to atom A) and AB″ (for the photon from atom A to atom B.) Then Alice will record the arrival of the photon from Tom's atom B at the

1) The resulting Equation (14) can be immediately obtained from Equation (10) if we note that $t'_2 - t'_1$ is simply the time events at a fixed point of the system K′ and apply to it the transformation Equation (47) in Section 2.8.

moment $t_A = \frac{L}{c}$. She has also found that the photon from her atom A will hit the receding atom B at the point B″ at a later moment $t_{B''} = \frac{AB''}{c}$. The simple geometry of Figure 5.2 tells her that $AB'' = L/\cos\alpha$, and since $\sin\alpha = v/c$, she gets for the arrival time $t_{B''}$

$$t_{B''} = \frac{L}{c\cos\alpha} = \frac{L}{c}\,\gamma(v) \tag{15}$$

She also notices that the x-coordinate of this event is

$$x_{B''} = v\,t_{B''} = L\frac{v}{c}\,\gamma(v) \tag{16}$$

Alice's conclusion is: "Both atoms fire their respective photons simultaneously, but are hit by these photons at different times! My atom A gets hit first, and Tom's atom B receives the photon from A later. And this can be *much* later if the Lorentz factor is sufficiently large."

The strange thing about this statement is that Tom will disagree with it, even though he observes the same set of events. He would maintain that he and his atom B are both stationary, while Alice with her atom moves to the left with the speed v relative to him. Then precisely the same treatment as Alice's, but applied in Tom's reference frame, will lead Tom to the result that his atom B will be hit first by the photon from atom A at the moment $t'_A = L/c$, and Alice's atom A will be hit by the photon from B at a much later time $t'_B = \gamma(v)\,\frac{L}{c}$. The only thing upon which Alice and Tom both agree is that their atoms fire simultaneously. But their conclusions about the arrival times of the two photons are dramatically different – they are just the opposite of each other. I want to emphsize again that now both observers speak about *the same* pair of events, and therefore it may seem that only *one* statement must be true. And yet, either observer is right. How can that be?

The resolution of this puzzle lies in the fact that both events are spatially separated in either reference frame. The diverging conclusions about their time ordering reflect the relative nature of time. We already know that two spacially separated events may be simultaneous for one observer and not simultaneous for another observer. The situation at hand illustrates that this relativity of time can be stretched a little further: of two events A and B, A may occur earlier than B for one observer, and later than B for another observer. If we go back to Section 2.9, we can see that such events are separated by a space-like interval. The time ordering for the end points of a space-like interval is interchangeable.

We can show this by applying Lorentz transformation to Alice's times t_A and $t_{B''}$. The time coordinates of the same events in Tom's reference frame are

$$\left.\begin{aligned} t'_{\mathrm{A}} &= \gamma(v)\left(t_{\mathrm{A}} - \frac{v}{c^2}\, x_{\mathrm{A}}\right) \\ t'_{\mathrm{B}''} &= \gamma(v)\left(t_{\mathrm{B}''} - \frac{v}{c^2}\, x_{\mathrm{B}''}\right) \end{aligned}\right\} \tag{17}$$

But Alice's atom A stays at the origin of her system A, so that $x_{\mathrm{A}} = 0$. As for Tom's atom B, Alice had found its x-coordinate at the moment $t_{\mathrm{B}''}$ to be given by Equation (16). Putting this into Equation (17) gives

$$\left.\begin{aligned} t'_{\mathrm{A}} &= \gamma(v) t_{\mathrm{A}} = \gamma(v)\,\frac{L}{c} \\ t'_{\mathrm{B}''} &= \gamma(v)\left[\frac{L}{c}\,\gamma(v) - \frac{v}{c^2}\,L\,\frac{v}{c}\,\gamma(v)\right] = \frac{L}{c} \end{aligned}\right\} \tag{18}$$

In Tom's reference frame, his atom B gets hit by the photon from A at the moment L/c, and Alice's atom A gets hit later, at the moment $\gamma(v)\,\frac{L}{c}$. This is precisely what Tom claimed from the very beginning. This claim, although opposite to that of Alice, does not contradict it logically, because it is related to another reference frame.

5.3 The longitudinal Doppler e ect

We have examined the case when the light from a source propagates in the direction perpendicular to the velocity of the source. Here we consider another imaginary paradox that arises in the case when these two directions are parallel, that is, when a source of light is moving along the line of sight of the observer. This is a typical situation in astronomy. We are living in the expanding universe, and the distant galaxies are running away from us the faster the further away they are. Even non-science students taking an introductory astronomy course know that this motion is accompanied by the "red shift" – the decrease in frequency and increase in the wavelengths of radiation from distant sources. The following is a typical argument illustrating a corresponding misconception:

„... the decrese in the light frequency from a receding source is accompanied by a corresponding *increase* in the light's wavelength, which secures the *invariability* of the speed of light. However, the statement about *contraction* of lengths in moving systems requires a *decrease* in the wavelength, which, given the decrease in frequency, inevitably gives us a value smaller than c for the speed of light, if this speed is defined as the product of frequency by the wavelength."

Let us look into these statements.

Imagine a light source at rest in a system K. A light beam which moves at a speed c from the source has a definite frequency $\omega = 2\pi f$ and wavelength λ, so that

$$c = \lambda f \tag{19}$$

Suppose we can measure the frequency and wavelength directly. We can use, for example, a set of small electric dipoles with rigidly fixed centers as a measuring device for frequency. The dipole charges are being held near the equilibrium position by an elastic force. Then the frequency of forced oscillations of a given dipole, i. e. the frequency *measured at a fixed point of our reference system*, will give us the frequency of light in this system.

Consider another system K′ moving at a speed v relative to K in the direction parallel to the light beam. Suppose it is equipped with a similar set of devices. They are stationary in K′, and thereby moving relative to K. What will *their* readings be?

First we note that a K′-observer will record a decrease in oscillation frequency of *the K-dipole* that is being watched by the observer in K. The frequency and period of the K-dipole observed from K′ are given by

$$\tilde{f} = \gamma^{-1} f\,; \quad \tilde{T} = \gamma T \tag{20}$$

However, this smaller frequency is *by no means the frequency of light* in system K′. Indeed, the chosen K-dipole is *moving* relative to K′, while the light frequency f' is the frequency of oscillations at a point *at rest* in K′.

Therefore, in order to determine f', we must watch the forced oscillations of a dipole, rigidly fixed in K′. It is easy to calculate the oscillation frequency of this dipole once the value of f is known in system K.

According to Equation (19), every new oscillation of the K-dipole begins in the system K′ $\tilde{T}$ s after the previous one. However, the emitted light signal will not reach the detector (the K′-dipole) $\tilde{T}$ s after the previous signal. Since the K-dipole is moving relative to K′, during the time $\tilde{T}$ it will travel the distance $\pm v\tilde{T}$, so that any two successive oscillations reaching the K′-dipole (and reproduced by it!) will be separated in time by the interval

$$T' = \tilde{T} \pm \frac{v\tilde{T}}{c} = \left(1 \pm \frac{v}{c}\right)\tilde{T} \tag{21}$$

where the "+" or "–" sign is chosen depending on whether K′ is moving away from the source or towards it. Thus, we obtain the frequency of light oscillations in K′:

$$f' = \frac{1}{T'} = \frac{1}{\left(1 \pm \frac{v}{c}\right)\tilde{T}} = \frac{1}{T}\,\frac{\sqrt{1 - \frac{v^2}{c^2}}}{1 \pm \frac{v}{c}} = f\sqrt{\frac{1 \mp \frac{v}{c}}{1 \pm \frac{v}{c}}} \tag{22}$$

We have derived a well-known equation for the relativistic Doppler effect for the case of relative motion parallel to the light beam. It can be seen from Equation (22) that the frequency of light oscillations in system K′ is not necessarily smaller than f. It may be *greater* than f if K′ is *approaching* the source of radiation.

This problem also admits a different approach, which elucidates the role of the Lorentz transformations. The electric field E observed at a point x at the moment t in reference frame K can be represented as

$$E = E_0 \cos(\omega t - kx) \tag{23}$$

where E_0 is the amplitude, $\omega = 2\pi f$, and

$$k = 2\pi \frac{f}{c} = \frac{2\pi}{cT} = \frac{2\pi}{\lambda} = \frac{\omega}{c} \tag{24}$$

The quantity k is called the propagation number. It shows the number of waves fitting into the segment of length 2π in the direction of propagation.

In frame K′ this disturbance has coordinates x' and t', which are related to x and t by the Lorentz transformations in Equations (43) in Section 2.6. Putting these relations for x and t into Equation (23), we can express the same disturbance in terms of primed coordinates:

$$E = E_0 \cos\gamma(v)\left[(\omega \mp vk)t' - \left(k \mp \frac{v}{c^2}\omega\right)x'\right] \tag{25}$$

Now we introduce the notations

$$\omega' = \gamma(v)(\omega \mp vk); \quad k' = \gamma(v)\left(k \mp \frac{v}{c^2}\omega\right) \tag{26}$$

Then Equation (25) will take the form

$$E = E_0 \cos(\omega' t' - k'x') \tag{27}$$

Equation (27) describes wave propagation as observed in system K′ (actually the electric field also undergoes Lorentz transformation, but this is immaterial for our purposes, since we are only interested in the field dependence on space and time coordinates.) The quantities ω' and k' in Equation (27) have a clear physical meaning: they are the wave frequency and propagation number in K′, respectively. Equations (26) are in fact the Lorentz transformation for these quantities, which relates their values in different reference frames. Using $k = \omega/c$ from Equation (24) allows us to rewrite Equation (26) as

$$\omega' = \gamma(v)\left(1 \mp \frac{v}{c}\right)\omega = \left(\frac{1 \mp \frac{v}{c}}{1 \pm \frac{v}{c}}\right)^{\frac{1}{2}}\omega; \quad k' = \gamma(v)\left(1 \mp \frac{v}{c}\right)k = \left(\frac{1 \mp \frac{v}{c}}{1 \pm \frac{v}{c}}\right)^{\frac{1}{2}}k \tag{28}$$

However, Equation (28) is just Equation (22), since the quantities f and ω differ merely by a factor 2π. As to the wavelength, its value in K′ can be found from Equations (28) and (24) as

$$\lambda' = \frac{2\pi}{k'} = \frac{2\pi}{k}\left(\frac{1 \pm \frac{v}{c}}{1 \mp \frac{v}{c}}\right)^{\frac{1}{2}} = \left(\frac{1 \pm \frac{v}{c}}{1 \mp \frac{v}{c}}\right)^{\frac{1}{2}} \lambda = \gamma(v)\left(1 \pm \frac{v}{c}\right)\lambda \tag{29}$$

The same also follows from Equation (22) by using $\lambda = \frac{c}{f}$.

Thus, there is no "wavelength contraction" of the kind described in the above statement. If the system K′ is moving in the direction of the light beam (that is, the source of light is moving away from us), we have the "+" sign in Equation (29), and the measured wavelength will be greater than that in K! When moving toward the source of ray (the source of light is moving toward us), we have the "–" sign in the same equation, and the observer in K′ will indeed find the light waves shortened. However, this "contraction" has little in common with the Lorentz contraction, and differs quantitatively from the latter by an additional factor $(1 + v/c)^{-1}$.

The conclusion about relativity of length that we have previously discussed pertains to sizes of material objects or distances between them. However, the size of a material system is not exactly the same thing as the wavelength, which is a characteristic of a periodic process. This difference manifests itself most clearly, for example, in the fact that the wavelength can be measured in a procedure (interference and diffraction experiments), which is different from measuring the size of a material body.[1)]

As a result of this difference, these two lengths are affected differently by the change of reference system: while the length of a rod *diminishes* in any system moving along the rod, *regardless* of the direction of motion, the wavelength in such a system can be either greater or smaller than in the original system, depending on the direction of relative motion of the source and detector.

Suppose we have a rod in system K, whose length l is equal to the wavelength of radiation propagating along the direction of the rod in this system. In any other system K′ moving relative to K, the length of the same rod will not coincide with the new wavelength. It can be said that the length of the rod and the wavelength λ transform differently when we switch from one reference frame to another. This means that the relation $\lambda = l$, which was established in system K, will not hold in other systems; this relation, it is said, is not covariant with respect to Lorentz transformations. Indeed, comparing Equations (28) here and Equations (51) in Section 2.8 (with primed and unprimed coordinates swapped!), we will not obtain in the system K′ the similar relation $\lambda' = l'$, but instead

$$\lambda' = \frac{\lambda}{\gamma(v)\left(1 \mp \frac{v}{c}\right)} = \frac{l}{\gamma(v)\left(1 \mp \frac{v}{c}\right)} = \frac{l'}{1 \mp \frac{v}{c}} \neq l' \tag{30}$$

1) According to quantum mechanics, all material objects possess wave properties, which become prominent for bodies with a sufficiently small mass. Diffraction of electrons, neutrons and even atoms is a well known phenomenon. This phenomenon can be only used to measure the corresponding wavelength which is associated with the motion of these particles. This so-called de Broglie wavelength has nothing to do with the size of a particle.

Finally, as to the speed of light, it is easy to show by calculating the product $\lambda' \nu'$ that no change in its value is observed in system K′:

$$c' = \lambda' \nu' = \frac{w'}{k'} = \frac{w}{k} = c \tag{31}$$

in complete agreement with the invariance of the speed of light.

5.4 Predicaments of relativistic train

In this section we shall discuss what had at first emerged as an apparently unsolvable paradox. The paradox is closely linked to the relativity of length, which we have already considered in some detail in Section 2.8. We will hereafter consider the paradox as a dispute between the two opposing sides.

... The general public was alerted to the implications of the length contraction paradox after a superpower on planet Rulia had come forth with an ambitious Project RT (Relativistic Train). The problem first popped up with the question: what would happen when such a train has to cross a canyon or river? Here I can give a very brief and simplified description of the problem, retaining only the most essential details. Imagine a train that has to pass a deep canyon. The train just fits across the canyon, so that its proper length L_t is equal to the proper length L_b of the bridge (Fig. 5.3):

$$L_t = L_b = L_0 \tag{32}$$

where L_0 stands for the comon proper length of the bridge and the train. Originally the bridge had been designed to sustain the train's weight $W_0 = m_0 g$, where m_0 is the rest mass of the train and g is the acceleration due to gravity on Rulia. The preliminary tests at low speeds were successful. The train had smoothly passed the bridge.

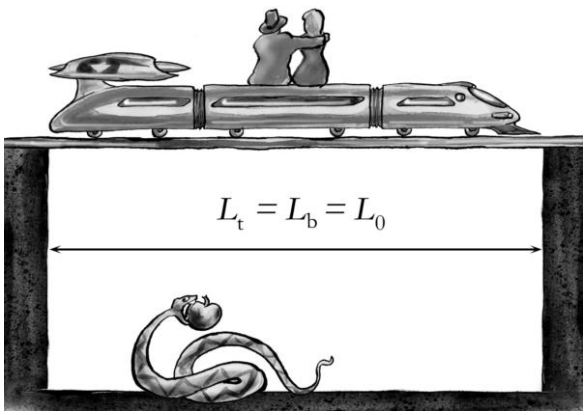


Fig. 5.3

Now, imagine this train moving with the speed close to c. Then it could fly over the canyon with even no bridge at all. The crossing time would be so small that the train would have practically no time to fall down by a tiniest increment in order to feel the obstacle. Let us, for example, estimate the distance the train on Earth would fall while flying across a canyon 1 km wide at a speed $V = 200\,000$ km s^{-1}, that is, two-thirds the speed of light. The time it takes for the front of the train to cross the canyon is $t = L_0/V = 5 \times 10^{-6}$ s. Denoting the vertical direction as y and entering the data into the equation $y = (1/2)\, g_E\, t^2$ with $g_E = 9.8$ m s^{-2}, we find $y = 1.23 \times 10^{-10}$ m, which is about the size of an atom. It is smaller than the finest irregularities of the rails' surface and definitely less than the distance the Earth's spherical surface curves away from the flat plane over a horizontal shift of 1 km. In other words, the curvature of the train's path would be less than the curvature of the Earth's surface. Such a train would, with even no bridge in place, fly off the Earth rather than go down, since its speed would by far exceed the escape speed for Earth (which is just 11.2 km s^{-1}) (Fig. 5.4). Apart from these, there is a finite *time factor* associated with the breakup of a bridge or any other system under an excessive force. A certain time is needed for a bridge to disintegrate and let the train down; this is the time necessary for a given force to do corresponding work. And this time for any real bridge is considerably greater than 5×10^{-6} s needed for the train to cross the bridge in our example.

But we are now concerned with the conceptual aspect of the problem rather than with technical one. Therefore, in order to avoid purely technical details obscuring the problem, we will represent a real system by its idealized model [19]. We will make the following assumptions: both the train and the deck of the bridge are infinitely thin, so that in the case of a crash it would take no time for the train to fall through the deck; the deck itself is ideally straight. In other words, Rulia is so huge that its surface is curved *less* than would be even the trajectory of a relativistic train over an unbridged canyon. We therefore need not bother about the escape preventing the crash, since there would be no escape for the train under these conditions. Astrophysicists would say that such a planet just falls short of becoming a huge black hole. For us, however, the assumption about Rulia's size just means that to a high accuracy we can consider the corresponding area of the planet's surface as flat, and the gravitational field within this area homogeneous. Accordingly, the train's trajectory in the case of a crash would be that of a projectile – a well known parabolic path from introductory college physics. We also suppose that the bridge is made of a highly idealized material that responds instantaneously to an applied force. It breaks instantaneously (in its own reference frame) under the whole train when the load

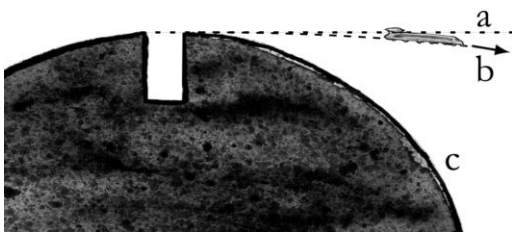


Fig. 5.4 An escape trajectory of a relativistic object. The trajectory's curvature is highly exaggerated. (a) The tangent to the planet's surface; (b) the escape trajectory; (c) one of the planet's meridians.

reaches a certain critical limit. We assume that this limit only slightly exceeds the weight of the stationary or slowly moving train.

We also assume that all technical problems associated with the design and launch of a relativistic train have been successively solved.

Now, after all these assumptions, the question is: how would the motion with relativistic velocity affect the train and the bridge? Will the bridge sustain the train or does it need to be reinforced? Will the train fly over the canyon or not?

An international team of experts from Earth had been invited to address the problem. Among the team members there was an engineer who had once worked at SSC (Superconducting Super-Collider) and then carried out some interesting experiments with Mr. O'Bryen during the latter's space missions. The engineer's opinion was unequivocal: the bridge under given conditions would not sustain the load, and the train would crash. The underlying reasoning is simple and straightforward.

"At a speed V close to c," the engineer said, "the train will undergo a longitudinal length contraction, so its length will be

$$l_{\mathrm{t}} = \gamma^{-1}(V)\, L_0 \ll L_0 \tag{33}$$

and its weight will accordingly increase due to the relativistic increase of the mass [Eq. 8, Sect. 2.1]:

$$W = mg = m_0 \gamma(V) g = W_0 \gamma(V) \gg W_0 \tag{34}$$

Therefore, the bridge must collapse, and the train will crash, smashed against the opposite wall of the abyss."

This conclusion is illustrated by Figure 5.5, in which, of course, the distance that the train would fall is highly exaggerated, to emphasize the result.

The engineer had encountered formidable opposition from the Chief Expert, who turned out to be one of the staunchest proponents of the Project. The Chief Expert was a very important man. He was so important that nobody has either heard or dared to ask about his real name. Everybody had respectfully referred of him as Mr. Ex, emphasizing his undisputable Expertise in the field of relativistic engineering.

"Ladies and gentlemen," Mr. Ex said amidst the awed hush, "if you want to see a real picture of the phenomenon, look at it from the viewpoint of the train's passenger. You then will see something different from the scenario conjured up by Mr. Fletcher (this was the engineer's name). The canyon with the bridge passes rushing by your train. The width of the canyon and the length of the bridge undergo relativistic length contraction:

$$l_{\mathrm{b}} = \gamma^{-1}(V)\, L_0 \ll L_0 \tag{35}$$

You can see from this equation that they are much shorter than the train."

Mr. Ex has illustrated this result by Figure 5.6 showing a ridiculously narrow crevice under the train.

"The train's weight," he went on after a significant pause, "will increase from its rest value W_0 up to the value $W = \gamma(V)\, W_0$, as stated by Mr. Fletcher, but for a different

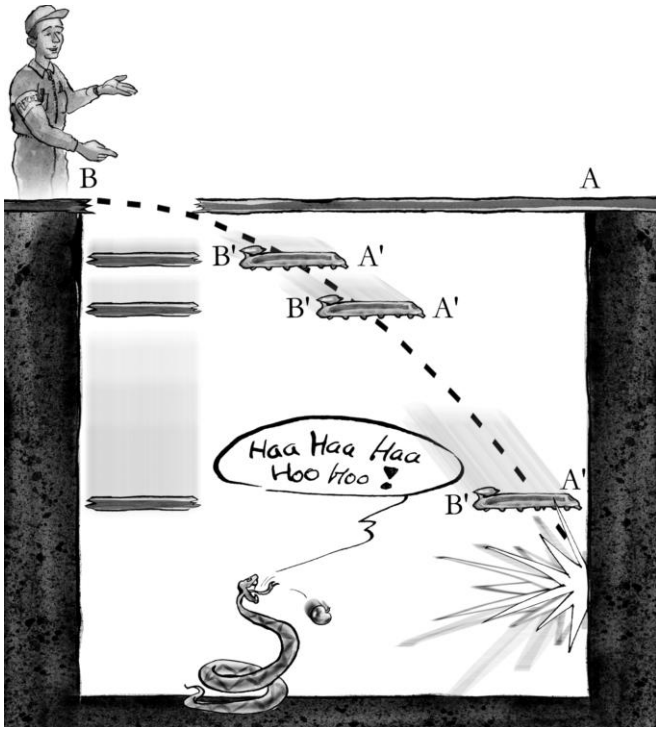


Fig. 5.5 The crash of relativistic train as expected to be observed by the engineer. The contracted and overweighted train collapses the underlying part of the bridge. The train slips down through the gap formed and traces out a parabola (the curvature of which is highly exaggerated).

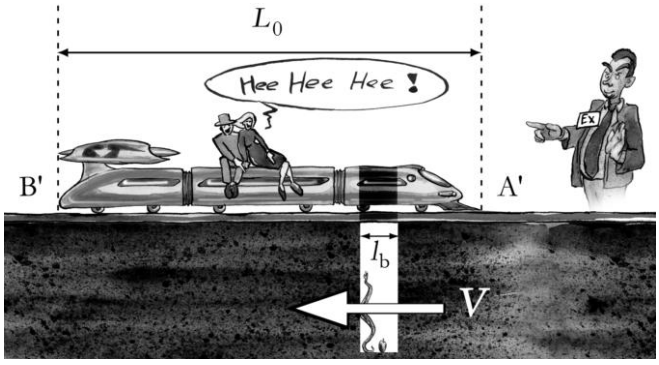


Fig. 5.6 The canyon with the bridge as observed by a train's passenger (in Mr. Ex's presentation).

reason. Since the planet is now moving relative to the train, the planet's mass M_0 undergoes a relativistic increase:

$$M = \gamma(V)\, M_0 \tag{36}$$

This will cause an increase in both the planet's gravitational pull and corresponding acceleration due to gravity:

$$g' = G\,\frac{M}{R_0^2} = G\,\frac{M_0}{R_0^2}\,\gamma(V) \tag{37}$$

where R_0 is the planet's radius. The resulting weight of the train will be

$$W' = m_0\, g' = m_0\,\gamma(V)\, g = W_0\,\gamma(V) = W \tag{38}$$

To Mr. Fletcher's credit, this particular piece of information about the train's weight in his report turned out to be correct. But, Ladies and Gentlemen," Mr. Ex concluded triumphantly after another significant pause, "the point is that only a small fraction of this weight will fall atop the bridge, the rest being supported by the ground. As you can see from Figure 5.6, this fraction is the same as the shaded fraction of the train's proper length L_0 that fits into the contracted bridge:

$$\frac{l_\mathrm{b}}{L_0} = \gamma^{-1}(V) \tag{39}$$

The resulting load on the bridge will be only $\gamma^{-1}(V)\, W = W_0$. It is the same load as that produced by the stationary or slowly moving train. But, as you all know, the slowly moving train has successfully crossed the bridge."

Mr. Ex made a third pause and fired his last victorious shot:

"We come *unavoidably*" (he stressed the word) "to the conclusion that the train will pass safely across the canyon – for two reasons. Geometrically, the train cannot go down because it just does not fit into the canyon's width; physically, it cannot go down because only a fraction of its full weight bears upon the bridge, and the fraction just small enough for the bridge to sustain. Geometry and physics, the two most general and established sciences about Nature, both give the same answer."

The final glorious scene followed after the engineer had been given the opportunity to reply.

"Sir," said the engineer, "as we all know, Relativity grants equal rights to all inertial observers. Therefore, a ground-based observer deserves the same respect as the passenger of the train, and his conclusion must be considered with equal attention. And if you do not see any error in my reasoning, then ..."

"Young man," Mr. Ex snapped, "it is not my business to look for errors in your reasoning. You had better find one in mine."

Mr. Ex probably tried to imitate a famous Russian physicist, Lev Landau, who had allegedly been the author of the above aphorism. And, although the engineer had sus-

pected that Mr. Ex was not a Landau, he could not, at the height and heat of the moment, spot any error in his opponent's argument. After a moment's silence, the audience burst into applause, and the project had been accepted. To make his victory more impressive, Mr. Ex even volunteered to board the experimental train (which was originally planned to be operated under remote control) and personally carry out all the measurements during the first relativistic test. It was decided to design a special cockpit for Mr. Ex. At the insistence of the Safety Board, the cockpit was to be installed at the rear of the train and equipped with an ultra-fast catapulting system.

The night after the meeting the engineer found what he believed to be an inconsistency in the Mr. Ex's treatment of the problem. He wrote down a detailed description of his solution and mailed it to Mr. Ex together with the letter of his resignation from the Project.

The essential parts of the engineer's solution ran as follows.

In the problem discussed we have a logical paradox that goes beyond simple relativity of length. The latter paradox is easily resolved by saying that the statement "The train is shorter than the bridge" is correct, and the statement "The bridge is shorter than the train" is also correct. If you ask how the two mutually exclusive statements may be both correct, the answer is that they are not mutually exclusive, because the physical property "length" is not absolute. It depends on the reference frame used when measuring the length of an object. The above two statements correspond to the two different reference frames. The first is associated with the bridge and the second with the train. Therefore, there is no paradox here, and both observers may be right – with the caution that each speaks for his or her own reference frame.

But when we speak about what *happens* to the train, we come to a statement of a different nature. It is a statement about an *occurrence*, not about a physical property that can be relative. An occurrence is something absolute, something that can be recorded by *any* observer. The reader can recall the examples with the number of passengers in a car or chemical composition of the bullet in the Introduction. Those were the examples of an absolute physical fact. Likewise, the train's crash (or its absence) is an absolute physical fact. If the train crashes, the event can be observed from any reference frame. If not, then nobody will observe any crash, no matter what the reference frame. The train either crashes or not, without any reference to a system of reference. Therefore, if one observer predicts that a certain condition will cause the train's crash, whereas the other holds that this condition insures safety, someone must be wrong. Who is?

In his report, the engineer indicates two errors in Mr. Ex's argument. The first was the statement that the gravity force in the train's reference frame was the same as in the engineer's one [Eqs. (37) and (38)]. That was wrong because it did not take into account the length contraction of the planet itself. Mr. Ex wrote Equation (37) (the Newtonian expression for g'), which holds only for the spherically symmetrical source of gravity. But because of the length contraction, the planet as observed from the relativistic train is an oblate ellipsoid (Fig. 5.7). Apart from the planet's mass being increased by a Lorentz factor $\gamma(V)$, its longitudinal diameter decreases by the same factor. This causes an additional increase in the gravity force exerted on the train. The parts of the planet symmetrical with respect to the equatorial plane P (e. g.

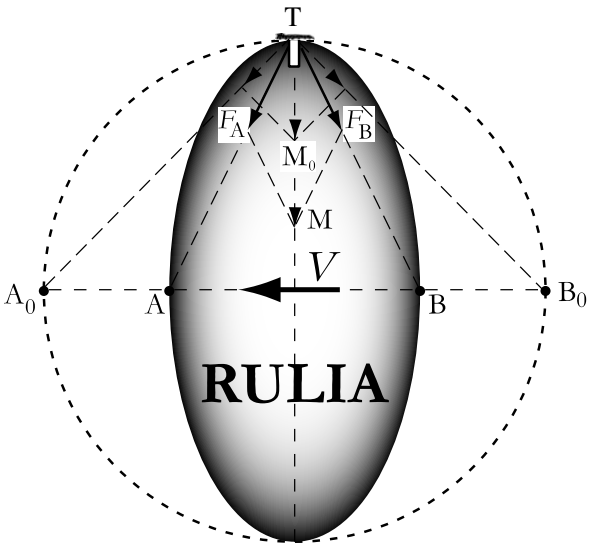


Fig. 5.7 The planet Rulia as observed from relativistic train (engineer's presentation). The planet is rushing to the left with a speed *V*. TM is a local vertical at T. A and B are the two opposite points on the contracted Rulia. The forces $\boldsymbol{F}_A$ and $\boldsymbol{F}_B$ exerted by masses at A and B on T are larger than in the reference frame associated with Rulia because: (1) the masses are larger and (2) they are closer to T. This picture ignores retardation of forces exerted by a moving object.

A and B) get closer to the train, which results in the increase of the gravitational force on the train. Also the angle subtended by the segment AB at T decreases, so that the corresponding forces F_A and F_B exerted on the train work in greater accord, which produces an additional increase in the resulting force.

The argument with individual forces as depicted in Figure 5.7 is not precisely accurate. It gives us only a feel as to why the resulting weight of the train measured by its passenger must be greater than *W*. According to the theory of relativity, a force exerted on an object by a system of *moving* masses at a given instant is not determined by the positions of the masses at the same instant. A proper account should be taken of the *retardation time* needed for a gravitational perturbation from each moving mass to reach the object. If we want to obtain a quantitative result, we have to write down the relativistic expressions for all retarded forces due to all small masses constituting the moving planet, and add them, that is, to perform the integration. However, we can get the result more easily by applying the general transformation rules to the force.

According to the relativistic equations of motion, we have for the gravity force

$$F_g = \frac{dP_y}{dt} \tag{40}$$

where P_y is the vertical component of the 4-momentum of the train. But this component, being transverse to the direction of train's motion, behaves as the transverse spatial dimension (see Sections 4.1 and 4.4!): it remains invariant under Lorentz transformation. The force F_g therefore transforms as the reciprocal of dt, that is

$$F'_g = \frac{dP'_y}{dt'} = \frac{dP_y}{dt'} = \gamma(V)\,\frac{dP_y}{dt} = \gamma(V)\,F_g \tag{41}$$

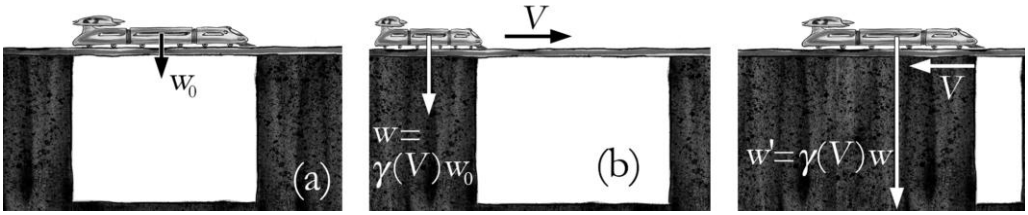


Fig. 5.8 The gravity force (a) on the stationary train, (b) on the moving train ($V \neq 0$) in the engineer's reference frame, and (c) on the train in Mr. Ex's reference frame ($V \neq 0$, Rulia is moving).

In other words, Equations (37) and (38) written by Mr. Ex are wrong. The transverse force is *not* the same in the two different inertial frames! This is precisely what Equation (33) in Section 4.2 on relativistic dynamics tells us: if the force on the train is $F_g = W = \gamma(V)\, W_0$ in the engineer's reference frame, it would be

$$F'_g = W' = \gamma(V)\, W = \gamma^2(V)\, W_0 \gg W \tag{42}$$

in Mr. Ex's reference frame (Fig. 5.8). It causes a corresponding increase in the acceleration of a free fall:

$$g' = \gamma^2(V)\, g \gg \gamma(V)\, g \tag{43}$$

There is deep irony in the fact that just when Mr. Ex had so condescendingly acknowledged the correctness of the engineer's conclusion about the gravity force being equal to $\gamma(V)\, W_0$, *he was himself wrong* in applying this conclusion to *his* reference frame. As we had found from geometry of an ellipsoid and, more generally, from Equation (40), both the acceleration g' and the gravity force in Mr. Ex's reference frame are much greater than in the engineer's. Therefore, even a small fraction of the train that fits into the bridge will, contrary to Mr. Ex's expectations, cause a much greater burden than the bridge can hold. This was the first of Mr. Ex's blunders. The second was about his "geometrical" treatment of the problem.

According to Mr. Ex, the fact that the train in his reference frame will not fit into the narrow crevice must prevent the train from crashing. But it does not. The train can crash piecemeal, bending and going down the crevice in small increments, one at a time (Fig. 5.9). From a passenger's perspective, the fall would be accompanied by a continuous deformation. First the front end goes down and is smashed against the opposite wall of the crevice; this gives room for the following parts to do the same; and this goes on until the whole train is swallowed up by the abyss. Instantaneous snapshots of the process show that the falling train consists nearly all the time of two different parts: one still horizontal, and one dangling above the crevice. The former is straight and the latter is curved down. The closer to the front, the steeper is the slope of the curve. After the front touches the opposite wall of the crevice, there emerges also a third part, common for all observers – the one smashed against the wall – but we do not consider it.

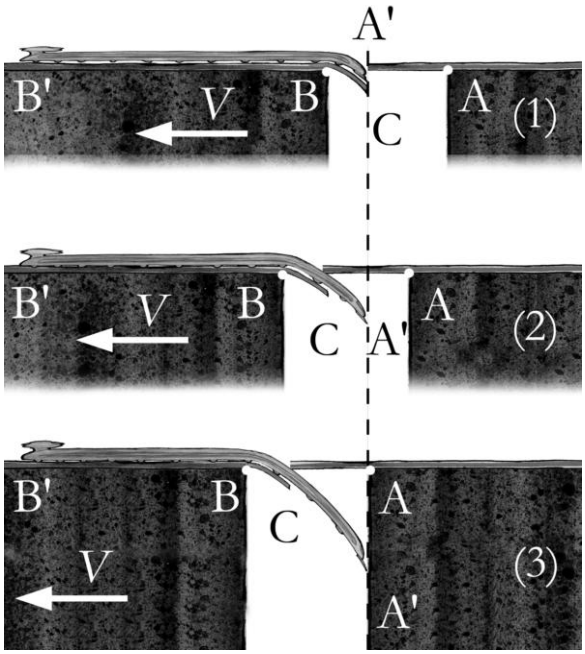


Fig. 5.9 The train crash to be observed by Mr. Ex (engineer's presentation). Three successive moments of crash are depicted here as (1), (2),and (3). The train is slipping down through the gap in the moving bridge. C is the broken fraction of the bridge.

From the moment when the train's front dives down, we can no longer speak about "the train's reference frame," because the train is no longer one rigid body. It is more like a fluent combination of rods and hooks, continuously changing from more "rod-dish" to more "hookish." We shall under these circumstances refer to Mr. Ex's reference frame, rather than the train's reference frame, since Mr. Ex, presumably located at the train's rear, would keep moving in a straight line nearly all the time until the very end (Fig. 5.9).

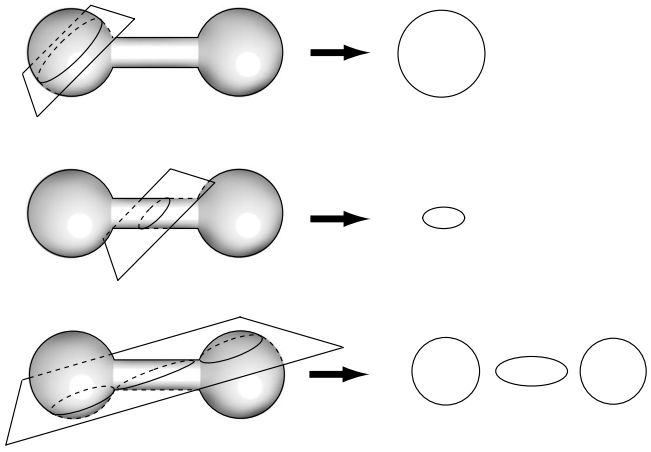


Fig. 5.10 Various cuts through a dumbbell illustrate its various possible two-dimensional pictures.

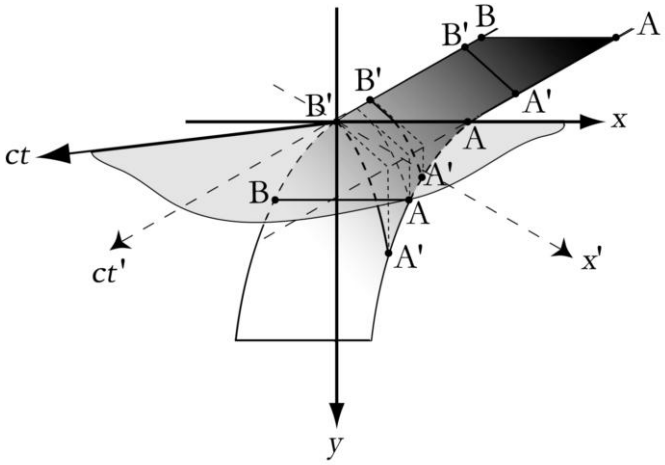


Fig. 5.11 The "world sheet" of the train. Its instantaneous cuts in the engineer's reference frame (K) and that of Mr. Ex's (K′) give the shapes of the train seen by the two observers. All cuts in K are straight segments parallel to the *x*-axis (three successive lines AB – before, at, and after the moment of gap formation in the bridge.) The cuts in K′ are intersections with vertical planes parallel to the *x′*-axis. Their shapes depend on time *t′*. The first line A′B′ (when looking in the direction of increasing *ct′*) is a straight segment; it represents the train's shape before the train reaches the bridge. The second line is composed of the straight and curved parts, and gives the shape of the train soon after the bridge had started to collapse (the collapse does not happen at one instant in K′!). The third cut is a curved line; it represents the train's shape when its rear passes the edge B.

Now, pit this picture against that observed by the engineer: the train going down all at once, in one piece, keeping the horizontal position all the time. One and the same thing is a totally straight rod for some observers, and a continuously changing combination of straight and curved lines for others. How can this be? Well, it is just one of those weird pictures of the world which, for all their strangeness, turn out to be correct and reasonably explainable. The shape of an object is a relative physical property. What we call shape is just a three-dimensional cross-section, cut trough an absolute four-dimensional structure in Minkowsky's world. Such sections are different for different "cuts" just as are two-dimensional sections cut through a three-dimensional object (Fig. 5.10). The way the object looks *now* is not the same for everyone because *now* is not one instant for everyone. In our case, the situation can be made clear if we represent the train by a moving segment and draw its *world sheet* in space–time (Fig. 5.11). It can be clearly seen from Figure 5.11 that what Mr. Ex would observe *now* are different moments in the train's history observed by the engineer, and vice versa. The two events A′ and B′ observed simultaneously by Mr. Ex occur at different moments to the engineer: the event A′ at the front happens *after* the event B′ at the rear. This means that the front was further down the crevice

than the rear when they were observed simultaneously by Mr. Ex. In other words, Mr. Ex observes the ends of the train at different horizontal levels. Therefore, the line going from A′ to B′, that is the form of the train, turns out to be tilted in his reference frame. The situation is in this respect similar to that with the rising water in Section 2.10. The difference is only in the initial conditions: first, the "water" (the train) is sinking now, and sinking with the acceleration; second, the train remains horizontal in the engineer's reference frame, rather than in Mr. Ex's reference frame (we also have Mr. Ex instead of Mr. O'Bryen, but this is irrelevant to physics). Now, if the point A′ is "captured" *after* point B′ in the engineer's reference frame, it is sinking faster. We know that the tilt is proportional to the velocity v of sinking. Therefore, the tilt at A′ is steeper than that at B′. Thus we obtain, at least on the qualitative level, an account of the observed features of the train's crash in Mr. Ex's reference frame.

The engineer's letter also contained a more detailed quantitative description of the transition process between the straight and the "hooked" parts of the train. I bring it along here for the needs of more sophisticated readers. Let us place the origins of the two discussed reference frames at points B and B′, and start to count time from the moment when the origins of both systems coincide. Then the train's vertical displacement in the ground-based reference frame K is

$$y = \begin{cases} 0\,, & t < 0 \\ -\dfrac{1}{2}\, g\, t^2\,, & t > 0 \end{cases} \tag{44}$$

In the passenger's reference frame K′ the same displacement can be obtained by just applying Lorentz transformation to y and t (note that transverse displacement is the same in both systems!):

$$y' = y = \begin{cases} 0\,, & \text{for } t' + \dfrac{V}{c^2}\, x' < 0 \\ -\dfrac{1}{2}\, g\, \gamma^2\,(V)\left(t' + \dfrac{V x'}{c^2}\right)^2\,, & \text{for } t' + \dfrac{V}{c^2}\, x' > 0 \end{cases} \tag{45}$$

As we see, an observer co-moving with the train's rear would measure the acceleration of the falling part of the train as $g' = \gamma^2\,(V)\, g$, in complete accord with Equation (43). The instantaneous velocity of the train's fall at a point x' is

$$v'\,(x', t') = \begin{cases} 0 & \text{for } t' + \dfrac{V}{c^2}\, x' < 0 \\ -g'\left(t' + \dfrac{V}{c^2}\, x'\right) & \text{for } t' + \dfrac{V}{c^2}\, x' > 0 \end{cases} \tag{46}$$

One might argue that Equation (46) cannot be true because it gives $v'\,(x',\, 0) = -\,g'\,(V/c^2)\, x' \neq 0$ for the initial velocity of fall at the zero moment $t' = 0$, whereas the fall in a given case begins from rest ($v' = 0$). But the zero moment $t' = 0$ is *not* the in-

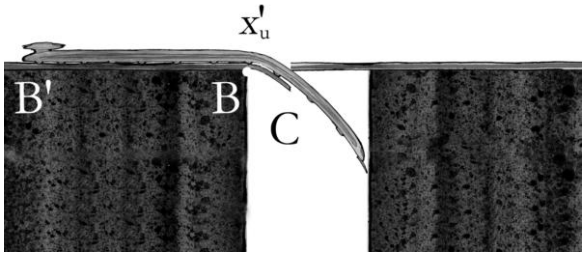


Fig. 5.12 An instant in the process of train crash depicted by Mr. Ex in his response to the engineer's letter. The separation point x'_u between the straight (horizontal) and curved part of the train appears to coincide with edge B of the bridge. They appear to slide *together* to the rear of the train in Mr. Ex's frame.

itial moment of fall for a point x' on the train. Recall that the zero moment has been determined as the moment when the *rear* of the train B′ coincides with the edge B of the bridge. By this time all the points x' to the front [that is, satisfying the second condition (45)] have already fallen down a certain distance and acquired non-zero vertical velocity. This velocity is given by the second term in Equation (46).

The instantaneous shape of the train in the system K′ is automatically described by Equations (45) as consisting of two parts. One is straight and the other one is curved down with its tilt increasing for greater x'. The expression for a local tilt at a moment t' is obtained from Equations (45) by taking the derivative

$$S'(x', t') \equiv \frac{\mathrm{d}y'}{\mathrm{d}x'} = \begin{cases} 0 & \text{for } t' + \dfrac{V}{c^2}\, x' < 0 \\ -g'\, \dfrac{V}{c^2}\left(t' + \dfrac{V}{c^2}\, x'\right) & \text{for } t' + \dfrac{V}{c^2}\, x' > 0 \end{cases} \tag{47}$$

Using the second of Equations (46), the last equation can be rewritten as

$$S'(x', t') = \tan\alpha = \begin{cases} 0 & \text{for } t' + \dfrac{V}{c^2}\, x' < 0 \\ -\dfrac{V\, v'}{c^2} & \text{for } t' + \dfrac{V}{c^2}\, x' > 0 \end{cases} \tag{48}$$

in direct agreement with the engineer's result for the tilt of moving water surface in Section 2.10.

The transition point x'_V between the "roddish" and "hookish" parts of the train is determined by condition $t' + (V/c^2)\, x'_V = 0$, that is

$$x'_V = -\frac{c^2}{V}\, t' \tag{49}$$

For each moment $t' < 0$ all x' to the left of x'_V belong to the straight segment; all x' to the right of this point belong to the "hooked" part. The succession of different snapshots of the process at different moments in both systems is shown in Figures 5.5 and 5.9. While the train falls as one horizontal piece in Figure 5.5, it slides gradually down in Figure 5.9 through the narrow slot in the bridge. In both reference frames,

the slot is the $\gamma^{-1}(V)$ fraction of the whole bridge. This picture completes the engineer's description of the process as expected to be observed from the train's cockpit.

A few days later the engineer received an official note of acceptance of his resignation. He was surprised to find an attachment written personally by Mr. Ex. The letter started as follows:

"Sir,

Since you have made an attempt to find an error in my calculations, I am now returning the favor by indicating one in yours."

"Hm", grinned the engineer, "so we are promoted from a young man to Sir along with our resignation. What a twist of a career!"

But what he read next withered his grin.

"Look at Figure 5.12. It is drawn following your own description of the train's fate as observed from its rear. Clearly, the separation point between the two parts of the train coincides with the point B (the left edge of the bridge). It is at this edge where the train starts bending down in your scenario. As time goes by, the point x'_V and the edge B both slide together towards the rear of the train. Your Equation (49) states that they both move down the track with a speed

$$u = \frac{x'_V}{t'} = \frac{c^2}{V} \tag{50}$$

This result of yours is wrong for two reasons. First, the speed u is greater than the speed of light! It thus follows from your treatment that the edge B moves faster than light, which is in flat contradiction with the Theory of Relativity. Second, this result is also in ridiculous contradiction with the initial conditions of the problem, according to which the speed of the edge is less than c and equals V. And, finally, your beautiful pictures in Figures 5.9 and 5.11 of the bent train may well represent a stick of putty, but not a real train. Our train is rigid, it presents stiff resistance against any deforming force. Therefore, your pictures contradict reality. This shows that all the rest of your reasoning was wrong, and you could have made better use of your free time by visiting a night club. I wish you all success in whatever career you choose."

All the rest of the day the engineer was deep in thought. Mr. Ex's argument about Equation (49) was irrefutable. The engineer must have made a grave error. And his pictures in Figures 5.9 and 5.11 indeed seem incompatible with the train's rigidity. Where could he have gone wrong? Up to this moment everything had seemed to fall so neatly in place, but after this moment ...

Instead of visiting a night club, the engineer spent the night over his papers. The next morning he mailed his second letter to Mr. Ex. It is not known for sure whether Mr. Ex had enough time to give full consideration to this letter.

A few months of preparation had passed with much fuss and fanfare. But on the evening of the day of the test the mass media reported about the miserable failure of the Project RT. The train had crashed into the canyon in the very first test. The only system that proved to be efficient was the catapult, that launched the cockpit with its passenger into space and then delivered it back to Earth. Mr. Ex was found to be safe and sound in all respects except that he was unable to speak for a considerable

length of time, during which an investigation had been carried out. The investigation found, among other things, the engineer's letters with the full account of the problem.

The engineer's account, together with the comments of other experts, was published in the "Final Report" of the investigating Committee. The following gives a simplified description of the engineer's last letter to Mr. Ex. The letter starts with the acknowledgement that Mr. Ex's statement about the superluminal velocity of the separation point x'_V (let us call it point D) was irrefutable. It follows directly from Lorentz transformations. The point D does move faster than light! But this does not in any way undermine the validity of the engineer's results, because point D does *not* coincide with the edge B, except for the very last moment when they both merge at the rear of the train. Let us take a close look at Figure 5.13. It corresponds to train's velocity $V = 0.968\,c$, for which $\gamma(V) = 4$. Accordingly, the train gets contracted to one-quarter of its proper length (and of the bridge's length) in the engineer's reference frame. The equal fraction of the bridge breaks under the train and falls into the canyon. In Mr. Ex's reference frame it is *the bridge* that shrinks down to one-quarter of its proper length; therefore, the gap in the bridge, being only one-quarter of the whole bridge, is just $(1/4)^2 = 1/16$ of the train's proper length. Figure 5.13a depicts the moment when the train is 1/16 of its full length above the canyon. Since the gravity force between the train and Rulia is, according to Equation (42), 16 times the weight of the stationary train, the load on the bridge is at its limit. This is the moment when the bridge is just about to break. The train begins to bend where the bridge begins to break. But it is not at the point B! The gap in the bridge forms instantaneously in the engineer's reference frame. But it is not one instant in Mr. Ex's reference frame! In his, the point on the bridge right under the train's front (point A′) gives in first, and it is *here* where the train starts to bend, shoving its head under

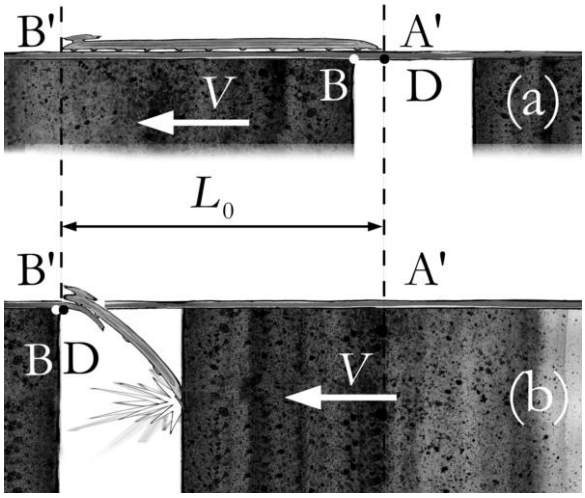


Fig. 5.13 The initial (a) and final (b) moments of the train crash to be observed in Mr. Ex's reference frame (engineer's presentation). The separation point between the straight and curved part of the train (point D) does *not* coincide with B except for the very last moment of train's crash. The edge B moves to the left slower than light (with the speed V), and separation point D moves faster than light, so that the product of their speeds is $Vu' = c^2$. It is instructive to compare this figure with Figure 5.5 (engineer's reference frame). There, both the train and the falling part of the bridge stay horizontal and remain on one (sinking) level.

the remaining part of the bridge on the right. The part on the left (which is already doomed to be broken), remains at this moment strictly horizontal, and so does the train, of course. There is thus an initial separation equal to the distance A′B (1/16 of the train's proper length) between points D and B. Only some time later (at the zero moment $t' = 0$) will the edge B of the bridge be observed as giving in. But, by this time, the edge B will reach B′, so that both points B and D will have merged only at the *rear* of the train. By the zero moment $t' = 0$ the process of train's bending in Mr. Ex's reference frame has been completed, and the train is converted entirely from a rod into a hook. This is where relativity of time reveals itself in its full sway! In order for D to catch up with B at the rear of the train, it must move faster than light while the edge B is, as it should, moving slower than light with the prescribed velocity V.

The fact that the separation point moves faster than light does not by itself contradict anything, since there is *no energy transfer* associated with this motion. We will see in Chapter 6 that this kind of motion is a rather common physical phenomenon.

Now, let us describe the whole process symbolically. The time t'_{D} that it takes the separation point to move along the train from its front to the rear is

$$t'_{\mathrm{D}} = \frac{L_0}{u} = \frac{V}{c^2}\, L_0 \tag{51}$$

The time t'_{B} that it takes the edge B to reach the train's rear is BB'/V, where BB′ is the original distance between them when the bridge started to break. From the above numerical example and Figure 5.12, this distance is readily found to be

$$BB' = L_0 - \gamma^{-2}(V)\, L_0 = \left[1 - \gamma^{-2}(V)\right] L_0 \tag{52}$$

Therefore,

$$t'_{\mathrm{B}} = \frac{BB'}{V} = \frac{V}{c^2}\, L_0 = t'_{\mathrm{D}} \tag{53}$$

It is therefore proved that, although the edge B moves, as any physical body should, slower than light, it needs the same time to reach the rear of the train as the superluminal point D, because it has to travel a shorter distance ($\mathrm{BB}' < L_0$). The boundary between the "roddish" and the "hookish" parts of the train starts from the train's front and moves to its rear faster than light. The product of these two velocities, according to Equation (50), is equal to $uV = c^2$.

Ironically, the superluminal speed of point x'_V, which, in Mr. Ex's opinion, was a fatal flaw of the engineer's description, actually resolves the problem with the "rigid" train resisting any bending force. No matter how rigid the object, the bending forces between its atoms cannot transfer information about atomic displacements faster than light. Consider a train's atom that starts to fall together with the collapsing part of the bridge. The adjacent atom finds itself also on the collapsing part and thereby already in a free fall *before* it knows that its neighbor had already started to do so. Because the process of falling is propagating along the train faster than light, there is

no time left for internal forces to respond. In this respect, the train, bent as it is in Mr. Ex's reference frame, remains physically non-deformed, so that the engineer's presentation accurately describes reality without contradicting anything.

This completes the description of a picture that the train's passenger would observe in the fleeting time interval before the crash if he or she were unlucky enough to remain there. Both physics and geometry in this description go hand in hand and see to it that there are no privileged observers, but *everybody* gets the same facts. The observed features and instantaneous physical characteristics of the phenomenon may be dramatically different in the two systems, but in either system they contrive to bring about the same result. The train that crashes in the engineer's reference frame crashes in the passenger's one. This is the way in which the relativity of certain physical quantities ensures the absoluteness of physical events.

5.5
Dramatic stop

Alice was watching a relativistic train of proper length $L_0 = 50$ m, approaching a narrow canyon of length $l_0 = 10$ m. The speed of the train was such that the Lorentz factor $\gamma(V)$ determining the length contraction

$$L = \frac{L_0}{\gamma(V)}, \quad \gamma(V) = \left(1 - \frac{V^2}{c^2}\right)^{-\frac{1}{2}} \tag{54}$$

was equal to 5, so that the train measured only $L = l_0 = 10$ m in the ground-based reference frame. Then at a certain moment the train must just fit into the bridge across the canyon (Fig. 5.14a).

"What happens," Alice thought, "if at this moment the train is suddenly stopped by a huge braking force, applied simultaneously to all its parts (assume that produced heat quickly dissipates)." The question seemed a real puzzle to Alice. On the one hand, the stopped train should, according to Equation (54), measure its proper length $L_0 = 50$ m in *Alice's* reference frame, which means that its front and rear would each stick 20 m from the respective edges of the bridge (Fig. 5.14b); but, on the other hand, if the equal forces are applied to equal parts of the train, then all the parts should have equal acceleration (or deceleration, for that matter), so that the length of the train cannot change (Fig. 5.14c).

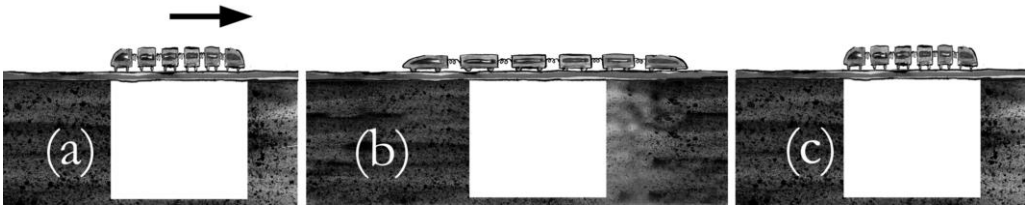


Fig. 5.14

Definitely, there must be something wrong with our logic here. A brief contemplation shows that when we talked about the applied force we were not careful enough with the words "*simultaneous*" and "*force*." We should have asked two crucial questions:

1. Simultaneous – in what reference frame?
2. Is the braking force the only actor in the play, or there are other participants?

The answers to these questions would put everything in place. It turns out that the outcome of the process dramatically depends on its mechanism [20–22].

So let us analyze the whole process as seen by the observers in two different inertial reference frames: one originally associated with the train, and the other associated with the bridge. We will first neglect the internal forces as compared with the braking force. Then we will take them into account and discuss qualitatively their role in the length contraction effect described by Equation (54).

We invite Alice and Tom as the observers. Let Alice stay on the ground (system A) and let Tom ride in the train before it stops (system T).

5.5.1 Braking uniformly in A

We consider first the mechanism in which the braking forces are dominant and are applied to all parts of the train *simultaneously in system A*. We start with Alice's account of the process. The moment Alice sees the rushing train all fit into the bridge, she records the readings (zero time) of her synchronized clocks at the two opposite sides of the canyon. The train's length measured by Alice after the stop turns out to be 10 m – just as it was before the stop. This is precisely what one would expect from the dynamics of motion: if we apply equal longitudinal forces simultaneously to equal parts of the train, then all the parts will stop synchronously; this cannot affect the size of the train. Thus, our assumption of the train slowing down simultaneously in all parts in system A is logically incompatible with the assumption of the train retaining its proper length (50 m) after the stop. If the moving train comes to a stop, *and* retains its proper length, then its length measured by Alice after the stop must be equal to its length measured by Tom before the stop. The length measured by Tom before the stop was 50 m. Therefore, Alice would expect to see the transition from the 10 m length-contracted moving train to the 50 m stationary train, just as it was depicted in Figure 5.14 b. But this is *not* what she actually observes. We therefore conclude that the proper length of an object may change after its state of motion had changed! Under the considered action of braking forces the proper length of the train must change from 50 m before the braking to only 10 m after the braking.

This change must be caused by an actual physical deformation that compensates for the *disappearance* of length contraction due to the stop. Accordingly, whereas Alice sees no net change in the length of the train, Tom must see the train contracted during the braking. If the theory of relativity is logically consistent, it must account for this discrepancy between two observations. There must be a physical reason for de-

formation of the train in Tom's reference frame. We thus turn our attention to system T – to Tom's account of *the same* process.

Owing to the relativity of time, events simultaneous in system A are generally not simultaneous in system T, and vice versa. The coincidences of the ends of the train with the sides of the canyon, which happen at the same moment (zero time) in A, happen at different moments in T. Tom sees the bridge moving to the left down the train (Fig. 5.15). The first event (the right edge of the bridge coincides with the front of the train) happens earlier in Tom's time than the second event (the left edge coincides with the train's rear). Applying the Lorentz transformation for time coordinates of the two events, taking into account that these events both happen at $t = 0$ in system A, and using Equation (54), we can express the time interval between the events in T in terms of the proper length of the train:

$$\Delta t' = \frac{V L_0}{c^2} \tag{55}$$

The length of the bridge as observed in system T must be much *smaller* than the length of the train: Tom sees the moving bridge contracted down to one-fifth of its proper length, that is, to $l = l_0/\gamma = L_0/\gamma^2 = 2$ m. This is similar to the picture of the canyon narrowed down to a crevice, seen by a passenger in the relativistic train in

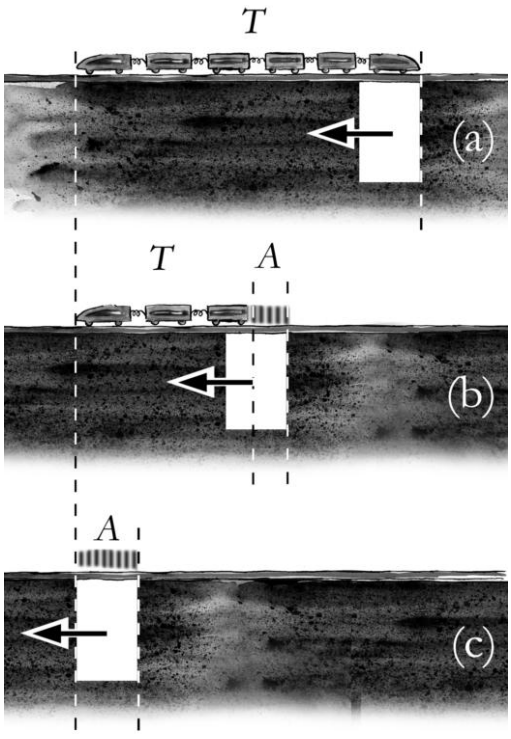


Fig. 5.15

the previous section. This is also consistent with the just-mentioned fact that the right edge of the bridge coincides with the front of the train *before* the left edge reaches the rear of the train.

In the first of these events the braking force stops the train's front relative to the bridge. But in Tom's reference frame the bridge is moving to the left. Therefore, Tom sees first the front of the train grip firmly with the moving bridge, and dragged by it to the left (Fig. 5.15 a) This compresses the train! As the new fractions of the braking train grip with the bridge, they become involved in motion to the left together with the bridge (Fig. 5.15 b). Thus the wave of compression forms and propagates to the left along the train. The front of the wave separates the train into two parts: one uncompressed to the left of the wave front, and the other compressed and moving together with the bridge, behind the wave front. So the wave front can be considered as the running separation between the two parts. Just as the separation point on the train in the previous section, here too the separation point runs down the train faster than light. Also, as in the previous case, the superluminal motion of the separation point does not contradict anything, because this point is just a moving boundary between the two regions rather than a physical particle. The compressed parts of the train move together with the bridge slower than light. The separation point outruns them because the new parts in front of it join this motion as they grip with the bridge.

As in the previous case, we can find the superluminal speed of the separation point between the two parts of the train. This point travels a distance L_0 in time $\Delta t'$ given by Equation (55) – essentially the same equation as used in Section 2.10. Thus,

$$u = \frac{L_0}{\Delta t'} = \frac{c^2}{V} \tag{56}$$

This superluminal wave of contraction stops when the separation point reaches the rear of the train. And, because the compressed parts of the train were also moving towards its rear, by the end of the process the train turns out to be contracted to the size of the bridge (Fig. 5.15 c).

During this process, we cannot consider the train as one reference frame. It belongs to two different frames of reference: one to the left of the advancing wave front (system T), and the other to the right of the wave front, but involved in motion to the left with the bridge and the canyon (system A). The separation into two systems starts from the whole train constituting one system T, and ends up with the whole train constituting (together with the bridge) system A. This also reminds us of the situation in the previous section, when the train, bent over the collapsing part of the bridge, no longer constitutes one rigid reference frame. The difference between the two cases is that the bent parts of the train in Section 5.4 were moving in the direction *perpendicular* to its straight part, whereas now the compressed parts of the train are moving *along* the train. In other words, the compressed parts of the train are moving in the direction in which we measure its length. However, if one part of the object is moving relative to the other in the direction of measurement, the two parts cannot have a common rest frame as it is defined for a uniformly moving rigid ob-

ject. The concept of a rest frame should in this case be generalized as a reference frame where the total momentum of the object is zero. In our case such a reference frame is not inertial because the total momentum of the train is not conserved owing to its interaction with the bridge. Only in the end of the contraction process, when the relative motion of the two parts of the train stops, can the train be assigned a constant proper length again. However, this new proper length is now different from the old one: the train shrinks down to the size of the bridge. The physical process responsible for this contraction is deformation of the material of the train under the applied frictional force.

The deformation in this case is so severe that no solid body could actually sustain it. A real train would be just destroyed in the described process, because it cannot be "accordioned" down to 20% of its original length and remain straight.

But here Alice interrupts the story.

"What deformation are you talking about?," she asks. "I do not see any deformation. All parts of the train are being stopped simultaneously. The distance between any two adjacent particles along the train cannot change in such a process. Therefore, there cannot be any change in the inter-atomic interactions, and if we manage to remove heat, the train cannot deform, let alone be destroyed."

But if Alice takes a closer look at her experiment, she will realize that even with heat instantly removed, what used to be the rigid train is now, at best, an exotic system with unusually high mass density, glued to the bridge across its surface.

We can understand this outcome by taking a closer look at the hitherto neglected atomic interactions and their role in the process.

Although it is true that we could, in principle, keep the inter-atomic distances constant during the braking, it is *not* true that inter-atomic forces will remain constant in this process [16]. The *proper* distance between two neighboring atoms in a typical solid is about $a_0 = 0.1$ nm (10^{-10} m). At this distance each atom is in a state of stable equilibrium (or vibrates about its equilibrium position). Imagine inter-atomic forces as springs connecting neighboring atoms. In equilibrium, the springs are neither stretched nor compressed.

According to the principle of relativity, an equilibrium state of a stationary system must be the same in any reference frame. Therefore, Alice and Tom must each observe such a state as a stationary chain of equidistant atoms separated by the distance a_0. Now, Tom, initially, has a train in such a state. Alice observes this state from her bridge, and measures the inter-atomic distance along the train to be only $0.2\,a_0$. It does not contradict anything, because the train is moving relative to her. If she manages to stop the train *without changing its length*, she has after the stop a *stationary* system with unchanged longitudinal inter-atomic distance $0.2a_0$. But this distance is five times shorter than the one characteristic of the state of equilibrium! Alice begins to realize that what she has is *not* an equilibrium state. It is a state with its atoms crushed and squeezed together. The springs representing inter-atomic forces are compressed to one-fifth of their normal length. This produces huge forces of repulsion between the atoms, which, if left unopposed, would destroy any real train. It is due only to the incredibly strong external forces that the atoms of the train are being kept in place after the stop.

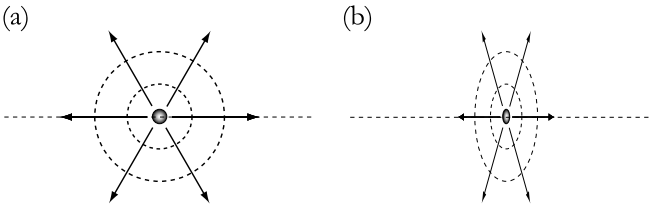


Fig. 5.16 Electric field of a point charge: (a) stationary; (b) moving.

In a model that represents atomic forces by springs obeying Hooke's law, a sudden disappearance of external forces will send the system into longitudinal vibrations around its natural proper size. If the vibrations are damped, the system will ultimately relax to this size, thus restoring its proper length.

Although springs model real internal forces only approximately, the conclusion that these forces act to some extent as the "keeper" of proper length does not depend on the model. For instance, if we solve the Maxwell equations for a moving point charge, we will find its field flattened in the direction of motion [16, 23]. Because the point charge has no size, this flattening is an intrinsic property of the field as such, depending *only* on relative motion between the charge and a reference frame, rather than on the shape of the charge. It can be considered as the length contraction of the field (Fig. 5.16). However, if the source of the field has a finite size, it *has* to possess the same property because the Maxwell's equations are self-consistent. The same is true for all other forces, since they all obey the relativistic equations, which are Lorentz-invariant. For a moving system, its particles and the *equilibrium distances* between them must be all Lorentz-contracted along the direction of motion (Fig. 5.17 a,b).

When the train stops, the individual fields of its particles (as observed in A) restore the shape characteristic for the stationary train, and the equilibrium distances increase. If the particles themselves are not allowed accordingly to shift apart, their fields become strongly overlapped (Fig. 5.17 c), which produces huge repulsive forces. If the external forces are suddenly removed, this repulsion may blow up the train. If external forces stay, we will have the "post-stop" picture observed by Alice – the dynamically compressed train with accordingly reduced proper length. However, if the external forces weaken sufficiently slowly, the internal forces can restore the

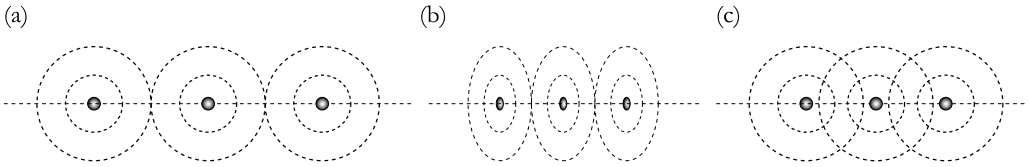


Fig. 5.17 Different states of an atomic chain. (a) Stationary, non-deformed, stable; (b) moving, non-deformed, stable; (c) after the uniform braking in A: stationary, outrageously deformed, extremely unstable.

original proper length of the train. In this respect, the internal fields act as a memory, keeping information about proper shape of an object.

Alice could summarize this part of the discussion in the following way. Before the stop, the train, although length-contracted, was *not* deformed, because the atomic distances matched the shape of the internal field produced by *moving* atoms. After the stop, even though the length of the train did not technically change in system A, the train *was* deformed because the atomic distances no longer matched the equilibrium shape of the *stationary* internal field. Our concept of deformation should be refined to describe adequately this kind of process in relativistic mechanics.

In this thought experiment, not only the train as a whole, but also each separate car, behaves first as a stick of putty, and only some time later as a rigid body. Therefore, in cases with huge external forces acting during a very short time (less than the travel time of the shock wave between the points of interest), we can first use the model of an infinitely deformable object (in essence, such an object can be represented by its two end points). Then we can try to find out how the internal forces change the results.

Whatever the model, we arrive at the description of the phenomenon, which, while appearing different to different observers, is consistent for all of them and predicts the final state upon which everybody agrees. To Alice, the train retains its initial contracted size because the equal external forces are applied simultaneously to its equal parts. Such forces can (and do) stop the train, but they cannot change its length. To Tom, the train has been compressed by external forces because they did not act synchronously. Such forces both stop *and* deform the train. Both agree that in the final state the train is physically deformed and just fits into the bridge.

(Tom, as Mr. Ex in the previous section, boards the rear cockpit of the train. By the end of the process, he is catapulted in the forward direction to continue his state of motion.)

5.5.2
Accelerating uniformly in T

Tom now suggests that in order to conserve the proper length of the train, the external forces should be applied to its ends *simultaneously in his reference frame T,* which is initially the rest frame of the train. Accordingly, we now start with the picture of the process as observed by Tom. So imagine two jet engines at the ends of the train, and Tom originally at the middle. Tom sees the canyon with the bridge rushing to the left. When the bridge passes the center of the train, both engines are switched on to *accelerate* the train in the same direction (Fig. 5.18). The left engine pulls to the left, stretching the train. The right engine pushes to the left, compressing the train. Both actions together accelerate the train to the left without changing its length in system T. If the engines are sufficiently powerful, the train acquires the speed of the bridge practically instantaneously. In the final state, the train and the bridge both form one single whole, and the edges of the train stick out of the bridge symmetrically on both sides.

This situation is "reciprocal" to the previous one. Tom and Alice exchange their roles. Alice had seen the moving train stopped. Now Tom sees the stationary train accelerated. Alice had claimed that the length of the train did not change. Now Tom claims

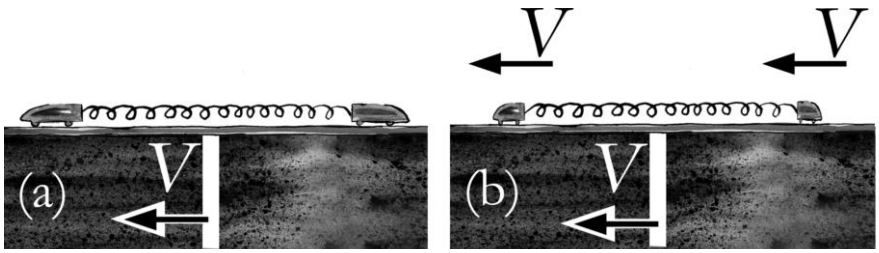


Fig. 5.18 Uniform acceleration in T: (a) right before the acceleration; (b) right after the acceleration.

the same. But since the length in system T had been the *proper* length, both now expect the proper length to be conserved.

However, very soon they notice that there is again something wrong with this result. Take a closer look at what happened. Tom is no longer on the train. The train started to rush from him the moment both engines roared in his system T. The engines instantaneously accelerated the train to the speed of the bridge. They transferred all the train at once from system T to system A. Tom, remaining in T, now observes the train with its unchanged original length 50 m moving together with the bridge whose length in system T is only 2 m. The ratio of these lengths is 25 : 1. Since both objects are now moving as a single whole, any other observer will measure the same ratio for their lengths. Therefore, Alice (who now sees the train and the bridge both at rest in her system A) also must see the train to be 25 times as long as the bridge. Because the proper length of the bridge is 10 m, she now measures the length of the train as 250 m, instead of the expected 50 m!

How can we explain this result? There remains only one possible explanation: this time the train must have undergone *physical deformation* (change in proper length), causing it to stretch! And, what is most surprising, the stretch must by far (by a factor of γ) exceed the original proper length of the train.

Let us try to imagine the same process as seen by Alice. She first sees the contracted 10-m train moving to the right. Then she observes two flashes indicating the ignition of the jet engines. The crucial point here is that these two events are *not simultaneous* in frame A (Fig. 5.19). First the rear engine of the train starts, immediately bringing the rear to a halt. The front car keeps on moving, thus *extending* the train! The engine at the right starts a certain time after it had passed by Alice. When it stopped, the train and the bridge form a single whole; the rear and front of the train are 250 m apart, and positioned symmetrically on either side of the canyon.

This qualitative account by Alice does not explain the actual amount of the stretch. We will outline here a more rigorous treatment using a simplified model mentioned above: two end points instead of a train.

The system A moves relative to system T in the $-x$ direction with a speed V. Let x_1 and x_2 be the coordinates of the left and right ends, respectively, of the train in system T. Tom observes two simultaneous engine flashes at these points. He denotes the times of these events as t_1 and t_2, respectively, and sets $t_1 = t_2 = 0$.

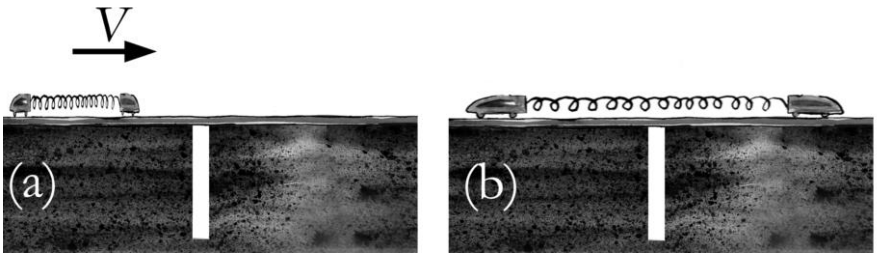


Fig. 5.19 The same process as in Figure 5.18 is, from the viewpoint of Alice, a non-uniform braking. (a) Right before the braking; (b) right after the braking.

Alice observes the same pair of events in her frame A. The space and time coordinates of these events measured by Alice are related to their coordinates x_1, x_2, t_1, t_2 in Tom's system by Lorentz transformations. Alice applies these transformations, and puts into them the zero values for both time coordinates t_1 and t_2. She then takes into account that the positions of the rear and front of the train in system T at the moment $t_1 = t_2 = 0$ are $-x_1 = x_2 = (1/2)\, L_0$. She puts these values into Lorentz transformation, and obtains the expressions for coordinates x'_1, x'_2, t'_1, t'_2 of the same events in her system A in terms of the proper length of the train and relative speed V.

According to her results, the end points of the stopped train are symmetrical with respect to the center of the bridge; they stop at different moments of time $t'_2 > 0 > t'_1$. The length of the stopped train is:

$$\Delta x' = x'_2 - x'_1 = \gamma\,(V)\, L_0 \tag{57}$$

For the time interval between the stops of the rear and the front of the train Alice obtains

$$\Delta t' = \gamma\,(V)\, \frac{L_0 V}{c^2} \tag{58}$$

Inserting the values $L_0 = 50$ m and $\gamma\,(V) = 5$, Alice obtains from Equation (57) $\Delta x' = 250$ m – the number that had been inferred from Tom's observations. For the time interval between the stops of the end points of the train she obtains $\Delta t' = 8 \cdot 10^{-7}$ s. The reader can check these results by carrying out all the calculations that Alice had done.

Now, imagine the train as a row of equidistant cars, each with its own engine, and all the engines fire simultaneously to stop the train at the zero moment in system T. Then Alice would observe a succession of consecutive flashes of the engines, each stopping its respective car, so that the pulse of flashes will run from the rear to the front of the train. The pulse starts at point x'_1 at the moment t'_1 and stops at point x'_2 at the moment t'_2 in Alice's reference frame. Therefore, its speed is

$$u = \frac{\Delta x'}{\Delta t'} = \frac{c^2}{V} \tag{59}$$

Again, the pulse propagates faster than light, but this does not violate any laws, because it is not associated with a signal or energy transfer.

Thus, simultaneous application of forces in T, with the atomic interactions turned off, also fails to preserve the proper length. And again, both observers come, although in different ways, to the same conclusion: in this case, contrary to the previous one, the train undergoes stretch.

How will this result change if there are internal forces? Let the end cars be connected by a spring. Some time after the engines stop (the cars are released), the stretched spring starts to contract. The system begins to vibrate around its equilibrium size that corresponds to the relaxed spring. By definition, the size of the stationary spring in equilibrium is the proper length of the train. In our particular case, the vibrations will not be symmetrical with respect to this size, because the stretch exceeds the proper length (the spring must be very elastic!). However, if the vibrations are damped, the system will ultimately relax to its natural size, thus restoring its proper length.

5.5.3 Non-uniform braking

After the two failed attempts to preserve the proper length of the train in its transition between the two inertial frames, Alice and Tom decide just to stop the ends of the train at the positions, shown in Fig. 5.14 b, where the proper length *is* conserved. As we had initially expected for this case, the rear must stop at the 25-m mark to the left of the center of the bridge, and the front must stop at the 25-m mark to the right of the center. The spacing between the marks will be the needed 50 m. But because the moving train was length-contracted, its ends cannot pass these markings simultaneously. First the rear of the train will pass the –25-m mark, and we should stop it at this very moment. Only *some time later,* the front of the train will pass the +25-m mark, and it must be stopped at this instant.

We now need to find these moments of time. We use again the model, in which the train is represented by its two end points. Tom is moving in the middle. Tom's position is the origin of system T. The center of the bridge is the origin of system A. Both observers set their personal clocks to zero when they pass each other. For the train with the proper length L_0, Alice would want to stop the rear and the front ends at the marks:

$$x_1 = -\frac{1}{2} L_0 , \quad x_2 = \frac{1}{2} L_0 \tag{60}$$

respectively. When the rear point reaches the mark x_1, Alice observes Tom being closer to the center of the bridge by half of the contracted train. The distance between Tom and Alice at this instant is

$$D = \frac{1}{2} L_0 - \frac{1}{2} \frac{L_0}{\gamma(V)} = \frac{1}{2} L_0 \frac{\gamma - 1}{\gamma} \tag{61}$$

The stopping of the rear point at mark x_1 does not affect Tom's motion. He passes by Alice D/V s. later. Since the moment of their passing each other is set to be zero, the moment when the rear end of the rod stops in A must be $-D/V$. Similarly, we find that the front end of the rod must be stopped at the moment D/V. Hence the instants of Alice's time corresponding to the marks x_1 and x_2, are, respectively,

$$t_1 = -\frac{1}{2}\,\frac{L_0}{V}\,\frac{\gamma-1}{\gamma}\,, \quad t_2 = \frac{1}{2}\,\frac{L_0}{V}\,\frac{\gamma-1}{\gamma} \tag{62}$$

If, instead of only two end points, we imagined the train as a chain of cars, each stopped by its individual engine one after another at equidistant moments, starting at t_1 and ending at t_2, then Alice could observe the pulse of engine flashes rushing from rear to front of the stopping train. The edge of the pulse separates two parts of the train – one stopped and the other still moving. The speed of the separation point can be easily found as

$$u = \frac{x_2 - x_1}{t_2 - t_1} = \frac{L_0}{t_2 - t_1} = \frac{V}{1-\gamma^{-1}} = \frac{c^2}{V}\left(1+\gamma^{-1}\right) \tag{63}$$

Again, the separation point moves faster than light. In the end, Alice sees the train stopped in the position depicted in Figure 5.14 b. In the process she observed the train being stretched from its Lorentz-contracted length L_0/γ to its proper length L_0. But she could only achieve this by stopping different parts of the train *at different moments.*

How does this process look from Tom's viewpoint? Applying Lorentz transformations to coordinates x_1, x_2, t_1, t_2 [Eqs. (60) and (62)], we find the coordinates of the same events in T:

$$x'_1 = -\frac{1}{2}\,L_0\,, \quad x'_2 = -\frac{1}{2}\,L_0 \tag{64}$$

and

$$t'_1 = \frac{1}{2}\,\frac{L_0}{V}\,\frac{\gamma-1}{\gamma}\,, \quad t'_2 = -\frac{1}{2}\,\frac{L_0}{V}\,\frac{\gamma-1}{\gamma} \tag{65}$$

Equations (65) give the coordinates of the end points of the train, measured by Tom at the moments when their engines fire. They are just what one would expect. But the corresponding moments of time are, according to Equations (64), observed in T in the reverse order: the moment t'_1 is *later* than the moment t'_2. This is another aspect of relativity of time. We can understand this if we recall the properties of space–time intervals in Minkowski's world (Section 2.9): if the time between two events is less than the time needed for light to travel from one to another (space-like interval!), the events are not in a cause and effect relationship. As is seen from Equation (63), the time between two engine flashes in Alice's reference frame is less than it takes light to travel between them (the pulse of flashes travels from one to another faster

than light.) Therefore, these events are connected by a space-like interval, and the time ordering of the flashes is not invariant.

Our next question is the dynamics of the process observed in system T. Tom sees Alice and the bridge moving to the left. By the time the engine on Tom's right fires, Alice has not yet passed by, because this time is *before* the zero moment. Alice's coordinate at this time is $x_A(t'_2) = -Vt'_2 = (1/2)\,L_0(\gamma - 1)/\gamma$, so the distance between her and the right end of the train is

$$D_R = \frac{1}{2}\,L_0 - x_A\,(t'_2) = \frac{1}{2}\,\frac{L_0}{\gamma} \tag{66}$$

At this moment the right end is accelerated to the speed V, it is "hurled" into Alice's reference frame and starts to move to the left. With the rear of the train still fixed, this means that Tom observes the train shrinking! After Alice had passed by Tom, at the moment t'_1, the left engine fires and hurls the rear of the train to Alice's reference frame. Now Alice's coordinate is $x_A\,(t'_1) = -Vt'_1 = -(1/2)\,L_0\,(\gamma - 1)/\gamma$, so the distance in system T between her and the left end is

$$D_L = x_A\,(t'_1) - x'_1 = x_A\,(t'_1) + \frac{1}{2}\,L_0 = D_R \tag{67}$$

The resulting distance between the edges of the train measured by Tom in the end of the process is $D_L + D_R = L_0/\gamma$. The ends of the train are symmetrical with respect to the center of the bridge. If we imagine again the train as consisting of equidistant cars, accelerating from rest to the speed V one after another, starting from the right, then Tom would observe the compression pulse running down the train from right to left. The velocity of this pulse would be

$$u' = -u \tag{68}$$

Both Alice and Tom observe the same pulse, but to Alice it is an expansion pulse moving to the right whereas to Tom it is a compression pulse moving to the left.

Tom observes that in the end of the process the train has shrunk from its proper length L_0 down to the Lorentz-contracted length L_0/γ. Because the train is now moving with speed V relative to Tom, he concludes that the proper length of the train measured by Alice is the same as the one originally measured by him – it is conserved. The train shrinks in length to retain its proper length! This statement appears no less crazy than those in Sections 5.5.1 and 5.5.2, where the respective observers retained the length of the train in their respective reference frames, only to realize that the train has undergone severe deformation with dramatic change of its proper length. And, just as in previous two cases, our latest "crazy" conclusion is quite rational and self-consistent. That the proper length here remains the same after compression is no contradiction, because the considered process of compression is relative. What is observed as compression by Tom is observed as expansion by Alice! *One and the same system appears to evolve in the opposite directions when viewed from two*

different reference frames. The direction of evolution of an accelerated object can be a relative property. This is due to the fact that the moments of the start and the end of this evolution have opposite ordering in these systems. In the final run, this is another manifestation of relativity of time. Here, as in many other situations, we find the relativity of time at the core of seeming paradoxes in the theory of relativity.

Let us summarize what we have learned from the thought experiments with Alice and Tom.

1. Contrary to a widespread misconception, the special theory of relativity describes accelerated motions in addition to uniform motions.
2. The accelerated motion of an extended system is a subtle process, involving the interplay of various forces, and the outcome depends critically on both the details of the process and the *physical structure* of the object.
3. If one stops a moving object by applying equal braking forces simultaneously to all its moving parts, it tends to compress the object in its original *rest* frame, thus decreasing its proper length.
4. If one accelerates an object by applying equal forces to its parts simultaneously in its *original* rest frame, it tends to stretch the object in its *final* rest frame, thus increasing its proper length.
5. If one stops an object by applying braking forces to its parts at different moments timed so that its proper length remains the same after the stopping, it tends to stretch the object in the reference frame where it had originally moved, and compress it in the reference frame where it had originally rested.
6. The binding *internal forces* within an accelerated object tend ultimately to restore its proper length perturbed by the external forces.
7. Accelerated extended objects may look very different (and sometimes even appear to evolve in the opposite directions) in different reference frames; but each observer has a consistent description, leading to the final state upon which everyone agrees.

5.6 The twin paradox

"Mom, who is that weird gentleman?"

"He is an astronaut, dear. A space traveler. He had visited some distant outer worlds, and now he's back."

"Look, he is behaving so funny!"

"Well, he is a man from the past. Let me see in my Quantum Reference ... Yes, this is the way people normally behaved in the 21st century. At that time a spaceship started from Earth to a star 300 light years away. It took them more than 600 years for a round trip."

"So he must be more than 600 years old?"

"No, dear, he is only 50."

"Are you kidding me?"

"Not at all. His proper age is just 50. Take a closer look at him. He does not look at all old. His biological age measured by his own watch is about 50 years."

"How can that be?"

"You see, when he was traveling, his time was running slower because he was speeding up and slowing down."

In the year 2650 from which I conjure up this possible dialog, the concept of relative time will probably seem to be almost as natural and simple as breathing. Imagine that once space travel is a common occurrence, you could encounter an astronaut who historical records indicate left the Earth 600 years ago. When he returns to the Earth, he is biologically, and by his own clocks, only 20 years older than when he left. So you can see in person real and relatively young people "hurled" into your historical period from the past via relativistic round trips. Then the concept of relative time would be no abstract matter for you; it would be the matter of everyday life experience. I have started this section with the above dialog in an attempt to bring the less sophisticated reader to a better psychological awareness of the paradox to be discussed here: the famous twin paradox. The paradox can be considered as a natural extension of our previous discussions of relativity of time. Remember, we considered the time of a process observed from two different reference frames and found it to be different for each frame despite the symmetry between all inertial frames (Sections 2.5, 2.8, and 5.2). We explained the difference as being due to the asymmetry of the observational procedure. In the twin paradox a situation is considered where the observational asymmetry seems to be eliminated, because it is now one and the same pair of events that is considered in either reference frame: the departure and return of a spaceship with one of the twins aboard. The paradox has been discussed so thoroughly and extensively, and its presentations are so elaborated, that I will just outline it following the way adopted in a good college textbook [24–26]. However, I will emphasize an important point that is either missing or left vague even in some good textbooks. I will also sharpen the paradox by increasing the time discrepancy beyond the biological lifetime for humans.

Consider first the conventional textbook version. Imagine two identical twins, Larry and Joe. Larry leaves the Earth to visit a distant star, while Joe remains at home. After a long journey, Larry returns home and turns out to have aged much less than Joe. This effect is predicted by the theory of relativity, and is confirmed experimentally in a laboratory version of the situation – experiments with a decay rate of μ-mesons moving in a circle in a magnetic field. The effect can be thought of as yet another manifestation of time dilation that we had studied earlier, in Sections 2.5 and 2.8. But now the effect seems to be a real paradox, because of the seemingly symmetrical role of both the twins and their respective observational procedures. According to this symmetry, Larry might claim that he could consider himself stationary and Joe moving together with the Earth first away from and then back towards Larry, in which case the same relativistic equations should give the result of Joe having had aged less than Larry. We thus come to two mutually contradictory statements: by the time of their reunion, Larry turns out to be younger than Joe, and Joe turns out to be younger than Larry. Both statements now refer to one situation when the twins are together in one place in one time in one common reference frame.

To sharpen the paradox, let us assume that the destination of the spaceship is hundreds of light years away. In this case there can be no possibility of a twins' reunion, unless there is some dramatic increase in human longevity, which we do not suppose to happen here. By the time of the ship's return, many centuries may have gone by on Earth and all Larry's contemporaries will have been dead for a long time. But Larry may still be alive! Suppose that the ship's speed is very close to the speed of light. Then Larry's proper time interval for the journey lasting many centuries by the Earth's time may, according to Equation (47) in Section 2.8, be just a few years. In other words, he will be just a few years older when he returns while many hundreds of years will have passed on Earth. He will thus find himself having been "hurled" into the distant future. This is the conclusion given by the special theory of relativity applied to the inertial reference frame associated with our Solar System. Things might look differently from Larry's perspective if he tried to apply the same theory to his ship's reference frame. Then he could claim that he is stationary in his ship while the Solar System with the rest of the Universe is moving relative to him. In which case, according to the same relativistic equations, it is *he* who will long have been dead inside the rusted ship, whereas Joe will be alive and kicking by the time the Earth "returns" to the ship. The paradox is thus restated as follows: when the ship returns, Larry is alive and sees Joe's ancient grave, and Joe is alive and sees Larry's skeleton in the remainder of the ship. The paradox is in that both statements refer to the same situation with the two objects (or subjects) brought together.

We now proceed with the solution of the paradox. First, although Larry *might* consider himself as stationary, his reference frame is *not* symmetrical to that of Joe. It is by no means inertial, because in order to return back to Earth, the ship must change its velocity, that is, it must accelerate. Larry might object that he observed *Joe* accelerated. But this dissention can be easily resolved by asking one simple question: who actually experienced the forces needed to produce the acceleration? Definitely it was not Joe. It was Larry who needed to turn the engines on to accelerate his ship towards its destination, then to decelerate it for landing, then to accelerate it this time back towards the Earth, then to decelerate it again for landing on Earth. The forces exerted on Larry and observed by Joe are real forces having a real source: for example, the jet stream of plasma from the ship's engines. The forces exerted on Joe (and the whole Earth) and observed by Larry are fictitious forces of inertia discussed in the Introduction. Larry cannot identify any real physical body responsible for these forces. These forces are due only to Larry's choice of his ship as a reference frame. From this fact alone, Larry could conclude that his is not an inertial reference frame. Therefore, the symmetry originally assumed to exist between the two reference frames turns out to be an illusion. *The two systems are not equivalent.*

Second, since Larry's reference frame turns out to be non-inertial, Larry must in all his calculations use the equations of the General Theory of Relativity.

Third, the General Theory of Relativity used in Larry's reference frame will give the same result as that of the Special Theory used by Joe: Larry will have aged less than Joe. In case of large time discrepancies, the common answer is: Larry will be alive and Joe will be dead.

Thus both possible approaches, if performed consistently, give results that are in agreement with each other and with the above-mentioned experiment with decaying mesons. Both twins (if they are educated enough) know these results and thus know in advance the only possible outcome of the anticipated space Odyssey: Larry will outlive Joe. They also know that there is no discrimination against Joe in this state of affairs, because Joe's biological life is not in any way physically affected by Larry's journey. Quite the contrary, it is Larry's life that has been so dramatically affected as to be extended into the distant future without the slightest change in his biological life-span. This is just another, and very sound, manifestation of the relativity of time.

We will now turn to a detailed description of the phenomenon and its possible observation. We will consider here an idealized situation in order to see how we can get the same answer for both parties using only the Special Theory of Relativity. Imagine a distant star M 2000 orbited by a planet that, according to recent data, might harbor extraterrestrial life. The star is 300 light years away from our solar system. Larry starts a journey to the star in a spaceship that accelerates practically instantaneously (this is our idealization!) up to 0.999 of the speed of light, then moves uniformly until he reaches the star, then turns also instantaneously and rushes back to Earth with the same speed. As in our previous discussions, we neglect technical details such as motions of the Earth and the other planets, the energy needed to accelerate the spaceship, the rate of energy supply to accelerate it that fast, and so on. These details, important as they are, will not change one essential feature of the anticipated journey: its dramatic effect on Larry's proper time.

We will try to understand the origin of the time discrepancy between the twins' lifespans by considering radio communication between the two reference frames. The Earth and the ship communicate by sending regular radio signals. Suppose they arrange to send each other one signal a year. Then each will know the other's age by just counting the total number of signals received between the moments of departure and return of the ship. Since time is a continuous variable, whereas the signals come in discrete lumps, let us divide each year into 10 equal time intervals, and mark each interval by sending a weaker signal; the number of such signals will give us the corresponding number in the first decimal place after the integral number of years; for instance, if we received five strong signals followed by only two weak signals, we can write the total number of signals as 5.2, which will correspond to 5.2 years of the sender's time. We can follow this procedure to mark ever smaller decimal fractions of the year. Thus the total number of signals in our treatment can be fractional, which enables us to specify time by the number of signals to an arbitrarily high precision.

Let us first evaluate the number of signals from Larry received by Earth. For the speed $v = 0.999\,c$, the round trip to a star 300 light years away will take about 600.6 years. Half of that time, 300.3 years, Larry will be receding from Earth. Owing to the Doppler effect, the signals sent to Earth by Larry while receding will be received on Earth at the lower rate. Applying Equation (22), we find for the low rate 0.0224 signal per year. This succession of low-rate signals will keep arriving on Earth during 600.3 years: 300.3 years of sending such signals plus 300 years it takes the last such signal sent from the destination point to reach the Earth. Multiplying 0.0224 by 600.3 gives

the total of 13.43 low-rate signals. In the remaining 0.3 years until Larry's return, the Earth will receive signals from the approaching ship at a higher rate. Applying again the same equation to the case of the source approaching the detector, we find the rate 44.7 signals per year. The total number of such signals is $44.7 \times 0.3 = 13.4$ signals. Hence the net number of signals received by Earth is $13.4 + 13.4 = 26.8$. Since each signal marks 1 year of Larry's life, we see that Larry has aged by only 26.8 years whereas 600.6 years have gone by on Earth. If Larry was 24 years old at the moment of start, he will be 50.8 years old when he returns. This conclusion is confirmed by compelling experimental evidence: the sight of smiling Larry emerging from the ship – the sight of a man slightly over 50!

Consider now the situation from Larry's perspective. In Larry's reference frame, both the Earth and the star M 2000 are moving with the speed $0.999\,c$ along the line connecting them. The distance between the Earth and the star is correspondingly contracted by a factor $\gamma(v) = 22.3$ and is equal to just 13.4 light years. The total time it takes him to travel this distance back and forth is $2\,(13.4/0.999) = 26.8$ years, in total agreement with the number of signals received from him by the Earth. Now, what is the number of signals from Earth received by Larry during his trip? In the first part of the trip the Earth is receding from Larry, and he receives, according to the same Equation (22), the signals at low rate 0.0224 signal per year. The total number of the low-rate signals received by Larry is $0.0224 \times 13.4 = 0.3$. The moment Larry reaches the star M 2000 and turns back (that is, the moment when the Earth stopped receding and started approaching his ship), he begins to receive signals at a higher rate. The rate is now given by Equation (22) (with changed signs) for the case of the approaching source and is again equal to 44.7 signals per year. The total number of high-rate signals received by Larry is $44.7 \times 13.4 = 600.3$ (one of these signals, say number 96, might have brought him the sad news about his twin brother Joe having passed away). Summing the high- and low-rate signals gives the net number of 600.6 signals received by Larry from Earth. Since each signal marks 1 year of Earth time, Larry knows that 600.6 years have gone by on Earth during his trip. Again, this is in agreement with the time it takes Larry's ship to make it to the star M 2000 and back to Earth.

For the mathematically curious, we will now present the analytical treatment of the problem. Assume our Solar System and another star M to be stationary. Let L_0 be the distance between them measured in their common inertial reference frame S_0. A spaceship starts from Earth at a speed v and moves to M. While moving to M the ship represents inertial system S. Upon reaching M, the ship turns and moves with the same speed but in the opposite direction, and therefore it now represents another inertial reference frame S′. The time it takes for a round trip is

$$T_0 = 2\,\frac{L_0}{v} \tag{69}$$

The proper time of the ship can be found in two different, but equivalent, ways: one can find the contracted distance L between the star M and the Sun, divide this distance by v and double the result for a round trip time:

$$L = \gamma^{-1}(v)\, L_0\,, \quad T = 2\,\frac{\gamma^{-1}(v)\, L_0}{v} \tag{70}$$

Or, one can apply Equation (46) in Section 2.8 for the time dilation to Equation (69):

$$T = \gamma^{-1}(v)\, T_0 = 2\,\frac{\gamma^{-1}(v)\, L_0}{v} \tag{71}$$

Thus simple treatment within the framework of the special theory of relativity gives us the unambiguous result that the proper time of the spaceship, Equation (71), is less than that of the Earth, Equation (69), by a factor $\gamma^{-1}(v)$. This factor can be made arbitrarily small for relativistic motions, so the proper time of the ship can be made arbitrarily small. Some people think that the equations of the Special Theory of Relativity cannot be applied to a spaceship in this kind of problem because the ship's motion is not inertial. This is a common misconception. The Special Theory of Relativity can be applied both to accelerated motions and to uniform motions; the only condition is that the system of reference where this motion is being considered be inertial, if we want a straightforward treatment. Insofar as we consider the ship's motion relative to an inertial reference frame S_0 associated with the Solar System, our treatment is not only valid, but also fairly simple. The difficulties may arise when we try to apply the equations of Special Relativity in the reference frame of the ship, which does not move inertially all the time. In the following treatment, we will avoid these difficulties by applying the equations of Special Relativity to the ship twice: first, when the ship moves uniformly toward the star M (the inertial system S), and second, when the ship moves toward the Earth (the inertial system S′).

With this in mind, we are now going to obtain the same results, Equations (69) and (71), by collecting the counts of radio signals received by Earth's and ship's detectors from their respective senders.

The number of signals from the ship received on Earth is the sum of the low- and high-rate successions of signals. The low-rate signals are from the receding ship and the high-rate signals are from the approaching ship. If the proper frequency is $f_0 = 1$ S/y (signal per year), then the low and high rates are, respectively,

$$f_{\mathrm{L}} = \sqrt{\frac{1-\beta}{1+\beta}}, \quad f_{\mathrm{H}} = \sqrt{\frac{1+\beta}{1-\beta}}, \quad \beta \equiv \frac{v}{c} \tag{72}$$

The time T_{L} on Earth for low-rate signals coming in consists of two intervals: the time $T_1 = L_0/v$ during which these signals are produced by the receding source, and the time $T_2 = L_0/c$ needed for the last of these signals to travel the distance L_0 between the Earth and the source. So $L_{\mathrm{L}} = T_1 + T_2 = (L_0/\beta c)\,(1+\beta)$. The number of low-rate signals is thus

$$N_{\mathrm{L}} = \sqrt{\frac{1-\beta}{1+\beta}} \cdot T_{\mathrm{L}} = \frac{L_0}{\beta c}\,(1+\beta)\sqrt{\frac{1-\beta}{1+\beta}} = \frac{\gamma^{-1}(v)\, L_0}{v} \tag{73}$$

The time T_H of high-rate signals coming to Earth from the ship is just the time interval between the moment of the first such signal's arrival and the moment of the ship's return to Earth:

$$T_H = \frac{L_0}{v} - \frac{L_0}{c} = \frac{L_0}{v}(1-\beta) \tag{74}$$

The corresponding number of high-rate signals is

$$N_H = \sqrt{\frac{1+\beta}{1-\beta}}\, T_H = \frac{\gamma^{-1}(v)\, L_0}{v} \tag{75}$$

The total number of signals from the ship and accordingly, the number of years passed there, is

$$N = T = N_L + N_H = 2\,\frac{\gamma^{-1}(v)\, L_0}{v} \tag{76}$$

This result is identical with Equation (71).

In a similar way we calculate the number of signals from Earth received on the ship. During the first part of the journey, the ship's detector counts low-rate signals; during the second part, which starts immediately after the turn, the ship receives high-rate signals. Both parts of the journey take the same time $\frac{1}{2}\,T = \gamma^{-1}(v)\,\frac{L_0}{v}$. Therefore, the total number of low- and high-rate counts in the ship's detectors is

$$N_0 = T_0 = \frac{\gamma^{-1}(v)\, L_0}{v}\left(\sqrt{\frac{1-\beta}{1+\beta}} + \sqrt{\frac{1+\beta}{1-\beta}}\right) = 2\,\frac{L_0}{v} \tag{77}$$

which is identical with Equation (69). Hence both the theoretical prediction of the time dilation effect and the described thought experiment with signal exchange give the same result: the Earth's history and the ship's history evolving during their separation are characterized by different times. The ship's time is less than the Earth's time. This property of split histories does not depend on the reference frame. The ship's crew observing the history of Earth from their rushing outpost get the same reading T_0 for Earth's time as do historians or physicists on Earth. The latter get the same reading T for the ship's round trip time as do the ship's clocks. And in everybody's account T turns out to be less than T_0 by the same amount. The time discrepancy between the two split and then reunited histories is their common physical characteristics.

There is, however, something missing in this account. It describes *how* all participants of the experiment get a common result for T_0 and then a common result for T. It does not explain *why* the results turn out to be the way they do. The attentive reader may have noticed an asymmetry in the counting procedure that was employed without much comment. When counting low- and high-rate signals coming

to Earth from the ship, we said that low-rate signals kept arriving for much longer time than high-rate signals. When counting the signals coming *to the ship from Earth* we said that both low- and high-rate signals had been received within the two *equal* time intervals. Why is it so?

The observers on Earth remain in place when the ship is on its way to the star M. If we had not calculated the ship's progress in advance, we would not know when the ship reached its destination. Actually, we have no other means to even know whether the ship has reached it at all, except for the radio signals coming from the ship. However, these signals need time to travel from the ship to Earth: zero time for the very first signal (sent at the start), and a very large time L_0/c for the last low-rate signal. The time between the arrivals of the first and the last low-rate signals is thus extended by this time interval L_0/c. When we receive high-rate signals from the returning ship, the last high-rate signal is received immediately (the ship is already here), while the first high rate signal is received only L_0/c years after the moment it was sent. Thus the time between the arrivals of the first and last of the high-rate signals is *shortened* by the same time interval. As you think of it, you find both processes to be just manifestations of the Doppler effect, which works for the first and last signals in a succession in the same way as it does for the two neighboring signals.

The crucial question is: why does the same not apply to the ship's reference frame? There, you remember, we found *the same* time intervals for the successions of low- and high-rate signals from Earth. Does the ship's crew not have to wait until the last low-rate signal and the first high-rate signal from Earth reached the ship? The answer is: it would have to, *had it remained in the same reference frame S*. But as the ship jerks back upon reaching the star M, it is equivalent to jumping from one reference frame S moving away from the Earth to another reference frame S′ moving towards the Earth. The moment the ship performs the jump, its detectors already start rushing towards the Earth (thereby making the Earth rush towards the ship). This *immediately* changes the extended succession of low-rate signals incident on the ship into a compressed succession of high-rate signals. The transition from low- to high-rate signals in the ship's reference frame occurs the moment the ship reaches the star M, whereas on Earth it occurs a long time after this event. You remember, we discussed the difference between Joe's and Larry's motions in our original example? We had emphasized that no real forces act on Joe whereas there are real forces exerted on Larry. It is these forces that cause the ship to jump and thereby change so dramatically its proper time. It is here where the crucial difference between the two systems comes into play.

The whole phenomenon becomes crystal clear if we draw the world lines of both systems (Fig. 5.20) and of the radio signals that they use. We will then see that the world line of the Earth is as straight as a laser beam, whereas the world line of the ship consists of two different segments of equal "lengths." The separation point C on the Earth's world line between the low- and high-rate signals from the ship is shifted from the middle of the line towards the future. There is not much room left for high-rate signals, which results in a relatively small net amount of signals received from the ship. On the ship's world line, the transition point between two successions of signals from Earth is at the vertex. The high-rate signals arrive dur-

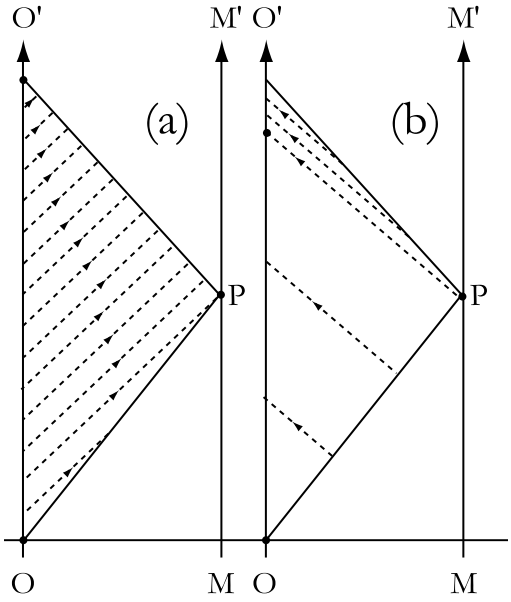


Fig. 5.20 Space–time diagram of Larry's round trip. OO′, the world line of Earth; OPO′, the world line of the spaceship between the same events; MM′, the world line of the planet system M; dashed lines, the world lines of radio signals between the Earth and the ship. (a) Signals from the Earth to the ship; (b) signals from the ship to Earth.

ing the same time as the low-rate signals, which results in a large total number of signals. Accordingly, a large time is recorded for Earth and a small time for the ship by the moment they meet again. Topologically, it is due to the vertex in the ship's world line, that is, to the singular point with an infinite curvature. This is a manifestation of a general law: out of two lines between the two points (events) in Minkowsky's world, the one with larger curvature has the shorter "length." Accordingly, an object moving along the line with a larger curvature has a shorter proper time between the two events.

This stands in a sharp contradiction with what we actually see in Figure 5.20. The path OPO′ is longer, not shorter, than the straight line OO′. It is because we see the *geometrical* path, in whose length the contributions from the horizontal and the vertical components of the segment OP, say, come with the same sign (Pythagorean theorem $OP^2 = OM^2 + MP^2$). The reader should recall that the world line being discussed here is the *kinematic* path in Minkowsky's world. Segment MP represents *time*, not distance. Contributions from the temporal and the spatial components of the interval in Minkowsky's world come, as we have learned in Section 2.9, with the *opposite* signs. The *kinematic* length of the OPO′ is smaller than that of OO′. Because the kinematic length (the space–time interval) is absolute (invariant under Lorentz transformations), this relationship between OPO′ and OO′ is true for *all* observers. Physically, the kinematic length of the OPO′ is Larry's proper time (multiplied by c) during his journey (recall Section 2.9). Thus, the proper time of anyone making a round trip relative to an inertial reference frame is always less than the proper time of the one remaining in place in this system. This statement is the shortest and the clearest explanation of the "twin paradox."

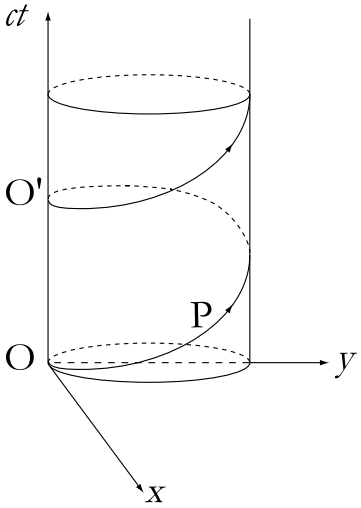


Fig. 5.21 The world lines of mesons A (OO′) and B (OP). Meson B is moving in a circle in plane XY, and meson A rests at a point O on this circle. The kinematic length OPO′ is shorter than the length OO′. This means that the proper time of B between two successive meetings with A is less than that of A. If B is moving fast enough, it will find A to have decayed long before B completes one cycle, while it will take many cycles for B to decay.

We conclude this section by looking back at the dialog at the beginning. This will turn our attention to something that has been half hidden in the intricacies of the above analysis and that actually forms its most essential part: *time travel is possible*. This exciting conclusion follows in the most straightforward way from the comparison of the world lines OO′ and OPO′ in Figure 5.20. The space traveler comes back to find himself in a distant future that is extended far beyond the lifetime of all his former contemporaries on Earth. He thereby becomes a time traveler. He has shifted from his epoch to another one. Such a possibility is not a speculation. It is a scientific prediction confirmed by experiment. The above-mentioned μ-mesons circling in a magnetic field provide a perfect laboratory model of time travel. Suppose you put a stationary μ-meson somewhere in the circular path and call this meson A. Call the circling meson B, and draw a space–time diagram for A and B (Fig. 5.21). Its only difference from the diagram in Figure 5.20 is that the world line of B forms a helix. Because of its curvature, the kinematic length of the helix is less than the world line of A between their two consequtive meetings. Accordingly, the proper time of B is less than that of A. Suppose that A and B are born simultaneously.

But if B is moving fast enough, it makes a full cycle within its lifetime while A will have decayed long before the cycle is completed. We call it the time dilation for B, but actually it is time travel: B finds itself in a distant future; its former "contemporary" had long ago decayed. Time travel is a real physical phenomenon.

The theory of relativity gave more than just a prediction of this phenomenon. It has shown practical means for its realization for humans. The means to travel in time (towards the future!) is to travel in space with a speed close to the speed of light. It is very difficult technically, but possible in principle. A couple decades before the first landing on the Moon only few believed that this would ever happen. Incredible as it may seem, time travel for humans may be just a question of time.

5.7
Circumnavigations with atomic clocks

More than half a century after the birth of the theory of relativity, the manifold and many-times told story about the twin paradox (or, more generally, the clock paradox) took on a new twist. By that time, the advances in experimental physics had led to the development of new types of clock – so-called atomic clocks – whose tickings were periods in radiation emitted by some atoms in the optical transitions between specified states with different, sharply defined energy levels. The precision and accuracy of the atomic clocks by far exceed those of any other clocks based on known macroscopic phenomena – mechanical, electromagnetic, or astronomical. For instance, the reported inherent drift of the hydrogen maser clock is less than 1 part in 10^{14} [27]. This allows us to extend the testing ground of the Special Theory of Relativity into the realm of non-relativistic velocities (see Section 3.3). Using an atomic clock, one can detect tiny changes of its proper time caused by motion with non-relativistic speed, such as that of satellites and even jet planes, by comparing it with the proper time of a similar clock that remained in place. Such an experiment had been suggested, designed, and carried out by Hafele and Keating and the crews of commercial flights in the early 1970s [28, 29]. The experimental scheme was fairly simple. Four cesium beam atomic clocks were flown on regular commercial jet flights around the Earth twice – once eastwards and once westwards. In both cases the clocks' speed was the same – equal to a typical jet speed of about 300 m s^{-1}. After one circumnavigation the flown clocks were compared with identical reference clocks at the US Naval Observatory. The results, within the margin of the experimental and computational errors, confirmed the predicition of the theory for a given situation. Before presenting the results, it is worthwhile to ask: what kind of prediction should we expect for this case?

According to the analysis in the previous section, we would expect that both flying clocks will lose a certain amount of their proper time compared with the reference clock that remained stationary, and inasmuch as they move with equal speeds, the lost amounts must be equal for the east- and westbound clocks.

Now, both statements here are wrong. The experiment showed that not only is there no symmetry in the lost proper time of the flown clocks, but there is even no apparent loss of proper time in one of them. The west-flying clock turned out to have *gained* proper time against the reference clock! These results stand in flat contradiction with the unambiguous prediction of the theory in the previous section.

Therefore, when the results were published, many people considered them as the experimental refutation of the theory of relativity. Here is a typical comment of the opponents of the theory of relativity:

"The theory of relativity clearly predicts that the time dilation should be equal for both flying clocks if they move with equal speeds. But in fact the westward-flying clock showed no time dilation at all – quite the contrary, it *gained* time against the reference clock. So there is a conflict between theoretical prediction and observation. If the observations made by Hafele and Keating are correct, we can consider the theory of relativity to have been empirically refuted."

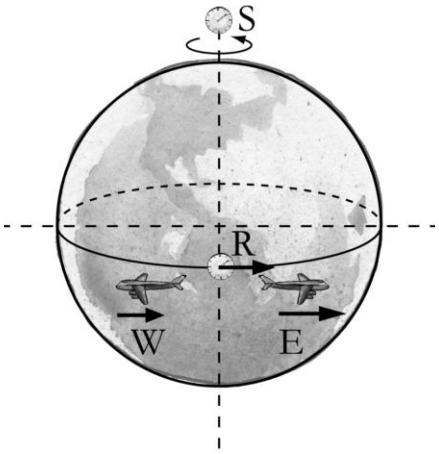


Fig. 5.22 The rotating Earth in the inertial (non-rotating!) reference frame. S, stationary clock; R, reference clock on the Earth's equator; E and W, the east- and west-bound clocks, respectively; arrows represent the speeds of the clocks. To the stationary observer, the west-bound plane appears to fly backwards because a typical speed of a commercial flight is less than the rotational speed of the Earth at the equator.

Now, where is the fallacy of these arguments?

It is in that the reference clock on Earth is confused with the stationary clock in an inertial reference frame. According to the theory, the loss in time, which is independent of the direction of flight, will be observed relative to a stationary clock in an *inertial reference frame*. In the experiment we discuss, the reference clock was stationary *relative to the rotating Earth*, which is by no means an inertial reference frame (Figure 5.22). There is an innate asymmetry between eastward and westward directions in a rotating system, which is imposed by the pre-existing direction of its rotation. This asymmetry can be described in three different but equivalent ways: first, on the qualitative level, one can draw and analyze a free-body space–time diagram for the world lines of all clocks (including the imaginary stationary clock on the extension of the Earth's rotational axis in the inertial reference frame); second, one can consider all the clocks involved (reference clocks included!) as non-stationary clocks moving with different velocities and accelerations in an inertial reference frame, and apply the correlations of the Special Theory of Relativity; third, one can use the known solution to Einstein's equations of the General Theory of Relativity for a gravitating mass representing the Earth, and apply it to clocks moved around the mass in circular orbits. We will use the first two options, and start with the simplest one – the space–time diagram.

To understand things better, consider first a hypothetical situation of a non-rotating Earth. The world lines of all clocks involved are shown in Figure 5.23 a. Here the cylindrical surface represents the world sheet of the Earth's equator. The vertical generatrix OO′ of the cylinder represents the world line of the reference clock. In the absence of rotation the reference clock is stationary and inertial. The two symmetric helixes represent the world lines of the flying clocks. As was the case in the previous sections, these world lines have the same curvature if the clocks fly at the same speed. They must accordingly have a smaller kinematic lengh than the line OO′. Because the kinematic length of a world line represents the proper time of an object tracing out this line, the proper times of the flying clocks after one circumnavigation

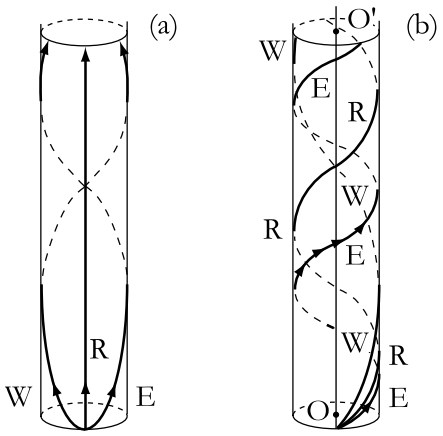


Fig. 5.23 The world lines of all the clocks involved (a) for non-rotating Earth and (b) for rotating Earth. In (b), the E clock is moving faster than the R clock, and its world line is curved more; this clock ages slower than the R clock. The W clock has the least curved world line; this clock ages faster than the R clock. All three moving clocks age faster (to different degrees) than the stationary clock S, whose world line OO′ is straight.

are both less than the proper time of the reference clock. If their speeds are equal, the loss of the proper time is the same for eastward- and westward-flying clocks – in total accord with the prediction of the theory of relativity for the non-rotating Earth.

Now we turn to the real rotating planet Earth. Because of rotation, the reference clock at the US Naval Observatory is no longer an inertial clock! Its world line is accordingly also a helix, and now has to be considered on a par with the world lines of the other two clocks. We can still consider the system of the three clocks within the framework of the Special Theory of Relativity. All we have to do is to compare their readings with those of a stationary clock in an inertial reference frame.

We will then immediately see that the world lines of all clocks fall neatly into two completely different categories (Fig. 5.23 b). The world line of the stationary inertial clock on the Earth's axis will be as straight as a laser beam, whereas the world lines of all other clocks (including the reference clock on the ground) will be twisted and curved into helixes. In terms of time intervals this means that the world line with no curvature corresponds to the maximum possible interval of the proper time between a given pair of events (the start and the finish of the circumnavigation trip, both signaled to the inertial clock); the world lines of the remaining three sets of clocks correspond to smaller proper times, that is, *all three sets of experimental clocks run slower than the stationary clock in the inertial reference frame*, in direct accordance with the prediction of the theory of relativity. Now, among these three experimental sets, the world line of the clock flying westwards has the least curvature, which corresponds to the greatest proper time for this category The world line of the reference clock on the ground has greater curvature (and smaller proper time) because this clock has greater acceleration than the westward-flying clock. Finally, the eastward-flying clock has the greatest curvature and the smallest proper time per circumnavigation. So there *must be an asymmetry* in the time readings in this situation. The actual flying clocks had, of course, traveled at some altitude (more than 10 000 ft) above the reference clock, and had accordingly experienced a different gravitational field strength than the reference clock. The gravitational field slows down the evolution of an ob-

ject. This effect can be rigorously described by the General Theory of Relativity, which lies beyond the scope of this book. The fact that the flying clocks had all been at a higher altitude caused a corresponding increase in their rate, which adds an additional contribution to the asymmetry in changes of proper time. This contribution partially accounts for the fact that the readings of the reference clocks are much closer to those of the eastward-flying clocks than to those of the westward-flying clocks. Even this fact has a full and clear explanation on the quantitative level. Thus the theory of relativity had predicted a new subtle and remarkable phenomenon, that was first pointed out by Hafele and Keating and soon after found experimental confirmation.

Now, for the mathematically trained, we will treat quantitatively the Hafele–Keating experiment. We here present a simplified version of the analysis, considering an experimental model in which all sets of clocks had all the time been at the same altitude.

The Special Theory of Relativity is equally well equipped for considering *all* possible kinds of motion – uniform or accelerated, rectilinear or in arbitrarily curved lines, subluminal or superluminal. The calculations are most straightforward if the motion to be discussed is considered relative to an inertial reference frame. This is precisely what Hafele and Keating did in their papers [28, 29]. The non-rotating system attached to the center of the Earth *is* inertial to a high degree of accuracy, because the centripetal acceleration associated with the Earth's orbital motion around the Sun is about 6×10^{-3} m s^{-2}, that is, less than one thousandth of the acceleration due to gravity on Earth. Also if, in addition, we take into account that this system is in a state of a free fall in the gravity field of the Sun, then, according to Einstein's principle of equivalence (which had been shown experimentally to be correct to an accuracy of 10^{-12} [1]), this system is inertial for all practical purposes.

We said that the situation at hand is non-relativistic in that all the clocks involved, as well as all the points on the rotating Earth, move much slower than light. But even in a non-relativistic situation, we can come to fundamentally different conclusions depending on whether we treat the problem according to Newtonian mechanics or according to relativistic mechanics. The motions in a rotational system that we are considering provide a good illustration. To see it, let us first find the results in a Newtonian approximation.

Place the reference clock at some spot on the Earth's equator. Let Ω be the angular velocity of the Earth's rotation and R be its equatorial radius. Relative to the inertial (non-rotating!) frame, the reference clock moves with a speed $v_R = \Omega R$. Now suppose two planes with the clocks depart simultaneously from the same spot and with the same speed v, one eastwards (clock E) and the other westwards (clock W). Relative to the inertial frame, the planes move with the speeds

$$v_E = v_R + v, \quad v_W = v_R - v \tag{78}$$

Now find the time t_E for one circumnavigation of the E-clock. To do it, we notice that *relative to the stationary frame* the E-clock makes precisely one rotation more than the reference clock (if you outrun your competitor on a racetrack, next time you catch up

with him you'll make one round more than him). Therefore the distance $v_E t_E$ traveled by the E-clock will be one equator longer than that $v_R t_E$ of the reference clock by the time they meet again:

$$(v_R + v)\, t_E = 2\,\pi R + v_R\, t_E \tag{79}$$

It follows that

$$t_E = \frac{2\,\pi R}{v} \tag{80}$$

The same reasoning applied to clock W shows that since it is moving slower than the reference clock relative to the inertial reference frame, by the time they meet again it will have traveled one equatorial length less than the reference clock, so we can write

$$(v_R - v)\, t_W = v_R\, t_W - 2\,\pi R\,, \quad t_W = \frac{2\,\pi R}{v} = t_E \tag{81}$$

The travel times are the same for both type of clocks if they move relative to Earth with the same speed.

The Earth-based observer can comment with a smile: "You could have saved your time if you ask me. I would have told the result to you immediately without any calculations. In Newtonian physics the time measured in a reference frame is the same in all other reference frames, therefore the times in Equations (80) and (81) found for the inertial frame must by definition be the same as the time of the Earth-based observer. Such an observer would obtain this result immediately without "looking out" beyond the Earth."

I agree. And now, before doing the *relativistic* treatment of the same problem, I want to follow this advice and ask: what outcome would one expect in this case?

We know that relativistic mechanics is more subtle than Newtonian mechanics, but *in this case* the result seems obvious. If I sit in a spot somewhere on the equator and dispatch two planes, one eastwards and one westwards, and the planes move with the same speed relative to me, they will surely circumnavigate the Earth in the same time by my clock, and after circumnavigation they will arrive simultaneously.

Well, this is wrong! We will now show this rigorously.

Let me remind again, that in Special Relativity, if we want to obtain the correct results in the most straightforward way, we need to consider the events relative to an *inertial* reference frame. So, we must first find the speeds of the flying clocks *relative to the non-rotating frame attached to the center of the Earth*. For this frame, we have, by applying, instead of Equation (78), the *relativistic* rule for addition of velocities (Section 3.1)

$$v_E = \frac{v_R + v}{1 + \dfrac{v_R\, v}{c^2}}\,, \quad v_W = \frac{v_R - v}{1 - \dfrac{v_R\, v}{c^2}} \tag{82}$$

Accordingly, Equations (79) and (80) for the travel times measured by the stationary (inertial!) observer will now take the form

$$\frac{v_R + v}{1 + \frac{v_R\, v}{c^2}}\, t_E = v_R\, t_E + 2\,\pi R\,; \quad \frac{v_R - v}{1 - \frac{v_R\, v}{c^2}}\, t_W = v_R\, t_W - 2\,\pi R \tag{83}$$

Solve these equations for t_E and t_W:

$$t_E = \frac{2\,\pi R}{v}\,\frac{1 + \frac{v_R\, v}{c^2}}{1 - \frac{v_R^2}{c^2}}\,, \quad t_W = \frac{2\,\pi R}{v}\,\frac{1 - \frac{v_R\, v}{c^2}}{1 - \frac{v_R^2}{c^2}} \tag{84}$$

Now we can find corresponding proper times for the two circumnavigations. According to our general result in Section 2.8, the proper time of a moving clock is related to the coordinate time read by the inertial system of clocks by the corresponding Lorentz factor depending on the clock's speed.

The speed of the reference clock is v_R. Therefore its proper times $\tau_R^{(E)}$ and $\tau_R^{(W)}$ are given by

$$\tau_R^{(E)} = t_E/\gamma\,(v_R) = t_E\,\sqrt{1 - \frac{v_R^2}{c^2}}\,; \quad \tau_R^{(W)} = t_W/\gamma\,(v_R) = t_W\,\sqrt{1 - \frac{v_R^2}{c^2}} \tag{85}$$

The travel times are different!

Where does this difference come from? The times were equal when we considered the problem in the framework of Newtonian mechanics and derived Equations (78)–(81). It is only because we have switched from the Newtonian rule in Equation (78) of addition of velocities to the relativistic rule in Equation (82) that the difference in times crept in. And the reason is that by rule (78) the velocities of the flying clocks remain symmetrical with respect to the speed v_R (one greater and the other smaller than v_R by the same amount), whereas by rule (82) this symmetry is broken. We see that the resulting difference in the travel times is a *purely relativistic effect*.

We will accordingly treat these two cases separately.

Case E: the coordinate time for one circumnavigation in the eastward direction is given by the first Equation (84). The speed of the E-clock is v_E. Therefore, *its* proper time τ_E is

$$\tau_E = t_E/\gamma\,(v_E) = t_E\,\sqrt{1 - \frac{v_E^2}{c^2}} \tag{86}$$

Now we are able to estimate the *experimentally observed quantity* – the relative time offset between the E-clock and the reference clock:

$$\delta_E \equiv \frac{\tau_E - \tau_R^{(E)}}{\tau_R^{(E)}} = \frac{\tau_E}{\tau_R^{(E)}} - 1 \tag{87}$$

Using the two previous equations and retaining only the terms of the order of c^{-2}, we obtain (you have to do some calculus to check it)

$$\delta_{\mathrm{E}} = -\frac{v\,(2\,v_{\mathrm{R}} + v)}{2\,c^2} \tag{88}$$

Case W: the coordinate time of one circumnavigation in the westward direction is given by the second Equation (84). Note that this equation, and also all other expressions for the W-clock, can be obtained from the corresponding equations for the E-clock by just changing the sign of v. Therefore, we can obtain the offset for the W-clock directly from Equation (88) as

$$\delta_{\mathrm{W}} = \frac{v\,(2\,v_{\mathrm{R}} - v)}{2\,c^2} \tag{89}$$

The offsets are different! For the E-clock the offset is negative, that is, it loses a certain fraction of its proper time with respect to the proper time of the reference clock. We can easily explain this by noting that this clock moves relative to the stationary frame faster than the reference clock, and therefore "ticks" slower. It turns out to have aged less when the clocks meet again.

For the W-clock, the sign of the offset depends on its speed. If the speed is greater than $2\,v_{\mathrm{R}}$, the offset is also negative, and this clock will also have aged less than the reference clock when they meet again. We can explain this if we note that at a speed exceding $2\,v_{\mathrm{R}}$ the W-clock is moving westwards *faster* relative to the stationary frame than the reference clock is being carried eastwards by the Earth's rotation. Again, the clock that moves faster ages slower. If the speed v is much larger than $2\,v_{\mathrm{R}}$, so that we can neglect the rotational motion, then Equations (88) and (89) tell us just what one would expect for this case – that both clocks will have aged less than the reference clock by approximately the same amount when they meet again. However, if the speed v is less than $2\,v_{\mathrm{R}}$, then the W-clock moves slower relative to the stationary system than the reference clock, and will after one circumnavigation have aged more than the reference clock. This will produce a positive offset.

(At speeds somewhere between v_{R} and $2\,v_{\mathrm{R}}$, the W-clock is moving westwards relative to the stationary frame. If the speed v of the W-clock is equal to v_{R}, it "cancels" the effect of the Eath's rotation, and the W-clock stands still in the stationary reference frame. In this case it has the maximum aging rate. If its speed is less than v_{R}, its is moving eastwards relative to the stationary frame.)

For typical conditions of international jet flights (altitude about 10 km and a speed $v \cong 300$ m s^{-1}) the theoretical prediction (including the above-mentioned contribution from the gravitational effect) was that the E-clock should have lost 40 ± 23 ns, and the W-clock should have gained 275 ± 21 ns compared with the reference clock. The experimental data showed that the E-clock actually lost 59 ± 10 ns and the W-clock gained 273 ± 7 n.

We see that the experimental results stand in quantitative agreement with the predictions of the theory of relativity if we take proper account of the motion of the refer-

ence clock. As Hafele and Keating put it, "These results provide an unambiguous empirical resolution of the famous clock 'paradox' with macroscopic clocks." So if we believe in the observations made by Hafele and Keating, we can then believe in Einstein's theory of relativity even more firmly then ever before.

And yet there remains one point mentioned above in passing [in comment to Eqs. (82) –(84)] which seems to constitute a real paradox. Here it is. We have proved that the two objects launched with *the same* speed in opposite directions must have different circumnavigation times. But this implies that they must have had *different* speeds! Indeed, imagine that you are sitting somewhere on the equator with the reference clock. At a certain moment of time (call it the zero moment) you record two planes departing with equal speeds, one eastwards and one westwards. If they maintain equal speeds during the whole flight, then how can it be that they return after circumnavigation at different times by your clock? Restate the argument. Suppose that the two planes take off simultaneously, fly around the Earth in opposite directions, and return back to the same airport at different times. Will you say that the planes had the same speed? The likely answer is: not unless I am crazy. And you confirm your answer with a simple calculation. If the radius of the Earth is R, then the length of the Earth's equator measured by an inertial observer is $L = 2\pi R$. Its proper length measured by people on Earth is $\Lambda = \gamma(\Omega R)\, L = 2\pi R \gamma(\Omega R)$. The speed of the E-clock is $\Lambda/\tau_{\mathrm{R}}^{(\mathrm{E})}$ and the speed of the W-clock is $\Lambda/\tau_{\mathrm{R}}^{(\mathrm{W})}$. Using Equations (84), (85), we easily obtain

$$v_{\mathrm{E}} = \frac{v}{1 + \dfrac{v_{\mathrm{R}}\, v}{c^2}}\,, \quad v_{\mathrm{W}} = \frac{v}{1 - \dfrac{v_{\mathrm{R}}\, v}{c^2}} \tag{90}$$

Thus, if the travel times are not equal, we obtain travel speeds that are *not equal*. This result appeals to our common sense but apparently contradicts the initial condition, according to which the speeds of both flying clocks relative to Earth are *the same*. The statement about the equal speeds resulting from unequal travel times over the same distance seems absolutely crazy. The equivalent statement that it takes different times for planes with the equal speeds to fly around the Earth must therefore also be crazy. As we mentioned in the comments to Equation (84), this result only appears when we switch from Newtonian mechanics to relativistic mechanics. Now we obtained Equations (90) also using relativistic mechanics. Does this mean that the theory of relativity contains contradicitions after all? If so, this would be the death sentence for the theory. Or does this mean that all our calculations have been fundamentally flawed? The whole situation appears to be not even a paradox, but just a huge nonsense. How can one reasonably explain this nonsense?

We will look for the answer in the next section.

5.8
Photon races in a centrifuge

Consider a disk with a circular path of length L around its center O. Pick a point P on the path and place two detectors there.

Suppose we launch two photons (considered as localized particles) from this point in opposite directions along the path. Each photon will circumnavigate the path and come back to P where it will be detected. Since both photons travel the same distance, and the speed of light does not depend on direction, both photons will return at the same time $T_0 = L/c$, and the detectors will fire simultaneously.

Now complicate the matter. Bring the whole platform with the pathway L into rapid rotation about its center O. Call it an optical centrifuge. Again launch two photons from P in two opposite directions along the path and wait for their return back to P after one circumnavigation around the path (Fig. 5.24). What would you expect to see? Will the two photons return to P simultaneously or not?

The answer to this question has been known for a long time, after the French physicist Sagnac had carried out the corresponding experiment [30]. The two photons no longer arrive back at the same time: the one launched in the direction of rotation arrives at P later than the one launched in the opposite direction.

How can we explain this result in terms of what we have learned?

According to the theory of relativity, the speed of light is a universal constant. It is the same in all directions and does not depend on the motion of its source. Therefore, each photon will start from P with the same speed c independent of direction even though the source P is moving itself. If we mark the position of P at the moment of the emission, both photons will return back to this position simultaneously. But the emitter itself (carrying also the detectors) is no longer there – it has progressed to another position P′ due to the rotation of the platform (Fig. 5.24 b). The photon that had been launched in the same direction has not yet reached this point; the photon launched in the opposite direction, however, has already passed this point and had accordingly been detected. Thus, the photons cannot arrive back at the emitter at the same time. Once the platform starts to rotate, the clockwise and counterclockwise motions in a circular path are no longer equivalent.

Just as we did in the previous section, we can calculate the arrival times, and, as before, it is easier to do it from the viewpoint of the stationary observer. Look at Figure

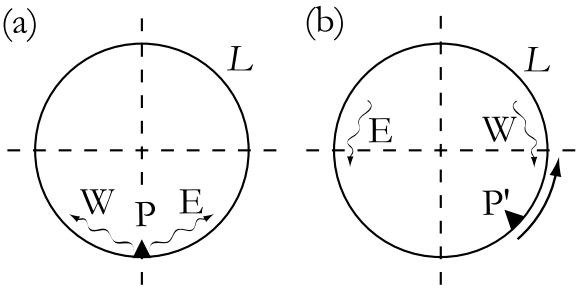


Fig. 5.24 A circular race-track (optical centrifuge) with a photon emitter and two detectors. The emitter launches two photons in the opposite directions ("east" and "west"), while the detectors record their return time. The photons do not complete their respective cycles simultaneously.

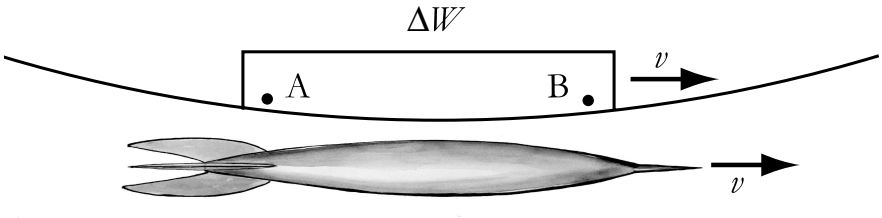


Fig. 5.25

5.24. As usual, we introduce some notations. Let R stand for the radius of the circular path, and let Ω be the angular velocity of the platform. Call the photon launched from P in the direction of the platform's rotation the east photon. The photon launched in the opposite direction will be called the west photon.

As can be seen from Figure 5.24, the west photon will meet with the detector again when it is at position P′. The east photon will hit the detector at a later time, when the detector will be at a position P″. The same treatment as in the previous section gives for the arrival times measured by a stationary observer

$$T_1 = \frac{2\pi R/c}{1+\dfrac{\Omega R}{c}} = \frac{T_0}{1+\dfrac{\Omega R}{c}}\,, \quad T_2 = \frac{2\pi R/c}{1-\dfrac{\Omega R}{c}} = \frac{T_0}{1-\dfrac{\Omega R}{c}} \tag{91}$$

where T_0 is the travel time along the loop for the stationary platform.

We see that $T_1 < T_0$, which is no surprise, as the west photon has traveled the shorter length between its emission and absorption. The time T_2, on the other hand, is greater than T_0, because the east photon has to travel a longer distance in the stationary system between its emission and absorption.

Consider now the same process from the viewpoint of the disk inhabitant Paul. He is sitting near P, and watches the motion of photons in his frame. Both photons travel the same distance along the circular loop. The length of the loop measured by Paul is its proper length given by

$$\Lambda = L\gamma(v) = 2\pi R/\sqrt{1-\frac{\Omega^2 R^2}{c^2}} \tag{92}$$

and the proper time of his clock is related to the corresponding time T in the stationary system by $\tau = T/\gamma(\Omega R)$, so that $\tau_1 = T_1/\gamma(\Omega R)$, $\tau_2 = T_2/\gamma(\Omega R)$. Thus, the clock at the detector measures two different circumnavigation times for the east and west photons.

Now, in the spirit of the concluding remark in the previous section, Paul can define the average speed of the photon circumnavigating the platform as the ratio Λ/τ. Applying this definition to our case and using Equations (91) and (92) gives

$$c_1 = \frac{\Lambda}{\tau_1} = \frac{L/T_0}{1-\dfrac{\Omega R}{c}}\,; \quad c_2 = \frac{\Lambda}{\tau_2} = \frac{L/T_0}{1+\dfrac{\Omega R}{c}} \tag{93}$$

But the ratio L/T_0 is local speed of light. Therefore

$$c_1 = \frac{c}{1 - \dfrac{\Omega R}{c}}\,; \quad c_2 = \frac{c}{1 + \dfrac{\Omega R}{c}} \tag{94}$$

Equations (94) can also be obtained directly from the general expressions (90) as a special case when $v = c$.

The photon speeds [Equations (94)] found by Paul from his measurements of their travel times are different from the speed of light! The west photon appears to move faster than light; if the rotation of the disk is rapid enough, such a photon travels arbitrarily fast. The east photon, on the other hand, appears to move slower than light, and on a sufficiently rapidly rotating disk it can slow down to half the speed of light.

So far these results pertain to the average speed of the photons. But we apparently cannot avoid the conclusion that the same must be also true for the local speed: it must be smaller than c if the light moves in the direction of rotation of the disk, and greater than c if light moves in the opposite direction. The results of the Sagnac experiment seem to provide unambiguous proof of this. Here is the argument presented by Paul.

"One of the goals of physics is to draw meaningful conclusions from experiments. As a platform-based observer I do not have to think much about its rotation and how it may affect the phenomena. I have to do the measurements. If the measurements tell me that both photons arrive back at P simultaneously, I conclude that the light propagates in two opposite directions with the same speed. If the oppositely traveling photons that had been emitted simultaneously do not return to me simultaneously, I have to conclude that they travel at different speeds. Now, I did perform the experiment and I see that when there is no rotation, the photons that were emitted simultaneously in the two opposite directions return simultaneously. I accordingly interpret this as another confirmation of Einstein's postulate about the constancy of the speed of light. However, when I repeat the experiment during rotation of the disk, the photons do not return simultaneously. The only conclusion I can draw from this is that the speed of light in a rotating system is different in different directions. And this must be true not only for the average speed, but also for local speed in any location."

This argument by Paul seems very strong indeed: since the conditions are the same at any point along the photon's circular path around the center, the average velocity along this path must be equal to velocity measured locally. Therefore, the Sagnac experiment can be considered as a direct measurement of the local speed of light separately for one and the other direction.

However, the last conclusion would be wrong. It does not follow from the Sagnac experiment! This experiment by itself only shows that apart from the local speed c, which describes the rate of photon motion from one point to another, one can also introduce two other speeds characterizing *complete cycles* of clockwise and counterclockwise motions of light around the center of a rotating system. These new speeds are different from the local speed c because the procedure of their measurement is

different from the local measurement. To show the difference, I will first present a purely physical argument in the form of a thought experiment, and then the arguments based on the definition of measurement and on the concept of time in rotating systems.

Consider on the rim of our disk an element ΔW with the segment of arch AB (Fig. 5.25). According to Paul's interpretation of Equations (94), the local speeds of light in the directions A → B and A ← B differ from one another. Let now the radius of the disk increase and its angular velocity decrease in such a way that the product $\Omega \cdot R$ and thereby the speeds c_1 and c_2 remain constant. The reference frame formed by the local region ΔW with ever increasing accuracy approximates the inertial reference frame in the limit $R \to \infty$, $\Omega \to 0$ (one can make sure of it by noticing that centripetal acceleration $a = \Omega^2 R = (\Omega R)\,\Omega = \text{const} \cdot \Omega$ of the element ΔW under given condition goes to zero). It follows then that Paul's conclusion about different local speeds of light in different directions has to be true for this inertial reference frame. And owing to the principle of relativity, it then has to be true in general!

To see the physical consequences of this, let us continue our thought experiment. In Figure 5.25 you see a spaceship traveling alongside with ΔW with the same speed v as the rim of the disk, that is $v = \Omega R$. Assume that $\Omega R = 0.99999999\,c$, so that c_1 is nearly infinite, and c_2 is nearly half of c. The spaceship moves in a straight line and therefore represents an inertial reference frame. The disk radius R is so huge that even the region ΔW that is small with respect to the whole disk is still huge by our standards. It is so huge that our spaceship is co-moving together with it for a considerable time required to perform our experiment, and all this time they practically touch each other. A crew member, Sam, who is an old friend of Paul, comes out to meet his friend. They greet each other and start to discuss problems of common interest. As they do so, they run together on cosmic roller-blades specially designed for space strolls and able to accelerate to nearly the speed of light. They are running shoulder by shoulder – Sam on a sideboard of his spaceship, and Paul along the rim of his disk – you remember, their platforms nearly touch each other. Suppose they start running in the direction from A to B. Nothing seems to be in the way of this run. For instance, Sam can step on Paul's platform without even noticing it, since the spaceship and the region ΔW are to the highest accuracy at rest relative to each other. But the moment Sam's roller-blades speed him up to half the speed of light, Paul yells:

"Hey, slow down, I cannot accompany you any further!"

"Why?"

"Because I cannot outrun light, and it propagates with the speed of only $0.5000001\,c$ in this direction."

"Come on, I also see this beam – it behaves quite normally."

"Yes, in your reference frame, but things are different on my disk."

"What are you talking about? There is no difference between our systems! They are stationary relative to each other. For all practical purposes they form one common reference frame. I do not know what is going on elsewhere on your disk, but this spot does not in any way differ from my spaceship. We are in the same conditions. But if you mind going further this way, let us turn back."

They turn back, accelerate up to nearly the speed of light c, and now Sam yells at Paul:

"Hey, slow down, I cannot follow you further!"

"What's wrong?"

"I cannot move faster than light, my blades are at their limit!"

"Come on, we are just crawling like the wretched photons in your spaceship, but here the speed of light is $10^7\,c$, we can frolic to our hearts' content!"

And Paul disappears from Sam's view – he crosses the light barrier without violating anything: even though he now runs much faster than c, it is still less than his limit of $10^7\,c$!

Now, if Paul is right in his conclusions, the above scene could in principle be possible. But this would flatly contradict the laws of Nature in the domain where they are established beyond any doubt. We have to conclude that we *cannot* identify the speeds c_1 and c_2 with the local speed of light, which remains c regardless of direction.

What, then, is the difference between them?

The best answer to this question would probably be the inspection of corresponding space–time diagrams. To measure the *local* speed, we mark, as usual, two events M and N along the trajectory of an object – its departure from one point and arrival at another point. Since each point is equipped with its own clock, the local speed of light is determined from the segment of the photon world line between the two world lines of these clocks (Fig. 5.26 a). To measure the speed c_1 or c_2 in the Sagnac experiment, we wait for the photon to return to *the same* clock after one circumnavigation. The resulting speed is determined in this case from the segment of the world line of this clock within one pitch of the helix representing the photon's world line (Fig. 5.26 b). In the more familiar terms of three-dimensional space, the local speed of light is measured along a small segment with two end points, whereas the speeds c_1 and c_2 are measured from two consecutive readings of only one clock on the closed loop. The fundamental difference between these two procedures is seen, for instance, from the fact that the former admits also the speed measurement for the "geodesic" photon moving in a straight line tangential to the rim of the disk, whereas the latter is unable to do this in principle. It is not surprising that the two fundamentally different procedures of measurement give generally different results – they actually measure two different physical characteristics. The speeds c_1 and c_2 in the Sagnac experiment have nothing to do with local speed of light.

From Figure 5.26 we can also clearly see why the "global" speed $c_{1,2}$ turns out to be double-valued. If the photons were moving along the rim of a *stationary* disk, the world line of the recording clock would be the vertical line OO′ – the generatrix of the cylindrical surface representing "the world pipe" of the rim. In this case the intersection points of OO′ with the world lines of the east and west photons coincide, which would yield the same travel time for both photons. In case of rotation, the helixes representing the world lines of the photons do not change because the local speed of light does not change. The world line of the clock, however, is now also twisted into a helix, and intersects with the photons' lines already at different points, which produces the difference between the travel times and thereby the different "global" speeds c_1, c_2.

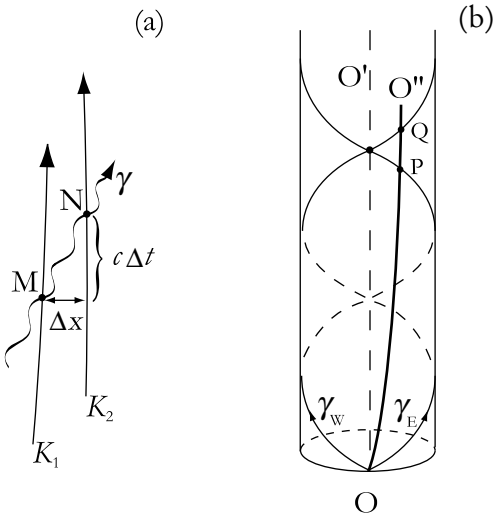


Fig. 5.26 The diagrams for local and "global" speed measurement on a closed race-track. (a) *Local* speed of an object can be measured as a slope of its world line MN between the world lines K_1 and K_2 of the two measuring clocks Δx m apart. (b) *Global* speed of the object (photons included) is determined by the world line of a *single* measuring clock between the departure and arrival of the object, and the circumference of the track. If the centrifuge is stationary, the clock's world line is vertical OO′, and the segment between the departure and arrival of the object with constant local speed will be the same for the east and west directions. However, if the centrifuge is rotating, the world line of the measuring clock is twisted (OO″) and accordingly intersects with the world lines of the east and west photons at *different* points (P and Q). This corresponds to different arrival times and thereby different global speeds of the two photons.

Therefore, if Paul asks: "Why does it take different times for a photon to travel along the same circle in opposite directions?" the answer would be: first, because the world line of the measuring device (the clock P) is twisted by the rotation of the disk; and second, this twist is mismatched with those of the photon lines. Were the photon lines affected by the rotation in the same way, as is the line of the clock, they would remain symmetrical *with respect to this line*, and all three would again intersect at the same point. This is precisely what happens in non-relativistic mechanics when we use the Newtonian law of addition of velocities. Recall our non-relativistic treatment of the circumnavigating clocks in the previous section, Equations (78)–(81). In that case, you remember, we obtained the same value for the travel times in the east and west directions (and accordingly for the local and global speeds). It was not for nothing, after all, that we had derived apparently unnecessary non-relativistic Equations (78)–(81). They now clearly illustrate that had the photons behaved in the same non-relativistic fashion, they would not display any time discrepancy either. Because the Sagnac experiment *does* record the discrepancy, it can be considered as yet further evidence of the relativistic nature of photons; and the measured magnitude of this discrepancy corresponds to the relativistic limit, when the photon speed added with any other speed (for instance that of the disk's rim) remains unchanged. In other words, the Sagnac experiment, which on the face of it appears to disprove Einstein's assertion about the invariance of the speed of light, gives an additional proof that the *local* speed of light is the same in *any* reference frame, rotating included!

Reversing this argument, we can obtain yet another perspective of the Sagnac experiment: the appearance of two additional speeds c_1 and c_2 indicates the rotation of the system. Thus, not only can we detect such rotation mechanically without "looking out" (recall the Introduction!), we can also do it optically. From this viewpoint, the Sagnac experiment can be considered as an optical analogue of Faucault's famous experiment with a pendulum [31].

The above analysis pertains not only to the speed of light – it has a general character. As another example, we illustrate the difference between the local and "global" speeds for the case opposite to that of light: the stationary particle. So, imagine a particle stationary in K and sitting close to the rim of the rotating disk. Its local speed relative to Paul is $v_L = \Omega R$. Let us now try to find this speed using the procedure analogous to Sagnac's experiment, that is, divide the circumference length $\Lambda = 2\pi R\gamma(\Omega R)$ of the circle in K′ by the time interval $\tau = T/\gamma(\Omega R)$ on the clock P between its two consecutive meetings with the particle (T is the rotation period of the disk). The result will be

$$v_G \equiv \frac{\Lambda}{\tau} = \frac{2\pi R\gamma(\Omega R)}{T/\gamma(\Omega R)} = \frac{2\pi R}{T}\gamma^2(\Omega R) = \frac{v_L}{1-\dfrac{\Omega^2 R^2}{c^2}} \neq v_L \tag{95}$$

This result is a special case of the general Equations (90) in the previous section, when the west-bound object moves relative to a rotating system with speed $v = v_R = \Omega R$ and therefore remains still in the stationary inertial system.

Suppose that we are unaware of the above analysis and naively believe that the result in Equation (95) yields the local speed of the particle. But upon closer inspection we will realize that the speed v_G has nothing to do with the local speed, and at sufficiently large ΩR it can become infinite – and this for a particle resting in K!

The speeds v_L and v_G in this example, as in the previous ones, have different physical meanings (Fig. 5.27). The local speed is determined from the readings of the two synchronized clocks in K′ at the end points of the segment MN of the world line of the particle. The global speed is determined by the proper time of the one pitch of the helix representing the world line of one clock K′. There is no reason to expect any equivalence between these two essentially different experimental procedures.

The following analysis of connections between the local and global speeds leads to another astounding discovery in the rotating wonder-world: there is no such thing there as one time for the whole space – even in one reference frame.

Let us apply the measurement procedure for the local speed of light to our case. The segment of the photon trajectory is along the circle of radius R (Fig. 5.28). In an inertial (non-rotating) system K associated with the center of the disk, the spatial separation between the events is dl, and the time interval between them is dt. Since the events are on the photon's world line, dl/d$t = \pm c$, depending on the direction of the photon's motion along the segment dl. The + and – signs relate to this direction. The speed itself is c regardless of direction. Now, perform the same measurement in the role of Paul – the observer on the rotating disk. If Paul measures the local speed, the procedure that he uses has nothing to do with the Sagnac experiment. He also marks

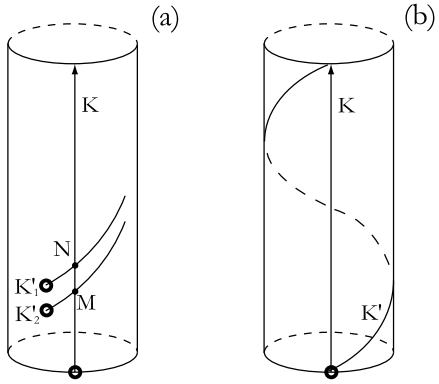


Fig. 5.27 The local and global speeds of a particle, measured in a rotating system K′ in a special case when the particle sits still in the stationary system. (a) The local speed is determined by the difference in readings of *pair* of clocks K'_1 and K'_2 at the intersections M and N with the world line of the particle, and the spatial separation between these clocks in K′. (b) The global speed is determined by the kinematic length of the world line of a single clock K′ and the circumference of the track.

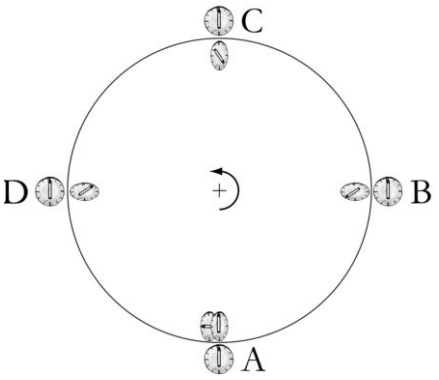


Fig. 5.28 The system of clocks on a rotating disk. If we synchronize the clocks, going step by step around the circle, then we realize that it is impossible to assign only *one* moment of time to an event. A, B, C, D: stationary clocks.

the end points of the directed segment dl' and the moments of the time interval dt' between the same events, and then calculates the ratio dl'/dt'. The result is given by the Lorentz transformation:

$$c' = \frac{dl'}{dt'} = \frac{dl - \Omega R\, dt}{dt - \dfrac{\Omega R}{c^2}\, dl} \tag{96}$$

Since $dl/dt = \pm c$, the last equation reduces to $c' = \pm c$.

How does this relate to the Sagnac experiment? Note that all the conditions at all points of the considered circular path are the same. Therefore, for each pair of close points along the path we can use the same inertial reference frame K, only at each locality its spatial axis along the direction of the motion of the photon at this locality is slightly tilted with respect to that of neighboring locality. Because the Lorentz transformations are linear and contain in our case the same parameter ΩR, their times dt' and distances dl' just add up algebraically, so we can apply the previous equation

to an arch of finite length l' and to corresponding finite travel time t' and obtain the same result

$$c' = \frac{l'}{t'} = \pm c \tag{97}$$

At each stage of this process we can use the "integral" Equation (97), which will always give $|c'| = c$ for both directions of the photon motion.

The situation changes radically when we apply Equation (97) to the whole rim of the disk so that the end points of the arch merge together at the opposite side of the rim, producing a complete circumference. In this case we can still use formally Equation (97) and obtain the result $c' = \pm c$. But now this result does not correspond to the real physical situation. The local times of the two events in the photon's life (its departure from and return to the detector) are now measured by *the same* clock. This by itself would be all right were the system not rotating. In a rotating system we cannot uniquely allocate one time to a point using a synchronization procedure around the closed loop. Look at Figure 5.28. At a certain moment in the stationary system all stationary clocks read the same time. On the disk, each clock ahead of the previous one reads an earlier time. I emphasize – each pair of clocks on the disk are synchronized. However, if we apply the same procedure until we return to the original clock, we realize that this clock should read two different times at once. (Go back to the chain of moving clocks in Figure 2.6 in Section 2.7 and try to imagine what happens if you wrap the chain around a circle!). We see that the clock synchronization procedure carried out along a closed loop around the center in a rotating system will allocate two (and more!) different times to the same point of space!

Peculiar properties of "global" time in rotating systems are accompanied by peculiar properties of space. First, Equation (92) tells us the same thing that we have mentioned in the previous section about the Earth's equator: that the rim of the rotating platform is not the same as the rim of the stationary platform. This has no practical consequences for Earth because the difference is negligible in this case, but generally the effect may be important. Owing to the Lorentz factor, the circumference Λ increases with increasing rotational speed of the disk. The disk radius, on the other hand, does not change. Therefore, the ratio of the circumference to the radius on the rotating disk is greater than 2π. What can this mean? Only one thing: the geometry of a rotating system is not the Eucledean geometry we learned in school! What we call space is now curved so that some of the axioms of the "regular" space no longer work there. Strange as this appears to be, it can still be understood if we refer to the space–time diagram. Consider such a diagram for a point P on the rotating disk (Fig. 5.29a). Its world line is a helix. What we perceive as space is the set of events simultaneous in a given reference frame. For the observer in this frame all three spatial axes are perpendicular to the time axis, which is the world line of a stationary particle in this frame. Unable to represent all three spatial axes on a two-dimensional sheet of paper, Paul limits himself to two axes, X' and Y', in the plane of the disk. He draws them perpendicular to the world line of P. But this world line is neither worse nor better than the world line of any other point on the disk. Therefore, the plane de-

fined by the X'- and Y'-axes must also be perpendicular to the world line of any other point on the disk. In an inertial reference frame we have no problem in satisfying this requirement, because the world lines of all the points of this frame are represented by straight lines all parallel to each other; so a plane perpendicular to one of them is automatically perpendicular to all the rest. Now, try to perform the same trick here together with Paul! You cannot! Because all the world lines of the particles on the rotating disk (except for that of its center) are twisted, a plane perpendicular to one of them will not be perpendicular to another (Fig. 5.29a). Accordingly, the events on such a plane are not simultaneous – they do not form the space of the disk. The only way for Paul to make the plane perpendicular to the world lines of all the particles of his disk is to twist the plane accordingly, so it is no longer a plane from our viewpoint. But it is a plane for Paul, with the distinction that it is no longer a Eucledean plane!

The considered effect of the twist is only manifest along a line of finite length. It has no effect on a local speed of light. The same is true about the whole space of the system. We define the space as a three-dimensional "hyper-surface" perpendicular to the world lines of the clocks stationary in a given system [32]. Rotation, while retain-

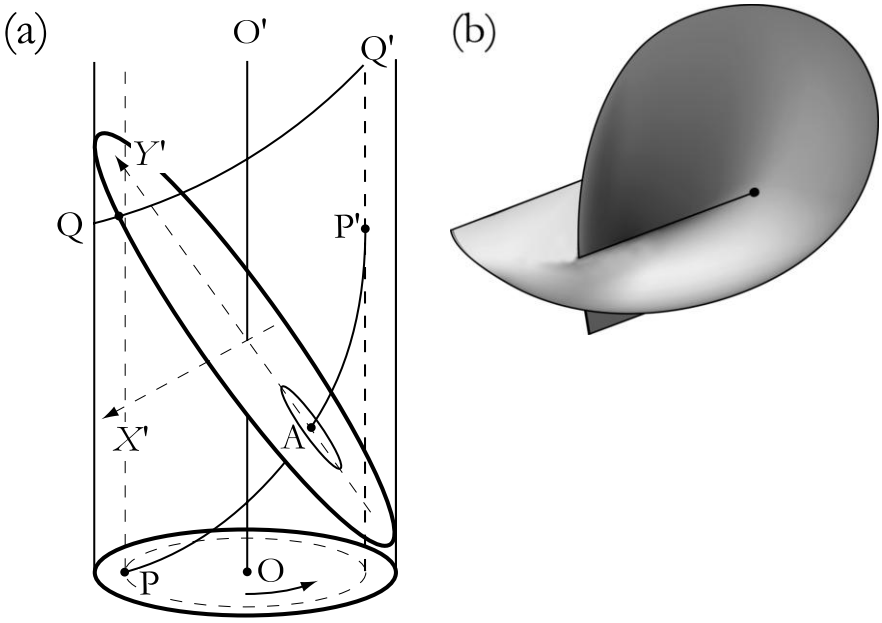


Fig. 5.29 (a) The plane X′Y′ is perpendicular to the world line PP′ of a disk particle. Therefore, its vicinity around intersection point A represents a small area on the rotating disk. However, this plane is not perpendicular to the world lines of other disk particles (for instance, line QQ′ or line OO′ of the center of the disk.) Corresponding elements cannot represent *spatial* areas around points O and Q. (b) If we twist the plane to make it perpendicular to the world lines of all the disk particles, its geometry departs from Eucleadean geometry, and the surface becomes self-intersecting.

ing the same structure of four-dimensional space, disrupts the whole fabric of the three-dimensional space. But it is well known that the change in topological properties of a space does not by itself affect its local properties. Therefore, synchronization of clocks and setting common simultaneity turn out to be impossible for the whole space if the system is rotating, but it remains always possible in any local region – in accordance with the postulate of the constancy of the local speed of light.

Applying this to the concluding part of the previous section, we see that there was no contradiction between Equation (90) and the initial condition that both flying clocks had the same speed. The speeds v_E and v_W found there are the "global" speeds characterizing complete cycles of motion of the clocks, whereas v is the local speed of the clocks. Both types of speed are *different* characteristics of motion in a rotating system, and both can be used for the full description of such motion.

Thus a close examination clarifies the "paradoxes" associated with the Sagnac experiment and rotational motion, and reveals some of its subtleties.

6
Superluminal Motions

6.1
Velocity, information, signal

After the appearance of the Special Theory of Relativity, many people came to believe that speeds exceeding the speed of light in a vacuum are impossible in principle. This statement, however, is wrong. The value of c is the limit for the speeds of material bodies or of the processes that could be used for the transmission of a signal. Under the term "signal" we mean the transmission of a certain amount of energy that carries information about an event at a point $\boldsymbol{r}_1$ at the moment t_1 and can change the state of a certain physical system at a point $\boldsymbol{r}_2$ at the moment t_2. According to what we know today from experiments, the nature of things is such that for all motions of this type

$$|\boldsymbol{v}| = \frac{|\boldsymbol{r}_2 - \boldsymbol{r}_1|}{t_2 - t_1} \leq c \tag{1}$$

And nevertheless, motions with $v \geq c$ also occur in nature. This does not contradict the above statement in Equation (1), since the latter type of motion is not associated with mass or energy transfer from $\boldsymbol{r}_1$ to $\boldsymbol{r}_2$. In other words, such motions cannot be used as a signal. The corresponding point representing this motion in the velocity space (recall Section 3.3) is not associated with the velocity of a real object. This point cannot be obtained from subluminal velocities in the framework of the usual Lorentz transformations. In what follows we will consider a number of examples of superluminal motions. A separate example of superluminal propagation of a light pulse in a specially "prepared" medium will be considered in Chapter 7. I have tried to make the pool of examples as representative as possible within the limited size of the book.

There are situations when superluminal motions of physical objects may appear in the phenomena associated with strong gravitational fields. Gravitation can be considered as a manifestation of curvature of space–time caused by stationary or moving matter. The apparent superluminal motions of an object in space–time curved in a special way are described in a book *Black Holes and Time Warps* by Kip Thorne [33].

Special Relativity and Motions Faster than Light. Moses Fayngold

ISBN: 3-527-40344-2

6.2 The scissors e ect

Let plane P be defined by the two intersecting straight lines AB and A′B′. The line AB is stationary while the line A′B′ rotates uniformly in the plane P with the angular velocity ω about the point O′, which is separated from AB by a distance a (Fig. 6.1). Owing to this rotation, the angle φ between the lines changes with time. We chose the initial moment $t = 0$ so that the lines are at this moment perpendicular to each other: $\varphi(0) = 0$. Then for an arbitrary moment t we will have $\varphi(t) = \pi/2 - \omega t$. The angle φ changes within the range $0 < \varphi < \pi/2$, so that $\omega t < \pi/2$. As time evolves, the point M of intersection slides along AB just like the intersection point of the scissor blades when at work. By analogy, we will call the associated phenomena the scissors effect. Now, what is the velocity of the intersection point? Qualitatively, we can see that as ωt approaches $\pi/2$ (that is, $\varphi \to 0$) the point M slides away to infinity within the finite time interval $0 < t < \dfrac{\pi}{2\omega}$. It means that in the end of the process point M acquires infinite speed. We can easily find the expression for the instantaneous velocity of M for any moment within the considered time interval.

Denoting the distance OM as x, we can readily see from Figure 6.1 that $x = a\tan\omega t$. Therefore,

$$v = \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{a\omega}{\cos^2\omega t} \tag{2}$$

At a certain moment equal to $t_c = \dfrac{1}{\omega}\arccos\sqrt{a\omega/c}$, the point M has instantaneous velocity c. Between the moments t_c and $t = \pi/2\omega$ it moves faster than light. At $t \to \pi/2\omega$ the velocity becomes infinite; at $t = \pi/2\omega$, point M jumps instantaneously from $x = +\infty$ to $x = -\infty$.

The superluminal motion of the point M does not in any way contradict the special theory of relativity, since we are dealing here with the mathematical point void of any physical content. This point does not represent a physical particle. Even if we could materialize segments AB and A′B′ as very long edges of the gigantic scissors, then at different moments t the point M would pass by the different physical particles constituting the edges. None of these particles moves together with M.

One can argue that such a particle will appear if we place at M a small ring through which both rods can pass freely. Then the rotation of A′B′ would cause the ring to

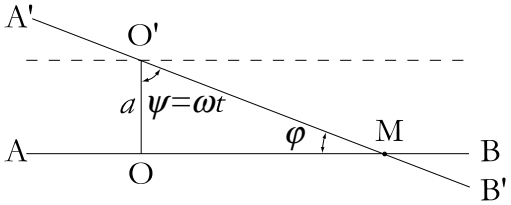


Fig. 6.1

move with the velocity given by Equation (2), and this would mean the superluminal velocity of a physical body!

The answer to this is that such an effect cannot be realized. Together with the ring, a new element will unavoidably enter the picture: the mass. The acceleration of the ring with even a very small mass requires the energy supply. As we know from Chapter 4, the kinetic energy of a body with a rest mass m and velocity v is $K = \gamma(v)\, mc^2$. As $v \to c$, the energy $K \to \infty$. The energy of the whole Universe would not be enough to accelerate the ring to the speed of light, and therefore such an experiment cannot be carried out in principle. It is just another way to say that the speed of light is an unattainable limit for any object with a non-zero rest mass.

6.3 The whirling swords

Let us consider another version of this thought experiment. The line AB is rotating uniformly about point A with an angular velocity c. Suppose that we are looking at a certain point M on this line. Let the distance OM be denoted as R. Then the linear velocity of M is

$$v = \omega R \tag{3}$$

and if R is large enough, then $v > c$. As in the previous case, the interpretation of this result is different depending on whether we regard AB as a mathematical abstraction or as a physical object (for example, a rod). In the first case the superluminal velocity of a point on the line does not contradict anything because there is no real motion of a physical object involved in this case. In the second case the rod possesses a certain mass and therefore the acceleration of any segment up to the speed $v > c$ is impossible. This confines the possible length to the limit determined by the condition $\omega R < c$, or

$$R < \frac{c}{\omega} \tag{4}$$

The value c/ω is the maximum possible limit for the radius of any rigid rotating material system.

Now there arises a question: how could a real rigid rod rotate if its length exceeds this limit? To answer this question, we must use the requirement that the physical speed of any, however distant, physical element on the rod cannot exceed the speed of light. But then we immediately come into contradiction with Equation (3), unless the angular velocity is zero. The only way out of this contradiction is to admit that the angular velocity is different for different points on the rod. This means that what we have assumed to be a rigid rod cannot actually rotate as a rigid body. Any attempt to realize such a rotation for a sufficiently long rod (there are no restrictions on length in relativity!) would result in deformation. Only a limited section of the rod

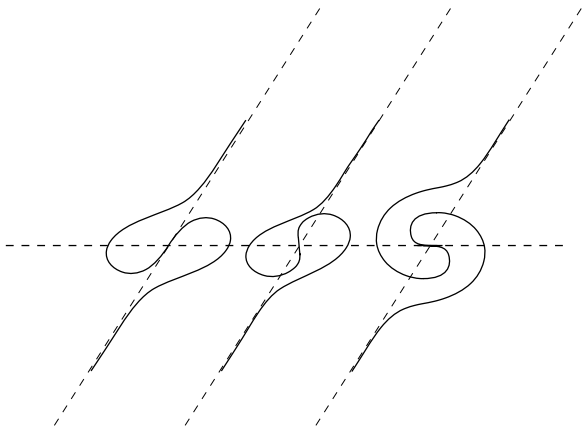


Fig. 6.2 Three consecutive snapshots of a rotating "rigid" rod (e. g. a whirling sword) of unlimited length.

would rotate approximately as a rigid body in accordance with Equation (3), while the more distant parts would lag behind. For any real rod this effect would be observed for distances much less than c/ω. In all practical cases this limiting length is very large. For example, a spoke of a bike wheel rotating at $\omega = 30$ rad s^{-1} can in principle be brought to a state of a "rigid" rotation only if its length is less than $3 \times 10^8/30$ m. The rod must be very long indeed in order for us to observe any deviation from the "rigid" rotation. This is why we have never observed this effect in everyday life.

In the general case, for very large ω or very long rods, the picture would appear as a continuously spiraling, twisting, and expanding rubber plate rather than rigid rod (Fig. 6.2). Hence in order to preserve the wholeness of the rod, we have to attribute to it unlimited elasticity. Thus the Special Relativity has again, as in Section 5.5, led us from a concept of an ideally rigid body to the directly opposite concept! Of course, any real body will just be destroyed under the given conditions; therefore, it is better in this case to consider the motion of a set of separate elements of finite size or point particles. This is what we will try to do in the next section.

6.4 Waltz in a magnetic field

There is a well known phenomenon which can be used as a model for a process under consideration: the motion of electrically charged particles in a uniform stationary magnetic field H. Let the field permeate the whole space and the charges be so small that their Coulomb interactions are negligible compared with the magnetic force. Under these conditions, a charge q of mass M moves in a circle with the angular velocity [16, 23]

$$\omega_0 = \frac{qH}{Mc} \tag{5}$$

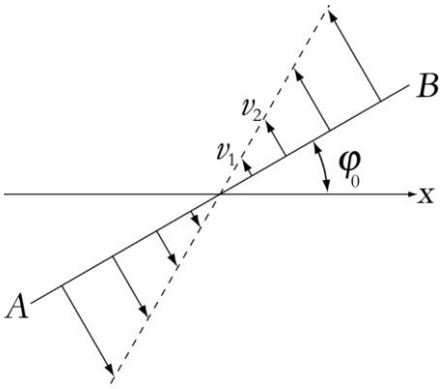


Fig. 6.3

If the ratio q/M is common for all particles, the angular velocity will also be common. The radius of each circle is proportional to the particle's speed: $R = \omega_0^{-1} v$. The particles with various speeds will trace out circles of various radii.

Suppose that at some moment the particles are aligned along the straight line AB, just as dancers on a stage. Suppose further that their velocities are all perpendicular to AB and proportional to their distances from a stationary point O (Fig. 6.3). We also assume that both AB and all velocity vectors are perpendicular to **H**. This arrangement ensures that all the particles rotate about a common center O in the same plane.

Because the angular velocity is the same for all particles, they must remain on one straight line while the line itself rotates about O. Thus our system, being actually a one-dimensional gas, realizes the model of a rotating rigid rod. But this model, too, can be realized only for relatively small distances AB. For sufficiently remote particles the existence of the light barrier will prevent superluminal motion, so that the proportionality between R and v cannot be maintained. The same relativistic Equation (8) in Section 4.1 that did not allow the mass to reach the light barrier comes in here again. The relativistic increase in the mass with increase in the speed must be taken into account in Equation (5) for the angular frequency. We have to put for mass in Equation (5) its expression given by Equation (8) from Section 4.1 to get the $\omega \leftrightarrow R$ dependence consistent with the requirements of the theory of relativity:

$$\omega = \frac{qH}{Mc}\,\gamma^{-1}(v) \tag{6}$$

The last equation precludes the possibility of our particles reaching or exceeding light's velocity. At $v \to c$ we have $\omega \to 0$, and at $v > c$ the angular velocity becomes imaginary.

Equation (6) describes the dependence $\omega(R)$ implicitly, since upon substituting ωR for v in the right-hand side, ω will appear on both sides of the equation. Solving this equation for ω, we obtain an explicit dependence:

$$\omega(R) = \omega_0\left(1 + \frac{\omega_0^2 R^2}{c^2}\right)^{-1/2}, \qquad \omega_0 \equiv \frac{qH}{MC} \tag{7}$$

It turns out that, contrary to our initial assumption, each particle rotates with individual angular velocity depending on the distance R. For the corresponding linear velocity we have

$$v(R) = \omega(R)R = \omega_0 R\left(1 + \frac{\omega_0^2 R^2}{c^2}\right)^{-1/2} \tag{8}$$

Assuming here that $R \to \infty$, we find that for very distant particles $\omega(R) \to 0$ and $v \to c$. We have automatically obtained the same result as for the rod. As was the case with the rod, the line AB which was originally straight, gradually bends into a spiral similar to that shown in Figure 6.4. Now, however, our knowledge of specific properties of the system enables us to obtain an exact form of the spiral for any moment. If all the particles were initially arranged along the straight line making an angle $\phi = \phi_0$ with the x-axis, then at later moments t the line will deform into a spiral:

$$\phi(t) = \phi_0 + \omega(R) \cdot t = \phi_0 + \frac{\omega_0 t}{\sqrt{1 + \frac{\omega_0^2 R^2}{c^2}}} \tag{9}$$

According to Equation (9), the closer a spiral segment is to the origin, the faster it twists.

A reader familiar with astronomy will possibly notice that Figure 6.4 reminds us of the picture of a spiral galaxy. At present there are only a few hypotheses to explain the origin of the spiral branches of such galaxies. One of these hypotheses is connected with the results of a study carried out by the astrophysicists Sophue and Fudsimoto [34], and is conceptually close to our model here. The difference is that in our model the magnetic field twists a system of particles into a spiral, whereas in the galaxies a system of charged particles (plasma) twists the magnetic field into a spiral. The latter mechanism can be understood in terms of the magnetic field lines penetrating plasma. If the plasma is sufficiently dense, the magnetic lines are "glued" to the substance of the plasma they penetrate. The moving plasma carries the magnetic lines along with its motion. In particular, it twists them while rotating. This mechanism might be responsible for the appearance of the spiral branches.

The observations by Sophue and Fudsimoto showed that the galactic magnetic field is aligned along the branches, where the interstellar gas and young stars are concentrated. The field in one branch is directed towards the center of the galaxy, and in the opposite branch away from the center. In other words, the field lines enter the galaxy through the end of one branch and exit through the end of the other one. This geometry of the field is consistent with the hypothesis according to which such a field has an external origin and had existed before the galaxy was formed.

Figure 6.4 shows successive stages in the formation of the spiral structure of the galactic magnetic field. At first there is a uniform magnetic field permeating a thinned,

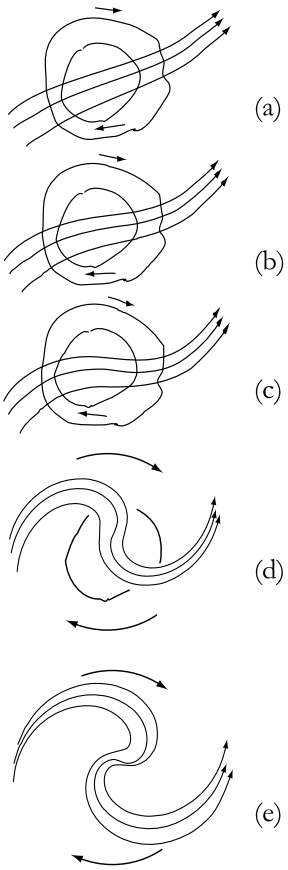


Fig. 6.4

slowly rotating cloud of plasma. As this cloud compresses due to gravity, it becomes denser, and the original magnetic field, following the plasma matter, also compresses and intesifies without losing its connection with its external sources outside the galaxy. Because of conservation of angular momentum, the compressed cloud accelerates its rotation, carrying along the magnetic field lines. But far away from the center of the galaxy, the substance and the field lines "frozen" in it cannot rotate at the same angular velocity as near the center (we can think of the field lines as elastic strings being "anchored" at the remote stationary sources). As a result, a spiral structure forms, looking much as the spiral shown in Figure 6.2. In this case, however, the process occurs naturally on a cosmic scale.

As in previous examples, the spirals are formed here because the remote parts of a system (in the last case, the intergalactic gas) cannot rotate with the same angular velocity as those in the galactic core. It is forbidden by the existence of the light barrier.

6.5 Spiraling ray

The next example deals with a rotating beam of light. There are two different ways of realizing such a beam. The first is to use a stationary source of monochromatic light at the center of a rotating opaque spherical shell with a hole at its equator. The second involves the direct rotation of the source itself, e. g. a rotating searchlight or laser.

Consider first the source with the shell. The aperture lets out a ray of light that propagates away from the center along a straight line connecting the source and the hole. If we place a screen far away from the shell but directly against the hole, and the shell does not rotate, we will see a bright spot right at the center of the screen. We describe the situation by saying that the ray projects the image of the hole onto the screen. Some people may say that the hole projects the image of the source onto the screen. We shall not worry about these terminological differences, they are a matter of choice. So far as we all agree that the source, the hole, and the image (illuminated spot on the screen) are on the same straight line, we have a common understanding of what "projection" means. Let us call the whole experimental setup "the projector." Now, set the shell rotating. How will it affect the ray and the image?

Well, the ray from the rotating projector will also rotate, and the bright spot on the screen will move. To observe this motion, we have to encircle the projector with a cylindrical screen of a large radius, whose symmetry axis coincides with the axis of rotation. The question now is: what are in this case the instantaneous form of the ray, the instantaneous position of the image, and the speed of its motion along the screen? It may seem at first that the ray must in this case realize the model of the rotating straight line, that is, a rotating infinite (or very long) rigid rod. After all, it is a well known fact that light in a homogeneous medium propagates along a straight line. It therefore seems impossible to bend a light beam in a vacuum. And indeed, ideally straight pencils of light from a searchlight scanning the night sky or sweeping across the horizon seem to present direct visual evidence of this conclusion.

And yet, the conclusion is wrong. The beam seems rectilinear only because the speed of the rotating aperture is much less than c. If the projector could rotate very fast, or if we could observe a beam of a sufficiently great length, we would see the beam bent. Since the speed of light is finite, it takes a while for it to form an image on the distant screen. While the photons from the aperture rush to the screen, the aperture itself shifts to another position, so the image, the hole, and the source are no longer on the same straight line. As a result, one would see the beam twisted as a spiral. However, this spiral differs drastically from those considered above: it rotates as a rigid body, that is, all its segments, no matter how distant from the center, have a common angular velocity ω which is equal to that of the shell. A remarkable distinction! A long rigid rod from Section 6.3 has to be deformable as a fluid when rotating, whereas such a soft matter as light forms a "rigid" structure. Therefore, the resulting form of the beam asymptotically (that is, at large distances) remains spiral rather than approaches a straight line (Fig. 6.5 a). In this case, the linear rotational velocity of a sufficiently distant segment of the ray (that is, the velocity of the bright spot traced out by the beam on the cylindrical screen) can be arbitrarily greater than c.

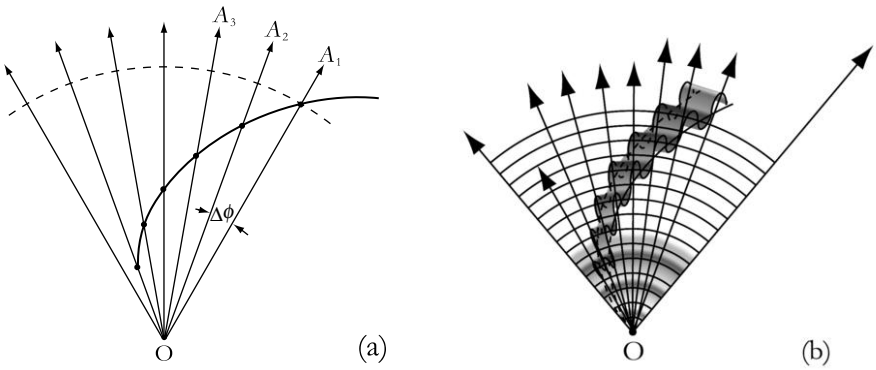


Fig. 6.5

We have apparently come to an absurdity: a light beam is traveling faster than light! A swarm of photons sweeps across the screen faster than each single photon can ever do. Apart from this logical absurdity, there is a physical one. An electromagnetic field stores energy. A bright spot on the screen can be considered as a localized lump of energy. If we plot the instantaneous energy density distribution over the screen, the illuminated spot can be represented graphically as an isolated hill in a desert. And, as we know, any energy is associated with corresponding mass according to $M = E/c^2$. So we have a massive "energy hill" in a desert. No mass can move faster than light, but this hill can. For a screen sufficiently distant, it sweeps over the "desert" with a superluminal speed. Let us find out what the solution to this paradox is.

The easiest way to understand the phenomenon is to consider light as a flux of photons – small clots of electromagnetic energy. From this viewpoint, the light source can be considered as a photon "machine-gun." With the shell rotating, the outcoming photons shoot across the entire plane perpendicular to the rotational axis. The trajectory of each individual photon is a straight line emerging from the source. This line is just a path followed by a photon. Because of the rotation of the aperture, each next photon that is let out will move along a radial line which makes a small angle with the previous one. If the time interval between two successive photon emissions from the aperture is dt, then the angle will be $d\phi = \omega\, dt$. Since each successive photon is emitted later than the previous one by the time interval dt, its distance from the center is at any moment less than that of its predecessor by $c\,dt$, that is $dr = -c\,dt$. As a result, all photons at any moment turn out to be aligned along a spiral. This spiral forms a true instantaneous picture of the rotating beam (Fig. 31.1 a).

Combining the last expressions for $d\phi$ and dr, we have for each small segment of the spiral

$$dr = -\frac{c}{\omega}\, d\phi \tag{10}$$

The minus sign here shows that larger ϕ correspond to smaller r. If the photons are being emitted continuously one after another, we obtain a continuous curve described by Equation (10). In the integral form the equation reads

$$r - r_0 = -\frac{c}{\omega}(\phi - \phi_0) \tag{11}$$

Here the constants r_0 and ϕ_0 depend on the moment we choose to start timing and on the size of the shell. We can express these constants in terms of the boundary and initial conditions. Let a be the radius of the shell, and the timing starts when the aperture intersects the direction $\phi = 0$. The photon exiting through the aperture at this moment is at the distance $r = a$ from the center. It keeps on moving along the direction $\phi = 0$, so that by the time t its distance from the center is $r(0) = a + ct$. By this moment the aperture will have swept through the angular range $0 < \phi \leq \omega t$, sending photons in the corresponding directions. Let us focus on a photon moving along a direction ϕ within this range. This photon was emitted later than the first one by the time τ it takes the aperture to turn through the angle ϕ:

$$\tau = \phi/\omega \tag{12}$$

Its distance from the origin will therefore be

$$r(\phi) = a + ct - c\tau = a + ct - \frac{c}{\omega}\phi \tag{13}$$

This equation is equivalent to Equation (11) with the constants t_0 and ϕ_0 expressed in terms of the shell's radius and the initial condition. This condition is actually a requirement that only the photons outside the shell are being considered, that is, $r \geq a$.

The curve described by Equations (11) or (13) was probably first considered by Archimedes and has ever since been called Archimedes' spiral. While each small element of this spiral (each photon) moves strictly radially, away from the center, the entire system (spiral beam) appears rotating as an ideally rigid body with an angular velocity equal to that of the shell. In other words, the spiral can be thought of as a rigid protuberance from the aperture. The linear rotational velocity of a spiral element at a distance r from the origin is equal to

$$v = \frac{r\,\mathrm{d}\phi}{\mathrm{d}t} = r\omega \tag{14}$$

This velocity, unlike those for spirals considered above, becomes greater than c at $r > r_c$, where

$$r_c = \frac{c}{\omega} \tag{15}$$

The corresponding superluminal bright spot, however, cannot carry any signal in the direction of its motion (that is, along the screen). The light spot at the point A′ at the moment t' does not originate from the light spot A at the previous moment t. It is caused by the photons shot independently in the direction OA′. These photons know nothing of the fate of photons emitted earlier along OA. Thus, the flux of energy (and information) is by no means directed along the screen. It flows *through* the screen – radially away from the source.

We can arrive at the same result if we use the picture of running electromagnetic waves rather than photons, and study the wave field structure within the spiral. We will then see that at each moment the crests and troughs of the waves do not at all form spirals, but rather concentric arcs (Fig. 6.5 b) just like the waves from a stone dropped into water. The wave carries energy in the direction of motion of its crests. This direction in our case is radial. The waves diverge radially from the origin with the speed of light. These radial directions are what we call the true rays. What, then, is the spiral ray? It is just an area *carved out* by the rotating aperture from the radially diverging waves. We can see from the Figure 6.5 b that it is only in the direct vicinity of the shell that the spiral beam is approximately coincident with the real radial ray.

A more advanced reader can arrive at these results using the concept of the *instantaneous light intensity* (the Poynting vector). The energy density in the electromagnetic field at any point in space is given by [23]

$$\eta = \frac{1}{2}\left(\varepsilon_0 E^2 + \frac{1}{\mu_0} B^2\right) \tag{16}$$

where $\boldsymbol{E}$ and $\boldsymbol{B}$ are the instantaneous electric and magnetic field vectors at this point. In the electromagnetic wave, this energy is transferred at the speed c in the direction $\boldsymbol{n}$ perpendicular to both $\boldsymbol{E}$ and $\boldsymbol{B}$. The amount of energy flowing per unit time through a unit area perpendicular to $\boldsymbol{n}$ is given by

$$\boldsymbol{S} = \frac{\boldsymbol{E} \times \boldsymbol{B}}{\mu_0} \tag{17}$$

where $\boldsymbol{E} \times \boldsymbol{B}$ is the cross product of the two vectors. In waves running in free space the vectors $\boldsymbol{E}$ and $\boldsymbol{B}$ are mutually perpendicular, and $E = c\,B$, so that

$$\boldsymbol{S} = \frac{E^2}{c\mu_0}\boldsymbol{n} \tag{18}$$

If we in addition recall that $\mu_0 \varepsilon_0 = c^{-2}$, we can easily verify by the direct inspection that the last expression is equal to $\eta c \boldsymbol{n}$. Thus, the instantaneous light intensity is equal to the product of the energy density and the velocity $c\boldsymbol{n}$ of the energy transport. Since the orientations of $\boldsymbol{E}$ and $\boldsymbol{B}$ in the spiral beam are the same as in a divergent spherical wave, the last equation immediately gives the already known result: whereas the illuminated spot runs *along* the screen, the energy is transported radially *through* the screen at the speed c. If the screen is opaque, the incident energy is ab-

sorbed and dissipates into heat. There is no energy transport with the bright spot along the screen, because the subsequent positions of the spot consist of the entirely different photons, arriving independently along the different radial lines and knowing nothing of one another. The time interval between the two different appearances of the spot on the screen may be arbitrarily short, just as the time difference between the two rain drops hitting the ground at two different points. It may be shorter than the time it takes light to travel from one point to another. This does not mean a superluminal transfer of mass or energy, because we have two different independent drops. We can, of course, connect the resulting wet spots by a straight line and say that the second spot has emerged after the first one faster that it would take light to travel between the spots. But it does not mean real superluminal motion of a drop between the places.

6.6
Star war games and neutron stars

A message from the Early Warning Center informed of an approaching target launched to imitate an alien military cruiser. Neither its exact position, nor distance from the station, nor velocity were known. The only information available was that the ship's orbit was in the same plane as the orbit of the station. The ship was coated with a special alloy absorbing nearly all incident electromagnetic radiation, which rendered it practically invisible in all regions of the spectrum. Captain Fletcher's task was to destroy the ship as soon as possible.

"In such a situation," the Captain thought, "there are two possible strategies. One is to shoot in random directions in the orbit's plane. Then the target will sooner or later be hit by a random shot. Another is to set the laser to a regime of continuous radiation and rotate it in the ship's orbit's plane, so that the whole plane be swept out by the laser beam. Then the target will be hit with certainty within the time needed for the laser to make one full rotation."

Now, let us estimate the strategies. In the first one there is a chance that the target will be hit with the first shot, but it is pretty slim. It may as well happen that the target will still remain unhit after a year of shooting. So the average "life expectancy" of the spaceship is very large. But they do not want the target to be hit some day. The order is to hit it as soon as possible. The second strategy does precisely this. The period of one rotation of a laser may be just a few seconds. For "as soon as possible," a few seconds will be acceptable. We decide in favor of the second strategy, and so did Captain Fletcher.

What is the difference between this situation and that of rotating perforated shell in the previous section? The shell's role was only that of a chisel – to carve out a spiral from a spherical wave radiated by a *stationary* source in all directions. Both the source and the observer of the ray belonged to one inertial reference frame. In contrast, now *the source itself* (the laser gun) is rotating, while Captain Fletcher with his team are at rest. The source and the Station embody two different reference frames. Accordingly, the resulting beam as observed from the Station will now have a differ-

ent structure. To find the difference we must recall Captain Fletcher's reasoning when he, as a shore-based engineer, observed the rising water in Mr. O'Bryen's ship (Section 2.10). The two different marks A and B that we observe *now* on the wave front are not simultaneous events in the laser's reference frame (Fig. 6.6). They belong to the same advancing front, but to different moments in its history. B is characterized by a later moment than A by the laser's clock $t'_B > t'_A$. By that moment B must have progressed further away along the laser's axis than A by the moment t'_A. Since it is *us* who observe *now* the different moments of the laser time, the line AB *as observed by us* is inclined to its direction A′B′ in the laser's reference frame. In the case of light the tilt angle α is given by Equation (68) in Section 2.10. The wave vector $\boldsymbol{k}$ makes the same angle $\theta = \alpha$ with the laser's axis. The current situation is more complicated, though, because the laser's motion is rotational, rather than translational. The different parts of the laser have different velocities. Accordingly, we must observe different tilts for different parts of the wave, as shown in Figure 6.6. The foremost part just leaving the laser has the largest tilt. Let us call it θ_m. Once free, the wave retains its original direction, propagating along a straight line making an angle θ_m with the corresponding radial line (for the rotating shell in the previous section the angle θ_m is equal to zero, and the photons would propagate strictly radially).

Consider again light as the flux of photons. As the laser rotates, it spills the photons, each photon making the same angle θ_m with the laser axis. At a certain moment t in a stationary reference frame, one can observe all the photons that have been fired since the rotating laser had started shooting. The set of these photons forms a curve

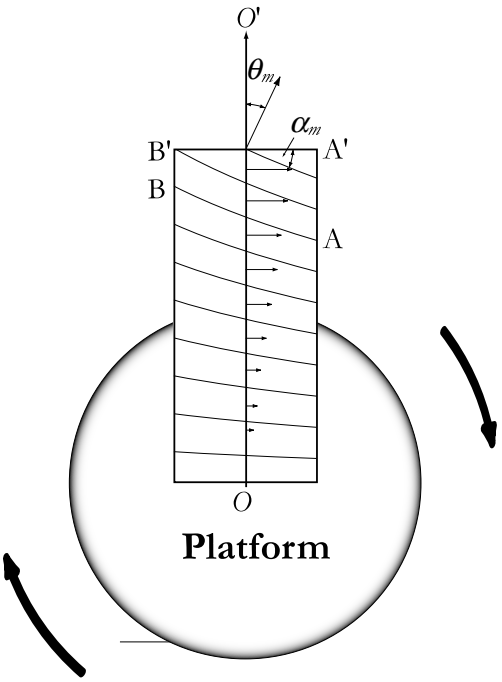


Fig. 6.6 The wave fronts of light propagating along the barrel of a rotating laser. OO′ is the axis of the laser barrel. The wave fronts as observed from a stationary reference frame are tilted to the axis because they are propagating in the rotating laser medium. The horizontal arrows represent corresponding local velocities of the medium.

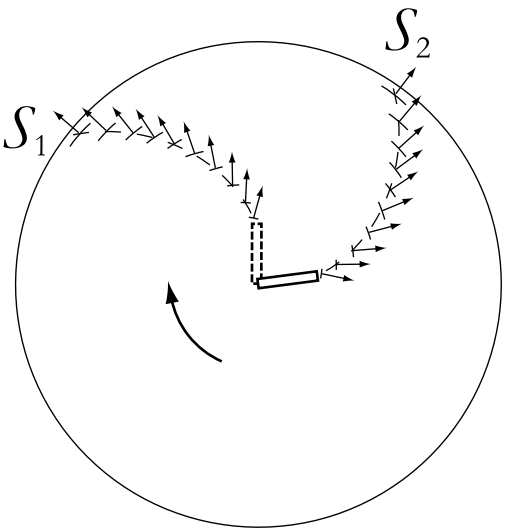


Fig. 6.7 Two positions of the laser beam at the moments t_1 and t_2. The second spiral can be obtained from the first one by rotating it through the angle $\Psi = \omega(t_2 - t_1)$, where ω is the angular velocity of the laser.

bent like a spiral around the rotational axis (Fig. 6.7). For a later moment we will have the same spiral turned around this axis through a certain angle. Qualitatively, we can regard the spiral as a rigid curved extension of the laser, rotating together with it. Imagine again a far away cylindrical screen centered about the rotational axis. The illuminated spot made by the spiral ray on this screen will run along it with an angular velocity ω equal to that of the spiral. If the screen is far enough, the spot will run across it with superluminal speed. This explains why just a few seconds may suffice for the laser beam to sweep out the unimaginably vast area with the lurking alien spaceship in it.

Now, to become better prepared for ever-changing situations in a star war game, we, in addition to Captain Fletcher, have to solve the whole problem quantitatively. We need to prove that the laser beam will indeed rotate synchronously with the laser as its huge rigid protuberance. And we also need to know the exact shape of the beam at any moment in order to determine how to move and aim the laser gun to hit an arbitrary moving target with certainty.

For a moment t, consider an instantaneous position of the spiral beam and single out an arbitrary point A on it (Fig. 6.8). It is a distance r away from the origin O formed by the center of the rotating platform. Take an instantaneous direction of the laser at the initial moment $t = 0$ as the reference direction OO′ in our inertial reference frame. Let the radial line OA to the chosen point make an angle ϕ with the reference direction. We need to find the equation relating r to ϕ for any moment t. Since the point A is arbitrary, this equation would hold for any point and thereby contain all information about the motion and shape of the whole beam.

To find the equation, note that the photon at A must have been fired at an earlier moment t' when the laser made an angle ϕ' with the direction OO′. If the laser rotates with angular velocity ω, then $\phi' = \omega t'$. Label as A′ the corresponding starting point

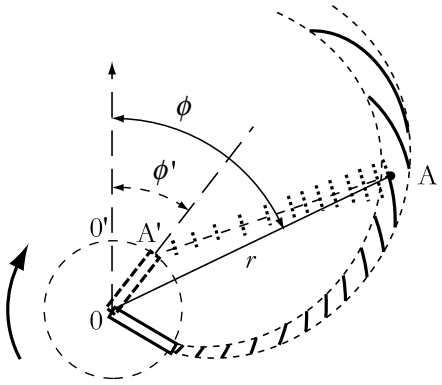


Fig. 6.8

of the photon that is now at A. It must be a distance l from the center, where l is the length of the laser. The distance A′A is equal to $c\,(t - t')$, and AA′ makes an angle θ_{m} with OA′. Now, applying high-school geometry (the sine theorem) to the triangle OAA′, we have two equations:

$$\frac{r}{\sin\theta_{\mathrm{m}}} = \frac{l}{\sin\left(\theta_{\mathrm{m}} - \phi + \phi'\right)} \tag{19}$$

and

$$\frac{r}{\sin\theta_{\mathrm{m}}} = \frac{c\,(t - \phi'/\omega)}{\sin\left(\phi - \phi'\right)} \tag{20}$$

We use the first equation to express ϕ' in terms of θ_{m} and ϕ:

$$\sin\left(\phi - \phi'\right) = \left(\sqrt{1 - \xi^2 \sin^2\theta_{\mathrm{m}}} - \xi\cos\theta_{\mathrm{m}}\right)\sin\theta_{\mathrm{m}} \tag{21}$$

$$\sin\phi' = \xi\sin\theta_{\mathrm{m}}\cos\left(\theta_{\mathrm{m}} - \phi\right) - \sqrt{1 - \xi^2\sin^2\theta_{\mathrm{m}}}\sin\left(\theta_{\mathrm{m}} - \phi\right) \tag{22}$$

where $\xi \equiv l/r < 1$. Putting this into the second equation gives

$$\sqrt{r^2 - l^2\sin^2\theta_{\mathrm{m}}} - l\cos\theta_{\mathrm{m}} =$$
$$c\,t - \frac{c}{\omega}\arcsin\left[\frac{l}{r}\sin\theta_{\mathrm{m}}\cos\left(\theta_{\mathrm{m}} - \phi\right) - \sqrt{1 - \left(\frac{l}{r}\right)^2\sin^2\theta_{\mathrm{m}}}\sin\left(\theta_{\mathrm{m}} - \phi\right)\right] \tag{23}$$

This is the sought relation between r, ϕ, and t. Let us play with it. Figure 6.7 shows two different positions of the beam calculated from Equation (23) with $\theta_{\mathrm{m}} = 30^\circ$, which corresponds to an extremely rapid rotation of ~1.5×10^6 cycles per second for

l = 15 m). The positions correspond to two different moments of time. The reader can easily obtain either one of them from another by simple rotation of the figure. Those who like mathematical rigorousness can set $r = R$, where R is any constant distance greater than l. One will in this case obtain the equation relating the angle ϕ to time for a point on the beam with a fixed distance R from the center. Differentiating the resulting equation with respect to time gives

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \omega \tag{24}$$

The result does not contain R, and therefore holds for any point on the beam. This proves that the whole beam rotates as one rigid body with the angular velocity ω.

Let us now compare the rotating laser beam (Fig. 6.8) with that of the rotating shell (Fig. 6.5). Since the photons from the shell propagate strictly radially from the very beginning, we might expect the initial segment of the corresponding spiral to be radial. Instead, it is inclined to the radial direction. In contrast, the laser photons leave sideways to the laser axis, and therefore we could expect the initial segment of the laser beam to be accordingly tilted. Instead, it is strictly radial. It looks at first as if Figures 6.5 and 6.8 have been mistakenly interchanged in the process of printing. But no, everything is correct. In the first case, the photons' velocities have no transverse component, while the aperture letting them through moves in the transverse direction. As a result, the line connecting the two neighboring photons must be tilted to the radial line. In the second case, the velocity of each emitted photon has a transverse component equal to the laser tip's velocity. Therefore, the segment connecting the two neighboring photons turns out to be radial. Thus the behavior of the *system* of photons does not imitate their individual behavior.

Now, consider some special cases. First of all, if $\theta_\mathrm{m} = 0$, the photons must propagate radially, and the situation must be the same as in the previous section. Indeed, if we set $\theta_\mathrm{m} = 0$ in Equation (23), it reduces to

$$r = l + ct - \frac{c}{\omega}\phi \tag{25}$$

and we recover the Equation (13). The same conclusion follows for very large distances, where deviation from the radial motion becomes insignificant, no matter the value of θ_m. Setting in Equation (23) $r \gg l$, that is, $\xi \ll 1$, we obtain

$$r = ct + \frac{c}{\omega}\theta_\mathrm{m} - \frac{c}{\omega}\phi \tag{26}$$

which differs only by a constant from the result in Equation (23) and also describes uniform rotation with angular velocity ω. The reader may have noticed from Figures 6.7 and 6.8 that the further away from the center, the smaller is the angle between the wave front and the axis of the beam. In this respect the peripheral parts of Captain Fletcher's beam are almost indistinguishable from those of the rotating perfo-

rated shell. In both cases the wave front becomes practically parallel to the axis of the beam and to the surface of the screen. This provides us with another explanation as to why the illuminated spot runs along the screen faster than the light signal. Suppose there are two alien ships A_1 and A_2 "on the screen," separated by some angular distance $\Delta\phi$ (Fig. 6.9). The arc between them would be $\Delta s = R\Delta\phi$ m long. Let γ be the angle the beam's axis makes with the cylindrical screen of radius R at the point of their intersection (Fig. 6.9). It is readily seen from Figure 6.9 that as the wave front advances outwards, the intersection point (the light spot) moves over the screen with the velocity

$$v \approx \frac{c}{\sin\gamma} \tag{27}$$

Differentiating Equation (23) with respect to ϕ and then setting $r = R$ gives for γ the expression $\tan\gamma = \mathrm{d}r/R\,\mathrm{d}\phi = c/\omega R$. When R is large enough, γ becomes so small that $\sin\gamma \to \tan\gamma \to \gamma \to 0$, and v exceeds the speed of light. This means that the ships would be vaporized one after another with a time interval much shorter than the time needed for the direct laser shot from A_1 to reach A_2. But this does not contradict anything, since in the case of the spiral beam there is no causal connection between the two events. The two blasts that have almost simultaneously converted two islands of life in the vastness of space into two clouds of atomic vapor were caused by two independent groups of photons knowing nothing of one another.

All this rolls over in Captain Fletcher's mind while he is giving orders and then waits for the results. Captain Fletcher knew that the command "as soon as possible" should be taken seriously and therefore he ordered the use of the smallest of his laser units, whose rotation rate is most rapid. It makes one complete circle in just 3 s, and is easier to start. Accordingly, the Captain expects to see a flash of blast in the sky within 3 s after the start. At the latest, the flash may come a fraction of a second past this time if the target happens to be close to the initial line of shooting but in the direction opposite to the rotation of the gun, and allowing for the time that the light signal of the blast needs to reach the station.

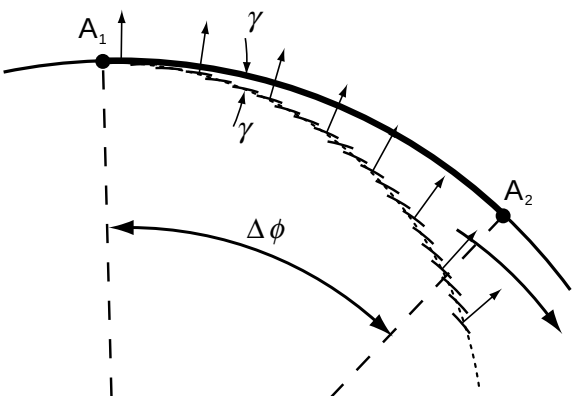


Fig. 6.9

But 3 s passed, then 4, then 6, ... and nothing happened. For a moment Captain Fletcher's face bore a puzzled expression. But then it became tranquil again. "Aha," he thought, "the superluminal motions do not overrule causality. With all these bright spots running faster than light along an imaginary screen, it is an outward-directed flux of energy that would hit the aliens. In the final run, I cannot hit the target faster than it takes a laser pulse from my gun to reach it. The fact that I did not get any signal of the target being destroyed indicates merely that it is further away from the Station than I had originally thought. I see now that I was not precisely correct when saying that *the whole plane* would be swept out by my laser beam in one period of rotation. What it actually meant was that an angular region ranging from 0 to 2π would be swept out by the beam. It does not mean covering of the whole plane, which would also involve ranging from 0 to ∞ for a radial variable *r*."

Captain Fletcher asks his computer to shade and display the area actually swept out by the laser beam in one rotation since the start. The printout from the computer is shown in Figure 6.10a. We see a sharp tooth and the indentation rather than a uniform circle. If an alien ship happened to be in this indentation, it would remain intact during the first rotation, but be hit in the second one. Since this did not happen either, the ship must have been further than 3 light seconds (about 900 000 km) away from the station (Fig. 6.10b). Beyond this distance, the laser's power may only suffice to damage the target, rather than destroy it. "My strategy was correct, but not implemented in its full sway," Captain Fletcher thought. "I won't be so bashful this time." He commanded to bring into operation the biggest firing laser facility in his unit, a 200 m monster with enough power do dig a canyon across the Moon's surface in the twinkling of an eye. It takes a full minute for one rotation, and Captain Fletcher figured out that the target would be hit, at worst, about 1.5 min after the shooting started. Actually, just 43 s after it, the optical scanners had registered the emergence of a new star in the sky on its intercept with the shooting plane. The star was so bright that its fleeting appearance was clearly seen by the naked eye through a large illuminator in the Station's battle compartment. The computer monitors of the Advance Defense Unit displayed a huge blast with the message: TARGET DESTROYED. Captain Fletcher's subordinates saw a shadow of a smile cross their commander's face.

The story of Captain Fletcher may or may not be found in the early records of the third millennium, but its essence is much closer to reality than most science fiction

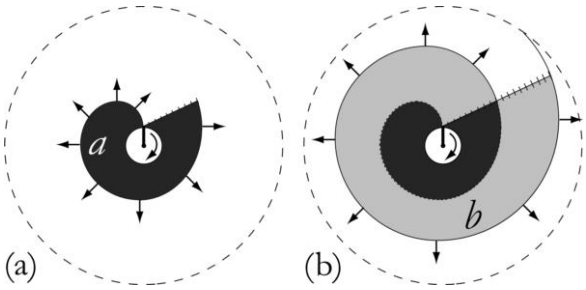


Fig. 6.10 An area actually swept out by the laser beam from the rotating laser gun. The beam forms an instantaneous boundary of the shaded area that expands outward in the directions given by the local arrows. (a) represents the area swept out in one rotation and (b) is the area swept out in two rotations.

stories. Events similar to those described above have been happening for millions of years, but only recently, in the late-1960s, did we become aware of them. The story about real events might have been much more dramatic, since in it our whole planet plays the role of an alien spaceship under the fusillades of laser beams whose sources are infinitely more powerful than the biggest of Captain Fletcher's puny monsters. Our planet's salvation lies only in the fact that these natural sources are too far away from us.

Their discovery is reminiscent of a detective story. In 1967 a very sensitive radio telescope was constructed near Cambridge in England by a team under Sir A. Hewish. The telescope was designed to detect and study weak sources of cosmic radio waves. One day the telescope started to receive very unusual signals: a rapid periodic sequence of sharp pulses of radiation [35].

Regularity as such is commonplace in astronomy. It was its period that astounded the researchers. The period was just about 1 s in contrast with hours, days, or years characteristic of the known periodic motions of celestial bodies. A 1-s period was more characteristic of the heartbeats or pulse of a living creature rather than of a cosmic object. Naturally, the first (and very appealing) thought was about the signals being a message from extraterrestrial civilization. But, on the other hand, the structure of the signals rendered this possibility very improbable. It would be natural to expect that a signal from intelligent life must carry some encoded meaningful message. Accordingly, a series of pulses should have had some patterns similar to those in speech or writing. But no such things were found. The sequence of almost identical spikes carried no information other than about the period of its source.

Soon after, a few more sources similar to the first one were detected in other regions of the sky. This has rendered the extraterrestrial intelligence hypothesis utterly implausible. It was improbable enough to find one such intelligence. To find a few that would simultaneously have taken special interest and pain to conduct a directed broadcast for us makes the idea highly incredible. Most probably, the received signals must have been a manifestation of a natural periodic motion associated with a new type of cosmic object. Further analysis showed that it might be an extremely compact rotating star. The smallest and most compact stars observed by that time were so called white dwarfs. Most stars evolve into white dwarfs after their nuclear fuel has burned out. We have sound reasons to believe that our Sun will also eventually shrink down to the size of a planet like our Earth and become a white dwarf (this will take a few billions years, so we need not worry about it). But not even a white dwarf can spin at a rate of one or more cycles per second. It would be torn apart by the centrifugal forces (which are actually the inertial forces, recall the Introduction!). It must be something much smaller and denser than even a white dwarf to sustain such a rapid rotation. But nobody has ever observed anything that small and compact. Then the researchers recalled a theoretical prediction made as early as the late-1930s of the possibility of the existence of *neutron stars*. Such a possibility followed from the theory of evolution of a certain type of stars considerably more massive than our Sun. When the nuclear oven inside such a star runs out of its fuel, the star's huge gravity can no longer be balanced by the internal pressure. The star keeps shrinking beyond a white dwarf's size, until all the electrons are pressed into the nu-

clei and merge with the protons, giving rise to the emergence of new neutrons in addition to those already present in nuclei. The whole star is thus converted into a congregate of neutrons. Hence the name of a newly born object – the neutron star. The whole mass of the former star is now compressed into a size of 5–50 km. The resulting mass density is stunning. If we would want to get something that dense here on earth, we would need to squeeze the whole of mankind into a raindrop. The unimaginably high density of nuclear matter produces pressure that can stop the advance of gravitational collapse, so that the neutron star is stable. The reader can find a brilliant account of the physics of neutron stars in the already mentioned book by Kip Thorne [33].

Now the question arises of why the neutron star should rotate so rapidly. This rotation is actually inherited from the parent star. The angular momentum of an isolated system is conserved just as is its linear momentum. When a star shrinks, its angular momentum must remain constant (assuming that its mass M does not change). Therefore, the reduction of the size must be compensated for by the increase in its angular velocity ω. The more a body shrinks, the faster it rotates.

This phenomenon is often used in skating. A skater first begins spinning with her arms stretched out. When she pulls the arms in, her rotation rate suddenly increases, producing an impressive finish.

If a star like our Sun (which makes one full rotation in about 27 days) were to turn into a neutron star of radius 100 km, its new frequency would be one rotation in just about 3×10^{-2} s.

Now we want to understand why and how a neutron star can radiate a directed flux of energy. The answer lies in the fact that together with rotation, the neutron star also inherits from the parent star its magnetic field. The field undergoes the same "procedure" as the rotation. As the star contracts, so does its magnetic field, until it becomes a billion times stronger than before. The resulting object is an incredibly powerful rotating magnet. The magnetic field lines connecting its magnetic poles can channel the motion of the charged particles and their radiation in the atmosphere of the neutron star [36, 37]. Imagine now that the magnetic axis of the magnet does not coincide with its rotational axis. Assume for simplicity's sake that these axes make an angle of 90° with one another. Then we have a rotating beacon (or a cosmic laser gun, for that matter), continuously emitting a beam of radiation sweeping out the equatorial plane. The resulting beam has a spiral shape similar to that of the Captain Fletcher's rotating laser. In the general case, when the rotational and magnetic axes of the neutron star make an arbitrary angle, the beam traces out a conical surface. Now, if our solar system happens to be close to such a surface, the beam periodically "thrashes" the Earth, and at these moments our detectors record splashes of radiation. The time interval between the two closest splashes is just the period of rotation of the source. Each new splash is seen as a spike on a recording tape. It is produced by a beacon that circumnavigates along a circle whose radius is a distance between the Earth and the rotational axis of the source. Typically, this distance (and accordingly, the length of the circle) is of the order of hundreds of light years. And the "light spot" makes it around the circle in just a fraction of a second! Obviously, it must be moving with practically infinite speed. Thus the lasers of Cap-

tain Fletcher and corresponding superluminal spots have been realized in Nature on the cosmic scale. But in both cases the apparent "rigid" motion of a spiral beam and superluminal velocities of its remote parts is a purely geometric effect. The *position* of the spiral changes *as if* it were a rotating rigid body. In reality, however, the spiral is not rigid and not even the same body. Its constituting material (electromagnetic field) is continuously renewing, spurting out of the rotating source as does water from a rotating nozzle of a garden hose. And, by the way, the argument presented in this section, is, in its essence, entirely applicable for a rotating water stream. If you water a sufficiently remote screen with such a stream, then the watered area can also move along the screen faster than light. However, the water in the stream itself (that is, each individual water particle) moves practically in a radial direction with normal speed $v \ll c$. The wet spot in each new place of the screen is formed of the new water particles from new parts of the stream. The motion of the wet spot is therefore not the motion of the same object, and its superluminal speed does not contradict anything.

6.7 Surprises of the surf

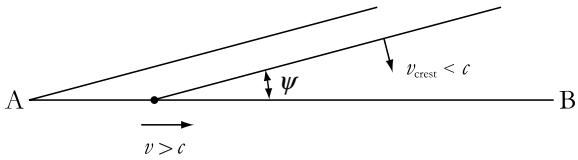
Having just mentioned water, let us consider another situation involving water: waves on the sea. Let them run at a tilted angle against a straight segment of the shore. If we compare the sea shore with the corresponding line on the screen, and the crest of a sea wave with the crests of the electromagnetic wave in the spiral beam in the previous section, we will find a certain difference between them. In the case of the electromagnetic waves in Section 6.6, the corresponding part of the spiral makes an acute angle with the screen, while the wave crests are almost parallel to the screen. Now, in the case of sea waves, the wave crests are inclined to the shore line (Fig. 6.11). Still, this difference is not essential. What is important is that both cases allow us to realize superluminal motion. Indeed, if ψ is the angle between the shore and the crest of a sea wave, then their point of intersection (which can be represented by a real breaker!) will run along the shore with a speed

$$v = \frac{v_{\text{crest}}}{\sin\psi} \tag{28}$$

If now ψ is sufficiently small (which is almost always the case!) then the breaker will run faster than light. In particular, for $\psi \to 0$ (all wave crests are parallel to the shore), $v \to \infty$ (all segments of the crest reach the shore line simultaneously), and the corresponding breaker instantaneously "traces out" the entire line AB (Fig. 6.11).

A reader who has observed the surf will possibly be surprised that superluminal motion can be realized in such a commonplace, everyday phenomenon. But in this situation, too, the superluminal breaker, for all its reality, cannot be used to transfer information faster than light. As in the above-considered cases, our breaker does not consist of one and the same body of water. If a sea wave casts ashore a bottle with a

Fig. 6.11



sealed message, this bottle will remain at rest on the shore (or will be washed away by another wave), but it will not rush at a superluminal speed along the shore. This phenomenon has another, more relevant, electromagnetic analogue apart from the spiral beam: the incidence of a plane wave front on the interface between the two different mediums. The intersection line between the wave front and the interface moves with a speed

$$v = \frac{c/N}{\sin\psi} \tag{29}$$

where N is the refractive index of the medium. This equation is completely analogous to Equation (28) and shows that at $\sin\psi < N^{-1}$ the "bright spot" (the optical analogue of the breaker) runs along the interface at a superluminal speed, which, however, cannot transport a signal.

6.8
The story of a superluminal electron

At the beginning of the 20th century Arnold Sommerfeld published a paper describing the electromagnetic field of an electron moving in a vacuum faster than light [38]. His basic result was that such an electron could not move uniformly with a constant speed even if it was in a free space with no external forces applied to it. The reason for this was the radiative force exerted on the electron by its own electromagnetic field. This force must be directed opposite to the electron's velocity, thus decreasing its momentum and energy. The energy lost must be carried away with the radiated field. In a nutshell: a superluminal electron would emanate light, radiating away its energy.

The physical nature and origin of such radiation are easy to understand from a well known analogy in hydrodynamics and acoustics.

Suppose you are in a boat that does not move. If you are jumping up and down, the boat bobs up and down, and there emerge diverging concentric waves. If the boat is moving, we can observe the Doppler effect, because the centers of the consecutive circular waves diverging from the boat are no longer in one place (Fig. 6.12 a). As a result, the waves are compressed in front of the boat and extended behind the boat. We can find quantitatively the amount of change in the wavelength if we write $\lambda = uT$, where u is the speed of the wave on the water surface and T is the period of the wave. When the boat is moving with a speed v, it will travel the distance $\Delta = vT$ during one period, so the distance between the two neighboring waves will be $\lambda \pm \Delta$

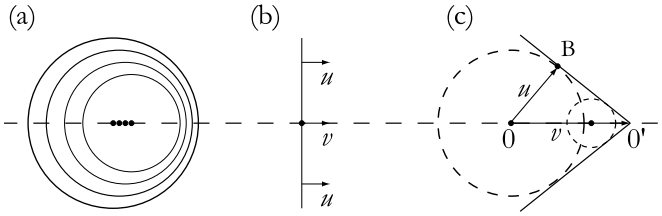


Fig. 6.12

(the + sign for the waves behind and the – sign for the waves in front). Hence we can write for the Doppler-shifted wavelength

$$\lambda' = \lambda \pm \Delta = \lambda \pm vT = \lambda \left(1 \pm \frac{v}{u}\right) \tag{30}$$

If the boat moves as fast as waves, it keeps abreast with the first wave, all the rest being piled up on top of one another, building a big splash. The same can be seen from Equation (30): when $v = u$, the distance between the two neighboring waves goes to zero, they are all compressed into one (Fig. 6.12 b). This one, straight wave front can be considered as an embryo of what happens when the boat moves faster then the surface waves. Equation (30) gives a negative wavelength in front of the boat for this situation. Physically, it means simply that there are no waves in front, since the boat outruns them. We then see two symmetrical breakers, trailing the boat on either side. They emerge as an envelope of concentric waves diverging from the boat (Fig. 6.12 c).

The moving boat does not have to bob to produce these waves. They are produced by the mere fact of its motion, because it stirs the water surface.

The faster the boat moves, the sharper is the "wedge" formed by the two trailing side waves. If the boat moves with an infinite speed, it traces out its trajectory all at once; we can consider this trajectory as an instantaneous linear source of circular waves, and their envelope will form two parallel crests receding symmetrically from the trajectory of the boat.

Once we have understood the simple mechanism of formation of waves from a source that outruns them (as envelopes of overlapping circular waves,) it is easy to describe quantitatively the shape of the envelope. The wave from an initial position O of the source will form in 1 second a circle of radius u centered about this position. By this time the source will be at a new position O′ a distance v from O. The two tangential lines from O′ to the first circle around O represent the envelope of all intermediate circular waves. From the triangle OO′B we have

$$\cos\theta' = \frac{u}{v} \tag{31}$$

Here the angle θ' gives the direction of motion of the envelope with respect to the direction of motion of the source. Consider a few special cases. When $v = u$, we expect the envelope to move with the boat in the same direction. Equation (31) gives just

that: $\cos\theta' = 1$, $\theta' = 0$. When $v = \infty$, we expect from the above qualitative treatment the angle θ' to be 90°. Equation (31) gives just that: $\cos\theta' = 0$, $\theta' = \pi/2$. The equation describes correctly the motion of the two envelopes.

A similar effect is produced by a moving supersonic jet. The only difference is that now we at each moment see a three-dimensional picture – an envelope of spherical waves in air instead of circular waves on a two-dimensional surface. The envelope of all the spherical waves forms a conical surface trailing the jet. Such a surface can be obtained by rotation of Figure 6.12c about the symmetry axis – the trajectory of the source – and is called the Mach cone. This surface of the cone moves perpendicularly to itself, and the direction of motion is given by the same Equation (31). As in the previous case, there are no waves outside the cone; but when the conical surface reaches the observer, he hears a sudden sharp "boom" because of a big perturbation there due to a large number of elementary waves piled up together. Therefore, the Mach cone is also called a shock wave.

Now we can get back to electrodynamics. Sommerfeld showed that the picture of motion of a superluminal source in vacuum is in many respects similar to the picture of the supersonic motion of a jet or projectile. The theory developed by Sommerfeld predicted the formation of the Mach cone around the trajectory of the corresponding source (a superluminal electron or any other superluminal charged particle). The only difference is that the cone is formed of the electromagnetic field rather than acoustic waves. Although the stationary electron produces a spherically symmetrical electric field around itself, this field becomes "Lorentz-contracted" for the uniformly moving electron (Section 5.5). This "contracted" field follows slavishly its "master" – the electron that produces it. When the speed of the electron reaches c, its field is flattened to a plane perpendicular to the direction of motion. The field in the plane is very strong because it is built up of all its former parts that had previously been spread throughout the whole space. If the electron moves faster than c, its field cannot keep up with its "master" any more, and forms instead a trailing Mach cone similar to one described above for a supersonic jet. As in the case of a jet, there is no electromagnetic field outside the cone. The field in front cannot form when it is outrun by its source. Therefore, the observer first does not observe any field, and then suddenly detects a "boom" – a rapid emergence of strong electromagnetic perturbation (electromagnetic shock). The direction of propagation of the perturbation is again given by the same Equation (31), if we put there c instead of u.

Sommerfeld showed that the electromagnetic field of the cone carries away energy and momentum. They are radiated from the electron. Accordingly, the electron must lose its energy and momentum. The loss of momentum is experienced by the electron as a braking force directed against its motion. This braking force must be proportional to the radiation rate. Sommerfeld obtained a simple equation relating the radiation rate (power) W to the radiation force f:

$$W = f v \tag{32}$$

It is similar to a familiar equation for a power consumed by a moving car experiencing a resistance force f.

However, when Sommerfeld actually tried to calculate the electromagnetic field on the cone, it turned out to be infinite! Accordingly, the radiative force and radiative losses also turned out to be infinite.

Sommerfeld understood that these infinities were not due to the superluminal speed of the electron. They resulted from the fact that he used the model of a point-like electron. Such a model was a notorious trouble-maker in classical physics, even in the well behaved mode of subliminal states. For instance, the electric field strength of a stationary electron becomes infinite as the observation point approaches the electron, and so is the amount of electromagnetic energy, stored by its field. Therefore, Sommerfeld modified his theory by assigning a finite size to his electron. He considered the simple model of a charged sphere of a finite radius a. But he did not know then what we know now – that such a sphere should undergo a certain distortion in the longitudinal (parallel to velocity) direction due to Lorentz contraction. He considered the sphere as absolutely rigid, which, as we know (see Section 5.5), is impossible. Therefore, the results obtained for this model also suffered from some inconsistencies. Soon after there appeared Einstein's theory of relativity, which showed that electrons cannot move faster than light. Therefore, the whole problem lost its appeal, and was abandoned for about half a century.

After that gap, the story of the superluminal electron experienced an unexpected rebirth. It started in Moscow, where a prominent Russian physicist, Sergey Vavilov, and his postgraduate student Pavel Cerenkov had studied optical effects accompanying the motion of high-energy charged particles in water or some other transparent medium. As early as 1934–36 they observed a new phenomenon: when such a particle (usually an electron or its counterpart – a positron) moves sufficiently fast, the medium becomes luminous. The intrigued researchers tried to determine the physical nature of this phenomenon. They tried various approaches and offered various explanations – all in vain. The effect that they observed was not the regular luminescence – the afterglow of some materials exposed to radiation. It did not look like anything familiar from previous experience. The researchers published their discovery with a detailed description of the experiments carried out and the behavior of the new kind of luminescence. The phenomenon was named the Vavilov–Cerenkov effect (in the West, just the Cerenkov effect.) An explanation came later: it was intuitively clear and totally unexpected. Two other Russian physicists, Igor Tamm and Ilia Frank, showed that the mysterious light observed by Vavilov and Cerenkov was coming from superluminal electrons! The "faster than light" charged particles, first introduced by Sommerfeld and banned by Einstein's theory of relativity, were gloriously resurrected into being. Was Einstein wrong? No. Nature just displayed another of her tricks: a particle in a medium can be superluminal *without violating Einstein's ban*. The resolution of the puzzle was very simple. The speed of light in a medium may differ from that in a vacuum by a factor of n. If n is greater than 1, which is the case for most media in a broad range of frequencies, then the speed of light in the medium is accordingly less than c. Thus, for water $n = 1.33$ in the visible range of the spectrum, so light in water moves with reduced speed $c/1.33 \approx 225\,000$ km s^{-1}. An electron can move faster than that, that is, it can be superluminal in water, without trespassing over the barrier c. And, because it outruns light waves in water, it produces the same trailing cone of

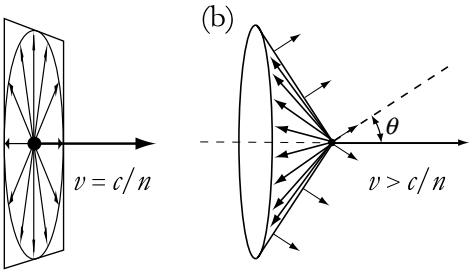


Fig. 6.13

electromagnetic radiation as described above (Fig. 6.13). This radiation makes the water shine and the electron lose its energy and slow down.

Tamm and Frank developed a detailed theory of this effect, which not only gave an explanation for all its observed features, but was also in excellent quantitative agreement with the results of measurements [39, 40]. Their theory was free from the flaws of Sommerfeld's previous treatment, in particular from the divergence of the radiative losses. The reason for this was that not all possible frequencies could participate in the formation of the Cerenkov cone. The high frequencies, as we shall see later, propagate in media with speeds inaccessible for particles; therefore, they do not form the cone, and without high frequencies the electromagnetic field of the cone, and thereby the radiation rate, remain finite. In 1958 Cerenkov, Tamm and Frank were awarded the Nobel Prize for the discovery, explanation, and theoretical description of the new physical phenomenon.

6.9
What do we see in the mirror?

Alice, Tom, and Peter were discussing relativity and, naturally, they started talking about light. For some reason, they came to the question of why light follows the reflection and refraction laws. They exchanged their knowledge on this topic, learned in college (they all attended different colleges). They were surprised to realize that in each college, their professors told them different things about the subject, and yet all arrived at the same laws. They decided that each should tell all the details about what he or she had learned in order for each of them to learn more.

Peter started first.

"Our professor told us that the behavior of light on the boundary between the two media can be entirely understood from Huygens' principle. Consider, for instance, a light wave incident on the water surface from air at an oblique angle (Peter drew Fig. 6.14 a). According to Huygens' principle, each point on the wave front can be considered as the center of the secondary waves propagating in all directions with the same universal speed c. We can find the position of the new wave front at a later moment as the envelope of all these waves. If we include the waves from the front's edge (where it intersects with the boundary) in this construction, we will obtain the additional front that appears to propagate from the water surface back into air. The

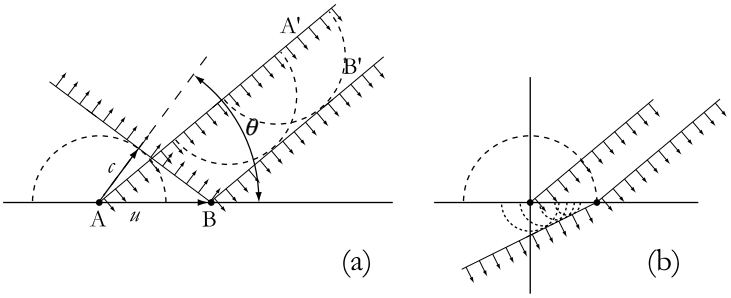


Fig. 6.14

directions of both fronts are symmetrical with respect to the normal to the water surface, which is the law of reflection.

We may also consider the spherical waves emitted from the edge of the incident front into the water (Fig. 6.14 b). In water these secondary waves propagate more slowly, with a speed c/n, where n is the refractive index of water. Therefore, the spheres in water are smaller than their counterparts in air. The envelope of the smaller spheres accordingly makes a smaller angle with the surface. The simple geometrical construction shows that the sine functions of these angles stay in the same proportion as the refractive indices for air and water ($n_r = n$, $n_i = 1$):

$$\frac{\sin\theta_i}{\sin\theta_r} = \frac{n_r}{n_i} \tag{33}$$

This is the Snell's law of refraction," Peter concluded.

"I did not know things can be explained so simply," said Tom, "our professor told us a different story. Here it is." Tom paused a while and continued.

"Imagine that you are on a sea beach. Suddenly, you see a beautiful girl (he looked at Alice) drowning. You rush to help her. If your way were all in air, you would naturally run along a straight line connecting you and the girl (Tom drew Fig. 6.15). But then you recall that you swim n times slower than you run. Knowing some physics, you understand that you would rather lose some time by running a longer distance in air, but in exchange, gain more time by swimming a shorter distance in water, so you would rather follow a broken line ACB than run straight towards the girl. The problem now is to find such a line that would give the least possible time."

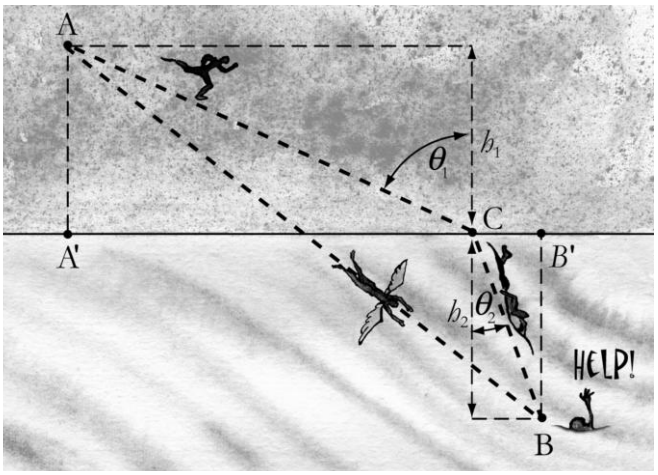
"I thought the problem was to save the girl," Alice broke in.

"Please, don't interrupt," Peter said, "we are doing physics now, and this is getting interesting. And one helps the other, anyway."

"Yes," Tom said, "and to do this we need some math." He looked at Alice again. "Be patient," he said, "the rescue is coming!"

"Look at Figure 6.15," he continued. "You see that the distance AC can be written as $h_1/\cos\theta_1$, and the time you need to run this distance with a speed v is $h_1/v\cos\theta_1$. Similarly, the swimming distance in water CB is $h_2/\cos\theta_2$, and the corresponding swimming time is $nh_2/v\cos\theta_2$. So the total time for you to reach the girl is

Fig. 6.15



$$T = \frac{h_1}{v\cos\theta_1} + \frac{n h_2}{v\cos\theta_2} \tag{34}$$

You can vary this time by changing the place where you enter the water, that is, by moving point C and thereby changing the angles θ_1 and θ_2. So the time T is a function of the two angles. We want this function to take its minimum value. We know that when a function is at its minimum (or maximum), its differential is zero. So we have to take the differential of T and set it to be zero:

$$\mathrm{d}T = \frac{h_1}{v\cos^2\theta_1}\sin\theta_1\,\mathrm{d}\theta_1 + \frac{n h_2}{v\cos^2\theta_2}\sin\theta_2\,\mathrm{d}\theta_2 = 0 \tag{35}$$

Now we notice that whatever the angles, they are always related by the condition

$$A'B' = h_1\tan\theta_1 + h_2\tan\theta_2 = \text{constant} \tag{36}$$

The differential of a constant is zero, therefore

$$\frac{h_1}{\cos^2\theta_1}\mathrm{d}\theta_1 = \frac{h_2}{\cos^2\theta_2}\mathrm{d}\theta_2 = 0 \tag{37}$$

If we use this, Equation (35) simplifies to

$$\sin\theta_1 = n\sin\theta_2 \tag{38}$$

which is Snell's law, Equation (33), for air and water. We see that Snell's law results from the tendency of light to select out of an infinite number of possible paths between the two points the one that requires the least possible time. Thus math can help us understand refraction better, and in addition save a girl's life."

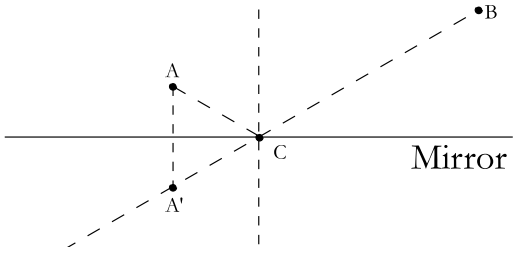


Fig. 6.16

"Quite the contrary," said Alice, "the poor creature would surely drown because of your math. She would rather have you choose any way and do running and swimming than do your math."

"Much depends on how curious the girl is and how fast you can do math," Tom answered.

"And how does this math explain the reflection law?," Peter asked.

"Oh, that's pretty simple. Look at Figure 6.16. Both the incident and reflected rays are in the same medium. The time to go from point A to the reflecting surface and then to point B is the shortest when the distance is the shortest. And the distance is the shortest when the image A′ of A is symmetrical to the original (that is AC = A′C), because in this case ACB = A′CB, and A′CB, being a straight line, is the shortest distance between A′ and B. Thus the principle of the least time explains both the refraction and reflection laws, and the reason why an image in a flat mirror is symmetrical to the original."

"The latter is simple, indeed," Peter said. "But how does light manage to pick up the path that takes the least time? It cannot stop for a while and do calculations."

"I think I could explain it," Alice said. "Our professor happened to have been a student of Richard Feynman himself. He said what Feynman had taught him, that light waves do calculations automatically by trying to propagate along all possible paths at once and allowing them to interfere. As a result, only those along the path of the least time survive."

Alice was very proud of her professor, so she tried to convey to her audience all she had learned from him.

"He also showed to us another way to explain the refraction and reflection. I remember him saying that there is no such thing as reflected light. The light coming from a mirror is not a marble bouncing off the wall. When you look in the mirror or on the water surface, you do not see the same light that was incident on it. That light is gone, it was either absorbed or passed straight into the medium."

"What do we see, then?" Tom asked.

"We see the secondary light re-radiated by the surface. The incident light excites the molecules on the interface, and they respond by emitting their own radiation."

"But how does this explain the refraction and reflection laws?"

"Oh, this has to do with the Cerenkov radiation from the superluminal sources. A source moving faster than light will radiate electromagnetic waves at such an angle that its cosine is the ratio of two speeds: the speed of light and the speed of the source." Alice wrote the equation

$$\cos\chi = \frac{c}{v} \tag{39}$$

from the previous section.

"Where do you see these sources?," asked Peter. "Such sources do not exist. Nothing can move faster than light. And the surface molecules do not move at all, unless you want to consider their thermal motion, which does not count here."

"It is not molecules themselves that move, but the waves of their excitations that propagate along the surface," said Alice. She pointed at Figure 6.14 drawn by Peter. "Look at this figure: the wave front incident on the surface excites the molecules at the intersection point B. The excited molecules emit secondary light. As the incident wave proceeds, the intersection point runs down the surface with the speed

$$v = \frac{c}{\sin\theta} \tag{40}$$

where θ is the angle of incidence. This is faster than light. Although the molecules themselves do stay in place, their excitation is being transferred from one to another with a superluminal speed. It is just a running illuminated spot, and this kind of superluminal motion is allowed. But the spot becomes the source of the secondary radiation! And it is this radiation that we see as the reflected light."

"How can you prove it?," asked Peter.

"Very simple! Consider the direction of this radiation given by Equation (39). Take into account that $\chi = 90^\circ - \theta'$, and use Equation (40) for v. Then you will get $\theta' = \theta$, which is just the law of reflection."

"The same superluminal spot," Alice continued, "radiates into the second medium. But light moves in this medium with a speed c/n instead of c. Therefore, Equation (39) becomes

$$\cos\chi_r = \frac{c}{nv} \tag{41}$$

Putting here again Equation (40) and taking into account that $\chi_r = 90^\circ - \theta_r$, we obtain

$$n\sin\theta_r = \sin\theta \tag{42}$$

which is Snell's law. So-called refracted light is nothing else but Cerenkov radiation from the running superluminal spot into the second medium."

Alice paused, and summarized:

"If we take a look at the Cerenkov cone as it is explained in the previous section, we will find that it is the envelope of the secondary spherical waves emitted by the moving source. Therefore, it is just another example of Huygens principle. In this respect, we come back to the original explanation of light behavior near an interface given by Peter."

"Yes," said Tom, "and if we take into account that interference of waves lies at the heart of the principle of the least time, as we found out today, we can conclude that

interference is the basic underlying concept explaining all the aspects of the discussed phenomena."

And everybody has agreed with Tom.

When I was writing this conversation among the three friends, I recalled a comprehensive review of related phenomena, published by Bolotovsky and Ginzburg [41] back in 1972. Speaking about refraction and reflection as a manifestation of Cerenkov radiation from superluminal sources, they wrote: "We can literally state that we did not know for a long time that 'we are speaking prose' and that the superluminal Cerenkov condition (as refraction and reflection laws – M.F.) ... has already been known for several centuries. The statements concerning the correspondence between the refraction and reflection laws, on the one hand, and Cerenkov radiation, on the other, are ... natural since all these relations are obtained from Huygens' principle in the same manner."

I also recall an article about parity violation in the weak interactions, with an impressive conclusion: "Next time, when you look in the mirror, think that maybe you see your own antiself." (A good description of the physical basis for such an overemphasized statement is given in [42].)

It is still not certain if there are or even can be the cosmic-scale regions of space with antimatter forming antigalaxies, antistars, antiplanets, and perhaps, antipeople. But it is surely clear that people had seen radiation from superluminal sources long before they started arguing whether superluminal motions are possible. So next time, when you look in the mirror, remember that you see the light from superluminal sources – waves of charges and currents – triggered by light coming from, among other things, your own self.

6.10
The starry merry-go-round

In this section we will discuss the apparent superluminal velocity which is related to a phenomenon known from time immemorial: the apparent rotation of the starry sky caused by the Earth's rotation on its axis. Neglecting a number of subtle points which are immaterial to us with regard to the questions in which we are interested, we can state that the Earth rotates as a rigid solid sphere about a fixed axis resting in space. Now we want to explore the motion of stars in this space. As always, we need a reference frame for this purpose. Let us use for such a frame what our ancestors had used for thousands of years: the Earth. But since the Earth is just a tiny speck in the vastness of space, we need to extend it in all directions. Of course, we can only do it as a thought experiment. So let us imagine a multitude of thin and arbitrarily long rigid rods extending from the Earth along radial directions, like the needles of a porcupine. To make it more rigid (ideally rigid, we want it!), let us add a system of rigid parallels and meridians similar to those on the Earth's surface (of course, the latter are also imaginary, but, imaginary as it is, we imagine it as ideally rigid). We obtain a spherical coordinate system which embraces the entire Universe and rotates relative to it. The rotation rate and direction are characterized by the vector $\boldsymbol{w}$, which is just

the angular velocity of the Earth's rotation. If some object is at rest in the inertial system K associated with the distant stars, it is moving in our rotating system with the velocity

$$\boldsymbol{v} = -\boldsymbol{w} \times \boldsymbol{r} \tag{43}$$

where $\boldsymbol{r}$ is the object's position vector originating from the Earth's center. If the object does not lie on the polar axis, then for sufficiently large $|\boldsymbol{r}|$, its velocity turns out to be greater than c. And, what is really disturbing, we are now talking about the motion of a real celestial body.

Stars give us the most direct confirmation of this argument. All of them do one complete revolution around the Earth in 24 h (more accurately, about 23 h and 56 min [17]), which means that they have one common angular velocity (that of $-\boldsymbol{w}$!) Therefore, their linear velocities are given by Equation (43), where $\boldsymbol{w}$ is independent of $\boldsymbol{r}$, and this inevitably implies superluminal velocities of stars. Even the nearest star, Proxima Centauri, is ~4.3 light years away from us. Its angular distance from the equatorial plane is relatively small, so that corresponding parallel circle in which it moves can be approximated by the equatorial line. I hope that all readers understand that this is not the equatorial line on the Earth's surface. This is the circle centered about the Earth's axis and drawn through the Proxima Centauri. Its length *in system K* is about $2\pi \times 4.3 \approx 27$ light years. In the *rotating* system associated with the Earth, this distance, as we know from Sections 5.7 and 5.8, is increased by a Lorentz factor. Even if we disregard this factor, it would take 27 years for light to travel along this gigantic circle. But Proxima Centauri makes this distance in 24 h! As a consequence, its speed is $27 \times 365 \approx 100\,000$ times greater than the speed of light. And this is the nearest star! What about more distant objects, other galaxies and quasars? Their velocities, calculated in the same way, will be infinite for all practical purposes.

The paradox thus obtained has to do with two important circumstances. First, any physically meaningful statement about properties of velocity must relate to a physical or *local* velocity (recall again Sections 5.7 and 5.8!). This velocity must admit, at least in principle, to its measurement with the meter-sticks and clocks positioned in the immediate vicinity of the given object. For instance, the reference clock in Section 5.7 and the detector in Section 5.8 for the measurement of circumnavigation times were positioned on the circumference, not at the center of the circle. Second, the mere possibility of such an arrangement of meter-sticks and clocks requires that the corresponding reference system in this area can be realized using material objects.

Neither of these conditions is fulfilled in our case. We have obtained the superluminal velocity of a star as a result of dividing the length of the path traversed by the star in some fantastically remote area by the Earth's time. In other words, the time is not being measured in the region where the given object is moving, and the distance traveled by this object is not being measured at all. We have just estimated it supposing that the rotation of the Earth can be extended continuously to infinity as a rigid-body rotation. But the latter assumption, as we had found in Sections 6.3 and 6.4, is wrong in principle: the clocks and meter-sticks positioned close to a star on a rigid

straight rod would have to move faster than light. Under such conditions, the Lorentz factor and thereby distance traveled and circumnavigation time of a star become imaginary. This reflects the impossibility of performing measurements in a rotating system sufficiently far from its center. It should come as no surprise that having assumed an impossible thing – superluminal velocities of measuring devices – we arrived at an impossible conclusion – that of superluminal velocities of the stars.

The superluminal velocities of stars, which we "observe" on Earth, is an artifact, because it is impossible to carry out a local experiment on an extended Earth to observe such a motion. The previously adduced estimate of stars' velocities is incorrect since it was based on the assumed extension of a rigid rotating system beyond its maximum possible radius $r_{max} = c/\omega$. Therefore, the computed superluminal values for the velocities of stars do not in any way relate to their real velocities. This example illustrates the pitfalls one should avoid when attempting to interpret our observations.

6.11
Weird dry spots, superluminal shadow, and exploding quasars

At the beginning of 1971 there came a report by a researcher at the Massachusetts Institute of Technology, Irvin Shapiro [43], which stunned many people like a bolt from the blue. Shapiro had observed the emission from the distant quasar 3C279 with the purpose of measuring an effect described by the General Theory of Relativity – the deflection of radio waves in the gravitational field of the Sun. He had used a high-precision device, a radio-interferometer with super-high resolution, which allowed one to distinguish the details of the quasar. Shapiro's observations, unparalleled for their precision at the time, suggested that the separate parts of the quasar 3C279 were flying apart with a speed about 10 times greater than the speed of light! Such "superluminal" expansion has since then been observed for at least a few other quasars.

This does seem to be a direct manifestation of the real superluminal motion of the physical bodies. But let us not jump to conclusions too fast. We had already realized more than once that an *appearance* may, upon closer examination, turn out to be something different from what it seems to be. So, let us suspend judgement and return from the distant galaxies back to Earth for a while. We will see that fairly simple college physics for undergraduates may help us to understand the message from far reaches of the Universe.

We shall start by considering a situation that may seem somewhat boring: it is raining. Since the mathematical description of any situation involves some idealization, the rain here is different from real rain: it consists of infinitesimally small droplets that fill the whole space, leaving no dry air gaps. They all fall vertically with a constant velocity u upon a horizontal plane α. This grim monotony is disturbed by only one single peculiarity – a waterproof object S (an umbrella), under which a continuous vertical chain of small air bubbles emerges within otherwise uniform watery shroud. We assume the size of the umbrella and thereby the size of the emerging

bubbles to be small. Since there is no water inside the bubbles, we shall call them dry holes. A corresponding chain of such holes extending from under the umbrella can be called the dry thread.

First let the umbrella S be at rest, hovering at a certain height above the plane α. Then the attitude of the dry thread q is determined only by the direction of the falling droplets, that is, the thread will be parallel to the rain streams. It is vertical in our case, and there will therefore appear a small dry spot S' right under S on the plane α. From a geometrical viewpoint, S' is just a projection of point S upon the plane α. From the practical point of view, S' is the only place where a microscopic traveller could find shelter in this utterly unfriendly surrounding (Fig. 6.17 a).

Suppose now that the umbrella starts to move at a constant velocity $\boldsymbol{v}$ which makes an angle θ with the vertical direction. This will cause both the thread q and corresponding dry spot S' to move (Fig. 6.17 b). For our traveller to remain dry, he will now have to move together with S'. At what speed and in what direction should he move?

An obvious answer comes to mind immediately: since the dry spot S' is just a projection of the umbrella upon α, it has to move with the speed equal to the projection of the umbrella's velocity $\boldsymbol{v}$ upon α, that is, $v' = v\sin\theta$. Accordingly, the speed v' can never exceed v. Also, it will certainly move in the same direction as does the umbrella (more accurately, in the direction of the horizontal component of the umbrella's velocity). Finally, since the speed of the umbrella is less than the speed of light, v' must always be less than c.

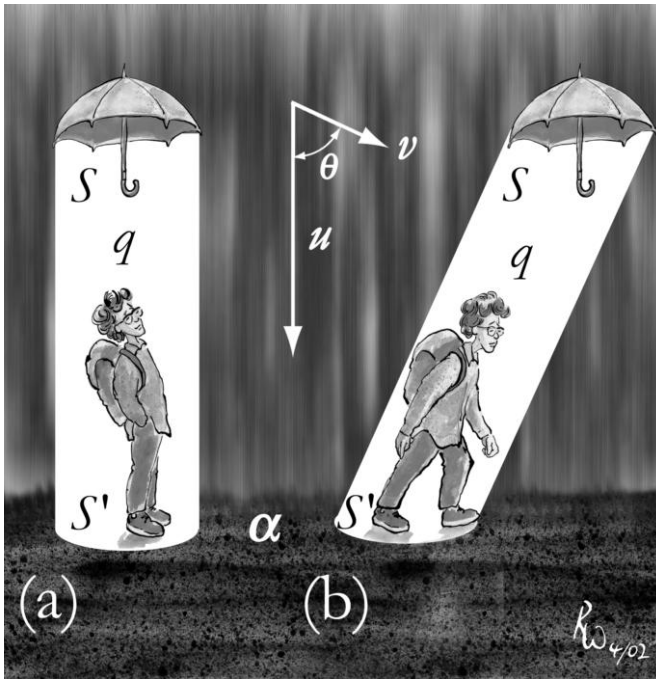


Fig. 6.17

Now, all the statements in this answer are wrong. They do not take into account the finite velocity of the projecting rain. Because of this, the projection is no longer an instantaneous event. It is a *process*, evolving with time. We could have ignored this in the static situation, but now the retardation time has to be introduced into our description. It is the time interval between the generation of a bubble right under the umbrella and its hitting the plane α. The retardation time is equal to h/u, where h is the umbrella's altitude at the moment when a given bubble is generated. Because of this retardation, each dry spot produced by the umbrella will no longer appear directly under it. It will rather lag behind. Further, since the umbrella's velocity is tipped below the horizon, its altitude h changes with time (i. e. is different for different bubbles); so is retardation, and, consequently, the lag. As a result, the speed of the dry spot will no longer be equal to $v \sin\theta$. The whole picture becomes more complicated and brings about unexpected consequences. To elucidate them, let us now solve the problem rigorously.

We need, first of all, to give a more rigorous definition of the velocity v'. We can do this using the above-introduced concept of a dry hole. The holes are being spilled out continuously from under the umbrella and fall down with the rain at a speed u. Consider two different positions of the umbrella S at the moments t_1 and t_2. A dry hole detaches itself from S at each of these moments and speeds straight down. Let t_1' and t_2' be the moments when these two holes hit the ground. Let $\Delta x'$ be the distance between the corresponding dry spots produced by the holes. We define the speed v' as

$$v' = \frac{\Delta x'}{\Delta t'} \tag{44}$$

where $\Delta t' = t_2' - t_1'$. Using this definition, we can find how v' depends on u, v, and θ (Fig. 6.18 a). The distance between the two positions of the umbrella at the moments t_1 and t_2 is $\Delta l = v \cdot \Delta t = v(t_2 - t_1)$, and therefore

$$\Delta x' = \Delta l \sin\theta = v \Delta t \sin\theta \tag{45}$$

The corresponding difference in the altitudes at these moments is $\Delta h = \Delta l \cos\theta = v \Delta t \cos\theta$. The time it takes a hole to travel this distance is $\Delta t_{12} = \Delta h/u = (v/u)\,\Delta t \cos\theta$. This is the difference between the corresponding retardation times. The time interval $\Delta t'$ differs from Δt by this amount:

$$\Delta t' = \Delta t - \Delta t_{12} = \Delta t \left(1 - \frac{v}{u}\cos\theta\right) \tag{46}$$

Equation (46) is equivalent to the well known expression for the classical (non-relativistic) Doppler effect: the period $\Delta t'$ between the arrivals of the two successive signals from a moving source changes by $(v_\parallel/u)\,\Delta t$ from its proper period. Here the quantity $v_\parallel = v\cos\theta$ is the longitudinal (parallel to the line of sight) component of the source velocity and u is the speed of the signal. In our case it is the dry holes,

with the particle S as their source, that play the role of the signal. Because $\Delta t'$ depends explicitly on the longitudinal motion of the source, the corresponding phenomenon is called the longitudinal Doppler effect (compare Section 5.3).

Putting now Equations (45) and (46) into Equation (44), we find

$$v' = \frac{v \sin\theta}{1 - \frac{u}{v}\cos\theta} \tag{47}$$

One can arrive at the same result using the concept of the dry thread. Notice that the point S' appears at the intersection of the thread with the plane α. The thread is carried down by the rain with velocity $\boldsymbol{u}$, whereas the source from which it is being "drawn out" moves with the velocity $\boldsymbol{v}$. As a result, the falling thread is parallel to the vector $\boldsymbol{w} = \boldsymbol{u} - \boldsymbol{v}$, and the problem reduces to determining the speed of advance of its intersection point with the plane α (Fig. 6.18 b). This can be done easily enough, and we leave it to the diligent reader to make sure that the solution of this elementary geometrical problem leads exactly to Equation (47).

Now, let us see what this equation tells us. First, we can see that if the denominator of Equation (47) is sufficiently small, the value of v' can become *arbitrarily large* and, in particular, it *can exceed the speed of light*. This conclusion does not contradict any laws of Nature, since it applies to an "object" that is not one and the same physical body. Two different positions of the moving dry spot are caused by the "hits" of two quite different dry holes. There are no limitations in Nature on the time interval between such events.

It can therefore happen that our traveller, in order to remain dry, would have to rush like a madman together with the dry spot, even if the umbrella is hardly moving. Furthermore, if v' does exceed the speed of light, then the traveller, no matter how hard he tries, can by no means keep up with the dry spot: the superluminal velocity cannot be reached by a physical body.

Second, if the denominator in Equation (47) becomes negative, the point S' will move in the direction opposite to that of the horizontal component of the umbrella's

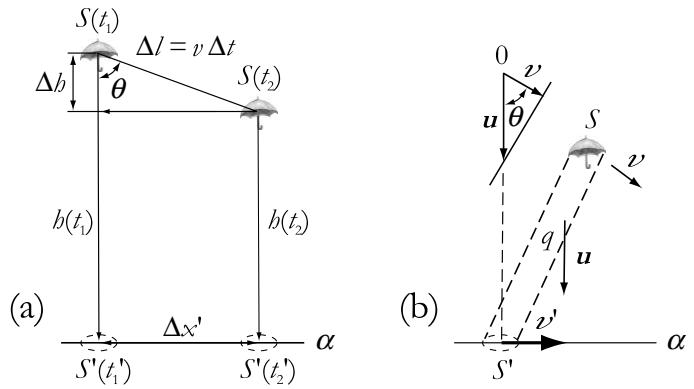


Fig. 6.18

velocity. For instance, whereas the umbrella will shift eastwards, the traveller, in order to avoid getting a running nose, will have to start running westwards.

Both results become obvious if we use the concept of a dry thread q falling on to the plane α. We already know that the intersection point between the thread and the plane can move arbitrarily fast. In particular, when $v\cos\theta = u$, the dry thread q is parallel to α, and all its points fall on to α simultaneously [Equation (46) then gives $\Delta t' = 0$]. This means that the point S' traverses its trajectory on the plane α instantaneously, i.e. $v' = \infty$ (Fig. 6.19). The reverse motion of the dry spot occurs when $v\cos\theta > u$; physically, this means that the umbrella is falling down *faster* than the rain, outrunning the rain drops, so that the associated dry thread, while falling down with the umbrella, remains above it (Fig 6.20 a). Equation (46) gives for this situation $\Delta t' < 0$. This means that of the two dry holes, the one that had detached from the umbrella first hits the ground last. It is readily seen from Figure 6.20 b that the first dry spot appears on the plane α at the moment when the umbrella crashes into the plane; it then moves away from the wreckage site in the direction from which the ill-fated craft had come.

If the vertical component of umbrella's velocity $v\cos\theta$ only slightly exceeds u, the denominator in Equation (47) is a small negative number. The value of v' is then negative and large. When its magnitude exceeds c, we will observe a negative superluminal velocity of the dry spot, which just means that it will move faster than light, and in the direction opposite to the umbrella's horizontal drift.

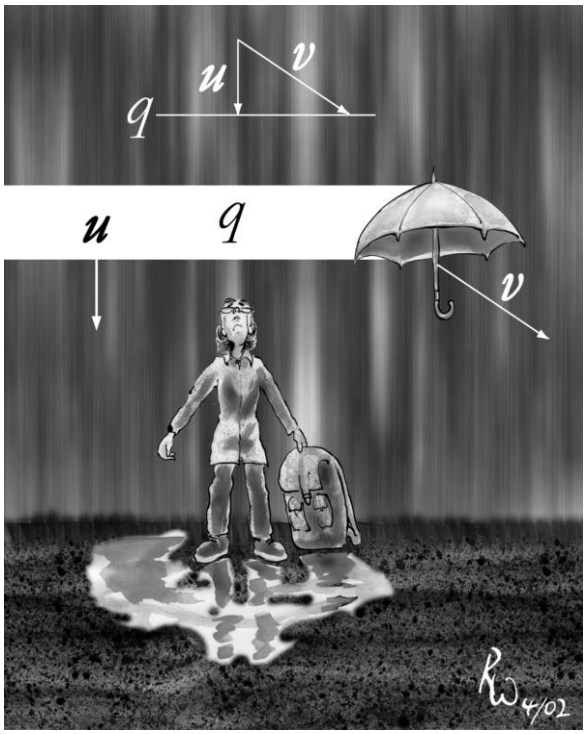
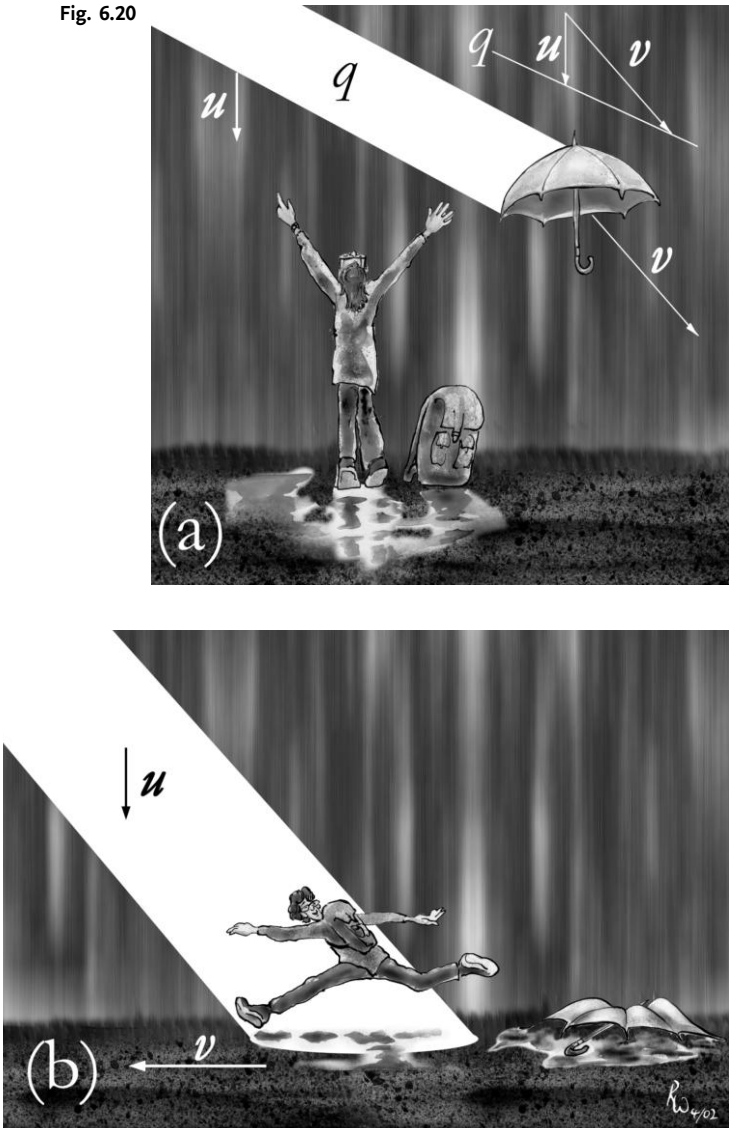


Fig. 6.19

Fig. 6.20



Let us make a brief preliminary summary of what we have learned. It amounts to a statement that one can observe superluminal velocities in such everyday situation as rainy weather on Earth. The dry image of a slowly moving object under slowly falling rain can move along the ground faster than light in a vacuum.

Some variations in the conditions of the problem are possible. For example, the rain can be replaced by broad beams of light streaming down. Then the velocity u in the above equations must be replaced by c, and Equation (47) for v' will take the form

$$v' = \frac{v \sin\theta}{1 - \frac{v}{c}\cos\theta} \tag{48}$$

Dry holes will in this case turn into dark holes, and the dry thread into dark thread (the shadow cast by the umbrella). The denominator in Equation (48) is now always positive since $v < c$. Therefore, the dark spot S' (the ray projection of the umbrella) can only move in the same direction as its geometrical projection. This motion, however, can also be superluminal if the right-hand side of Equation (48) is greater than c, that is, if

$$\frac{\beta \sin\theta}{1 - \beta\cos\theta} > 1 \tag{49}$$

where $\beta \equiv v/c$.

The interested reader can easily find the range of the values v and θ for which the above inequality holds. The corresponding area in the (v, θ) plane forms a shaded segment (Fig. 6.21). If the direction and the magnitude of the umbrella's velocity are such that the point (v, θ) lies inside the segment, the umbrella's shadow will move with superluminal velocity. We will call the corresponding phenomenon "the superluminal shadow." As can be seen from Figure 6.21 [or be obtained from Equation (49)], such a shadow can emerge under the given conditions if the umbrella's velocity exceeds the value $v_c = c/\sqrt{2}$.

The mathematical formulation of the problem will not change if we "reverse the signs" in the physical conditions: replace the incident light with darkness, the umbrella with the light source, and the tilted dark thread (the shadow line q) with the light ray composed of photons. These photons are being emitted by the source in the downward direction in exactly the same way as the dry holes (the air bubbles) had been "emitted" by the umbrella in the previous example (the concept of photons as point particles is not precisely correct, but this will not cause any errors in the following discussion.) The ray q can now be considered as the locus of all photons emitted

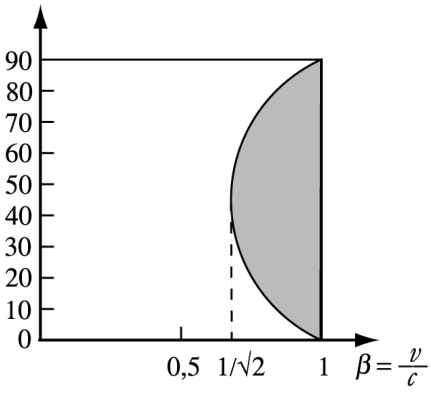


Fig. 6.21

by the source S in the downward direction at different moments of time. The ray strikes plane α, producing bright spot S'. This spot can be considered as the image of the source. More precisely, it is the image of a certain previous moment in the source's history, the moment when the given photon producing the bright spot was emitted by the source.

Well, after we have made all these substitutions, our simple mathematical problem suddenly turns from a funny trifle into a model of a grandiose astrophysical phenomenon. With this model, we will be far better equipped to interpret observations described at the beginning of this section. Imagine that there has occurred an explosion of a distant cosmic object, and the clots of hot plasma are flying apart in all directions from the sight of the cataclysm. One of the clots (denote it S) is moving towards the Earth with a velocity $\boldsymbol{v}$ making an angle θ with the line of sight to the object. Then the image of the clot projected by the ray q on the plane α is moving with a speed given by Equation (48). According to this equation, for the values of v and θ satisfying the inequality (49), the speed v' will be greater than c. And if we judge the phenomenon only by its appearance to an observer (by the moving spot S'), we will arrive at the wrong conclusion, that of the superluminal velocity of the source itself. This observational artifact has been called "the relativistic cannonball." If, for example, the relativistic cannonball is moving almost directly towards the observer, the latter will see the ball's optical image S' moving in the plane α perpendicular to the line of sight (let us call it a picture plane) at a superluminal speed.[1)]

Another example may elucidate the nature of this effect. Let a relativistic clot of plasma move toward the Earth with a speed equal to $0.99\,c$, but slightly sideways rather than directly toward us, so that in 300 years it will shift in the picture plane by 25 light years from its original position. But in the longitudinal direction it will come closer to us by $\sqrt{(300 \times 0.99)^2 - 25^2} \approx 297$ light years. The light emitted by the clot from its new position will have to travel 297 light years less than its predecessor. It will therefore reach us only 300 – 297 = 3 years after the arrival of the first signal. Consequently, it will appear to the observer on the Earth that the clot has shifted by 25 light years in the picture plane in only 3 years, in other words, it was moving with the speed of about $8\,c$! But this result is illusory, since it was obtained using the time interval $\Delta t'$ between the *receptions* of the two consecutive signals. This time interval can be much smaller than the time Δt between the *emissions* of the signals because of the change in the distance to the source (longitudinal Doppler effect).

Thus we arrive at somewhat paradoxical conclusion that the apparent superluminal *transverse* velocity is actually a manifestation of a *longitudinal* Doppler effect. The possibility of such a manifestation was predicted by a British astronomer, Martin Rees [44], as far back as 1966, yet for a long time it had not attracted much attention.

1) The described model is not precisely accurate. The real source emits light in all directions (this is why we can observe its different positions). Because of the immense distance to the source, its image on the photographic film may be shifted by only a fraction of millimeter per year even for the above-considered conditions. But the speed of the corresponding motion in the picture plane α (which is assumed to lie very far away from the Earth and close to the source) can be computed from the shift of the secondary image on the photographic film if proper account is taken of the distance to the source. This computed speed is the above speed v', and it can exceed the value of c.

But publication of Shapiro's results alerted the astronomers to the problem. Of course, the overwhelming majority of physicists and astronomers are now convinced that what we observe is not a real superluminal motion of physical objects, but in all likelihood merely the manifestation of the effect of the relativistic cannonball. This effect, as we have seen, can be well understood and simply described quantitatively within the framework of known physics.

Those readers who had anticipated the Shapiro's observations becoming one of the greatest discoveries shattering the very foundations of science may feel disappointed at seeing the would-be scientific revolution undercut by Rees's crystal clear model of the relativistic cannonball. They are wrong. The exploding quasars *are* a big discovery. *And* it may shatter foundations. Just think what kind of cannonball has been discovered here. Physicists of the 20th century were proud of being able to accelerate elementary particles such as a proton or an electron up to relativistic velocities. But we cannot accelerate *clots* of matter. There are two main obstacles in the way. First is the amount of energy needed. Suppose we want to accelerate a 1-mg grain of matter to $v = 0.8\,c \approx 240\,000\ \mathrm{km\ s^{-1}}$. Then the energy input into this speck must be $\sim 1.5 \times 10^{11}$ J. It is definitely greater than the total energy stored in the speck according to Einstein's famous equation $E = mc^2$, and is enough to lift 1.5×10^3 tons to the peak of Mount Everest. Because of the very low efficiency of the acceleration process, the actual amount of energy needed would be much more than that. Second, even if the energy needed is available, there is a problem of how to convert it into kinetic energy of the clot. The only efficient way that we know of now is to charge the body electrically by stripping its atoms of at least one electron each, and then accelerate the obtained charged clot in an external electromagnetic field. This scheme works fine for one atom (or the molecule, for that matter). It will not work for a macroscopic body, because the moment you have charged it in the above-described way (suppose you *know* how to do so, to begin with!), it would blow up into constituent particles flying apart faster than the whole thing could be accelerated. It would also blow up all around itself as if in a thermonuclear blast, since the original energy input needed for this preliminary charging is greater than that needed for the would-be acceleration. And this tendency for the initial electric energy to exceed the final energy of the accelerated clot becomes ever more impressive as you go to heavier clots. For a 1-g clot the ratio will be about 10^6:1, for 1 kg it jumps to 10^{11}:1, and so on. And this is with the kinetic energy of the accelerated clot itself increasing proportionally to its mass.

So, with all our physics and all the sophisticated, state-of-the-art paraphernalia associated with it, we do not know how to accelerate to relativistic velocities just a 1-mg speck of matter. And with all that, here we are, observing relativistic velocities for clots of matter – relativistic cannonballs – with a mass perhaps a million times greater than that of the Sun! This drastic contrast shows that Shapiro's observations *do* constitute a discovery although, perhaps, one of a special kind. It is not the discovery of superluminal velocities for physical bodies, so it does not seem to contradict any of the fundamental laws of physics, and we therefore do not see any necessity for a radical change of our picture of the world. It is rather a discovery of *new mechanisms* for acceleration of huge cosmic objects up to relativistic velocities. What are these mechanisms? What forces could accelerate to relativistic speeds gigantic,

almost unimaginable, masses of matter? Definitely, it is not an electromagnetic mechanism of the type described above. Nature is more inventive than we are. Most probably, the catastrophic explosions are caused by a peculiar combination of the electromagnetic and the gravitational forces associated with gigantic black holes supposedly lurking at the heart of quasars. But we are very far from a full understanding of what is going on there, let alone having a comprehensive theory describing it.

Thus, the discovery, and then a seemingly simple explanation, of the apparent superluminal velocities in the exploding quasars may put us at the beginning of a chain of new exciting discoveries that may well extend far into the new millennium.

6.12
Phase and group velocities

Up to this point we have considered the phenomena associated with light propagation in a vacuum. New phenomena occur when electromagnetic waves pass through a medium. One such phenomenon – refraction of waves – is caused by the fact that the speed of a wave in a medium is different from that in vacuum. It is no longer equal to c. Denote it as u. We know that the speed of a wave equals the product of its frequency f and the wavelength λ. The frequency f is not affected by the transition from vacuum to the medium. Therefore, the change in the wave's speed causes a corresponding change in the wavelength only, and in the same proportion. Generally, this change is different for diffent frequencies, and so both u and λ in a medium are functions of f (or of ω, for that matter). We can therefore write $u = u(\omega)$ and $\lambda = \lambda(\omega)$. The ratio of c to the light wave's speed u in a given medium is an important characteristic of this medium; it is called the refractive index and, according to the above note, is also the function of frequency:

$$n(\omega) = \frac{c}{u(\omega)} = \frac{\lambda_0}{\lambda(\omega)} = \frac{k(\omega)}{k_0} \tag{50}$$

where $k = 2\pi/\lambda$ is the wavenumber and the subscript 0 indicates the vacuum value of corresponding quantity. Since $k_0 = \omega/c$, we derive from Equation (50) a general relation between the wavenumber in a medium and corresponding refractive index:

$$k(\omega) = k_0\, n(\omega) = \frac{\omega}{c} n(\omega)\, c^2 \tag{51}$$

Since light is the coupled oscillations of the electric and magnetic fields, we can expect the optical properties of a medium to be determined by its electric and magnetic properties. Readers versed in electromagnetism will recall that the refractive index is completely determined by a medium's permittivity ε and permiability μ:

$$n^2(\omega) = \mu(\omega)\, \varepsilon(\omega)\, c^2 \tag{52}$$

In many cases the frequency dependence (we call it dispersion) is not very pronounced; the quantities involved change very slowly with the frequency within a certain frequency range. In such cases we can, as a first approximation, consider these quantities as constants. For instance, the refractive index of water in the optical range of the spectrum can, to good precision, be described by a constant number $n = 1.33$. Glass in the same range is characterized by $n = 1.5$. If you put these numbers in Equation (50) you will see that the speed of light in these two media is less than c: $\sim 2.2 \times 10^8$ m s^{-1} in water and $\sim 2 \times 10^8$ m s^{-1} in glass: nothing to be excited about.

The situation becomes different, however, if we shift to a higher frequency range or consider other states of matter, such as plasma. In the X-ray region the refractive index experiences a noticable change with frequency. This change can be described by

$$n(\omega) = \sqrt{1 - \frac{\omega_p^2}{\omega^2}}, \quad \omega > \omega_p \tag{53}$$

where the constant parameter ω_p depending on the properties of the medium, is called the plasma frequency. The term refers to the fact that at very high frequencies of electromagnetic radiation the electrons of the medium react as if they were free, as in plasma. Equation (53) holds for plasma also. In all cases it describes the propagation of radiation through the medium for frequencies greater than ω_p.

Now, if you take a closer look at this equation under the above conditions, you notice that the refractive index becomes less than 1. According to the definition of the refractive index, it means that the propagation speed of electromagnetic waves becomes greater than c at the high enough frequencies:

$$u(\omega) = \frac{c}{n(\omega)} = \frac{c}{\sqrt{1 - \frac{\omega_p^2}{\omega^2}}} > c \tag{54}$$

In other words, light of high enough frequency can propagate in matter faster than in a vacuum. It should be stressed that, unlike the cases with the "light spots" considered in the previous sections, we have here a superluminal velocity along the direction of propagation of the wave. And since it is *one and the same* portion of the electromagnetic energy advancing with the wave, it could be used as a proper carrier for a signal. With such a carrier, faster-than-light signaling seems possible with radiowaves in plasma or using radiation with sufficiently high frequencies through some other media.

This conclusion, however, is wrong, because a wave with the sharply defined frequency (monochromatic wave) cannot carry a signal. Such a wave exists for ever. It is moving permanently and fills the entire space with ever repeating uniform ripples. The electric charges in any detector exposed to such a wave would perform interminable harmonic oscillations (see Section 5.3). The oscillations, and with them the state of the detector, would never change. What a boring picture! No change – no message

– no information! The wave does not carry any signal and, consequently, its superluminal velocity does not violate any known physical laws. It is merely the velocity with which the wave's phase is moving in space. It can be envisioned, for instance, as a sinusoid [a wave profile represented by a sine function $y = y_0 \sin(kx - \omega t)$] sliding along the x-axis at a speed

$$u = \omega/k \tag{55}$$

Generally, it can be either smaller or greater than c, remaining in either case under the jurisdiction of the theory of relativity.

If we want to send a signal with the wave, we must somehow violate wave's uniformity, making a kind of a "notch" or "incision" in it, change its amplitude or the wavelength at some place, i. e. to disturb it in any way within some segment Δx. The arrival of the distorted segment of the wave at some place will change the state of oscillation of the charges in the local detector. The detector will record this change – it will receive a signal.

What is the velocity of this signal? An obvious answer is that since the signal is carried by a sinusoidal wave, it moves with the velocity of this wave. The conclusion is definitely correct for empty space. The light pulse from a laser is not a sinusoid, its instantaneous profile looks more like a spike, and this spike moves through the emptiness of space at precisely the speed given by Equation (55) for an infinite sinusoidal wave. Generally, however, Equation (55) does not hold for the the signal's velocity. In order to understand why it only holds for light in a vacuum, and does not hold in general, we must learn more about signals.

As we have already mentioned, any signal introduces a certain distortion in the original sinusoidal profile. Once distorted, it is no longer a sinusoid. The new profile (for instance, the just mentioned "spike" of a laser pulse) is described by a new function $y(x,t) \neq y_0 \sin(kx - \omega t)$. Any such function can be thought of as a combination of the whole *group* of monochromatic waves with slightly different frequencies within a certain region $\omega \pm \Delta\omega$ around the central frequency ω. In the case of a single pulse, these waves, adding together, reinforce each other in the corresponding region of space, producing a big splash there, because they all oscillate in phase; and they cancel each other outside this region, producing darkness since their oscillations there fall out of phase and cancel each other out. In the general case, a combination of sinusoidal waves with different periods can form an arbitrary profile. Its shape will depend on the amplitudes and initial phases of the individual waves. This remarkable feature of sinusoidal waves is exploited, for instance, in electrical engineering to shape out pulses with a desired form. The associated group of the individual waves (and also the resulting pulse) is called a "wave packet", and the velocity of the pulse is called the group velocity.

Now, consider a packet of electromagnetic waves in a vacuum. Each individual wave differs from the others in its frequency, but they all have one common velocity c. They form a sort of a "rigid" system moving as a whole. In this case, there is no other option for a resulting pulse (that is, for a signal) but to move in pace with all consti-

tuent waves (that is, without changing its shape). The conclusion is that when all the constituent waves of different frequencies move with the same velocity (no dispersion!), the resulting group also moves with the same velocity. In other words, the group velocity must in this case be equal to the phase velocity. This situation is described by Equation (55).

What happens when there is dispersion? The phase velocity will now be different for different waves. The system of waves is no longer "rigid": the individual sinusoids slide relative to one another. Correspondingly, the resulting pulse cannot retain its shape. Rather, it will spread out as it propagates. Now, what is its propagation speed? An obvious answer is that it must be somewhere close to the average phase velocity of the group. If the group propagates in a plasma, where all phase velocities are greater than c, then the group velocity must also be greater than c. This would suggest the possibility of faster-than-light communication through a plasma, or under water when using X-rays. If this conclusion is correct, then the theory of relativity is in real trouble, since it predicts that superluminal signal transfer is impossible.

Well, the above conclusion, obvious as it may seem, is not correct. It has been based on the false analogy between the group of waves and, say, a group of cyclists on a highway. For the cyclists, their group velocity can be roughly defined as the velocity of the guy riding somewhat slower than the first cyclist in the group, but faster than the last one. This will be close to the average velocity of the group. The situation with the waves is more complicated because a wave is not a localized grain of matter; it repeats itself interminably. To obtain correct result for waves, we must find a model retaining their periodicity.

Well, here is the model. Consider two very long trains on parallel tracks. They move in the same direction, but with different velocities u_1 and u_2. Each train has a row of windows through which bright beams of light shine on the wall parallel to the tracks (Fig. 6.22). The row models periodicity; the distance between the centers of the two closest windows is analogous to the wavelength; let these lengths be λ_1 and λ_2 for the first and second train, respectively; each window models a wave crest and each spacing between the two closest windows represents a trough. Suppose now that at a certain moment a window in one train is just opposite that of the second train. Then the beams out of both windows combine together, producing a spot on the wall twice as bright as each beam taken separately. The other windows in both trains, closest to the first two, will overlap only partially, producing dimmer spots on the wall. Let us call the system of spots a *group*, and the brightest spot its center. Consider this bright center as representative of the whole group, and call its velocity the group velocity. This completes the description of our model. Now, what does this model tell us about the group velocity?

If both trains moved at the same speed, the whole group would run along the wall with the speed of the trains. But the speeds are not the same. How will this affect the motion of the bright spot? We can visualize what happens before doing calculations. The original two windows producing the bright spot will move apart owing to the difference in velocities, so that at some later moment this spot will fade. Instead, the two other windows in the trains, closest to the original ones, will line up with each other, producing a bright spot of their own at some other place on the wall.

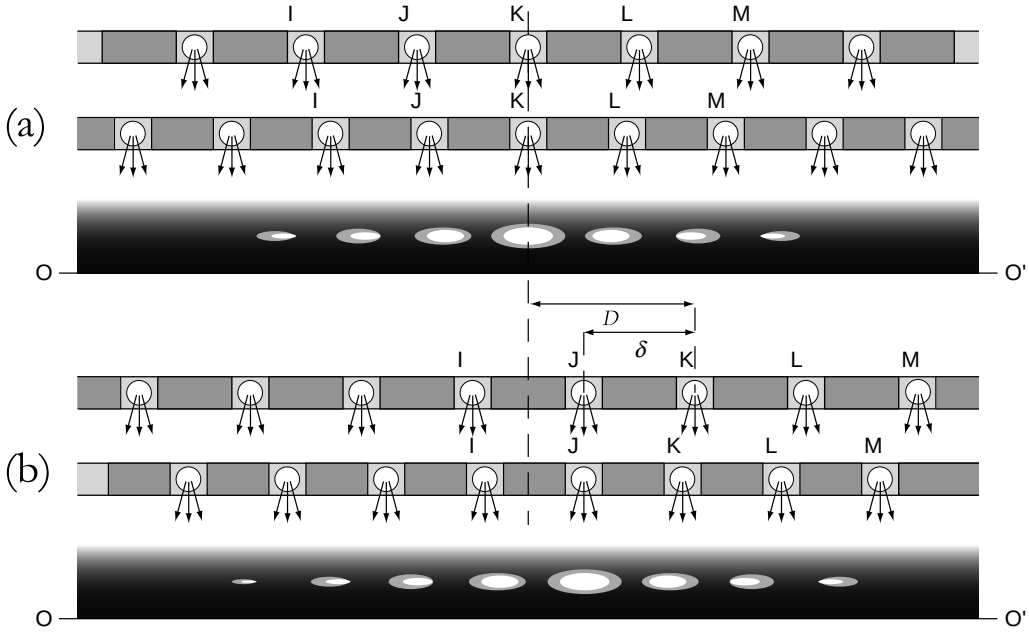


Fig. 6.22 A model illustrating the group velocity. The train on the upper track models the wave train with longer waves than that on the lower track. The train windows are labeled with capital letters. (a) The windows K of both trains are lined up, producing a bright spot on the screen OO′. (b) The windows J are lined up, producing the bright spot on the screen. The position of the second spot is a result of two motions: the shift D in the forward direction due to motion of both trains, and jump δ from K to J (for the situation depicted here – with the long waves moving faster). The group velocity is in this case smaller than the phase velocity (the bright spot moves slower than either train). This suggests that for waves moving faster than light, longer waves must move faster than the shorter waves, in order for the group to move slower than light. We will see later whether Nature complies with this requirement.

This spot is in all respects identical to the original one except for its location, so we can consider it as the old one just having jumped into the new place. Hence the resulting motion of the bright spot can be thought of as the combination of the continuous translational motion with the average velocity of the trains and the jumps between their windows. Because of these jumps, the resulting velocity v of the group may be quite different from the average velocity u. The difference must be proportional to the range of one jump, that is, to the average spacing between the windows. Since these spacings represent the wavelength, it must be proportional to λ. It also must be proportional to the difference Δu in the trains' velocities: the greater the difference, the faster will the bright spot fade and then resurrect in another place: the faster the jump. Also, it must be inversely proportional to the difference in the spacings (that is, the wavelengths): the greater this difference, the longer it takes for the next two windows to line up with one another: the slower the jump. Combining all this information, we can express the rather lengthy verbal description of the process by one short equation:

$$v = u - \lambda \frac{\Delta u}{\Delta \lambda} \tag{56}$$

For more sophisticated readers we can give a few more derivations. Let $u_1 > u_2$ and $\lambda_1 > \lambda_2$. Suppose you run with train No. 1, so that it stands still relative to you. The second train will move slowly relative to you with a velocity $\Delta u = u_1 - u_2$ in the opposite direction. Originally you are in the bright spot. However, since the second train is moving, some time later the spot will jump from you by a distance λ_1 to the next window of the first train. This jump will occur within the time interval it takes the second train to move a distance $\Delta\lambda$ (for its next window to line up with the next window of the first train). That is, $\Delta t = \Delta\lambda/\Delta u$. Thus, the velocity of the jump relative to you is $\lambda/\Delta t = \lambda\,(\Delta u/\Delta\lambda)$. Relative to the platform, we obtain for the group velocity Equation (56).

You may notice an inconsistency in this derivation: when commuting between the two reference frames, we have used non-relativistic transformation rules for the length, time, and velocity. In the relativistic domain, would this not cause a large error? The answer is: it might, if we were to remain in the moving reference frame. But we have not remained there: we boarded this frame for a while, wrote wrong expressions there, and then returned to the platform, writing wrong transformations once again. The errors made in these transformations, when performed in the two opposite directions, canceled each other out, producing the correct output. This kind of "error compensation" always happens when we actually remain in one reference frame. Similarly, we may not bother that it is physically impossible for a real train or observer to run together with the light wave, whose speed in plasma exceeds c. This is just a *thought* experiment. We cannot outrun light, but we *can imagine* doing so.

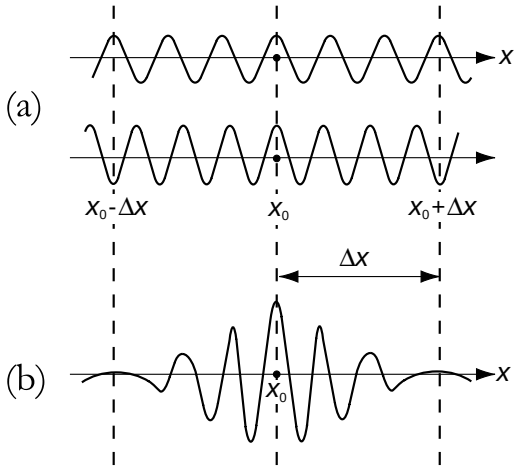
Now that our model has served its purpose, we can get rid of it, and go on to consider a real wave picture of the process. We want to see how individual waves in the group interfere with one another. Because each wave moves with its own speed $u(\omega)$, the relationships between their phases are always changing; for example, if at some moment the crest of one wave coincides with the crest of another wave, producing a peak there, then the next moment the crests slide apart, and the big splash at this point will decrease; the maximum will jump one wavelength to where the two neighboring crests become coincident. As a result, the envelope of all the waves and, in particular, its maximum, ride along the whole system at a speed that is completely different from those of individual monochromatic components.

To find the velocity of this maximum, consider two waves with close frequencies ω and ω'. At some moment t_0 the two crests of these waves coincide, producing the maximum at the corresponding point x_0 (Fig. 6.23). Consider the region $x_0 - \Delta x$, $x_0 + \Delta x$ on whose borders the two waves have opposite phases (Fig. 6.23).

$$\Delta\phi = \phi - \phi' = 2\pi\,\Delta x \left(\frac{1}{\lambda} - \frac{1}{\lambda'}\right) = \Delta x\,(k - k') = \Delta x\,\Delta k = \pi \tag{57}$$

The envelope of the two waves falls off from the peak value at x_0 to zero within the distance $\Delta x = \pi/\Delta k$ on either side of x_0. What will we see at some later moment

Fig. 6.23



$t_0 + \Delta t$? The upper and the lower train of waves in Fig. 6.23 will move to the right by the slightly different distances $u(\omega)\Delta t$ and $u(\omega')\Delta t$, respectively. Their crests that originally coincided will slide apart by $(u - u')\Delta t$. The maximum that they initially produced will blur and eventually disappear. Instead, the crests neighboring them which were initially separated by the distance $\lambda - \lambda'$ will now line up, producing a new maximum. Since both distances must be the same, $\lambda - \lambda' = (u - u')\Delta t$, we have

$$\Delta t = \frac{\lambda - \lambda'}{u - u'} = \frac{\Delta\lambda}{\Delta u} \tag{58}$$

The point where the new wave crests coincide will be shifted from x_0 by:

$$\Delta s = u\Delta t - \lambda \tag{59}$$

The corresponding group velocity is

$$v = \frac{\Delta s}{\Delta t} = u - \lambda\frac{\Delta u}{\Delta\lambda} \tag{60}$$

We see that v is different from the phase velocity u.

It is convenient to express v in terms of the wavenumber k rather than λ. This can easily be done by rewriting Equation (60) as

$$v = \frac{u\Delta\lambda - \lambda\Delta u}{\Delta\lambda} = \frac{u'\lambda - u\lambda'}{\lambda - \lambda'} = \frac{u'/\lambda' - u/\lambda}{1/\lambda' - 1/\lambda} = \frac{\omega' - \omega}{k' - k} = \frac{\Delta\omega}{\Delta k} \tag{61}$$

Concluding this set of derivations, we will obtain the same result in a more formal way similar to that used when analyzing the Michelson experiment. We just add together two sinusoids with the same amplitude E_0 but slightly different wavenumbers

k_1, k_2 and frequencies ω_1, ω_2, and apply the equation for the sum of the two sine functions:

$$E_0 [\sin(k_1 x - \omega_1 t) + \sin(k_2 x - \omega_2 t)] =$$

$$2\, E_0 \cos\left(\frac{k_1 - k_2}{2} x - \frac{\omega_1 - \omega_2}{2} t\right) \sin\left(\frac{k_1 + k_2}{2} x - \frac{\omega_1 + \omega_2}{2} t\right) =$$

$$2\, E_0 \cos\frac{1}{2}(\Delta k x - \Delta\omega t) \sin(k x - \omega t) \tag{62}$$

Here we put $k = \frac{1}{2}(k_1 + k_2)$ and $\omega = \frac{1}{2}(\omega_1 + \omega_2)$ for the average values of the corresponding quantities k_1, k_2 and ω_1, ω_2.

The last multiplier in Equation (62) describes a running wave with frequency ω, the wave number k, and phase velocity $u = \omega/k$. However, the amplitude of this wave is itself changing in space and time – it is modulated with the frequency $\frac{1}{2}\Delta\omega$ and the wavenumber $\frac{1}{2}\Delta k$ [compare with Equation (57)!]. It therefore has a maximum whose location is at any moment defined by the condition

$$\Delta k x - \Delta\omega t = 0 \tag{63}$$

It is moving at the speed $v = x/t = \Delta\omega/\Delta k$. This expression is identical with Equation (61). At $\Delta\omega \to 0$, $\Delta k \to 0$, both expressions become

$$v = \frac{\mathrm{d}\omega}{\mathrm{d}k} \tag{64}$$

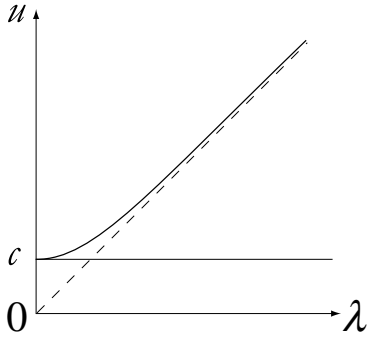
Thus, the phase velocity is equal to the ratio ω/k, whereas the group velocity at which the signal is transported, equals the derivative $\mathrm{d}\omega/\mathrm{d}k$.

Now we can explain the "paradox" with the superluminal propagation of the radiowaves in plasma. This propagation pertains entirely to separated monochromatic waves, which cannot transmit a signal. The latter can only be transferred with a *group* of waves. It remains for us to find the group velocity for the case in Equation (53) using Equations (60) and (51), and see if the result complies with the theory of relativity. Putting in Equation (51) the expression (53), we obtain

$$k(\omega) = \frac{1}{c}\sqrt{\omega^2 - \omega_\mathrm{p}^2} \tag{65}$$

Using Equation (65) and $k = 2\pi/\lambda$, we can express ω in terms of λ. Putting this expression into Equation (54) gives the phase velocity in terms of the wavelength. The corresponding dependence $u(\lambda)$ is shown in Fig. 6.24. We see that very short waves ($\lambda \to 0$) move with a speed slightly above c, and the speed increases with increasing wave-

Fig. 6.24



length. Beyond the light barrier, the longer the waves the faster they run! This is precisely the kind of behavior predicted by our simple model illustrated in Fig. 6.22.

Due to this behavior, the derivative $du/d\lambda$ is positive, and the Equation (60) gives for this case a group velocity, which is not only just *less* than the phase velocity, but less than c. Everything is tuned so neatly in Nature, that the greater the phase velocity, the grater the subtracted term in Equation (60). It can be readily seen from Fig. 6.24 that far from the origin, the derivative $du/d\lambda \to u/\lambda$. As a result, the Equation (60) not only brings the group velocity within the subliminal domain, but it even makes it approach zero when λ and u go to infinity. In this limit the corresponding frequency approaches the plasma frequency ω_p, and the wave's energy goes nowhere.

In general, it is easy to show that the considered group velocity is always less than c by applying its equivalent definition (64) to Equation (65).

It follows for v:

$$v = \frac{d\omega}{dk} = \left(\frac{dk}{d\omega}\right)^{-1} = c\sqrt{1 - \frac{\omega_p^2}{\omega^2}} \tag{66}$$

This is less than c! The paradox has disappeared.

For those who like to see things from different perspectives, it may be instructive to explore another approach to the problem. Recall that there is an energy flow associated with any wave. One can thus talk of the energy transfer even for one monochromatic wave, when there is no signal transfer. How fast, then, is energy transferred? We had already defined the corresponding velocity in Section 6.5. It is the ratio of the energy flux density to the energy density. The flux S in a medium is determined by Equation (17) that holds in general case. The only distinction is that the values E and B are now related to one another by

$$B = \mu H, \quad H = \sqrt{\frac{\varepsilon}{\mu}}\, E \tag{67}$$

Concerning the energy density, the expression for it changes in a dispersive medium. It follows from Maxwell's equations (see, for instance, Ref. [46]) that in such a medium

$$\eta = \frac{1}{2}\left[\frac{\mathrm{d}(\omega\varepsilon)}{\mathrm{d}\omega}\langle E^2\rangle + \frac{\mathrm{d}(\omega\mu)}{\mathrm{d}\omega}\langle H^2\rangle\right] \tag{68}$$

Here a quantity in brackets means its time average over the period.

If we express, using Equation (67), the quantities S and η in terms of only the electric field, we will obtain

$$S = \langle EH\rangle = \sqrt{\frac{\varepsilon}{\mu}}\langle E^2\rangle \tag{69}$$

$$\eta = \frac{1}{2}\left[\frac{\mathrm{d}(\omega\varepsilon)}{\mathrm{d}\omega} + \frac{\varepsilon}{\mu}\frac{\mathrm{d}(\omega\mu)}{\mathrm{d}\omega}\right]\langle E^2\rangle \tag{70}$$

We therefore find for the velocity of energy transfer (ET)

$$v_{\mathrm{ET}} = \frac{S}{\eta} = \frac{2}{\sqrt{\frac{\mu}{\varepsilon}}\frac{\mathrm{d}(\omega\varepsilon)}{\mathrm{d}\omega} + \sqrt{\frac{\varepsilon}{\mu}}\frac{\mathrm{d}(\omega\mu)}{\mathrm{d}\omega}} \tag{71}$$

This symmetrical expression can easily be transformed to

$$v_{\mathrm{ET}} = \frac{2}{2\frac{\mathrm{d}(\omega\sqrt{\mu\varepsilon})}{\mathrm{d}\omega} - \omega\left(\varepsilon\frac{\mathrm{d}\sqrt{\mu/\varepsilon}}{\mathrm{d}\omega} + \mu\frac{\mathrm{d}\sqrt{\varepsilon/\mu}}{\mathrm{d}\omega}\right)} \tag{72}$$

The expression in parentheses in the denominator is equal to zero and $\sqrt{\mu\varepsilon} = \frac{n}{c}$. Hence

$$v_{\mathrm{ET}} = c\frac{\mathrm{d}\omega}{\mathrm{d}(\omega n)} = \frac{\mathrm{d}\omega}{\mathrm{d}(\omega n/c)} \tag{73}$$

According to Equation (5.2), the quantity $\frac{\omega}{c}n$ is the magnitude of the wave vector in the medium. We finally have

$$v_{\mathrm{ET}} = \frac{\mathrm{d}\omega}{\mathrm{d}k} = v \tag{74}$$

Thus, energy is transferred with the group velocity! And the latter, as we had seen, is less than c even in the situations with the superluminal phase velocities.

In the simplest possible case, when motion occurs in vacuum, we have $\omega = ck$, and the group velocity is $\mathrm{d}\omega/\mathrm{d}k = \omega/k = c$, that is equal to the phase velocity. This is natural: if all monochromatic waves in a group have a common speed, the signal can only have the same speed. This is what allows us to talk about the same velocity c for both distributed electromagnetic waves and localized, ultrashort laser pulses! In other words, it is the absence of dispersion in vacuum that brings about velocity c as a fundamental constant of nature.

6.13 The de Broglie waves

Another unusual example of superluminal phase velocities occurs in quantum mechanics. We have already mentioned that light which seems to be continuous can also be represented as a flux of particles, each carrying a sharply defined portion (quantum) of the electromagnetic energy. The corpuscular characteristics of an individual quantum – its energy E and momentum p – are determined by the frequency and the wavenumber of the corresponding wave:

$$E = \hbar\omega\,, \quad p = \hbar k \tag{75}$$

where $\hbar$ is Planck's constant equal to ~10^{-34} J s. The idea that light possesses both wave and particle properties can be traced back to Planck and Einstein. A powerful complementary idea was introduced by de Broglie. He suggested that there must be a deep symmetry in Nature. If a light wave can turn out to an observer to be a particle, then for any particle of matter there also exists a certain kind of an associated wave, whose frequency and the wavenumber are determined by the particle's energy and momentum:

$$\omega = \frac{E}{\hbar}\,, \quad k = \frac{p}{\hbar} \tag{76}$$

The waves associated with the particles according to this equation have come to be known as the de Broglie waves. These waves have a probabilistic nature. Their intensity at a given point $\boldsymbol{r}$ at the moment t determines the probability to find the particle at the vicinity of this point if you look at it at this moment of time. Once the energy and momentum of the particle have been determined with an ideal accuracy, the frequency and the wavenumber of the associated de Broglie wave are determined with the same precision according to Equation (76). That is, we will have a monochromatic wave with a constant amplitude, and therefore this particle is equally likely to be found at any point in space. In other words, its location prior to observation is totally undefined. If, however, we carry out a measurement and find the particle localized within certain place, then, according to the probabilistic nature of the de Broglie waves, they must now have finite intensity at this place and the zero intensity elsewhere. However, this is only possible when there is a group of monochromatic waves with different frequencies, producing a splash inside the volume with the particle and canceling each other outside this volume. The localized particle can be identified with such a group. But this means that a localized particle does not possess a definite momentum; we have seen in the previous section [see Equation (57)] that a group of waves producing splash within a region Δx must have different wavenumbers within an interval Δk such that $\Delta k\,\Delta x = \pi$. It follows then from the de Broglie Equation (76) that a particle's momentum has different values within an interval

$$\Delta p \geq \frac{\hbar}{\Delta x} \tag{77}$$

whenever a particle is localized within a region of size Δx. This relationship between the uncertainties Δx and Δp in the particle's position and momentum follows immediately from the de Broglie postulate and the probabilistic interpretation of the de Broglie waves. This was first pointed out by Heisenberg and had since come to be known as his uncertainty, or indeterminacy, principle.

Now, for all these concepts to be consistent, the group velocity of the de Broglie waves must equal the real velocity of the associated particle. Let us check if this requirement is met. In both classical and relativistic mechanics, the velocity of a particle is related to its kinetic energy and momentum by

$$v = \frac{\mathrm{d}E}{\mathrm{d}p} \tag{78}$$

([Sect. 4.1, Eq. 13])

Taking the equation for the group velocity $v_\mathrm{g} = \mathrm{d}\omega/\mathrm{d}k$, and expressing in it ω and k in terms of E and p according to the de Broglie relation, we immediately obtain

$$v_\mathrm{g} = \frac{\mathrm{d}E}{\mathrm{d}p} = v \tag{79}$$

We see that, at least in this respect, the concepts of quantum mechanics do make sense! However, if we ask about the *phase* velocity of the de Broglie waves, the answer would be different. The phase velocity is given by the ratio ω/k, which, according to Equation (76), is equal to E/p. In relativistic mechanics, the energy and the momentum of a particle with the rest mass m are

$$E = mc^2\gamma(v)\,; \quad p = mv\gamma(v) \tag{80}$$

This gives for the phase velocity

$$u = \frac{E}{p} = \frac{c^2}{v} \tag{81}$$

that is,

$$uv = c^2 \tag{82}$$

Equation (82) displays a remarkable symmetry between the phase and group velocities. They are on different sides of the light barrier, and in such a way that the speed of light is always given by their product. Because the velocities of the material particles are less than c, the phase velocities of their de Broglie waves are always superluminal. In particular, for the particle at rest the associated de Broglie wave has infi-

nite phase velocity! But we are already mature enough not to get too excited about it. We know that the energy of the particle is transferred at the group velocity, which is subluminal, and for the resting particle is just zero. There is another thing worth mentioning, however: if the particle is really at rest – if it knows that it has precisely zero velocity – then, according to the uncertainty principle, it has no idea whatsoever about *where* in the world it is resting!

6.14 What happens at crossing of rays?

We consider here another example of superluminal velocity that can be realized in a vacuum and may be important for further discussions [45]. Recall a superluminal light spot running along the interface between two media 1 and 2 when a plane wave is incident upon it at a slanting angle. The velocity of the corresponding light spot is then given by Equation (40). We now want to get rid of the two media separated by a stationary plane interface. Replace this plane interface by a new wave front moving at the same speed as the first one. We obtain two crossed plane waves in the same medium. This medium can also be a vacuum, then $n = 1$, and it is just the case we will study. Both waves can be realized in the form of plane wave packets. Think of either packet as a very thin, luminous sheet. The two intersecting sheets rush each with a speed c in the direction of its normal $\boldsymbol{n}_1 = \boldsymbol{k}_1/k$ or $\boldsymbol{n}_2 = \boldsymbol{k}_2/k$, respectively (Fig. 6.25). Denote the angle between $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ as 2θ. The intersection line q between the sheets is sliding perpendicular to itself along the bisector of this angle with a speed

$$v = \frac{c}{\cos\theta} \tag{83}$$

The maximum energy density is concentrated along the line q. Think of it as of a rod that is brighter and moves faster than the two individual sheets producing it. Now we

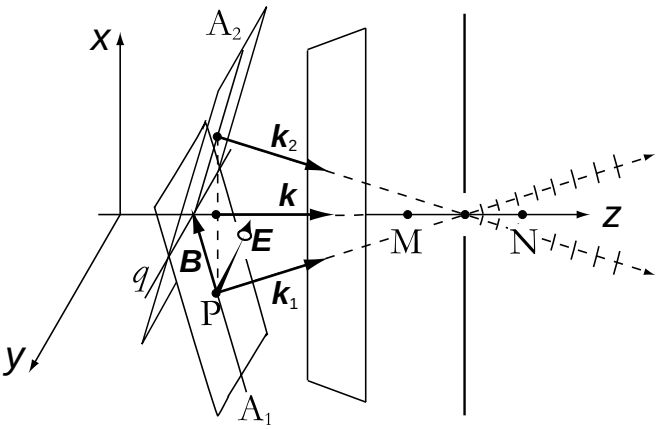


Fig. 6.25

have a region with high energy density (a "lump" of energy) that outruns light in a vacuum. We had found earlier that moving lumps of energy may constitute a signal whose velocity in a stable medium is just the lump's velocity. We therefore cannot avoid the conclusion that here we have a genuine case of superluminal signal transfer.

Our doubts will increase if we consider the surfaces of constant phase in a system of cross-waves. Let us first consider two monochromatic waves with the same amplitude $\boldsymbol{E}_0$ and the common frequency ω, each wave belonging to its packet 1 or 2, respectively. Let light in both packets be polarized parallel to the intersection line. Then the electric fields $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ in both waves oscillate parallel to the rod, and we can add them algebraically:

$$E(\boldsymbol{r}, t) = E_1(\boldsymbol{r}, t) + E_2(\boldsymbol{r}, t) = E_0[\sin(\boldsymbol{k}_1\boldsymbol{r} - \omega t) + \sin(\boldsymbol{k}_2\boldsymbol{r} - \omega t)] \tag{84}$$

where $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$ are the propagation vectors of the two waves, that is, their propagation number k multiplied by the unit vector $\boldsymbol{n}_1$ or $\boldsymbol{n}_2$, respectively. Adding the two sine functions in the same way as we did before, we obtain

$$E(\boldsymbol{r}, t) = 2E_0 \cos\left(\frac{1}{2}\Delta\boldsymbol{k}\boldsymbol{r}\right) \cdot \sin(\boldsymbol{k}\boldsymbol{r} - \omega t) \tag{85}$$

where

$$\Delta\boldsymbol{k} \equiv \boldsymbol{k}_1 - \boldsymbol{k}_2\,, \quad \boldsymbol{k} \equiv \frac{1}{2}(\boldsymbol{k}_1 + \boldsymbol{k}_2) \tag{86}$$

Thus vector $\boldsymbol{k}$ points along the direction of advance of the rod and vector $\Delta\boldsymbol{k}$ perpendicular to it and to the rod.

When computing the resultant magnetic field, we have to be more careful, because the magnetic field vectors $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ under given conditions do not oscillate along the same direction (Fig. 6.25). The summation is therefore to be carried out separately for the different components. Call the direction of propagation of the rod the positive z direction. Direct y along the rod, out of the page. Then the positive x will point up the page (Fig. 6.25). With these notations we have

$$B_x(\boldsymbol{r}, t) = B_{1x}(\boldsymbol{r}, t) + B_{2x}(\boldsymbol{r}, t) = B_{1x}\sin(\boldsymbol{k}_1\boldsymbol{r} - \omega t) + B_{2x}\sin(\boldsymbol{k}_2\boldsymbol{r} - \omega t) \tag{87a}$$

$$B_z(\boldsymbol{r}, t) = B_{1z}(\boldsymbol{r}, t) + B_{2z}(\boldsymbol{r}, t) = B_{1z}\sin(\boldsymbol{k}_1\boldsymbol{r} - \omega t) + B_{2z}\sin(\boldsymbol{k}_2\boldsymbol{r} - \omega t) \tag{87b}$$

But

$$B_{1x} = B_{2x} = \frac{E_0}{c}\cos\theta\,, \quad B_{1z} = -B_{2z} = \frac{E_0}{c}\sin\theta \tag{88}$$

This leads to the sum of the two sines in Equation (87 a) and to their difference in Equation (87 b), and we obtain

$$B_x(\boldsymbol{r}, t) = 2\frac{E_0}{c}\cos\theta \cdot \cos\frac{1}{2}\Delta\boldsymbol{k}\boldsymbol{r} \cdot \sin(\boldsymbol{k}\boldsymbol{r} - \omega t) \tag{89a}$$

$$B_z(\boldsymbol{r}, t) = 2\frac{E_0}{c}\sin\theta \cdot \sin\frac{1}{2}\Delta\boldsymbol{k}\boldsymbol{r} \cdot \cos(\boldsymbol{k}\boldsymbol{r} - \omega t) \tag{89b}$$

The surface of constant phase in Equations (85) and (89) is defined by the condition

$$\psi \equiv \boldsymbol{k}\boldsymbol{r} - \omega t = \text{constant} \tag{90}$$

This condition determines, for any given moment t, a family of the phase planes that are all perpendicular to the vector $\boldsymbol{k}$ (see Fig. 6.25). We can rewrite the last equations as $kr\cos\varphi = \text{const} + \omega t$, where φ is the angle between $\boldsymbol{r}$ and $\boldsymbol{k}$. The product $r\cos\varphi$ is just the distance z of a given phase plane from the origin. The above equation thus says that this distance changes with time as

$$z = \frac{\text{constant} + \omega t}{k} = \text{constant} + \frac{\omega}{k}t \tag{91}$$

According to Equation (91), any chosen plane of the set and thereby the whole set moves in the direction $\boldsymbol{k}$ with the speed

$$u = \frac{\omega}{k} \tag{92}$$

Now, it is readily seen from Figure 6.25 that

$$k = k_1\cos\theta = k_2\cos\theta = \frac{\omega}{c}\cos\theta \tag{93}$$

and therefore

$$u = \frac{c}{\cos\theta} \tag{94}$$

which is just Equation (83). We have carried out more detailed and rigorous calculations to emphasize two points here. First, there emerges a family of phase planes that are quite different from the original ones. Second, this family moves as a whole with the phase velocity [Equation (83)] exceeding c. And this seems to be more than just a superluminal motion of a mathematical point! Here we have come across a situation where the phase planes in a system of electromagnetic waves in vacuum move with the superluminal velocity!

We know that, generally, the superluminal phase velocities are possible in certain media and for certain frequencies. However, for an electromagnetic wave *in vacuum*,

both phase and group velocities are always equal to c. And yet, here we have the case that contradicts this law: a genuine paradox!

This paradox can be resolved if we look more attentively at the system of crossed waves. We then will notice that neither of the resulting waves [Equations (85) and (89)] is actually a plane wave. A wave is called plane if its surfaces of constant phase are infinite planes *and* if it has a constant amplitude in such a plane. The amplitudes of the waves in Equations (85) and (89) are different at different points of the wave surface [Equation (90)]. In particular, the values of E and B_x become zero on the "node planes" determined by the equation

$$\frac{1}{2}\Delta \boldsymbol{kr} = \pm\left(j+\frac{1}{2}\right)\pi \tag{95}$$

and the value B_z remains equal to zero in the planes

$$\frac{1}{2}\Delta \boldsymbol{kr} = \pm j\pi\,, \quad j = 0, 1, 2, \ldots \tag{96}$$

These planes divide each of the phase surfaces in Equation (90) into a system of strips parallel to the vector $\boldsymbol{E}$ so that amplitudes in any two neighboring strips have opposite signs. However, the sign of the amplitude, by definition, is always positive. The difference in sign simply means that the field components in the adjacent strips oscillate out of phase (Fig. 6.26). Thus the "phase surfaces" in Equation (90) turn out not to be what they "pretend" to be. They are not at all the surfaces of constant phase. It is true that the phase remains constant within each separate strip – after all, the phase does satisfy the condition (90)! – but it is not the same in the two adjacent strips. The phases in each such two strips differ by π. Also, this change of phase at the boundary between the two strips cannot even be called discontinuity. There is no actual jump here because the boundary is where the corresponding field component has a node. Now, if we look at the amplitude itself, we will see that it changes even within each strip, from zero on the edges to maximum in the middle.

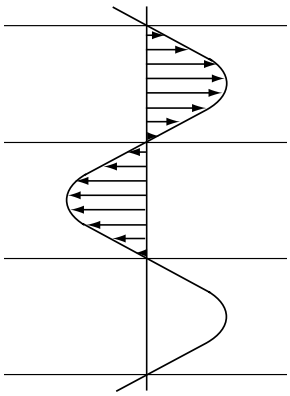


Fig. 6.26

We can find the width Δx of the strip, noticing that it is just the distance between the two neighboring node lines. From Equation (96) we have

$$\frac{1}{2}\Delta k \cdot \Delta x = \pi\,, \quad \Delta x = \frac{2\pi}{\Delta k} \tag{97}$$

Now, from Fig. 6.25 and Equation (86):

$$\Delta k = 2\,k\tan\theta = 2\left(\frac{\omega}{\mathrm{c}}\cos\theta\right)\tan\theta = 2\,\frac{\omega}{c}\sin\theta \tag{98}$$

Finally,

$$\Delta x = \frac{\pi c}{\omega\sin\theta} = \frac{\lambda}{2\sin\theta} \tag{99}$$

We conclude that the resulting wave propagating along $\boldsymbol{k}$ is by no means a plane wave. Accordingly, its velocity $u = c/\cos\theta$ cannot be called a phase velocity. It is just the velocity of a certain mathematical construction – system of strips. This is not very much different from the motion of the intersection point of scissor blades. And once we have "lost" the phase velocity, the notion of the group velocity in such a system also loses its meaning. The group velocity had initially been defined in terms of the phase velocity [see Equation (56)!], and therefore disappears with it. Putting it another way: we have defined a group as a set of plane waves *moving in the same direction*. Accordingly, their wave vectors were different in magnitude, but all collinear. What we have now is a set of *crossed* plane waves. Their corresponding wave vectors are equal in magnitude, but not collinear. This is a different physical system, which we can no longer call a group (at least not in the above-considered sense). Only in the limit $\boldsymbol{k}_1 \to \boldsymbol{k}_2 \to \boldsymbol{k}$ (that is, $\theta \to 0$) is the strip of constant phase broadened to sweep across the whole plane, and the system of waves becomes what we had called a group; but then, according to Equation (83), $v \to c$. In the general case $\theta \neq 0$, we actually have the two different groups propagating each with speed c in one of the two different directions symmetrical with respect to the z-axis. The superluminal rod that appears along their intersection slides along z *without carrying any signal in this direction*. To prove this, let us set up an opaque screen with a slit in it just large enough for the rod to pass through it and transfer its energy from the point M to point N on the opposite side of the screen (Fig. 6.25). The screen will absorb all the energy of the waves that "feed" the rod except for those parts that pass through the slit. After passing through it, the original rod will split into two new rods, which will now each move with velocity c at angles θ to the z-axis and, in addition, spread in the transverse directions owing to diffraction on the slit. There will be no energy transport from M to N since the original rod will just disintegrate [45].

Someone may want to argue that this negative result is caused by the introduction of the screen which destroys the conditions for the stable motion of the rod, whereas with no screen the rod would carry a superluminal signal. This argument is self-dismissive: by assuming that the screen destroys the conditions for the stable motion of the rod

you admit that the energy does *not* flow *together* with the rod whose path MN remains unobstructed. But even granting the validity of the objection, we will now show that there is no superluminal transport of energy even in a free space with no screen. Consider the two different positions M and N in the undisturbed motion of the rod (Fig. 6.25). The velocity of the rod is $MN/t = c/\cos\theta$. However, this velocity has nothing to do with the actual energy flow. The energy feeding the rod at N has come from the regions around the points P and Q on the constituting wave fronts. It is the distance NP or NQ that has to be divided by t to obtain the signal's velocity at N, and the result will be equal to c. As to its component along MN, it will be $c \cdot \cos\theta$, that is, less that c!

It may be instructive for the advanced reader to find the velocity of energy transport by using the formal algorithm. Let us find the energy density in a system of the crossed waves:

$$\eta = \frac{1}{2}\left(\varepsilon_0 E^2 + \frac{B^2}{\mu_0}\right) = \frac{1}{2}\varepsilon_0\left[E^2 + c^2\left(B_x^2 + B_z^2\right)\right] \tag{100}$$

(we have used here a textbook equation, $c^2 = 1/\mu_0\varepsilon_0$ [23]). Now use Equations (85) and (89) and carry out all the calculations. You will come up with

$$\eta = 2\varepsilon_0 E_0^2\left[(1+\cos^2\theta)\cos^2\left(\frac{1}{2}\Delta\boldsymbol{k}\cdot\boldsymbol{r}\right)\sin^2(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t) + \sin^2\theta \cdot \sin^2\left(\frac{1}{2}\Delta\boldsymbol{k}\cdot\boldsymbol{r}\right)\cos^2\left(\frac{1}{2}\Delta\boldsymbol{k}\cdot\boldsymbol{r}-\omega t\right)\right] \tag{101}$$

Similarly, we determine the density of *energy flow* $\boldsymbol{S} = \boldsymbol{E}\times\boldsymbol{B}/\mu_0$ (the Poynting vector). This time it is more complicated than in Section 6.12, because we are now dealing with two crossed waves. Using the rules for the cross product of the two vectors, and doing all the calculations, we find

$$S_x = \frac{1}{\mu_0 c} E_0^2 \sin\theta \cdot \sin\left(\Delta\boldsymbol{k}\cdot\boldsymbol{r}\right)\sin 2\left(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t\right) \tag{102a}$$

$$S_z = \frac{4}{\mu_0 c} E_0^2 \cos\theta \cdot \cos^2\left(\frac{1}{2}\Delta\boldsymbol{k}\cdot\boldsymbol{r}\right) \cdot \sin^2\left(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t\right) \tag{102b}$$

We see that the energy density and both components of the energy flow density are distributed periodically in space, and change periodically with time at each point. It is left for the reader to verify that at each point an instantaneous velocity of the energy transport does not exceed c:

$$v(\boldsymbol{r},t) = \frac{S(\boldsymbol{r},t)}{\eta(\boldsymbol{r},t)} = \frac{\sqrt{S_x^2(\boldsymbol{r},t) + S_y^2(\boldsymbol{r},t)}}{\eta(\boldsymbol{r},t)} \le c \tag{103}$$

However, in actual experiments we usually observe average values of the quantities in question. Let us now find the average velocity of energy flow $\eta(\boldsymbol{r},t)$. The averaging

of η, S_x, and S_z over variables $\boldsymbol{r}$ and t is carried out independently. The expression for the x-component of $\boldsymbol{S}$ contains the factor $\sin 2(\boldsymbol{kr} - \omega t)$, and its averaging over time gives zero. This is just what one would expect, since vectors $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$ have the opposite x-components. Thus, the *averaged* energy flow moves precisely along the z-axis. We now turn to its magnitude, i. e. the z-component of $\boldsymbol{S}$ [Equation (102 b). The factors $\sin^2(\boldsymbol{kr} - \omega t)$ and $\cos^2\left(\frac{1}{2}\Delta\boldsymbol{k}\cdot\boldsymbol{r}\right)$ give 1/2 after averaging over t and $\boldsymbol{r}$, respectively. Therefore,

$$\langle S\rangle = \langle S_z\rangle = \frac{E_0^2}{\mu_0 c}\cos\theta \tag{104}$$

The same procedure for η gives

$$\langle \eta\rangle = \varepsilon_0 E_0^2 \tag{105}$$

We now define the average velocity of energy transport by the equation [46]

$$v_{av} = \frac{\langle S\rangle}{\langle \eta\rangle} \tag{106}$$

Then, using Equations (104) and (105), we obtain

$$v_{av} = \frac{\cos\theta}{\mu_0 \varepsilon_0 c} \tag{107}$$

Finally, since $1/\mu_0\varepsilon_0 = c^2$:

$$v_{av} = c\cos\theta \tag{108}$$

This is precisely the result that was obtained from simple geometrical considerations.

If, instead of the two monochromatic waves, we now consider two wave packets propagating in the directions $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$, then the fields $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ in each packet will be the sum of many monochromatic waves with various frequencies, and the same is true with regard to $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$. When composing the products $\boldsymbol{E}\times\boldsymbol{B}$, $\boldsymbol{E}^2$, $\boldsymbol{B}^2$ we will get mixed terms of the form $\sin\omega t\cdot\sin\omega' t$, $\sin\omega t\cdot\cos\omega' t$, etc. The average value of all such terms is zero, and expressions for $\langle\eta\rangle$ and $\langle S\rangle$ will reduce to the sum of corresponding expressions for the separate monochromatic components. Because for each such component the ratio $\frac{\langle S(\omega)\rangle}{\langle\eta(\omega)\rangle}$ does not depend on ω and is equal to $c\cos\theta$, the ratio of their sums is also equal to

$$v_{av} = \frac{\langle S\rangle}{\langle\eta\rangle} = \frac{\int\langle S(\omega)\rangle\,\mathrm{d}\omega}{\int\langle\eta(\omega)\rangle\,\mathrm{d}\omega} = c\cos\theta \tag{109}$$

Hence our result has a general character and is to the effect that in any symmetrical system of crossing waves the velocity of the "phase planes" and of the corresponding electromagnetic "clot of energy" is greater than c and equal to $c/\cos\theta$, whereas the velocity of energy transfer is less than c and equal to $c/\cos\theta$. The product of these two velocities is equal to

$$uv = c^2 \tag{110}$$

A remarkable result! The velocities u and v here relate to c in precisely the same way as the phase and group velocities of the de Broglie waves do in Equation (82). What is it: a formal similarity, or something deeper? We don't know as yet ...

6.15
The mystery of quantum telecommunication

We conclude this chapter with a discussion of one of the most fascinating phenomena that, on the face of it, appears to demonstrate beyond any doubt the existence of superluminal (and even instanteneous) communication between distant objects. The phenomenon was called quantum non-locality, and is so impressive and difficult to accept that it appears to be on the verge of mystical. It has been widely discussed from the early days of quantum mechanics up to this day [47]. Its most characteristic features were recently confirmed in experiments by Aspect *et al.* [48]. These experiments aroused great interest in the physics and optics communities. They also caused numerous speculations in which their results have been interpreted as a proven demonstration of instantaneous signal transfer.

As we have seen so far, none of the discussed superluminal motions can be used for faster-than-light communication. Here we want to show that this applies also to the manifestations of non-locality demonstrated in the mentioned experiments. The demonstrated non-locality is just another aspect of the probabilistic nature of the wave function describing the behavior of physical objects (see Section 6.13). Even for a single electron with initially undefined position (that is in a state described by a plane wave embracing the entire universe), its wave function collapses instantaneously at the moment of measurement of its position to a well-defined wave packet within a small region of space. Such a collapse, or reduction of the wave packet from initially infinite to a finite size, occurs with an infinite speed (imagine something shrinking within a twinkling of an eye from the size of the visible universe down to the size of an atom!). And there is no contradiction with the theory of relativity in it, because already prior to the act of measurement there was a non-zero chance for the electron to happen to be at precisely the same atom where it was found to be after the measurement. On the other hand, even though there was a much greater chance for it to find itself some place within the Andromeda nebula, it was only a chance, not a certainty! One cannot therefore describe the collapse of the wave function in terms of cause and effect, for instance, as the convergence of a certain compressible fluid whose tractable parts occupy at any moment a well defined place in space. Therefore,

these quantum mechanical phenomena, subtle as they are, do not in any way undermine the foundation of the theory of relativity. There is an extensive physical literature on all these questions, e. g. see Refs. [49], [50] and [51].

Quantum non-locality is closely connected with the quantum mechanical description of subatomic phenomena. It has received wide attention after Einstein, Podolsky, and Rosen [EPR] wrote their famous paper [52] intended to show that the conventional theoretical scheme of quantum mechanics was incomplete.

Let me first describe the basic idea of this paper. Imagine that two high-energy photons (γ quanta) move from opposite directions towards each other. There is a chance that when they collide they will both disappear and produce instead a pair of particle and antiparticle. Suppose that this chance has been realized, and they produce an electron–positron (e + p) pair. The energy needed to create the pair comes from the energy of the two photons, and all other conservation laws are equally satisfied.

We will now focus on one important physical characteristic of micro-particles: angular momentum. Electrons have an intrinsic angular momentum called spin, and the same is true, of course, for positrons. In macrophysics, we represent the angular momentum by an arrow. The direction of the arrow gives the orientation of the rotational axis and the sense of rotation around it. We can use the same representation for spin in microphysics.

Suppose that the initial photons were so polarized that their total angular momentum was zero. Then the total spin of the system (e + p) must also be zero. This means that the arrows representing the individual spins of the two particles must point in opposite directions. If, for instance, the electron is in the state with its spin up (e↑), then the positron must have its spin down (p↓). The combined state of the pair would then be (↑↓) (here and below the first arrow in the double is for the electron and the second for the positron). Let us call this state A. Were the electron to have its spin down (e↓), then the positron by necessity would have its spin up (p↑). The combined state of the system of two particles would then be (↓↑). Call this state B. Since in either state the individual arrows point in opposite directions, the total spin of the pair is in both cases zero, as it should be. Nevertheless, the states A and B are physically different, because the individual spins are different in the two cases. For instance, the electron's spin in state A points up, whereas in state B it points down.

If the initial conditions are such that the system is in a sharply defined state A or in a sharply defined state B, no problems arise. However, it can so happen that the initial condition at the moment of the creation of the pair did not specify spin direction of either particle. Then all we can say about the physical state of the pair can be expressed in two statements:

1. The electron has its spin either up or down.
2. If the electron spin happens to point down, then the positron spin has to be up, and vice versa.

We can compress these two statements into a very short symbolic expression. We are not allowed to consider state A only, or state B only, because the particles of the pair have not been created in a state with definite spin. The system can be in either of

states A or B and, accordingly, we must consider both of them at once. The physical state of the pair is a specific combination (superposition) of both states. Let us call this new physical state Ψ. Then we can write[1)]

$$\Psi = A + B = \uparrow\downarrow + \downarrow\uparrow \tag{111}$$

Here we come to the crucial point, which is the source of many apparent paradoxes in quantum mechanics. The superposition (111) describes a very interesting state of the two particles. Either of the particles here shares a certain part of its individuality with the other one. Neither of them has a physical state of its own, because its state is not separated from the rest of the Universe (in our case, from the state of its counterpart.) The particles are entangled with one another, and the state of the pair is called the entangled state. The basic property of the entangled particle is that one or several of its characteristics depend on analogous characteristics of another particle. Measurements of any such characteristics in one would accordingly change the corresponding characteristics of another. After the measurement, the particles will be disentangled with respect to this characteristic. To disentangle our system (e + p), we must perform the measurement of an individual spin. Suppose we measure the spin of the electron, and the measurement has shown that the electron spin is up. Then, according to our basic statement about the properties of the pair, it is immediately known that the spin of the positron must be down. Hence the measurement of the spin of the electron also determines the spin of the positron. The resulting state of the system is A. The system has collapsed from being in both state A and state B to being only in state A. If the measurement shows that the spin of the electron is down, it automatically determines the spin of the positron to be up. The corresponding state of the whole pair is now state B. The wave function has collapsed from a superposition of A and B to B only.

Things can be understood better through comparison. To understand the nature of the entangled state (111) better, let us compare it with another possible state of the electron–positron pair. Consider again a situation when the direction of an electron spin is not specified by the initial physical conditions, but this time it does not depend on the spin direction of any other particle. Then it has a state of its own – a superposition of the two basic states – one with its spin up and the other with its spin down. Denoting this superposition as Ψ_e, we can write

$$\Psi_e = \uparrow + \downarrow \tag{112}$$

[The superposition (112) of the two arrows is *not* the double ($\uparrow\downarrow$). Both arrows in superposition pertain to the same particle, whereas in the double one arrow pertains to the electron and the other pertains to the positron.]

Similarly, we can consider a positron whose spin variable can be measured without disturbing other particles. Such a positron also can be described by its individual

1) To simplify the story, we ignore the coefficients, with which individual states enter the superposition.

wave function Ψ_p. If the spin direction of the positron has not been determined, this wave function is a superposition of two basic states – spin up and spin down:

$$\Psi_p = \uparrow + \downarrow \tag{113}$$

We can consider the system of the electron and the positron described by expressions (112) and (113) as one physical system. This system will be represented by a wave function

$$\tilde{\Psi} = \Psi_e \Psi_p = (\uparrow + \downarrow)(\uparrow + \downarrow) \tag{114}$$

Carrying out multiplication, we obtain

$$\tilde{\Psi} = \uparrow\uparrow + \uparrow\downarrow + \downarrow\uparrow + \downarrow\downarrow \tag{115}$$

Apart from the familiar double-terms $\uparrow\downarrow$ and $\downarrow\uparrow$, describing the (e + p) states with opposite (antiparallel) individual spins, we also see here the states $\uparrow\uparrow$ and $\downarrow\downarrow$, describing two particles with parallel spins – both up or both down. Physically, the presence of the additional terms means that there is *no correlation* between the particles – they were not born together, and are generally independent of one another. Measuring the spin of the electron bears no effect on the state of positron, and vice versa. *Both spins are to be measured independently,* and their measurements can produce any possible outcome. In the state (115) there is an equal 25 % chance for any one of the four possible outcomes of measurement.

The entangled state (111), on the other hand, describes a rigid correlation between the particles – their spins must always come up opposite, so that the spin measurements can only produce the anti-parallel doubles, and *measurement of only one spin automatically gives the result for another one.*

Now let the members of the entangled pair fly apart, so that some time later they are far away from one another. But, if they do not interact with anything else, they keep their common entangled state Ψ. Much later, even though they may be separated by millions of light years of space, they remember their initial conditions. Neither particle knows the direction of its spin until the measurement is performed. And yet either particle knows that this direction, whatever it might turn out to be, must be opposite of that of its partner.

Now, here comes the crunch. Suppose that our electron is now on Earth, and its twin antiparticle is in another galaxy, on planet Rulia. The physicists on Earth measure the spin of the electron and find it to point up. Immediately it becomes certain that the spin of the positron on Rulia must be down. If the Rulian physicists do a measurement (which is no longer necessary!), the measurement will only confirm this fact. Apparently, there must have been some agent that instantaneously transferred to Rulia the information about the measurement outcome on Earth, which helped the positron to decide upon the direction of its own spin.

The intention of EPR was, first, to demonstrate that the spin of a particle can sometimes be measured without performing an actual measurement on it and, second, to

point out that such measurement demonstrates an instantaneous physical action at a distance, which would be incompatible with Special Relativity. Einstein, Podolsky, and Rosen had interpreted this contradiction with Relativity as an indication that the theoretical description of the world given by quantum mechanics was incomplete (EPR paradox).

The first of these statements is true. The second one is false.

But before moving further and explaining why it is false, let us analyze a possible objection to the interpretation of this thought experiment. Suppose that someone on Earth prepared two bottles of wine. One bottle is Burgundy and the other is Chardonnet. One is for himself, the other for a friend on Rulia. It does not matter who gets which wine, so the sender just puts one of them in his refrigerator and the other one into the cargo spaceship due Rulia, – without looking – and then goes into hibernation for a few million years. After a long trip, the spaceship arrives at its destination. The Earth-based physicist wakes up completely unaware of which wine is on which planet. He opens the refrigerator and immediately knows that the wine on Rulia is Chardonnet.

How could he in an instant get information about an object in another galaxy?

The answer is very simple in this case. The earthling did not receive any signal from Rulia. The only signal he has got is the one from his own refrigerator. The signal has changed his knowledge about the refrigerator's contents from total uncertainty (50 : 50 chance of the wine there being of a definite kind) to complete certainty (100 %) that it is Burgundy. Together with preliminary information about the two bottles that he had, this enables him to conclude with the same dead certainty that the wine on Rulia is Chardonnet.

The act of observation in this case did not (and could not) physically change the type of wine in either package. The Burgundy that was taken out from the refrigerator had been Burgundy before the observation. The Chardonnet on Rulia had been Chardonnet long before it arrived there. We can distinguish here between the physical state of the observed object and the physical state of the observer. Only the latter has changed in the act of observation. There was no distant communication in this case.

This might tempt some readers to draw the same conclusion about the experiment with the entangled particles. But such a conclusion would be wrong. The situation with our pair of particles is fundamentally different from the two bottles of wine described above. When the observer on Earth performs the experiment and finds that his particle has spin up, this does *not* mean that the particle had had its spin up before the experiment. No! The particle undergoes a dramatic change of its physical state in the process of measurement. It converts from the entanglement with its distant partner into a disentangled state of its own. In the former state the particle, even though separated from its partner by the vastness of space, did not have its full identity totally independent of the partner's identity. Their identities remained intimately shared. In the final state each particle has its own full identity and can be described by a wave function of its own, independent of the rest of the world.

Thus the measurement made on Earth changes instantaneously the situation not only on Earth, but also on Rulia (and vice versa). In the language of the wave func-

tions we can say that the wave function of the whole entangled system instantly collapses into one of the two possible independent wave

$$\Psi = \uparrow\downarrow + \downarrow\uparrow \Rightarrow \begin{cases} \text{either} & \uparrow\downarrow \\ \text{or} & \downarrow\uparrow \end{cases} \tag{116}$$

It appears that some physical agent does indeed carry the information about the change on Earth, and this communication occurs instantaneously, changing immediately the situation on Rulia. It does look like we are facing a completely new phenomenon – a superluminal (instant) quantum telecommunication.

Yet this conclusion would be also wrong. Relativity is not violated, because the changes of states that we discuss are inherently statistical. By its original definition, the event N can be considered to be causally affected by the event M only if the change in M changes an observable property of N from one uniquely defined value to another uniquely defined value, for instance, if the influence of M changes the spin of N from up (↑) to down (↓). In our case, however, the original spin direction of the positron on Rulia was not sharply defined. The positron was in a state with an indefinite direction of spin. There was from the very beginning a 50% chance of finding its spin up, and a 50% chance of finding it down. Therefore, in any individual measurement, when you find the positron spin up, you can always say: so what? If I could have known with certainty that the spin before the measurement was down, and now I find it up, I could interpret this change as the effect of some outside agent. But when there had already been a 50% chance of finding it up, and I do find it up, why should I attribute it to some external influence? I will rather say that this outcome is just the realization of pre-existing potentiality.

Thus, no individual outcome can be the evidence of superluminal or any other telecommunication. If the measurement results are statistical by nature, the only way we can make sure that the instantaneous telecommunication does take place is to find the difference in statistical distributions of measurements at N with and without corresponding measurements at M.

Following this program, we must perform the set of individual measurements in two different conditions and compare their results.

Condition 1. Prepare a big ensemble of (e + p) pairs all in the same state Ψ (111). For each pair, measure only the positron spin on Rulia without disturbing its counterpart on Earth, and record the results.
Condition 2. Prepare again the same ensemble. For each pair, measure first the electron spin on Earth, and immediately after the electron spin on Rulia, and record the results.

Next we compare the records obtained for the two conditions. Here they are:

1. *Results for condition 1*: the measurement outcomes are distributed uniformly – 50:50. Within the margin of statistical fluctuations, half the positrons collapse to the state up and the other half collapse to the state down.

2. *Results for condition 2*: the measurement outcomes are distributed uniformly – 50 : 50. Within the margin of statistical fluctuations, half the positrons collapse to the state up and the other half collapse to the state down.

The net results are identical. The measurements on Earth *do not* cause any changes in the collective experimental data for measurements on Rulia. The whole experiment does not reveal any evidence whatsoever of any communications or signaling between the distant objects.

At the same time, if we compare in case 2 the *individual* results for members of each pair, rather than only the averaged data, we obtain the absolutely irrefutable evidence of the long-distance (non-local) correlations between individual events. For each pair, if the electron spin on Earth collapses to the state up, the positron spin on Rulia, measured at nearly the same moment, collapses to the state down. If the electron on Earth is found in the state down, the positron on Rulia is found in the state up. The positron appears to know instantaneously what happened to its counterpart millions of light years away. And yet it does not prove the existence of a signal exchange between them, because in each case the positron (or we) can say that the correlation is a mere coincidence. Its chance was 50 %, so there is nothing extraordinary that it happens. But why do these coincidences happen 10, 20, ..., 1000 times in a row? Well, this is really weird, the chances of N coincidences in a row are small indeed, $(1/2)^{-N}$, but they remain finite for any finite N.

These nearly absolutely improbable multiple coincidences reveal a new physical phenomenon – non-local quantum correlations (or quantum non-locality). They have no classical analogue and show that a quantum system can keep certain correlations between its parts even when these parts are separated by voids of space. These correlations are manifest in the fact that the spatially separated parts of an entangled system collapse simultaneously to correlated states even when the measurement is performed on only one part. As you think more of it, this collapse seems no more (and no less!) surprising than the collapse of the wave function for a single particle from the state of the uniform and infinitely large probability cloud to the state of the infinite density at a single point surrounded by emptiness with no clouds.

The reader will understand that the situation described above is only a thought experiment. The real experiments are much more difficult, because it is extremely difficult to maintain the "purity" of the system – to protect its parts from random perturbations. Such perturbations (or fluctuations) from the surroundings destroy the quantum correlations (quantum coherence) within the system, and the further apart its constituents, the harder it is to preclude them from "decoherence." But, despite numerous difficulties, the experiments [48] have been successfully carried out. They had been performed on a much smaller scale than the thought experiment considered here, but their ideas and results are essentially the same. They showed the ability of quantum systems to keep a memory about a common past. They did not show in either individual or collective experimental data any evidence of faster-than-light communications.

The whole situation also shows how tricky Nature is. She seems to tease us. In the above experiments, she appears to violate blatantly the relativity postulates, but with each new trial attempting to catch her in action she comes clean of any violations.

7
Slow Light and Fast Light

> The Queen propped her up against a tree, and said kindly, "You may rest a little now."
> Alice looked round her in great surprise. "Why, I do believe we've been under this tree the whole time! Everything's just as it was!"
> "Of course it is," said the Queen, "what would you have it?"
> "Well, in *OUR* country," said Alice, still panting a little, "you'd generally get to somewhere else – if you run very fast for a long time, as we've been doing."
> "A slow sort of country!" said the Queen. "Now, *HERE*, you see, it takes all the running *YOU* can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!"
>
> Lewis Carroll
> *Through the Looking Glass*

7.1
Monitoring the speed of light

We start this chapter with a brief summary of what we have learnt after more than two centuries of intense studies of electromagnetic phenomena: the speed of light in a vacuum is a fundamental constant of Nature; its speed in a medium is an intrinsic characteristic of this medium, specified by the refractive index n.

Having made these statements, I now want to describe some new experiments with light demonstrating how tricky Nature is in her every turn, and how physicists unravel her tricks. We will see how researchers have managed to influence light's propagation rate in different media; in other words, how they have succeeded in effectively changing a medium's refractive index. They now can fool a medium into guiding light much slower than it would ordinarily do, or, even more surprisingly, much faster. They can create conditions under which a light pulse in a medium travels faster than light in a vacuum! We want to see how it is possible and whether or not we can use it for superluminal signal transmission.

Let me first describe some experiments with retardation of light. They can be divided into two different categories according to two different physical phenomena used to tackle the natural optical properties of matter. In the first category a medium's ability to transmit or absorb light is influenced by shining on it an additional light beam of

Special Relativity and Motions Faster than Light. Moses Fayngold

ISBN: 3-527-40344-2

different color. One of the recent experiments of this type was reported in a 1999 issue of the journal *Nature* [53]. A group of researchers cooled a cloud of sodium atoms to ultra-low temperatures (50 nK–2 K). At such temperatures the sodium cloud experiences a transition to the so called Bose–Einstein condensate – a state of matter where nearly all constituent atoms are in the same physical state. Whereas at higher temperatures motion of an individual atom in a gas is random and independent of others, in Bose–Einstein condensate all the atoms share the same state of motion. Imagine first a crowd of people rushing each on his or her own business in a hectic marketplace; then a military unit standing still and only breathing synchronously, or marching in a highly organized fashion during a parade – and you will have a glimpse of the difference between the ordinary state of matter and Bose–Einstein condensate. A stable monochromatic laser beam is an example of Bose–Einstein condensate of photons as opposed to the glow of a light bulb. The identical motions of all the particles in a Bose–Einstein condensate make it possible to influence all of them by an external agent (e. g. a beam of laser light) in the same predictable way.

For a cloud of atoms one would expect the refractive index to be only slightly larger than 1, which corresponds to light propagation just slightly slower than in a vacuum. However, the light of certain colors (frequencies) may be absorbed by the cloud. The range of frequencies at which the light is absorbed in a given medium is called the absorption band, and each medium has its own specific set of absorption bands. Normally this also holds true for a Bose–Einstein condensate. However, in the described experiment the propagation conditions in the condensate were carefully designed to produce an apparently "abnormal" situation. First, the major frequency of the light pulse sent through the sample was chosen to lie in the absorption band of sodium vapor. This does seem weird, for how are you going to study the pulse propagation in a medium that is opaque in a chosen frequency range? Second, simultaneously with the probing pulse, the researchers beamed the cloud with *additional* light of a slightly different frequency (called the coupling frequency), as if they had not had enough trouble already with the doomed probe pulse. It turned out, however, that the *combination* of the two conditions worked miraculously towards the success of the experiment. The combined electromagnetic field of the probe and the coupling beam canceled absorption, thus making the vapor effectively transparent for the probe. This phenomenon is called *electromagnetically induced transparency* (EIT) because the medium is made effectively transparent by the "electromagnetic intervention" of the coupling beam.

What is important for us here is the fact that in a frequency range where light is absorbed, the refractive index changes rapidly (Fig. 7.1). This variation is effectively enhanced by the coupling beam. The rapid change of the refractive index with frequency may result in the dramatic change of the group velocity even though the index itself is close to 1. You will remember that we found in Section 6.12 that the group velocity of the pulse is determined by the derivative $d\omega/dk$. When we applied this result to the propagation of a light pulse in a plasma, we found its *group* velocity to be less than the speed of light in a vacuum even though the *phase* velocities of constituent waves were all greater than the speed of light in a vacuum. We may expect the retardation of a light pulse in a condensate to be even more impressive because the constituent waves here already move slower that the speed of light in a vacuum.

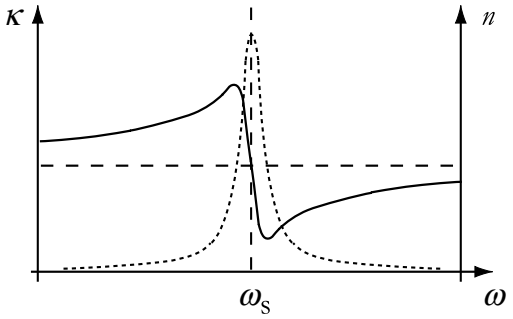


Fig. 7.1 Two optical properties of a medium – the coefficient of absorption (dashed curve) and refractive index(solid curve) are plotted against frequency w. All three quantities are expressed in arbitrary units. The value of w_s is the frequency at which the absorption reaches its peak. Within the absorption band (the frequency range with high absorption) the refractive index undergoes a rapid change. In mathematical terms, the derivative dn/dw and thereby dk/dw become very large in this region. According to results in Section 6.12, this corresponds to small group velocities.

In the described experiment the observed group velocity of the pulse was just 17 m s^{-1} – the speed of light retarded by about 20 million times!

So rapid is the pace of discovery at times that it makes it difficult for anybody to write about it. Soon after the news about that astoundingly slow light, there appeared a report in the January 19, 2001 issue of *Nature* [5] announcing that the same group of researchers had managed to bring a laser beam to a complete stop!

A stopped light is such an extraodinary phenomenon that the original authors' account of it in *Nature* was preceded by a special science update column, written by Philip Ball under the title "Stop that light beam, I want to get off," and with a tag "E-mail this story to a friend."

The detailed explanation of how the coupling beams can slow light in a probe lies beyond the scope of this book (interested readers are referred to, e. g., Refs.[53] and [54]).

However, there is another category of experiments with light retardation. The experiments in this category are conducted under less exotic conditions – at normal temperatures and with no coupling beams. Only one pulse at a time is passed through a medium. But the pulse's frequency range is also selected to lie in the medium's absorption band. How can one compensate for absorption of light in this case? The experimenters achieve such compensation by using pulses of very high intensity. Light of high intensity striking an opaque medium causes a non-linear optical effect called *saturation of absorption*, thereby making the medium transparent to this light. This kind of transparency, imposed on the medium by extremely bright light passing through it, has been called *self-induced transparency* (SID). The researchers demonstrated that self-induced transparency can be accompanied by a dramatic delay in pulse propagation (*slow light*). The pulse velocity is in this case neither phase velocity nor group velocity; it depends not only on the physical properties of the medium, but also on the brightness of the pulse itself. Thus, it can be changed by varying the pulse intensity. Already in the early 1970s there appeared descriptions of experiments on slowing the intense pulses of laser light [7, 8].

A similar technique, but with some variations, has been used for the "manufacturing" of most bizarre and puzzling objects – superluminal pulses. The experimenters shone a laser pulse on a specially prepared medium and observed rapid acceleration

of the pulse (*fast light*). One of the latest experiments of this kind was described in the July 20, 2000 issue of *Nature* [3, 4]. A light pulse with a specially shaped profile was directed towards the front face of a cell containing cesium (Cs) vapor. The researchers observed the emergence of the pulse from the back face of the cell *before* it entered its front face! The effect appeared to precede the cause! This astounding result made a big splash in the scientific community, although the professional physicists received the news much more quietly. Most of those working in the field knew about much earlier experiments with superluminal light pulses. As far back as 1969 there appeared an article and a review in Russian journals [9, 10]. The authors not only described experiments with superluminal group velocities in certain media, but also gave a detailed explanation of the observed effects. Readers can find a description of these phenomena from a very general point of view in a monograph [55].

Now the time has also come for us to look for a clear explanation of the phenomena described in this section. We will try to address all the basic questions raised in the above experiments, especially ones about superluminal pulses. We know already about superluminal phase velocities in plasmas, and we know that we cannot use them for superluminal transmission of a signal. We have learned how a *group* of waves with such velocities always contrives to sweep across the medium much slower than any of constituent monochromatic waves. We have proved that the group velocity of any group carrying a signal is always less than *c*. We visualize a group as a pulse of a certain shape with a maximum, shaped something like a tortoise, or a bell, or a bump, for that matter (Fig. 7.2). Now we consider experiments

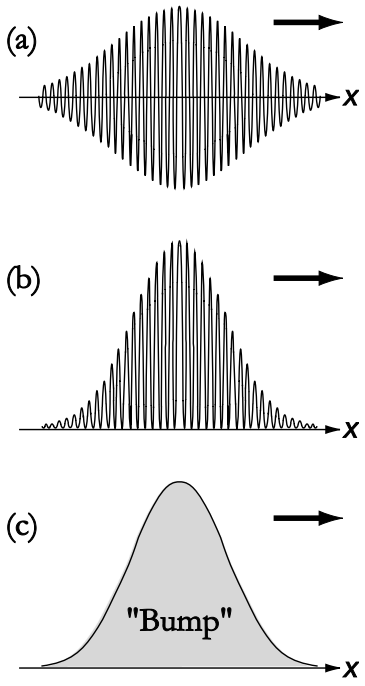


Fig. 7.2 (a) A snapshot of the instantaneous distribution of the electric field strength in a light pulse moving along the *x*-direction. (b) Corresponding instantaneous light intensity distribution. (c) Light intensity distribution (a "bump") as observed with lower resolution.

with a light pulse, or a "luminal" *bump*, propagating faster than *c*. Because such propagation seems to violate causality, we must consider these experimental results with all seriousness.

We will focus on the following questions:

1. How is it possible for a light pulse in an opaque medium to induce transparency?
2. Why and how does the induced transparency cause the pulse to slow down?
3. How is it possible to switch from slowing a pulse to accelerating it?
4. Does or does not a superluminal pulse carry a signal, and if it does not, why?

I want to proceed with a somewhat fantastic and apparently unrelated episode inspired by Lewis Carroll's famous books *Alice in Wonderland* and *Through the Looking Glass* [56, 57]. Physicists often refer to these books, because the world of Carroll's creative imagination appears to reveal in a most artistic and poetic way intimate connections to hidden intricacies of the real world. I invite you to read again the epigraph to this chapter before plunging into the next section.

7.2
Adventures of the Bump

Alice wandered away from the Garden of Live Flowers deep in thought.

"How is it possible to stay in the same place while running as fast as you can?," she kept asking herself. "And how can you run TWICE as fast as THAT? This is the most weird thing of all I have ever met here!"

She thought it over for a while and decided: "It probably means that in this world, if you really run as fast as you can, you sometimes lag behind yourself, and sometimes outrun yourself."

Alice was very proud of her discovery, even though she could not clearly see how it was possible for anyone to run either behind or ahead of oneself. While she was turning it over in her head, she stepped right in a big hole probably made by rabbits or moles. Down she went into a deep well. After a long fall, she crashed with a deafening "Bang!," and found herself having bumped into a big Rabbit with the glasses on his nose. The Rabbit held a book, *Alice in Wonderland*, in one hand and a clock in another.

"Hi, Alice," he said, "you came right on time. We are just starting our show with the Bump."

Alice greeted him politely and looked around. They were in a long, brightly lit corridor with a shelf running horizontally all the way down high on one of the walls. And then she saw what the Rabbit has referred to as the Bump. It was just a big heap of earth piled on top of a flat wheeled platform. The platform was pushed down the corridor at a fast, steady rate by a few creatures resembling small chimpanzees in shape (Fig. 7.3).

"These chimps are the fastest runners known in our world. Therefore nothing can outrun the Bump," the Rabbit said proudly.

"Is this a show? It does not at all make any sense to me," Alice thought.

"It will make more sense to you," said the Rabbit, "when you see it all."

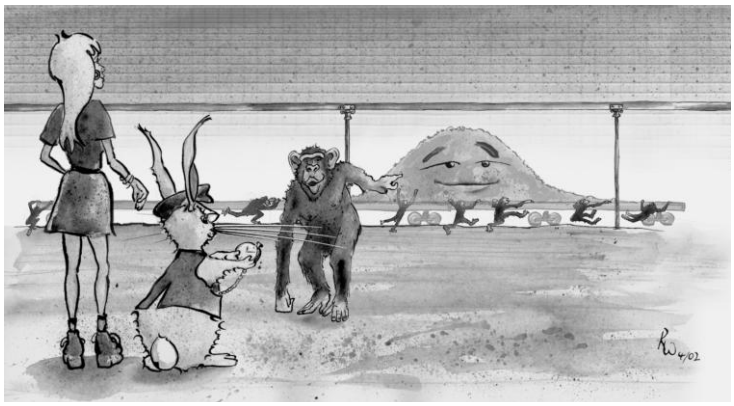


Fig. 7.3

At this moment Alice noticed a row of moles down the track, each with a spade in its paws (Fig. 7.4). What followed next was very strange indeed. Just when the front slope of the Bump was well inside the corridor, the moles, one by one, started scooping the earth from it and hurling it up on to the shelf (Fig. 7.5). As the platform with the Bump progressed down the corridor, the higher and higher parts of the slope passed each mole; accordingly, each mole worked faster in proportion to the height of the slope passing by. When the top of the Bump was abreast of the moles, they handled their spades so furiously that the top was almost immediately flattened, with huge amounts of earth hurled on to the shelf. Now Alice noticed that the shelf had been perforated, and the earth that had been piled up on it started to spill down through the holes right on the back slope of the Bump. The net result of these activities was that the earth was being continuously transferred from the front of the Bump to its rear, thus impeding the Bump's progress (Fig. 7.6).

Alice watched in amazement how the platform supporting the Bump kept on advancing forward at the original fast rate, while the Bump itself was now hardly moving past the row of working moles. Because of the moles' work, it was continuously displaced to the back of the platform at almost the same rate as the platform's advance

Fig. 7.4 The moles are preparing for their job on the approaching Bump.

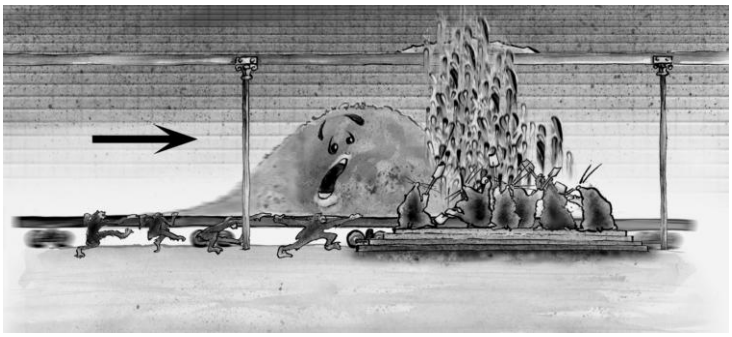


Fig. 7.5 The moles start their work. Each mole works at a rate proportional to the instantaneous hight of the slope passing by.

Fig. 7.6 The moles are slowing the Bump by taking the earth from its leading slope and hurling it up on to the shelf.

in the forward direction. The overall motion of the Bump became so slow that it appeared to Alice almost stationary.

"The slow Bump," whispered the Rabbit. And the Queen's voice rang again in the Alice's ears:

"Now, HERE, you see, it takes all the running you can do, to keep in the same place."

"How wonderful," Alice thought. "It seems to me, I now understand how one can lag behind oneself while running at top speed. But," Alice sighed, "I am still a little foggy about how one could run AHEAD of oneself."

"Well, you will see more of it now. Here it comes to the next stage," said the Rabbit.

The platform, with the Bump having been recycled to its very rear, had passed the row of moles. Now that the moles were left behind, the Bump was again moving fast together with the platform.

A minute later, the scene changed. A new platform with the Bump on its rear was approaching the entrance to the corridor. There were moles along the track again, and the same long horizontal shelf running higher than the Bump's top. But the mo-

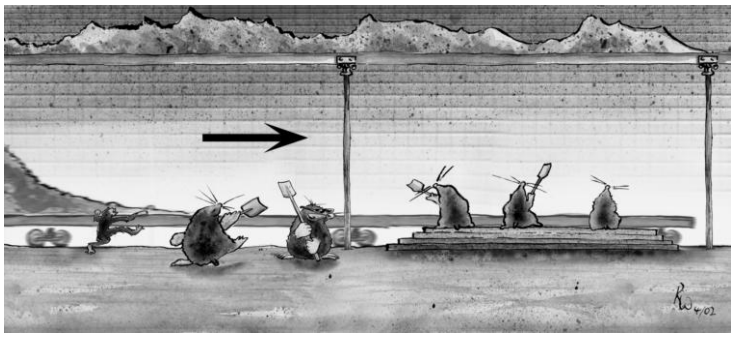


Fig. 7.7 The heaps of earth on the shelf are prepared in advance to be spilled on the front slope of the approaching Bump.

les were fewer than before, and over the shelf, across its whole length, there were already spread in advance large piles of earth (Fig. 7.7).

As the front slope of the Bump was entering the corridor, the moles began, as before, to throw the earth from the slope on to the shelf. But owing to vibration from the moving platform, much more earth started to spill down on to the slope through the shelf's holes. For each handful of earth hurled up by the moles there were a dozen or more streaming down from the shelf. Alice stopped paying any attention to the moles at all, and focused instead on what was happening to the Bump. And what she saw looked pretty weird. Because so much earth was being spilled on to the front slope of the Bump, the front became higher and steeper, and soon it had grown nearly as high as the Bump's top! Less earth remained on the shelf for the parts of the slope closer to the original top of the Bump, and these parts grew slower than the front. The overall result of this remarkable play was that the Bump appeared to

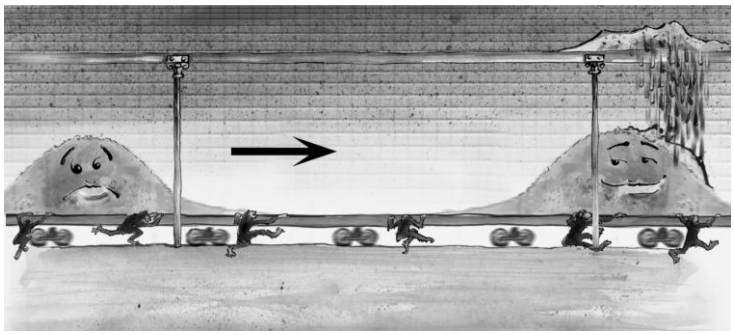


Fig. 7.8 The Bump outruns itself by producing vibrations which cause the earth on the shelf to spill onto the Bump's leading slope. The slope grows into a new Bump that propagates faster than its original self.

move faster than its own head; and soon the newly grown front peak had run so far ahead that it emerged from the end of the corridor before the original Bump entered its beginning (Fig. 7.8)!

"The fast Bump!" screamed the Rabbit, and the whole crowd of participants exploded in cheers. This was the very essence and the culmination of the ceremony.

The ensuing noise and tumult helped Alice shake off her spell. She recalled vividly the Queen's last words: "If you want to get somewhere else, you must run at least twice as fast as that!"

"Well, I think that now ..." Alice started thinking.

"Well, I think now you understand a little better how one can outrun oneself," echoed the Rabbit to Alice's unfinished thought.

"Even if ..." Alice started thinking again, struggling to get everything straight.

"Yes," said the Rabbit and winked significantly, "even if one is already running at TOP speed!"

7.3 Slow light

Strange as it may seem, the fantastic scene in the previous section provides a pretty close analogy for a qualitative understanding of slow and fast light propagation in some media. Here we will use the analogy to gain more insight into the physics of slow light in the regime of self-induced transparency. We had emphasized that in this regime the medium is normally opaque for the chosen frequency: the light is being absorbed by the medium. The absorption process can be understood on the microscopic level if we imagine the light pulse as a bunch of photons. In absorbing material the atoms are tuned to the photon frequency so that each photon is absorbed almost immediately upon entering the medium. An absorbing atom gains the photon's energy – gets "excited" in physicists' jargon. We illustrate this in the following way: represent each atom by a dot, and the atom's energy by the dot's position on the energy scale plotted vertically. Normally an atom is in a state with the lowest energy – the ground state. We put the dot representing such an atom on the lowest level – the ground level. The excited atom has a higher energy, and therefore the corresponding dot should be put on a higher level on the energy scale, just as we find ourselves on a higher level after gaining some potential energy when climbing upstairs (Fig. 7.9). The excited atomic states are usually short-lived: the atom either squanders its energy into heat or re-radiates it in a random direction. The last process is called *spontaneous emission.* In either case the incident photons are removed from the original beam and the beam dies out rapidly – the light is absorbed.

What happens if we increase the light intensity so that more and more photons are rushing by? It turns out that the atoms' ability to absorb photons increases with the abundance of photons. A thick crowd of photons "invigorates" atoms, enabling them to do their job faster. Moreover, this "invigoration" works both ways: once in the excited state, an atom is quickly coerced by passing photons (which are all indistinguishable from one another in a laser beam) to release its energy back in the form of

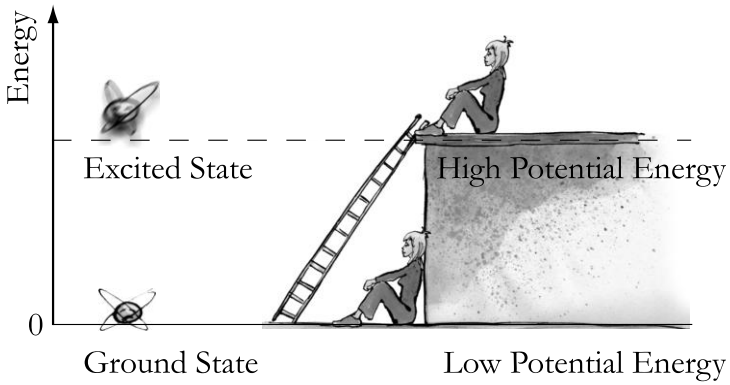


Fig. 7.9

the original photon. This coerced emission of a photon into the same state as that of photons whizzing by is called the *stimulated emission*. The more crowded the photons' beam, the greater is the chance that an absorbed photon would be returned to the beam rather than spewed in a random direction or lost to heat; accordingly, the higher is the rate of stimulated emission. At sufficiently high intensities practically all absorbed photons are being returned back to the beam! At this state we have at each moment the same number of atoms on the ground and on the excited levels. The former absorb photons and get excited; the latter re-emit them in the direction of the beam and return back into the ground state. Absorption is balanced by stimulated emission. The photons are not being taken away from the beam permanently, but rather are "borrowed," soon to be returned. The beam therefore proceeds practically without attenuation, even though the medium is intrinsically opaque! This peculiar state of permanent recycling is what we referred to in Section 7.1 as saturation of absorption and self-induced transparency.

Now imagine that instead of a continuous beam we have a powerful but short pulse of laser light. How would it affect the light propagation? There is no single answer to this question, because the outcome would depend on the pulse's shape. But there are certain shapes for which the pulse propagation becomes very similar to the retarded motion of the Bump in the previous section. In that motion, the earth taken from the Bump by the moles represented the absorption of light; the earth spilled back down on to the Bump represented stimulated emission, returning light back into the pulse; and the return time between the two processes was tuned to the motion of the platform so that the earth taken from the leading slope of the Bump started to spill back down precisely when the trailing slope was passing under it – no sooner, no later.

Now, the same trick can be performed with light! Suppose the light *intensity* in the pulse is such that the time between the absorption and subsequent stimulated emission of a photon is 10^{-11} s (normally this time is longer, about 10^{-9} s, but here atoms are invigorated by the intense pulse!). Suppose, further, that with the given intensity we adjust the *width* of the light pulse so that it travels a distance equal to this width in the same time, 10^{-11} s. As a result, a photon absorbed by the atom from the front slope will be re-emitted precisely when the rear slope will pass by it. The photons

will be transferred from the front to the rear of the pulse in the same way as the earth had been shuffled from the front to the rear slope of the Bump. This will result in effective retardation of the pulse. It turns out that one can always find self-consistent conditions under which the pulse will move through the medium with decreased speed without changing its shape. Such a pulse represents an example of so-called solitons – solitary waves with stabilizied shape.

There is one more thing to emphasize here. Although the motion of the above luminous soliton is indeed dramatically slowed, it does not mean that the constituent photons have become more sluggish. Recall the scene in the previous section with the Bump being recycled towards the back of the moving platform. While this was happening, every lump of earth in the Bump kept on moving together with the platform with the initial speed. The earth making up the Bump was not sliding back relative to the platform.

The Bump appeared to move slower because of its continuous reshaping produced by the moles. This reshaping effectively slows the Bump as a whole, but it does not affect the motion of individual particles of earth that remain inside. Similarly, the continuous reshaping of the pulse in slow light does not change the speed of individual photons between the acts of their interaction with the absorbing atoms. The light slows without impeding the rush of its photons. It takes all the running the photons can do to keep them all in place.

7.4 Fast light

After a thought trip to the land of moles, followed by the discussion of slow light, we must be better prepared for understanding the most intriguing part of the propagation of light. We will now consider experiments in which researchers observed *superluminal* group velocities of the light pulses. As in our previous discussions, we define the group velocity as the velocity of the maximum of the wave packet (that is, the top of the Bump!). We will first describe qualitatively a way to create such a "Bump," and then consider whether superluminal bumps would allow superluminal communications, thereby overthrowing limitations imposed by the theory of relativity.

Following this outline, I want first to refer to the concluding part of the ceremony observed by Alice at the end of Section 7.2, the fabrication of the fast Bump. What do you think was *the* most important feature in this process? It was the fact that this time the busy moles *had not been employed* for the purpose (the few ones seen by Alice had worked, as before, for retardation of the Bump, rather than its acceleration, and their role in the whole process was insignificant). One might be tempted to think that acceleration of the Bump's motion could be achieved by just instructing the moles to do the opposite of what they had done before: to throw earth from the rear slope of the Bump to its front slope. But this is impossible! In order to hurl a lump of earth from rear to front one must impart to it a speed greater than that of the platform. But the speed of the platform represented the speed of light in a vacuum, which could not be exceeded by a physical body. By the same token, one could not use the upper shelf as a vehicle for the earth transfer: in order to transport the

earth from the Bump's rear to its front, the shelf should itself be moving faster than the platform. We see that the acceleration of the Bump cannot be achieved by just regrouping pieces of earth of which it is made up. However, it is possible to grow its front by adding to it new earth prepared in advance along the way! Hence the heaps of earth on the shelf, waiting to be shaken off down on to the front slope by vibrations, as the Bump makes its way down the track. In this arrangement, the Bump represents the original light pulse, and the earth on the shelf represents the excited atoms ready to release their energy to feed the leading slope of the pulse.

We can describe it graphically in the following way. Suppose that an input pulse at an initial moment t_1 has a shape described by a function $E(x, t_1)$ with a maximum at point x_1 (Fig. 7.10). By the moment t_2 the pulse shifts some distance in the x direction. Its new profile is now described by a new function $E(x, t_2)$ with a sharper and possibly higher maximum at a point x_2. If we measure the velocity of this maximum as $u = (x_2 - x_1)/(t_2 - t_1)$, it turns out that u exceeds the speed c. The superluminal velocity of the maximum is in this case achieved not so much because of its own shift, but because of the specific distortion of its shape. Its front slope has become steeper, which has caused an additional advance of the maximum in the forward direction. Physically, this advance is caused by the increase in the intensity in the front region of the propagating pulse. This increase is produced by the influx of energy that has been stored in the excited molecules of the medium prior to arrival of the pulse. The excited medium is unstable. As the pulse sweeps through it, even a relatively weak electromagnetic field of its leading slope triggers light emission from the excited molecules there. This produces additional quanta of light (photons) identical with those already present in the front side of the pulse. The newly acquired photons intensify the front field in the pulse. What had previously been the front slope of the pulse gradually becomes its new "peak". The lagging sides of the pulse have to pass through the "exhausted" medium that had already been "skimmed" by the front. If the external source cannot restore the population of excited molecules (the rate of stimulated emission exceeds the pumping rate), then the center and the back of the original pulse cannot increase significantly. As a result, we observe a preferential increase of the leading slope of the pulse at the expense of the medium's energy. If this increase is sufficiently large, the field magnitude in the front may reach and even exceed that in the original peak as shown in Fig. 7.10. This results in the additional advance of the peak that manifests itself as the acceleration of the pulse. Because of the

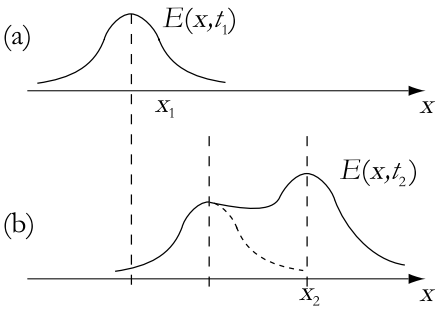


Fig. 7.10 (a) A disturbance at an initial moment t_1. (b) The same disturbance in a non-equilibrious medium at a later moment t_2.

specific distortion of its shape, the pulse "outruns itself," so that its velocity can become superluminal. If the original pulse is broad enough (broader than the experimental cell), we can even observe it exiting the cell before it enters it – precisely what Alice had seen in the land of moles!

Recall now the previous chapter where we have shown that a group of waves transfers both energy *and* signal, and the group velocity can only be subluminal. This appears to stand in flat contradiction with the result obtained right now. Because this result is well established experimentally, it seems to overthrow the very foundations of the theory of relativity. This, however, is not the case, and here is why. All theoretical and experimental evidence that the group velocity cannot exceed the value of c refers to media close to a state of thermodynamic equilibrium. The medium that we are dealing with now is *not* in a state of thermodynamic equilibrium. It contains many more atoms in the excited states than in the ground states. Using the procedure described in Section 7.3, we represent this situation graphically in Figure 7.11. It looks pretty much like the shelf with prepared earth heaps in the fast Bump ceremony. We call it a system with inversely populated energy levels, and, as is readily seen from Figure 7.7 (and Figure 7.11!), the name is self-explanatory. Such a system is unstable, and it feeds the front slope of the pulse to grow a runaway top just as the heaps of earth in the fast Bump ceremony fell down from the shelf to form the Bump's runaway head. We see that owing to the instability of such a system it is possible for the top of the pulse and thereby a group velocity to exceed the speed of light!

Now we will discuss whether or not this phenomenon constitutes a contradiction with the established physical principles. We will show that a superluminal disturbance in a non-equilibrious medium cannot be used for a superluminal transport of a signal.

Consider a pulse observed at a point A at a moment t_A. Let this be a moment *before* the pulse enters the experimental cell. Suppose that a detector at point B at the *opposite* side of the cell records the disturbance (also shaped as a pulse) at *the same moment* t_A (Fig. 7.12). Since the event at B occurs *before* the original pulse at A entirely enters the cell, it appears that the effect in this experiment precedes its cause. But this can only be true if the change in the detector's state is indeed caused by the *disturbance that we have observed at A*. It may well happen, however, that the forerunner of the initial disturbance – its leading slope – has been intensified by the medium in the cell, and the detector will have felt its front edge much earlier than it would nor-

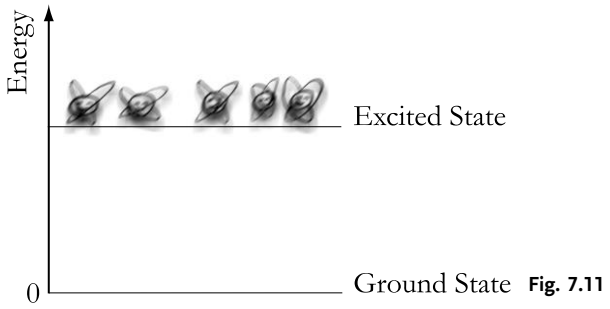


Fig. 7.11

Fig. 7.12

mally do. In this situation, the detector is prematurely triggered by the arriving disturbance because of the influence of the medium between points A and B rather than because of the event at A. Precisely such a phenomenon occurs when the medium is not in an equilibrium state.

Let us now consider the propagation of a disturbance in such a medium in more detail.

When we observe the motion of a pulse, we record the advance of its maximum. In the process we have just considered, the maximum at point B originates not from the maximum at A, but from the amplified field in the vicinity of point B (in the case of Fig. 7.12, in the layers of the cell adjacent to its exit face.) As we have just mentioned, the increased field here results from the stimulated emission caused by the "forerunner" – the pulse's leading slope that enters the cell when the top is still outside. This, in turn, stimulates new emission from the excited atoms there and further down the way. As a result, the propagating pulse is being distorted by a nonuniform amplification favoring its front slope. The amplification wave thus formed can, in principle, accelerate itself up to infinite speed at the expense of the medium's energy. The illusion of superluminal velocity of the signal appears just because the speed of the amplified wave is added to the velocity of the actual signal.

To see it, imagine that a certain message is encoded at the original top of the light pulse (Fig. 7.13). Can the detector at B read this message? The answer is, yes, but only at a later time when the specific feature at the original top will reach it. Can the detector read it *now*, when the feature has not yet entered the cell? The answer is, *no*. The detector does record the arrival of a pulse at the moment depicted in Figure 7.13, but this pulse is just a runaway head produced by the amplification wave. It contains information only about physical conditions in the immediate vicinity of the point B, having nothing to do with the awaited message from A.

There is another, more rigorous (and perhaps more elegant), way to show that the signal velocity cannot exceed c even in the "fast light" experiments. To see it, note that all the confusion about superluminal communications is associated with the amplification wave being allowed to propagate unimpeded over large distances because we use a very broad initial pulse. We have assumed in our discussions that the slopes of a pulse only gradually (*asymptotically*, in mathematical language) approach the zero level. Under these conditions, the front slope of an approaching pulse is spread far (theoretically, infinitely far) ahead of its top, and accordingly, the amplification starts long before the center of the pulse with the encrypted message enters

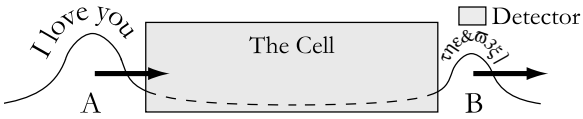
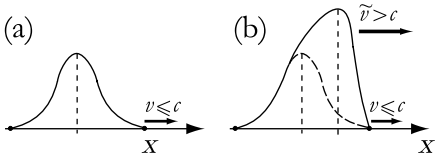


Fig. 7.13

Fig. 7.14



the medium. It is not surprising that we can detect a "signal" before the arrival of the original message. It takes some scrutiny to realize that the signal is false.

Consider now a more realistic model – a pulse of a finite width with a sharply defined leading edge (Fig. 7.14 a). There is no perturbation before the edge, and therefore the detector there records no signal at all (noise neglected!) – no matter whether the medium is inverted or not. The detector only starts to read a regular signal at the moment when the leading edge arrives at its location. It is natural, therefore, to define the signal velocity as the velocity of the leading edge. Once we have made this definition, it becomes crystal clear that the signal velocity cannot be affected by the amplification wave. The latter can only be originated by the initial perturbation, which is non-existent in front of the pulse with the sharply defined edge. The amplification wave starts within the span of the pulse, distorts and accelerates its top, but is stopped at the edge (Fig. 7.14 b). The distorted top can become superluminal, but it does not carry a signal; the edge does, but its velocity remains equal to or less than c.

The last question: what is the actual speed of the photons *inside* the superluminal pulse?

Let us again use the analogy with the Bump in Section 7.2. In the Bump's fast stage, when it appeared to move *faster* than the platform, not a single speck of earth shared this faster motion. The earth inside the Bump was not sliding forward relative to the platform; it was moving together with it as a passenger in a car. The apparent acceleration of the Bump was associated with *new* earth spilled down on to its front slope. The added earth contributed to the Bump's width, whose expansion in the forward direction appeared as acceleration of the Bump itself. Actually, each new lump of earth, once settled down on the Bump's surface, acquired its speed of motion together with the platform.

Similarly, the fast light's energy is not being carried along with superluminal speed. Like the new earth in the previous example, it is being donated to the pulse from the excesses of the excited medium as the pulse sweeps across. The donated energy is released in the form of photons that pop into existence from the excited atoms in front of the approaching pulse. Once there, the new photons start their life moving forward with a speed equal to or less than c. The apparent superluminal speed of the fast light results from the expansion of the pulse's width in the forward direction by acquiring more and more new photons in front. This expansion by gaining new particles is not the same thing as motion of particles themselves. The fast light is made up of normal photons moving together with a common speed not exceeding c. Since the light energy is carried by photons, we conclude that the energy in fast light also flows no faster than c.

Thus, as we take a close look at the fast light, there is no room left for the notion of a superluminal communication or superluminal energy transfer.

8
Tachyons and Tachyon-like Objects

To be, or not to be: that is the question ...
Shakespeare
Hamlet

8.1
Superluminal motions and causality

In previous chapters we have realized that the apparently simple concept of velocity turned out to be not that simple after all. A few quite different velocities can be associated with the same process. Here are some of them:

1. phase velocity (the propagation rate of a surface of constant phase)
2. the phase velocity of a bounded area at the crossing of rays (Sect. 6.14)
3. the group velocity
4. the velocity of signal (and energy) transfer.

We have seen that the first three types of velocity can take on any value (for the group velocity, recall Sect. 7.4), and it would not contradict anything. But there is one velocity – that of a signal transport – that does not exceed c in any observations. This special status of the signal velocity is attributed to the fact that signal (and thereby energy) exchange carry out the *causal connections* between spatially separated events. In order to see how the ban on superluminal signal transfer between causally connected events emerges from the existing theory, we have to discuss *causality* – one of the most important scientific concepts.

In the physical world, not a single event is isolated from others. One of the most important manifestations of causality is that the world's events always influence one another in a certain way. Namely, for any event (the *effect*) it is always possible to find at least one other event that has brought it into being – its *cause*. (There is one remarkable exception that does not fall into the scheme: the Big Bang, that brought our Universe into being. The Big Bang can be considered as the ultimate cause of everything in existence today; but what caused the Big Bang itself, or whether it had any cause at all, remains a murky issue at the time of writing this book.)

All observable events are governed by a fundamental principle: the cause precedes the effect. We call this principle the retarding causality (an effect occurs later than its cause). We introduce this principle here as an *additional element* in the description of

Special Relativity and Motions Faster than Light. Moses Fayngold

ISBN: 3-527-40344-2

the world. This additional element, combined with relativity, restricts the speed of any interactions transferring a signal. Let us see how it works.

Suppose that the signal velocity u can take on *any* value and consider an event A at a point $\boldsymbol{r}_A$ at moment t_A; let A cause another event B to happen at a position $\boldsymbol{r}_B$ and moment t_B. Draw the x-axis through points $\boldsymbol{r}_A$ and $\boldsymbol{r}_B$. Then the y and z coordinates of the two events are zero, and the positions of the events are characterized by their x-coordinates x_A and x_B so that separation between the events will be $\Delta x \equiv x_B - x_A$.

According to the principle of retarding causality, B happens later than A, that is

$$\Delta t \equiv t_B - t_A > 0 \tag{1}$$

Since the events are connected with the signal traveling at a speed u, we have

$$\Delta x = u\Delta t \tag{2}$$

Consider now another system K′, moving uniformly along the x-direction with velocity V. What time interval between the same events will be measured by an observer in K′? Assuming the axes x', y', z' in K′ running parallel to x, y, z, and using Lorentz transformations (44) in Chapter 2, we have

$$\Delta t' \equiv t'_B - t'_A = \gamma(V)\left(\Delta t - \frac{V}{c^2}\Delta x\right) \tag{3}$$

Now, using Equation (2), Equation (3) gives

$$\Delta t' = \gamma(V)\left(1 - \frac{uV}{c^2}\right)\Delta t \tag{4}$$

Thus, $\Delta t'$ is proportional to Δt. The factor $\gamma(V)$ is, according to its definition, positive for all $V < c$. As to the second factor $(1 - uV/c^2)$, it can generally have any sign depending on the signal velocity u. However, according to the principle of relativity, the causality relation for the considered pair of events must hold in *all* reference frames. Therefore there must be $\Delta t' > 0$. As is clearly seen from Equation (3), this means that the factor $(1 - uV/c^2)$ must always remain positive no matter the relative velocity V between the reference frames. And this can only be the case if the signal velocity u does not exceed c. Indeed, if it were possible for events A and B to be connected by a superluminal signal, that is $u > c$, one could always find a reference frame K′, moving relative to K with a speed

$$V > \frac{c^2}{u} \tag{5}$$

for which the factor $uV > c^2$, that is, $1 - uV/c^2 < 0$, and accordingly, $\Delta t' < 0$. This means that for the pair of causally connected events A and B the effect would be observed before its cause.

So here is the logical chain restricting the speed of causal interactions: the invariance of the speed of light requires the relativity of time; the relativity of time makes it possible for a succession of events to be different in different reference frames: an event A can precede B in a reference frame K and follow B in a reference frame K′ (recall, for instance, the phenomena discussed in Sections 5.4 and 5.5). However, if A and B are *causally connected*, then, according to retarding causality, their ordering must be the same for all observers, *despite* relativity of time. This requires the speed of any causal interaction between them not to exceed c. If this requirement is not met, then the time ordering of A and B can be reversed for an observer in some other reference frame K′. In the framework of the above reasoning, this would be violation of causality. To prevent this from happening, it seems to be necessary to exclude the possibility of superluminal signals.

8.2 The physics of imaginary quantities

The essence of almost all "bans" for material objects to move faster than light lies in the algebraic structure of the Lorentz factor:

$$\gamma(v) = \left(1 - \frac{v^2}{c^2}\right)^{-1/2} \tag{6}$$

The value of v here is either the relative velocity of the two inertial reference frames or the speed of an object. Provided that this speed remains less than c, everything runs smoothly. The problems arise when we set $v \geq c$. Let us discuss some of them.

We start with the Lorentz transformations (43) and (44) in Section 2.6. Consider an event with coordinates (x, y, z, t) in a reference system K. Then in another reference frame K′, which would move along the x-axis of K with velocity $v = c$, we would have $x' = \infty$, $t' = \infty$. This means that in the reference frame K′ all points of physical space and the times of all the events would be infinitely far away from the event O′ at the origin. In a conventional sense, they would not exist in real space–time. All physical concepts lose their conventional meaning in such a system. We therefore say that no reference frames (that is, material bodies carrying clocks and meter-sticks) can move with a speed c.

Consider now the case $v > c$. Then the Lorentz factor becomes imaginary:

$$\gamma(v) = \left(1 - \frac{v^2}{c^2}\right)^{-1/2} = -i\hat{\gamma}(v), \quad \hat{\gamma}(v) \equiv \left(\frac{v^2}{c^2} - 1\right)^{-1/2} \tag{7}$$

and we obtain

$$y' = y; \; z' = z; \; x' = -i\hat{\gamma}(v)(x - vt); \; t' = -i\hat{\gamma}(v)\left(t - \frac{v}{c^2}x\right) \tag{8}$$

Now the coordinates x' and t' are both finite, but they have imaginary values! Since all directly measurable physical quantities can be only real, we have to conclude that the space and time coordinates of events cannot be directly measured in superluminal reference frames. This, in turn, may lead us to conclude that such systems are impossible.

But from the mathematical viewpoint, the transformations (8) are as good for $v > c$ as they are for $v < c$. The main requirement of the invariance of an interval under the Lorentz transformations is satisfied in both cases. Let us check it for a relative velocity $v > c$. Putting Equations (8) into the expression for the interval of an event (Sect. 2.9), we perform somewhat tedious but straightforward manipulations:

$$
\begin{aligned}
c^2 t'^2 - x'^2 &= -\hat{\gamma}^2(v)\, c^2 \left(t - \frac{v}{c^2} x\right)^2 + \hat{\gamma}^2(v)\,(x - vt)^2 \\
&= \hat{\gamma}^2(v) \left[-\left(c^2 t^2 - 2\,vtx + \frac{v^2}{c^2} x^2\right) + (x^2 - 2\,vtx + v^2 t^2) \right] \\
&= \hat{\gamma}^2(v) \left[(v^2 - c^2)\, t^2 - \left(\frac{v^2}{c^2} - 1\right) x^2 \right] = c^2 t^2 - x^2
\end{aligned}
\tag{9}
$$

Thus, one can formally speak about superluminal Lorentz transformations (SLT) and superluminal reference frames [58]. However, such systems cannot be formed by ordinary matter.

What specifically is it in the physical properties of material bodies that does not allow it to form a superluminal reference frame? Let us consider first a material particle of a mass m_0, radius r_0, and a proper lifetime τ_0. According to Equation (8) in Section 4.1, the total mass depends on the particle's speed:

$$
m(v) = m_0 \gamma(v) \tag{10}
$$

When $v \to c$, the mass $m(v) \to \infty$, and so do the particle's kinetic energy and momentum. We have already emphasized that because of this no material body of finite mass m_0 can reach the limit of light velocity.

Now, apply the relativistic Equations (47) and (51) in Section 2.8 to the particle's size and lifetime:

$$
r_{\updownarrow}(v) = r_0\, \gamma^{-1}(v), \quad \tau(v) = \tau_0\, \gamma(v) \tag{11}
$$

where $r_{\updownarrow}$ is the longitudinal size along the direction of the particle's motion. Again, we see that when $v \to c$, the particle's longitudinal dimension goes to zero, and its lifetime goes to infinity. In other words, the particle moving with the speed of light would "lose" one of its spatial dimensions (it would degenerate into an infinitesimally thin disk perpendicular to the direction of motion) and "freeze" in its internal evolution.

The latter conclusion can be visualized in terms of the Doppler effect. Imagine an excited atom emitting light. Since light carries energy, the atom's excited state lasts

only a short time. If the atom is moving, the waves emitted in the forward direction become "compressed," while the waves emitted in the backward direction become extended (Fig. 6.12 with $u = c$). This decreases the rate of the energy output, thus increasing the atom's lifetime in the stationary reference frame. If the atom's speed could reach the limiting speed c, it would "ride on its own waves," the waves would not be able to depart from it, and there would be no energy loss. As a result, the excited state would last forever – it would freeze in time.

Now, what would happen if the particle could move faster than light in a vacuum?

Setting in Euations (10) and (11) $v > c$ and using the definition in Equation (1), we obtain

$$m(v) = -i\, m_0\, \hat{\gamma}(v)\,; \quad r_{\updownarrow}(v) = i\, r_0\, \hat{\gamma}^{-1}(v)\,; \quad \tau(v) = i\, \tau_0\, \hat{\gamma}(v) \tag{12}$$

The equations tell us that beyond the light speed barrier, the particle's mass, and thereby its energy and momentum, become imaginary. The same result follows for its longitudinal size and the lifetime.

Because all the observable properties of material objects are real, the appearance of the imaginary values in the theory indicates that corresponding quantities cannot be observed and measured. But what cannot, in principle, be observed does not exist. In other words, there cannot be any superluminal particles.

Thus, apart from the general requirement of retarding causality, the requirement for the observable physical quantities to be real excludes the possibility of the superluminal motions of physical objects. This conclusion had for a long time been regarded as absolutely clear, and had not been subjected to serious doubt. Not until recently.

8.3 The reversal of causality

There is a fascinating story written by an outstanding popularizer of science, Camille Flammarion, long before the appearance of the theory of relativity [59]. The main character of the story leaves the Earth and starts receding from it with a superluminal velocity. In this way, he outruns the electromagnetic waves from Earth, in which is encoded information of all the events of the Earth's history. Our hero catches up first with the waves that were emitted recently, and then with the waves having started earlier. Accordingly, he observes the whole historical process in reverse succession, as in a movie run backwards. For example, in the battle of Waterloo he sees first the battlefield soaked with blood and covered with corpses. The blood then gets absorbed back into the corpses of the dead soldiers, they come back to life, jump up, grab at the weapons having flown into their hands, and run backwards to form their original units. The cannon balls burst out of the earth pits and fly into cannon barrels. Then the columns of the hostile armies, marching backwards, diverge in different directions.

This is a very unusual world, where people would live their lives backwards, first emerging from their graves, then changing into babies and returning into their mothers' wombs. The amount of disorder in such a world would decrease, and the

amount of order would increase. According to thermodynamics, that studies subtle connections between the observable macroscopic phenomena and the motions of the constituent micro-particles, the probability of such a world is zero. But in all other respects, this reversed world, for all its apparent weirdness, would be subordinated to laws that are intrinsically consistent. It would follow the rule of cause and effect. The only difference is that compared with our usual world, the cause and effect switch roles. What is the cause in our world is the effect in the described one, and vice versa. For instance, the cause of a cup of tea jumping on to the table would be its self-assembling from the splinters on the floor, absorbing moisture from it and collecting heat, part of which would accumulate into kinetic energy. Although some of the laws of nature appear to be turned inside out, causality not only conserves but, strangely enough, even retains its retarding character. This is due to the fact that simultaneously with the reversal of time, the cause and effect *change roles*.

Most of the laws of nature are invariant with respect to the time reversal. This means that, unlike the macroscopic world, which seems different and strange when run backwards, in the micro-world of single particles there is often no difference between the direct and reversed flow of time. Let, for instance, an excited atom A_1 radiate a photon at a moment t_1 and return to its ground (normal) state. The emitted photon becomes absorbed by another atom A_2 at a later moment t_2, and causes its transition from the ground state to the excited state. Clearly, the cause of the excitation of the atom A_2 was the photon emission from the atom A_1. Let us now reverse the process. Then we will first observe the radiation of the photon by atom A_2 due to its optical transition to the lower state at the moment t_2. This will cause the excitation of the atom A_1 due to its absorption of the photon at the moment t_1. Because time is reversed, the moment t_1 now occurs *later* than the moment t_2, so that again, the cause precedes the effect. Despite the reversal of time, there is nothing unusual in the resulting process.

In contrast, the reversing of macroscopic phenomena seems unusual, but the laws of nature remain self-consistent, because synchronously, the cause and effect also change their roles. Ordinarily, if a hare is shot dead by a hunter, the hunter's shot is the cause and happens earlier in time, while the death of the hare is the effect and happens later. In the time-reversed world, the dead hare would suddenly resurrect, with the bullet emerging out of it, and then this bullet, moving backwards, would whack into the barrel of the hunter's gun. One would now call the first event (the emission of the bullet from the hare) the cause, and the second one (the "absorption" of this bullet by the gun) the effect. This *reinterpretation* of the cause and the effect saves the principle of the retarded causality in the time-reversed world.

Let us now apply a similar trick to the problem of superluminal signals. Suppose that our atoms exchange superluminal signals instead of photons. Let such a signal be represented by a fictitious superluminal particle. Imagine observing such a particle emitted by an atom A_1 at a moment t_1 and then absorbed by another atom A_2 at a later moment t_2 in an inertial reference frame K. It is clear that the first event is the cause of the second. But we know already that for a superluminal signaling the interval between the corresponding events is space-like, and one can always find such inertial reference frame K′, in which the time ordering of the events changes. This

seems to contradict retarded causality. However, one can avoid this contradiction in a way similar to that described for the time reverse, but in a more limited sense: one might reinterpret the cause and effect *only* for the events along the space-like intervals and their end points, when their time ordering is reversed under the corresponding Lorentz transformation (that is, when we transfer to another reference frame moving sufficiently fast). The superluminal agent moving from A to B and causing some change in B would be observed from another reference frame as moving from B to A and causing a corresponding change in A.

"What's the big deal?," one might think. "This is a familiar effect, I often see it during driving, when I happen to outrun a pedestrian strolling in the same direction on the sidewalk. Relative to my car, the pedestrian then appears to move in the opposite direction."

But this would be a false analogy. If the pedestrian first crossed 6th Street, and then 7th Street, you will from your car see him doing this in the same succession. You will *not* see him crossing 7th street first and 6th street after that, no matter how fast you drive.

The situation with a superluminal particle is totally different. You do not (and cannot) outrun such a particle. And yet you can see its motion in reverse – literally in reverse, that is – crossing 7th Street first, and only then 6th Street. This is a purely relativistic effect, when the two events are interchanged *in time* for an observer in another reference frame.

We now can describe some implications of the above properties of the space-like trajectories on a macroscopic scale. Imagine that tachyons do exist and people have learned how to manufacture superluminal bullets out of them. Imagine that the hunter Tom fires such a bullet and kills a hare. Because the bullet is superluminal, these two events (the shot and resulting death of the hare) are connected by a space-like interval.

Now consider the same process from the viewpoint of Alice flying by in a spaceship. Traveling in a spaceship does not produce any *global* time reverse of the type described in the beginning of this section, so Alice will observe Tom's and the hare's lives in their normal course. In Alice's reference frame, as in Tom's, Tom first aims, then shoots; the hare first grazes, then dies. And yet in the shooting episode she will see something strange (it so happens that Alice often gets into strange situations). Here is her account.

"I flew by and watched a hare frolicking on a forest meadow. Then all of a sudden the hare dropped dead. A bullet burst out of it and zipped away with a stupendous speed. Then I saw my friend Tom hunting. His behavior was a little weird. He noticed the hare and took good aim at it as if the hare were not dead. At this moment the bullet from the hare struck Tom's gun right in the barrel, and a tiniest fraction of a second later Tom pulled the trigger. Then he ran to see what had happened to the hare. It appears to me from what I saw that the hare died by itself and produced that horrible bullet aimed at Tom, and the recoil of Tom's gun was the effect of this event."

As you compare Alice's and Tom's accounts, you will see obvious contradictions between them. Tom insists that he has fired first and killed the animal with his bullet.

His shot was the cause and the hare's death was the effect. Alice witnessed that the hare has died first, and its death was accompanied by the emergence of the bullet that caused the recoil of Tom's gun.

Who is right?

Both are, because the time ordering of the events separated by the space-like interval is relative, and so may be the designation – which event is the cause and which is the effect. Alice's reinterpretation of what is the cause and what is the effect is logically consistent and helps save the principle of retarded causality. By using the reinterpretation, the principle holds in either reference frame.

The possibility of such reinterpretation would mean that superluminal communications do not by themselves contradict retarded causality. One can therefore speculate about the possibility of the existence of the superluminal particles and superluminal communications.

It would be much more difficult for Alice to explain why Tom's aiming and triggering his gun were so remarkably accurately timed with the arrival of the bullet from the hare. In Alice's reference frame, the triggering of the gun is not the cause of the bullet having flown into it. Nor is it its effect. At the same time there is an obvious correlation between them – a non-causal correlation. It is manifest in the time coincidence between them. A possible explanation is that this is just a chance coincidence. Such a coincidence would be, of course, extremely unlikely, but logically possible.

And what about Equations (12), which prohibit real particles from moving faster than light? We will discuss these questions in more detail in the following sections.

8.4
Once again the physics of imaginary quantities

"Suppose that someone studying the distribution of population on the Hindustan Peninsula cockshuredly believes that there are no people north of the Himalayas, because nobody can pass through the mountain ranges! That would be an absurd conclusion. The inhabitants of Central Asia have been born there; they are not obliged to be born in India and then cross the mountain ranges. The same can be said about superluminal particles."

These lines belong to an Indian physicist, Sudarshan, who was one of the first to revive the concept of superluminal particles [60, 61]. They answer the question at the end of the previous section.

Indeed, as we know, Equations (10) and (11) prohibit the values $v \geqslant c$ for a massive object. If such an object at a certain moment moves slower than light, then it cannot acquire a speed faster than light. Not only cannot such objects cross the light barrier, they cannot even reach it because this would require an infinite amount of energy and momentum.

And yet the equations do not rule out the possibility of the existence of the objects that *always* move faster than light. After all, we know of the existence of photons which are thriving and can only live at the speed $v = c$, whereas Equation (10) prohi-

bits this speed! And nothing horrible happens to photons, they all have decent finite energies and momentums.

How do the photons get around the ban? Very simple! The photon's energy and momentum – the quantities that we can measure! – remain finite because the infinity of its Lorentz factor is multiplied by zero. The photon's *rest mass* is equal to zero. Zero rest mass of an object means the absence of the resting object itself. In other words, a stationary photon in a vacuum is impossible. We have come again to the known result that one cannot stop a photon in a vacuum. This result means that one cannot slow down a photon to a non-zero speed $v < c$ either, because we could always find a co-moving reference frame in which such a photon would be stationary. The photons can only exist by balancing on the razor's edge – by moving with the speed c, which is unattainable for any "massive" particle.

Thus, the divergence of the Lorentz factor $[\gamma(v) \to \infty]$ at $v \to c$ means only that it is impossible to accelerate a "massive" ($m_0 \neq 0$) particle up to the speed of light; it does not exclude the objects with the zero rest mass, for which always $v = c$. And this is consistent with the fact that the value of c does not depend on the choice of a reference frame – it does not change under the Lorentz transformations – it is absolute.

Now, we can apply the same reasoning to motions faster than light!

Just as the divergence of the Lorentz factor at $v \to c$ is compensated for by the zero rest mass for a photon, the imaginary value of this factor at $v > c$ for a superluminal particle can be compensated for by an *imaginary* value of its rest mass. The same can be said about a proper longitudinal size and a proper time of such a particle. According to Equations (12), they all have to be imaginary to compensate for the imaginary value of $\gamma(v)$. Let us write this down in symbols:

$$m_0 \Rightarrow \tilde{m}_0 \equiv i\, m_0\,; \quad r_0 \Rightarrow \tilde{r}_0 \equiv i\, r_0\,; \quad \tau_0 \Rightarrow \tilde{\tau}_0 \equiv i\, \tau_0 \tag{13}$$

Here and hereafter we will frequently denote quantities related to tachyons by symbols with the "tilde" symbol, ~. Equation (13) states that tachyon's "rest mass" $\tilde{m}_0$, "proper radius" $\tilde{r}_0$, and "proper lifetime" $\tilde{\tau}_0$ are all imaginary. This conclusion does not contradict anything, because the proper values $\tilde{m}_0$, $\tilde{r}_0$, and $\tilde{\tau}_0$ are not observable physical quantities for superluminal motions. They are characteristics of the stationary state, but the particles moving faster than light cannot be stationary relative to ordinary matter. To bring a superluminal particle to rest, we must board a spaceship moving faster than light and catch up with the particle; but no spaceship made of the ordinary matter can move faster than light. The superluminal reference frame made of ordinary matter co-moving with a superluminal particle is in principle impossible; therefore, it is impossible to make any direct measurement of their *proper* characteristics – which is manifest in the fact that their values are imaginary. At the same time the observable (not proper!) values of the energy (and thereby the total mass $\tilde{m} = \tilde{E}/c^2$), momentum, size, and the lifetime, which can be measured during the passing of a superluminal particle, turn out to be real when we make the transition (13) in Equations (12), so that we have a self-consistent picture.

We could also, in principle, measure $\tilde{m}_0$, $\tilde{r}_0$, and $\tilde{\tau}_0$ indirectly in a fairly simple way. Consider, for instance, measuring the rest mass of a superluminal particle. To em-

phasize that the rest mass is imaginary, let us write it according to Equation (13) as $\tilde{m}_0 = i\,m_0$, where m_0 is a real number. We could measure it, for instance, by measuring the total energy and then using $\tilde{m} = \tilde{E}/c^2$. We can measure simultaneously the speed $\tilde{v}$. Then, knowing $\tilde{m}$ and $\tilde{v}$, we can calculate m_0 using Equation (12).

Alternatively, we can use the relativistic energy–momentum relation:

$$\tilde{E}^2 = \tilde{p}^2 c^2 + \tilde{m}_0^2 c^4 = \tilde{p}^2 c^2 - m_0^2 c^4 \tag{14}$$

and measure the energy and momentum of a superluminal particle. Then we can calculate m_0 directly from Equation (14).

An even more "exotic" remedy can be found for the problem of imaginary proper times and distances measured in superluminal reference frames (Sect. 8.2). We will illustrate this remedy first graphically, and then analytically.

Look at Figure 8.1. It represents a moving reference frame K′ from the viewpoint of a frame K, which is considered stationary. The coordinate axes of K′ are skewed with respect to K. We know the physical meaning of this geometrical distortion (Sect. 2.9): the events along the spatial axis x', which are all instantaneous in system K′, are not instantaneous in K, so that the world line connecting these events has a time component in K. Similarly, the consecutive events along ct', which all happen at one place in K′, are observed at different points of space in K, so the world line connecting these events has a spatial component to it. However, although the two axes are skewed in K′, the x'-axis remains space-like, and the ct'-axis remains time-like.

Imagine now the system K′ moving faster than light. Then a strange thing happens (Fig. 8.1 b). The spatial axis x' of K′ will lie in the "time-like" domain of space–time,

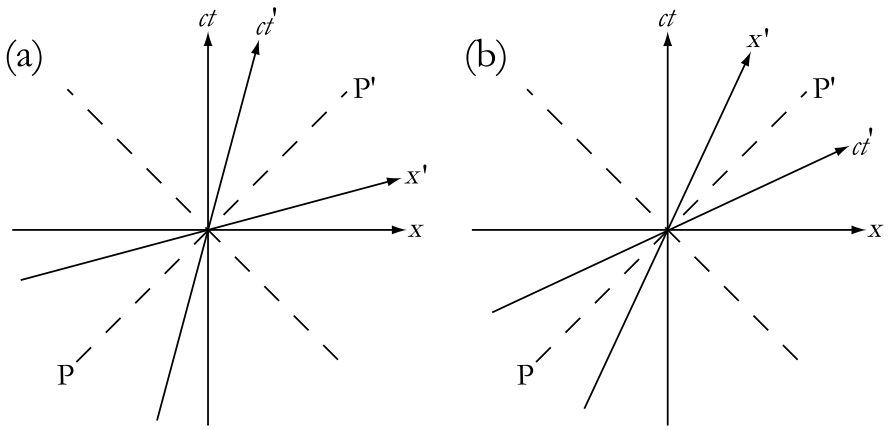


Fig. 8.1 (a) The axes of reference frame K′ represented in the reference fame K. From the viewpoint of K, the axes ct', x' are rotated toward each other (to the photon world line PP′). Were system K′ able to move with the speed of light, the axes ct', x' would both merge with the line PP′, and there would be no difference between time and space in this system. (b) The same for a hypothetical reference frame K′ moving faster than light. The x'-axis would then lie in the "time domain" of K, and the ct'-axis would lie in the "space domain".

and the temporal axis ct' will lie in the "space-like" domain! The axes exchange their roles. What is time for K is space for K′, and vice versa! More accurately: of the three space dimensions, any one along the direction of relative supeluminal motion of the two reference frames is interchangeable with time. If material objects could move faster than light, then by selecting the direction of relative motion, we could make time interchangeable with any one of spatial dimensions. This astounding conclusion follows directly from the diagrams in Figure 8.1. In terms of the space–time physics, we can understand this in the following way.

If the two events at the points x_1 and x_2 on the x-axis of system K are separated by a space-like interval, this interval is time-like in K′. In particular, if K′ is moving at such a speed that its origin passes points x_1 and x_2 at the respective moments when the above two events happen there, then both events occur at the origin of system K′, so that the interval between them in K′ is a pure time $c\Delta t'$. If the two events happen simultaneously in K, so that the interval between them is purely spatial distance $\Delta x = x_2 - x_1$, and K′ is moving infinitely fast, then a pure space interval Δx in K is changed to a pure time interval $\Delta t'$ in K′.

Similarly, if the two consecutive events occur at the origin of K at the moments t_1 and t_2, so that the interval between them is just a pure time $c\Delta t = c(t_2 - t_1)$, this interval is space-like in K′. Indeed, the origin of K slides down the axis x' of system K′ to the left with a speed v. Because this speed is superluminal, the two events at the origin of K will be separated in K′ by the distance $\Delta x' = v\Delta t'$ greater than $c\Delta t'$. Thus, the space-like component of the interval between the events is larger than the time-like component, so the total interval is space-like. If again the speed v is much greater than c, then $\Delta x' \gg c\Delta t'$, the temporal component can be neglected, and the interval in K′ would be almost pure distance in space.

These results follow directly from the superluminal Lorentz transformations (8). For instance, setting there $ct' = 0$ (set of simultaneous events in K′ , forming the space in K′) yields $x = (c/v)\,ct < ct$ (the set of the same events turns out to form a time-like world line in K). If we set in Equations (8) $x' = y' = z' = 0$ (all the events occurring at the origin of K′, that is the world line of the origin – a pure time interval in K′), the equation yields $x = (v/c)\,ct > ct$ (the same events turn out to form a space-like world line in K).

Thus, the basic assertion of Einstein, that time and space are relative properties and can be mixed together to form a more general entity, space–time, can potentially be extended still further. We can now say that if superluminal objects exist, then space and time can be converted into one another via superluminal Lorentz transformations.

Realizing this allows us to interpret the imaginary values of the transformed coordinates x' and ct' in system K′. Squaring the imaginary time coordinate ct' will give the negative contribution to s'^2. Also recall that by definition of the interval, the squares of the spatial coordinates are *subtracted* from the square of the time coordinate. Therefore, when the corresponding coordinate (in our case x') is imaginary, its square will give the *positive* contribution to the s'^2. Now, according to the same basic definition of the square of the interval, the coordinate whose square enters s'^2 with the plus sign is the time coordinate, and the one whose square enters with the minus sign is the spatial coordinate. Hence ct' now plays the role of a space coordi-

nate, and x' plays the role of the time coordinate. We have already realized that this is precisely what happens under superluminal Lorentz transformation. Now we come to the same conclusion from the analysis of the imaginary values of the transformed coordinates. These values actually tell us that we should *reassign* the notations for coordinates in a superluminal reference frame: $ct' = -i\tilde{x}'$, $x' = -ic\tilde{t}'$. Then the expression for the interval in system K′ will take the form

$$s'^2 = c^2t'^2 - x'^2 = c^2\tilde{t}'^2 - \tilde{x}'^2 \tag{15}$$

While the variables t' and x' here are imaginary and have a meaning opposite to their original notations, the variables $\tilde{t}'$ and $\tilde{x}'$ are real and have the physical meaning of time and space coordinates, respectively, in system K′.

Hence we come to a conclusion that it is possible to introduce into physics a new type of particle, which is different from all the others in that it always moves faster than light. This kind of particle has been named, at the suggestion of John Feinberg [62], the *tachyon* – after the Greek word meaning "fast". The concept of tachyons makes the world more symmetrical by allowing the existence of natural objects on *both* sides of the light barrier, so that the latter becomes the two-sided limit for all possible speeds. The fuller symmetry makes the world appear more perfect, which appeals to our esthetic feelings. Therefore, the curious reader may attempt to venture into the new uncharted waters. The ideas we are going to describe in the following sections may appear to be unusual, and some of them may be controversial, but this is always the case when we cross the boundaries of established knowledge.

8.5 Tachyons and tardyons

Once we have realized that the existence of tachyons can be a logical possibility within the framework of the special theory of relativity, we can explore the emerging new domain. We will then find a striking symmetry between the world of tachyons and our conventional physical world.

First let us introduce new terms. If we have given a name to the superluminal particles, it is reasonable to do the same for the subluminal particles. Using the same arsenal from ancient Greek, physicists have dubbed regular, well behaved subliminal particles the *tardyons* (the English words *retardation, retarded* stem from this root.) Thus, all the particles that can possibly exist in the Universe fall into three different categories according to their speed: tardyons ($v < c$), photons and gravitons ($v = c$), and tachyons ($v > c$).

It is easy to see that these three categories correspond to the three different domains of Minkowski's world (Fig. 8.2). Pick an event O in space–time. Let it be a reference event. Consider all possible world lines passing through the event O. *Each* line can be a four-dimensional path of a particle. The world lines of tardyons are all time-like and fill out the interior of the light cone with the apex at event O. The world lines of tachyons are space-like and fill out the exterior of this cone. The world lines of

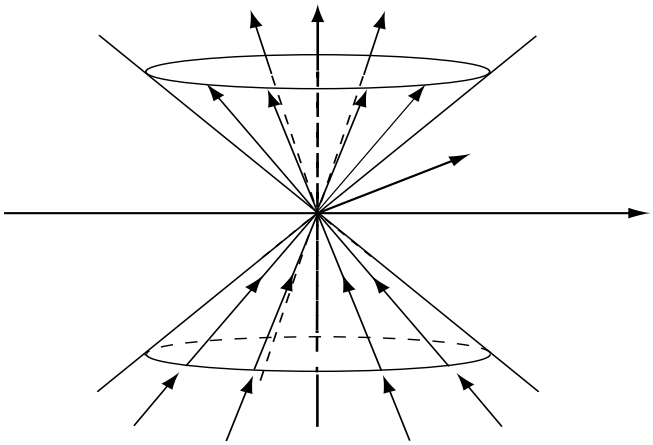


Fig. 8.2 The world lines of tachyons, tardyons, and photons.

photons (and gravitons) are isotropic (have zero kinematic length!) and form the generatrices of the light cone.

Next, let us look again at some basic concepts of relativistic kinematics – the kinematics of tardyons and photons. We will see how easily it can incorporate the tachyons.

We start with the basic expression for the interval ds between two close events on the world line of a particle moving with a speed v:

$$\mathrm{d}s^2 = c^2\,\mathrm{d}t^2 - \mathrm{d}r^2 = c^2\,\mathrm{d}t^2/\gamma^2\,(v) \tag{16}$$

Recall that according to its definition in Equation (25) in Section 2.5, the Lorentz factor γ is positive for the tardyons, infinite for the photons, and imaginary for the tachyons. Therefore, the interval ds is real for a tardyon, and for the photons it is zero,- just as one would expect from the definition of the photon's world line. As for the tachyons, the value of ds turns out to be imaginary. But this does not by itself preclude tachyons from existence, because the interval is not directly measurable quantity. It is only the mathematical expression formed from dt and d$\boldsymbol{r}$, which are all real for all kinds of particle.

The same can be said about the components of 4-velocity of a tachyon. Although the magnitude of 4-velocity, according to its definition [see Eqs. (1) and (2) in Section 4.1], is always equal to 1 for *any* kind of particle, the components of 4-velocity for tachyons

$$u_0 = \gamma\,(\tilde{v})\,; \quad u_a = \frac{\tilde{v}_a}{c}\gamma\,(\tilde{v})\,, \quad \alpha = 1, 2, 3 \tag{17}$$

turn out to be imaginary [Eq. (4 b) in Section 4.1]. But this does not mean that these components cannot be observed. Unlike the usual coordinates and velocities, as well

as energy and momentum, the components of *4-velocity* are not directly measurable quantities. But they can be measured indirectly, by computing from directly measured components of $\tilde{\boldsymbol{v}}$, which are always real.

In the special theory of relativity the product of the rest mass by c and by the 4-velocity gives the 4-momentum [Eq. (6) in Sect. 4.1]. For an ordinary particle

$$p_j = m_0\, c\, u_j\,, \quad u_j = 0, 1, 2, 3 \tag{18a}$$

For a tachyon

$$\tilde{p}_j = \tilde{m}_0\, c\, \tilde{u}_j \tag{18b}$$

Since both factors $\tilde{m}_0$ and $\tilde{u}_j$ are imaginary for a tachyon, its momentum is real.

All the components of 4-momentum have a simple physical interpretation. Setting in Equation (18 b) $j = 0$, one has

$$\tilde{p}_0 = \tilde{m}_0\, c\, \gamma\,(\tilde{v}) = \frac{\tilde{E}}{c} \tag{19}$$

and for the spatial components $\alpha = 1, 2, 3$ (that is, x, y, z):

$$\tilde{p}_a = \tilde{m}_0\, \tilde{v}_a\, \gamma\,(\tilde{v}) = \tilde{m}\, \tilde{v}_a \tag{20}$$

Thus, the zeroth component (in analogy with the time coordinate in Minkowski's space–time) is just the energy divided by c, and the spatial components are just the components of the regular relativistic 3-momentum. Thus, the components of the 4-momentum for a tachyon have the same meaning as the components of the 4-momentum [Eqs. (9)–(11) in Sect. 4.1] for a tardyon, obtained in Section 4.1.

Based on this interpretation, we obtain the relation between the energy and momentum of a tachyon:

$$\frac{\tilde{E}^2}{c^2} - \tilde{\boldsymbol{p}}^2 = \tilde{m}_0^2\, c^2 \tag{21}$$

in the same way as we did in Equation (11) in Section 4.1 for a tardyon.. The reader should keep in mind that while the tachyon's rest mass $\tilde{m}_0$ is imaginary (it cannot be brought to rest in *our* space), its energy and momentum are real (and therefore measurable) physical quantities.

The left-hand sides of Equation (11) in Section 4.1 and Equation (21) here look exactly the same as the square of the four-dimensional interval between two events (recall Sections 2.9 and 4.1). They show that all the properties of an interval apply to any four-dimensional vector.

With this in mind, we see from Equation (11) in Section 4.1 and Equation (21) here that the 4-momentum of a tardyon is time-like (has real magnitude $m_0 c$), and the 4-momentum of a tachyon is space-like (has imaginary magnitude $\tilde{m}_0 c$). This is just

another way to say that tardyons always move slower that light and reside in the interior of the light cones, and tachyons, if in existence, move faster than light and reside in their exteriors.

For a photon ($m_0 = 0$) the 4-momentum is isotropic (has a zero length). This is a natural result of the fact that the photons' world lines form the generatrices of light cones.

These statements are expressed analytically in three simple relations:

$$\left.\begin{aligned} E &> pc \\ E &= pc \\ \tilde{E} &< \tilde{p}c \end{aligned}\right\} \tag{22}$$

for a tardyon, a photon, and a tachyon, respectively.

We can now express the speed of a particle in terms of its energy. Using Equation (21), we have

$$v = \frac{\mathrm{d}E}{\mathrm{d}p} = \frac{p}{E}c^2 = c\sqrt{1 - \frac{m_0^2 c^4}{E^2}} < c \tag{23a}$$

$$\tilde{v} = \frac{\mathrm{d}\tilde{E}}{\mathrm{d}\tilde{p}} = \frac{\tilde{p}}{\tilde{E}}c^2 = c\sqrt{1 - \frac{\tilde{m}_0^2 c^4}{\tilde{E}^2}} > c \tag{23b}$$

Now we want to depict graphically Equation (21) and compare the graphs for different kinds of particles. In order to remove inessential details, consider the case when the vectors $\boldsymbol{p}$ and $\tilde{\boldsymbol{p}}$ are collinear, and the *absolute values* of masses are equal (m_0 has the same value for both kinds of particle). Then the 3-vectors of momentum are reduced to only one component along their common direction, and the equations take the form

$$E^2 - p^2 c^2 = m_0^2 c^4 \tag{24a}$$

$$\tilde{E}^2 - \tilde{p}^2 c^2 = -m_0^2 c^4 \tag{24b}$$

In a coordinate system where the energy and momentum of a particle are plotted along the axes (Fig. 8.3), Equations (24) describe a hyperbola whose slope at each point (for each pair of variables p and E) determines the speed of a particle with given values of the variables. For a regular tardyon (the branch of the hyperbola in the upper part of the plane) this slope is everywhere less than 1. No matter how much we increase the energy and momentum of the particle, the corresponding point in the plane will slide along the curve ever further from the origin, and the speed of the particle, although approaching ever closer the speed of light, will remain less than c.

Of special interest is the point where $p = 0$. It corresponds to a particle at rest with the minimum possible energy. This minimum, as is seen from Equation (24a) at $p = 0$, is equal to $E_0 = m_0 c^2$. It is called the rest energy of the particle.

If the rest mass of the particle is zero (a photon or graviton), so is its rest energy. This means that if you try to stop such a particle, you are left with nothing. Particles

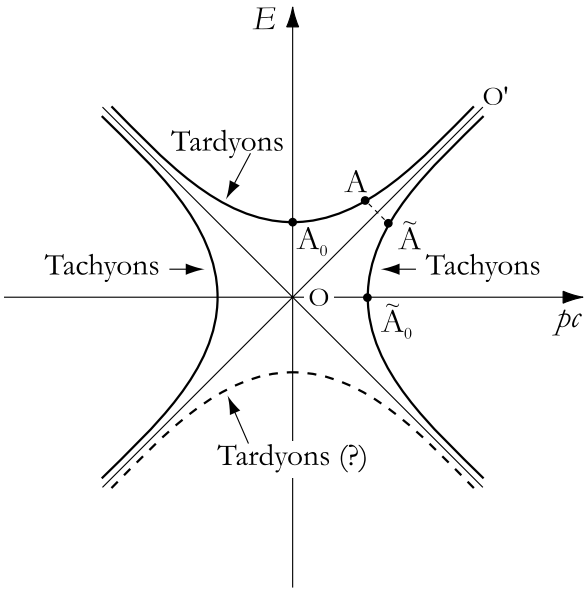


Fig. 8.3 $E \leftrightarrow p$ relations in the momentum space.

of this kind do not exist at rest. The corresponding Equation (24a) splits into two simpler equations:

$$E - pc = 0 \tag{25a}$$
$$E + pc = 0 \tag{25b}$$

that is, $E = \pm pc$. They describe a hyperbola, degenerated into two intersecting straight lines passing through the origin. Physically they correspond to the propagation of photons, whose energies, as we know, satisfy Equations (25). The straight lines represented by Equations (25) lie on the generatrices of a light cone in "momentum space" and are the asymptotes of the hyperbolas represented by Equations (24).

In geometry, the set of hyperbolas described by the equation $x^2 - y^2 =$ constant consists not only of the curves for which the constant is positive or zero, but also includes the curves for which the constant is negative. A pair of such curves is shown in Figure 8.3 – they are the right and the left branches of a hyperbola, with their apexes on the pc-axis. These branches are described by Equation (24b) and correspond to tachyons. Thus, by allowing tachyons to exist, we give a physical meaning to this group of curves in the hyperbolas family, that is, introduce an element completing the picture to the full symmetry.[1)]

The hyperbolas of this group, as is clearly seen from Figure 8.3, have at any point a slope (that is, $d\tilde{E}/c\,d\tilde{p}$) *larger* than 1. But the difference from tardyons is not restricted

1) In fact, the full symmetry is not achieved even in this case, because the tardions with negative energies corresponding to the lower branch of the hyperbola are unknown. We will not discuss the related topics here.

to this distinction only. The dependence of the tachyon's energy and momentum on its speed is also dramatically different from that of the tardyon, namely, the energy and momentum of a tachyon *decrease* with the increase in its speed! Look at a branch of the hyperbola corresponding to tachyons. As a point on this branch slides away from its apex, both $\tilde{E}$ and $\tilde{p}$ increase, while the slope of the curve approaches 1, that is, the speed of the tachyon decreases, approaching c. Thus, to slow down a tachyon with $\tilde{E} > 0$, one has to pump it with additional energy, and to speed up the tachyon one has to subtract energy from it!

Such behavior at first seems paradoxical, but if we give it more thought, this behavior is natural and even necessary in the world of superluminal velocities. The mere term "superluminal" means that the speed of light is the *lower limit* for this type of particle. To approach this limit, the particle must slow down, and for the limit to be unattainable, the slow down must require unlimited energy supply. This "paradoxical" behavior of the tachyons ensures the symmetry of the light barrier: no matter from which side an object approaches the barrier (the speed of a particle approaches c), this is accompanied by unlimited growth of energy and momentum of the object.

The curves $\tilde{E}(v)$ and $\tilde{p}(v)$ illustrate another weird feature in the behavior of tachyons: as the tachyon is accelerated to an infinite speed, its energy does not just decrease, but goes to zero independently of its proper mass $\tilde{m}_0$, and its momentum remains finite and goes to $m_0 c$.

The speeds $v = 0$ and $v = \infty$ can be considered symmetrical with respect to c in that either of them is maximally remote (in the corresponding domain) from the light barrier. We can say that the speed $v = \infty$ plays the same role for a tachyon, as the speed $v = 0$ for a tardyon. However, if we compare the energy and momentum of the tardyon at $v = 0$ with those of the tachyon at $v = \infty$, we will notice another peculiarity. Whereas for a tardyon

$$E(v = 0) = m_0 c^2 , \quad p(v = 0) = 0 \tag{26}$$

for an "equivalent" tachyon with "symmetrical" speed $v = \infty$ one has

$$\tilde{E} = 0 , \quad \tilde{p} = m_0 c \tag{27}$$

The quantities E and p change roles: the energy of the tachyon behaves more like the momentum, and the momentum behaves more like the energy. This is a natural consequence of the "reinterpretation" of the meaning of the temporal and spatial coordinates for tachyons, as discussed in Section 8.4.

We call the particle with $v = 0$ the stationary particle. The tachyon with the infinite speed also deserves a special name. It has been named transcendent. The properties of transcendent tachyons are very unusual. Such a tachyon, tracing out the whole space in an instant, is observed at all points of its trajectory at once. But this observation lasts only an infinitesimally short moment, because owing to the infinite speed of the transcendent tachyon, it emerges and momentarily disappears simultaneously at all points on its track. Therefore, a strange phenomenon is observed in the corre-

sponding reference frame: at first there is nothing there, and then there suddenly appears and momentarily disappears an infinitely long rigid "rod" consisting of the tachyon "smeared out" along its whole length.

In this respect, a stationary tardyon and a transcendent tachyon act as certain kind of antipodes: the former stays at one point in space throughout the whole time; the latter stays for only an instant at all points of a spatial line. The former has zero momentum and finite energy $E = m_0 c^2$, whereas the latter has zero energy and final momentum $\tilde{p}c = m_0 c^2$. The former is represented by an apex A_0 of the hyperbola (24 a), and the latter by the point $\tilde{A}_0$ of the hyperbola (54 b), which is symmetrical to A_0 with respect to a photon line OO′. The equations

$$\tilde{E} = pc\,, \quad \tilde{p}c = E \tag{28}$$

are the analytical expression of this symmetry. In some respects, the resting tardyon and the transient tachyon of the same mass m_0 are the symmetrical counterparts of one another.

But there is more to it! If we take a closer look at Figure 8.3, it hints at an obvious generalization of this symmetry. Suppose we pick up a tardyon with some *arbitrary* speed $v < c$. It will be represented by a point A with corresponding "coordinates" E and pc on branch 1 of the hyperbola in Figure 8.3. Can we find a tachyon symmetrical to this tardyon? The graph suggests a positive answer. It would be the tachyon represented by a point symmetrical to A with respect to the asymptote (generatrix) OO′. The "coordinates" of this point are related to coordinates E and pc by Equations (28). The corresponding vectors (E, pc) and $(\tilde{E}, \tilde{p}c)$ satisfy the equation

$$E\tilde{E} - p\tilde{p}c^2 = 0 \tag{29}$$

which is equivalent to Equations (28). In the geometry of Minkowski's world, the expression $E^2 - p^2c^2$ determines the square of the 4-vector (E, pc), and thereby its magnitude (kinematic length). Similarly, the expression $E\tilde{E} - p\tilde{p}c^2$ determines the scalar product of the two different vectors. In our case this product is equal to zero. As we know from geometry, it means that the two vectors are perpendicular. The fact that the vectors OA and OÃ in Figure 8.3 do not look mutually perpendicular, is caused by inadequacy of the graphical representation used: we have to represent the relations of the *pseudo-Euclidian* geometry on the ordinary Euclidian plane. For the pseudo-Euclidian space we say that these vectors are *dual* to one another. Let us also call the tardyon and symmetrical tachyon represented by mutually dual vectors on the diagram in Figure 8.3, mutually dual. The property of two particles to be mutually dual is Lorentz invariant. If this property is found for a pair of particles in one inertial reference frame, then owing to Lorentz invariance of the scalar product of the representing vectors, it will be found in any other inertial reference frame.

Now, here is an interesting question: how are the speeds of mutually dual tardyon and tachyon related to each other? According to the general definition of speed [Equation (14) in Section 4.2], we have

$$\tilde{v} = \frac{\tilde{p}}{\tilde{E}}c^2\,, \quad v = \frac{p}{E}c^2 \tag{30}$$

so that

$$v\tilde{v} = \frac{p\tilde{p}}{E\tilde{E}}c^4 = c^2 \tag{31}$$

or

$$\tilde{v} = \frac{c^2}{v} \tag{32}$$

We have obtained this same relation between two velocities a few times, on different occasions, and now we have it again! Recall, for instance, that the same relation connects the phase and group velocities of the de Broglie waves of a free particle.

Hence the speed of a tachyon dual to a given tardyon is equal to the phase velocity of the de Broglie wave associated with this tardyon! This allows us to suggest that the concept of a tachyon not only is logically possible within the framework of the special theory of relativity, but also has some physical meaning. As we think of it, there appears an impression that we have caught a glimpse of something deep. But today we can only say that *if* the tachyons do exist, then tardyons and tachyons can come in dual pairs whose characteristics are described by Equations (28) and (29) or by the equivalent equations

$$\tilde{m}_0 = im_0\,, \quad \tilde{v} = c^2/v \tag{33}$$

We want to emphasize once again that the transition from one member of a dual pair to another described by the transformations in Equations (28), (29), and (33) cannot be realized by a continuous change from v to $\tilde{v}$ or vice versa through the light barrier. We have already shown this using the law of conservation of energy (the transition through the light barrier would require an infinite energy). It will be instructive to show the same thing using the law of addition of velocities.

The impossibility of reaching the barrier from its subluminal side has already been shown in Chapter 3. It is clearly seen from Equation (27) there that if one of the two input speeds is c, the output does not depend on the second speed. If we admit the possibility of superluminal reference frames consisting of tachyons, then the speed of light relative to such reference frames would also be constant and equal to c.

Consider now a tardyon and a tachyon whose velocities are collinear and differ from c by the same amount δv, so that

$$v = c - \delta v\,, \quad \tilde{v} = c + \delta v \tag{34}$$

Their relative speed, according to the rule in Equation (5) in Section 3.1, is

$$\tilde{v}' = \frac{\tilde{v} - v}{1 - \frac{v\tilde{v}}{c^2}} = 2\,\frac{c^2}{\delta v} \tag{35}$$

Let the tardyon speed up and the tachyon slow down, so that $\delta v \to 0$, and their speeds approach from the opposite sides their common limit – the speed of light. If we plot the two velocities as points on the velocity axis, then corresponding points approach closer and closer to each other, eventually merging together at the common point representing c. One would be tempted to imagine the corresponding two particles eventually moving together at one common speed – the speed of light. However, their *relative* speed, given by Equation (35), will go to infinity! Here the impossibility of crossing the light barrier from either side is manifested in the most dramatic and impressive way.

When the two particles approach the light barrier from opposite sides, they must remain in different realms separated by the barrier. If you sit on the tardyon, the tachyon on the other side of the barrier should move relative to you faster than light – its relative speed must exceed c, no matter how close to the light barrier you both are. If the special theory of relativity is logically consistent, it must meet this requirement. And this is precisely what it does – with astounding efficiency: the infinite relative speed in the limit $\delta v \to 0$ is definitely larger than c!

On the other hand, if the tardyon with the same speed $v = c - \delta v$ is moving *towards* the tachyon with a speed $\tilde{v} = c + \delta v$, their relative speed *decreases*; applying the rule in Equation (5) in Section 3.1 to this case yields

$$\tilde{v}' = \frac{\tilde{v} + v}{1 + \frac{v\tilde{v}}{c^2}} = \frac{c}{1 - \frac{\delta v^2}{2\,c^2}} < c + \delta v \tag{36}$$

The reader can check that this inequality holds for all $\delta v < c$. It tells us that the relative speed, while remaining larger than c, is *smaller* than the tachyon speed in the initial reference frame. This result, although not as dramatic as the previous one, still contradicts our intuition, according to which, if I go 5 km h^{-1} to meet a friend who runs towards me at 10 km h^{-1}, our relative speed will increase to 15 km h^{-1}, rather than decrease.

Note that the speed of the transcendent tachyon relative to a stationary tardyon is infinite. On the other hand, we know that such a pair is a special case of mutually dual particles. In this connection there arises an interesting question: what is the relative velocity between two dual particles in the general case? We can obtain an answer if we recall that duality between the two particles is an invariant property. Paradoxical as it may sound, the same is true for the value of relative velocity, by mere definition of this quantity (to measure *relative velocity* of the two objects, *any* observer has to transfer to the rest frame of one of the objects; recall Section 3.3). Therefore, the relative velocity between an arbitrary tardyon and the dual tachyon will not change, nor will they stop being dual to each other, if we switch to the rest frame of this tardyon. By doing this we come back to the special case of the tachyon dual to the stationary

tardyon. But such a tachyon is trancendent, it moves with an infinite speed. Because the relative speed is invariant, it must have been infinite in the original reference frame also.

We can obtain the same result directly from the law of addition of velocities. Putting in Equation (5) in Section 3.1 mutually dual velocities v and $\tilde{v} = c^2/v$, we will obtain

$$\tilde{v}' = \frac{\tilde{v} - v}{1 - \frac{v\tilde{v}}{c^2}} = \infty \tag{37}$$

Reversing this argument, we can give another definition of dual particles: two particles with equal absolute values of the rest mass are mutually dual if their relative speed is infinite. Indeed, the infinite relative speed requires that the denominator of the above expression be zero, from which there immediately follows $v\tilde{v} = c^2$, the definition of dual particles.

Another interesting question is: what is the relative velocity between the two tachyons? Once we admit that such particles can, at least in principle, exist, then one could, at least in principle, admit a reference frame and clocks connected with each of them (probably made of the same kind of particle.) That would allow us to measure the velocity of one tachyon relative to another. Let the tachyons move in one direction with speeds $\tilde{v}_1$ and $\tilde{v}_2$. The relativistic law of addition of velocities applies equally well to *any* velocity – subluminal or superluminal. Applying it to two superluminal velocities $\tilde{v}_1 = c + \delta_1$, $\tilde{v}_2 = c + \delta_2$, $\delta_1, \delta_2 > 0$, gives

$$\tilde{v}_{12} = \frac{\tilde{v}_2 - \tilde{v}_1}{1 - \frac{\tilde{v}_1\tilde{v}_2}{c^2}} = \frac{\delta_2 - \delta_1}{-\frac{\delta_1 + \delta_2}{c} + \frac{\delta_1\delta_2}{c^2}} < c \tag{38}$$

Let us make the tachyons move in the opposite directions ($\tilde{v}_2 = -\tilde{v}_1$), to increase their relative velocity. Then the same addition law gives

$$\tilde{v}_{12} = c\,\frac{2 + \frac{\delta_1 + \delta_2}{c}}{2 + \frac{\delta_1 + \delta_2}{c} + \frac{\delta_1\delta_2}{c^2}} < c \tag{39}$$

Thus, any two tachyons move relative to one another with a speed less than the speed of light! In other words, they behave like tardyons with respect to each other!

Let us summarize our conclusions. In the same physical space and time (in Minkowski's world) there can exist two equivalent worlds. The objects of either world move relative to each other slower than light. But the relative velocity of the two objects belonging to different worlds is always more than c. No object can by continuous change of speed transfer from one world to another. The worlds are impenetrable. They are separated by the impenetrable barrier – the speed of light – the speed of photons and gravitons. The latter particles form the third world, all the particles

of which move relative to all other particles with the same fundamental speed – the speed of light.

There emerges a picture so complete in its symmetry that one might start to wish that tachyons really exist!

But not everything is that simple in Nature. Hypotheses about tachyons can be considered with full seriousness only under the condition of their observability. The latter is possible only if tachyons can interact with the known matter – tardyons, photons, and gravitons. But the moment we admit the possibility of such interactions, we run into major difficulties and contradictions. We consider some of them in the next two sections.

8.6
Tachyon–tardyon interactions

The difficulties begin the moment we look at the lower part of the tachyon–tardyon diagram (Fig. 8.3), where the tachyon energy is negative. Let us call such tachyons negative tachyons. Applying the relation $\tilde{p} = (\tilde{E}/c^2)\,\tilde{v}$ to negative tachyons, we see that for these $\tilde{p}$ and $\tilde{v}$ are of the opposite signs. This means that momentum carried by a negative tachyon points opposite to its velocity. If the tachyon moves, say, to the right, the momentum assotiated with this motion, points to the left! How are we to understand this?

Momentum is a property that can be transferred from one object to another, if the objects can interact. We have admitted the possibility of tachyon–tardyon interactions. So imagine what happens if a negative tachyon moving to the right collides with a stationary tardyon. Normally one would expect the tardyon to get a kick to the right. But now it will acquire the momentum of the negative tachyon, which points to the left. Accordingly, it will start moving to the left!

Now, what happens if we push a negative tachyon in the direction of its velocity, trying to accelerate it? We can see the result by analyzing Figure 8.3. The negative tachyon, moving to the right, has its momentum pointing to the left. This is represented by a point on the lower left branch. By pushing it to the right (that is, by applying a force pointing to the right), we add to it the right-directed momentum $\Delta\tilde{p}$. The resulting momentum will be smaller in magnitude than the original one by $\Delta\tilde{p}$. Therefore, the point representing our tachyon will slide along the curve closer to the origin, where its slope is steeper. Accordingly, the magnitude of its velocity increases. Thus, it behaves in this respect like a regular tardyon.

Consider another type of interaction: that of tachyons and photons. Suppose that a tachyon can radiate light. The photons will then be emitted into the frontal hemisphere of the moving tachyon. Actually this emission is the Cerenkov radiation considered in the Chapter 6. Then we can see from the same diagram in Figure 8.3 that radiating positive tachyons ($\tilde{E} > 0$) lose their energy (approach the state with $\tilde{E} = 0$) and thereby accelerate. For radiating negative tachyons, the loss of energy also results in sliding down the curve. But in this case the sliding takes them further down from the state $\tilde{E} = 0$. They become more and more negative, and their speed decreases in magnitude, approaching the speed of light! The beautiful explanation of

why tachyons cannot reach the speed of light, which we developed in the previous section, works here in the opposite direction! Whereas positive tachyons need an infinite energy input to reach the speed of light, negative tachyons can spontaneously release an infinite energy and approach the light barrier!

If this result is true, and tachyons exist, we should observe huge spontaneous outbursts of energy in the form of electromagnetic and gravitational radiation. Astronomers do observe grandiose phenomena in remote parts of the Universe. One of them is so called gamma-bursts – unimaginably powerful explosions resulting in flashes of γ-radiation. It is tempting to speculate whether tachyon–tardyon interactions can provide a plausible explanation of some of these effects. For instance, the Big Bang itself: can it be the result of just one tachyon having decayed into huge numbers of photons and gravitons?

But it is not that simple. For instance, if a tachyon can emit a photon, it should also be able to absorb a photon. Then we should observe the occurrences of spontaneous absorption of radiation. We do not observe such phenomena. This negative observational result can be considered as indirect evidence against negative tachyons.

Some physicists have long been trying to eliminate negative particles from the picture of the world. They partially succeeded with negative tardyons, at least in the domain of classical physics. It is easy to see from Figure 8.3 why they could have done this. The negative branch for tardyons is separated from the positive branch by a gap that cannot be classically transcended. This allows one to say that even though there exists a possibility of negative tardyons, which is reflected in the existence of the negative branch in Figure 8.3, the initial conditions might have been such that no negative tardyons had been created. If this was the case, this condition must persist because positive tardyons cannot cross the gap between the two branches: all interactions in classical physics involve only continuous energy exchange.

This argument, of course, does not hold in quantum mechanics. Also, for the tachyons, it does not hold at all, because either one of the two negative branches for tachyons is just the continuous extension of the corresponding positive one. Therefore, nothing precludes even a classical tachyon from sliding down the curve to its negative branch – to the lower half of the (E, p) plane.

A glance at the plane shows that the situation is even worse than that. Both right and left tachyonic branches lie in the space-like domain of the plane, where the sign of the energy is not invariant, nor is the sign of the time coordinate [see Eq. (44) in Section 2.6 and Eq. (12) in Section 4.1]. Therefore, one and the same tachyon, while being positive in one reference frame, can be negative in another reference frame!

Consider a situation: a stationary atom in the ground state is approached by a positive tachyon with energy $\tilde{\varepsilon}$ and momentum $\tilde{p}$. Let the total energy of the atom be E_A. Then the total initial energy and momentum of the whole system are

$$\left.\begin{aligned} E_{tot} &= E_A + \tilde{\varepsilon} \\ p_{tot} &= \tilde{p} \end{aligned}\right\} \tag{40}$$

Suppose that the tachyon energy $\tilde{\varepsilon}$ is tuned to an allowed atomic transition, so that it is absorbed by the atom at zero time. We can describe the absorption by saying that

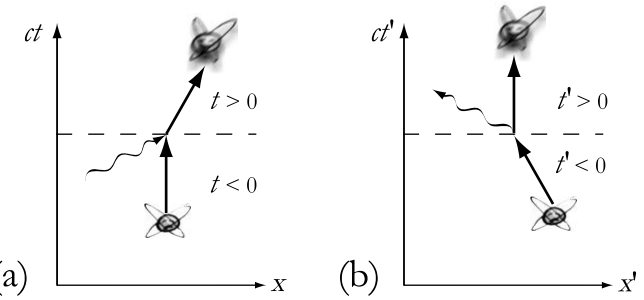


Fig. 8.4 The world lines of a tachyon and an atom interacting with each other. (a) In system K: The atom in its ground state absorbs the tachyon, which causes its transition to the excited state. (b) In system K′: The moving atom in its ground state emits the tachyon, which causes it to slow down and transfer to the excited state.

in the past ($t < 0$) there were two objects, the atom and the tachyon, and in the future ($t > 0$) there is only one object, the slowly moving excited atom (Fig. 8.4a). The atom is moving because it must have inherited the momentum of the tachyon. The final energy and momentum of the atom are equal to the initial energy and momentum of the system:

$$\left.\begin{aligned} E_{\text{final}} &= E_{\text{A}} + \tilde{\varepsilon} \\ p_{\text{final}} &= \tilde{p} \end{aligned}\right\} \tag{41}$$

The absorbed tachyon energy $\tilde{\varepsilon}$ is shared between the internal energy of the atom and the kinetic energy of its motion. Everything seems O.K., nothing unusual.

Consider now the same process from another reference frame K′ that carries the experienced observer Alice to the right with velocity V. The synchronized clocks in K′ are set to show time $t' = 0$ when the origin of K′ coincides with the origin of the system K. What does the above picture of absorption look like to Alice? The answer depends on how fast she is moving. If she moves so fast that $V\tilde{v} > c^2$, the ordering of any two events along the direction of motion, connected by the tachyon world line, is reversed. The world line of a tachyon is space-like. Accordingly, the tachyon's life must be seen in reverse by Alice if $V\tilde{v} > c^2$. It is important to emphasize that this does not pertain to the atom's life, whose chronology is, of course, Lorentz-invariant (Fig. 8.4b). Therefore, Alice sees first (in her past, $t' < 0$) *only one object*, the atom in the ground state moving relative to her with velocity $-V$. In her future ($t' > 0$) she sees two objects, the excited atom and the tachyon moving away from it to her left. From her viewpoint, the tachyon has been emitted from the atom. Indeed, at zero time Alice sees the atom performing a sudden transition from the ground state to the excited state, with the emerging tachyon that subsequently recedes away. What is seen as absorption in K is seen as emission in K′! But how can the atom become excited, that is, increase its internal energy, and simultaneously emit a particle, which also must carry some energy? The reader can already anticipate the answer: in

Lorentz transformations, the energy transforms in the same way as time does. For any object with a space-like world line, if its time coordinate changes its sign, so does its energy. Therefore, the energy of the emitted particle in our case must be negative. If an object emits a particle with positive energy, the energy of the object decreases. If the object emits a particle with a negative energy, the energy of the object increases. Therefore, the emission of a negative tachyon is perfectly consistent with the excitation of the emitting atom.

Let us check whether this nice scheme really works. All we have to do is to apply the Lorentz transformation to the energy and momentum of the system. Alice in her past ($t' < 0$) sees only the atom and no tachyon. The energy and momentum of the atom are, respectively

$$\left.\begin{aligned} E'_{\mathrm{A}}\,(\text{initial}) &= \gamma\,(V)\,(E_{\mathrm{A}} - V p_{\mathrm{A}}) = \gamma\,(V)\,E_{\mathrm{A}} \\ p'_{\mathrm{A}}\,(\text{initial}) &= \gamma\,(V)\left(p_{\mathrm{A}} - \frac{V}{c^2}\,E_{\mathrm{A}}\right) = -\,\gamma\,(V)\,\frac{E_{\mathrm{A}}}{c^2}\,V \end{aligned}\right\} \tag{42}$$

Since there is nothing else but an atom before the zero moment, Equations (42) give the total energy and momentum of the system in the initial state.

After the zero moment we have the same atom in the excited state and the receding tachyon. The energy and momentum of the atom are

$$\left.\begin{aligned} E'_{\mathrm{A}}\,(\text{final}) &= \gamma\,(V)\,(E_{\mathrm{A}} + \varepsilon - V p_{\mathrm{A}}\,(\text{final})) = \gamma\,(V)\,(E_{\mathrm{A}} + \varepsilon - V\tilde{p}) \\ p'_{\mathrm{A}}\,(\text{final}) &= \gamma\,(V)\left[\tilde{p} - \frac{V}{c^2}\,(E_{\mathrm{A}} + \varepsilon)\right] \end{aligned}\right\} \tag{43}$$

(recall that the atom after the absorption in K carries the tachyon momentum $\tilde{p}$). The energy and momentum of the tachyon in K′ are, respectively

$$\left.\begin{aligned} \tilde{\varepsilon}'\,(\text{final}) &= \gamma\,(V)\,(\tilde{\varepsilon} - V\tilde{p}) = \gamma\,(V)\left(1 - \frac{V\tilde{v}}{c^2}\right)\tilde{\varepsilon} \\ \tilde{p}'\,(\text{final}) &= \gamma\,(V)\left(\tilde{p} - \frac{V}{c^2}\,\tilde{\varepsilon}\right) = \gamma\,(V)\,(\tilde{v} - V)\,\frac{\tilde{\varepsilon}}{c^2} \end{aligned}\right\} \tag{44}$$

Now we can find the total final energy and momentum of the system in K′ by summing Equations (43) and (44) and compare them with the total initial energy:

$$\left.\begin{aligned} E'\,(\text{final}) &= \gamma\,(V)\,(E_{\mathrm{A}} + 2\tilde{\varepsilon} - 2\,V\tilde{p}) \\ p'\,(\text{final}) &= \gamma\,(V)\left[2\,\tilde{p} - \frac{V}{c^2}\,(E_{\mathrm{A}} + 2\tilde{\varepsilon})\right] \end{aligned}\right\} \tag{45}$$

Now compare this with Equations (42). The initial and final energies and momenta are not the same in K′ – they do not conserve!

The reverse of the ordering at the transition from K to K′, that "hurls" the tachyon from Alice's past ($t' < 0$) into her future ($t' > 0$) and converts the absorption into emission has disastrous consequences!

This would be an emergency situation in physics. It shows that either the original assumption about the possibility of tachyons was wrong, or such particles cannot interact with tardyons, or else some additional hypotheses about tachyons are needed.

The first two assumptions are physically equivalent. If a certain entity does not interact in any observable way with the ordinary matter, this entity is, in all practical terms, as good as non-existent. Thus, if one still wants to save the concept of tachyons that brings about an additional symmetry in the world, one should invent an additional hypothesis to save the conservation laws.

Such a hypothesis had been suggested by Bilaniuk, Deshpande, and Sudarshan [60]. It was called the "reinterpretation principle." Its essence can be described in the following way. The authors had noticed that the same condition $V\tilde{v} > c^2$ that swaps a tachyon's past and future automatically reverses the sign of its energy and momentum. They suggested that in all such cases we should make an additional sign reverse of the two latter dynamic quantities. In other words, together with reinterpreting the cause in K as the effect in K′ and vice versa, we should reinterpret the negative tachyon in K′ as positive (the sign of momentum will then be reversed automatically). So, if we do,

$$\tilde{\varepsilon}' \Rightarrow -\tilde{\varepsilon}' > 0 \tag{46}$$

then, automatically,

$$\tilde{p}' \Rightarrow -\tilde{p}' < 0 \tag{47}$$

This would be mathematically equivalent to mechanically (without changing signs!) transferring these variables to the other side of Equations (43), (45). Physically, this is equivalent to swapping the tachyon back between the initial and final states, thus restoring the status quo, characteristic of ordinary particles. Then, as if by magic, everything will fall into place. If the curious reader goes back to Equations (43), (45) and performs the described operations, the terms with the "extra" tachyonic variables will cancel, and the conservation laws will be restored.

But here Alice intervenes.

"Excuse me, but I cannot believe your math," she says. "It contradicts all I see. I first see the atom in its ground state, and moving to the left. Then I see this atom performing a transition to an excited state, and emitting the tachyon. I could understand this when the tachyon energy was negative. Now you are saying that it must be positive, because only in this way can the total energy be conserved. But this seems ridiculous. How can the atom become excited, that is, increase its internal energy, and simultaneously emit a *positive* tachyon, which requires additional energy? Where does all this energy come from?"

How can we answer her? By taking into account that the atom in her system K′ had been moving to the left with a speed V. Then it emitted the tachyon to the left. Since

the tachyon has been declared to be positive, its momentum must also point to the left. This left-directed momentum comes from the atom. The magnitude of the atomic momentum, and thereby its speed and kinetic energy, must accordingly decrease. Thus, the energy of the atomic excitation and the positive tachyon energy both come from the kinetic energy of the atom.

"Well," says Alice after turning it over in her head, "now I think I can understand a little better why the poor hare in Section 8.3 was killed when the superluminal bullet had formed inside its body and burst out of it. The bullet must have consisted of a huge number of tachyons. Each tachyon was born in a spontaneous excitation of an atom, so a huge number of atoms in the tachyon's way must have been excited or even ionized. The energy for both atom excitation and the tachyon production must have come from the kinetic energy of the hare and the planet, which had been both moving together relative to my spaceship. Definitely, the simultaneous exitation of so many atoms must be fatal for any organism. This is how the law of conservation of energy can include tachyons, and work on the macroscopic scale."

Alice thought a little more, then sighed and said, "There is one thing here that seems quite mysterious to me. How could so many atoms be spontaneously excited practically all at once? And, for goodness sake, how could their excitations be so accurately arranged as to form a bullet aimed directly at the barrel of Tom's gun, at precisely the moment when Tom was about to trigger it?"

Alice's questions pose a few difficulties in the hypothesis of tachyons.

First, the restoration of the conservation laws has been achieved at a high price. The operations in Equations (46) and (47) are essentially the statement that, physically, the tachyon with a negative energy is equivalent in its behavior to the tachyon with the same proper mass and positive energy – to its anti-tachyon. But the arbitrary change of sign in the Lorentz transformations is not a legitimate operation. It is arbitrary because it is performed only on tachyons and never on tardyons. And for tachyons, it is performed only when $V\tilde{v} > c^2$, and never when it is otherwise. Hence changing V or $\tilde{v}$, or both, will change the criteria for applicability of the reinterpretation. In other words, the same tachyon may be a particle for one observer and its anti-particle for another observer. If such an interpretation represents reality, then a tachyon cannot carry any of the known physical charges (for instance, electric charge), because a charge of a particle does not change under Lorentz transformations.

Second, the way things would look on a macroscopic scale if tachyons are involved and behave as described (e. g. in the scene with the hare) is definitely different from normal physical behavior. Spontaneous atomic transitions, whose individual times are in principle unpredictable, cannot produce a macroscopically ordered motion, which is in addition highly correlated with motion of other macroscopic bodies a long distance away. It would contradict one of the most fundamental and firmly established laws of nature – the second law of thermodynamics, already mentioned in Section 8.2.

Such a state of affairs is, of course, unsatisfactory. Either the whole concept of tachyons, despite its tempting attractiveness, is fundamentally incompatible with relativity, or it should be reintroduced with some additional ideas that so far seem to be missing.

8.7
Flickering phantoms

Suppose you stand in front of a mirror with a source of tachyons. What happens if you fire a tachyon at the mirror? Suppose that tachyons can interact with the mirror as photons do. Then we will first see a tachyon with energy $\tilde{E}$ and momentum $\tilde{p}$ approaching the mirror, and then the tachyon with the same energy and momentum $-\tilde{p}$ moving away from the mirror after reflection (Fig. 8.5).

Consider the same process from the viewpoint of an observer Peter in another reference frame K′, which is moving to the right with a speed V along the x-axis of the original system K. We can reconstruct his observations qualitatively by considering the world lines of all the parties involved (Fig. 8.5 a, b). If Peter's motion is sufficiently fast (his x'-axis runs below the tachyon's world line), then he sees the tachyon's history differently than we do. The moment of reflection occurs in his system *before* all other moments of the tachyon's history, so that both approaching and receding branches of its world line start at the moment of reflection and move towards the future. The word "approaching" in this case becomes a misnomer, since both branches are now receding. The tachyon moving from left to right in the original system K is moving from right to left in system K′. Peter first sees nothing; then, at some moment of his time the mirror emits two tachyons at once, which then both whiz to the left at different speeds. The difference in speeds is due to the fact that to each of the two tachyons in K′ corresponds one and the same tachyon in K but at different times t_1 (before the reflection) and t_2 (after reflection); at the moment t_1 the tachyon was moving with the speed $\tilde{v}$ to the right, and after the reflection it was moving with the same speed to the left. Applying the law of addition of velocities to these two cases, we obtain two different speeds in system K′.

Now, if this can be true, what about conservation laws? If created tachyons are "the real things," they must each possess some energy and momentum. Who pays for them? They appear to pop out of nothing. However, knowing some physics, Peter understands that to obtain the correct conclusion, all the objects involved have to be considered, mirror included. Conservation laws must hold for combined system tachyons + the mirror. As the created tachyons are fired to the left, the moving mirror gets a kick to the right, which slightly decreases its momentum and thereby its kinetic energy. The lost energy of the mirror goes to the tachyons.

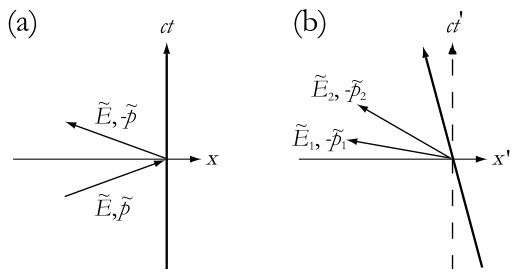


Fig. 8.5 The world lines of a tachyon interacting with the mirror. (a) In the rest frame of the mirror; (b) in Paul's reference frame $\left(V > \frac{\tilde{v}^2}{c}\right)$.

If mathematical framework of relativity can incorporate the tachyons, then the combined systems tardyons + tachyons must obey the laws of relativistic kinematics in all reference frames. We had found that with respect to conservation laws this is true only for the reference frames with relative speed $V \leq c^2/\tilde{v}$. If the relative speed exceeds this limit, we can only save the conservation laws by performing a pretty ugly trick – an arbitrary change of the sign of energy. Some readers may find it instructive to see how it works in the case of the mirror. For these readers we consider the case quantitatively, and on the way we will find a couple of additional interesting details.

Let us focus first on the energy; once it is known, we can always find the momentum using the universal relativistic relation

$$\tilde{p} = \frac{\tilde{E}}{c^2}\tilde{v} \tag{48}$$

In reference frame K we have a stationary mirror and one tachyon. The tachyon's energy and momentum are $\tilde{E}$, $\tilde{p}$, respectively, for the approaching (incoming) branch of its world line, and $\tilde{E}$, $-\tilde{p}$, respectively, for its receding (outgoing) branch. Label these two branches 1 and 2, respectively. Although the property of being incoming or outgoing is not generally invariant, each branch itself is a geometrical object independent of a reference frame, so that the label 1 or 2 uniquely specifies the branch. We can now determine characteristics of each branch in the reference frame K′ by applying the Lorentz transformation:

$$\tilde{E}'_1 = \gamma(V)(\tilde{E} - V\tilde{p}); \quad \tilde{E}'_2 = \gamma(V)(\tilde{E} + V\tilde{p}) \tag{49}$$

Using Equation (48), one can express this in terms of only the energy and speeds:

$$\tilde{E}'_1 = \gamma(V)\left(1 - \frac{V\tilde{v}}{c^2}\right)\tilde{E}; \quad \tilde{E}'_2 = \gamma(V)\left(1 + \frac{V\tilde{v}}{c^2}\right)\tilde{E} \tag{50}$$

If $V\tilde{v} < c^2$, the ordering of all the events on the whole world line is conserved. Peter observes the same stages of the reflection process as we do. The only difference is that he sees the mirror moving to the left rather than stationary, and this is the physical reason why the tachyon in his system has greater energy after the reflection than before ($\tilde{E}'_2 > \tilde{E}'_1$). The kinetic energy of a baseball also increases when it is hit by an oncoming bat, at the expense of the bat's energy.

If $V\tilde{v} > c^2$, the ordering of events on branch 1 is reversed in Peter's reference frame; to him, both branches become outgoing, and he observes *two* tachyons, emitted to the left by the moving mirror. And the energy of the tachyon moving along branch 1 ("tachyon 1") turns out to be negative. This tachyon belongs to the lower branch of hyperbola in Figure 8.3. But we know that in such cases we must forcibly "hurl it back" on to the upper branch: we must perform the operation in Equation (46), $\tilde{E}'_1 \Rightarrow -\tilde{E}'_1$. Then the total energy of the pair of tachyons will be

$$\tilde{E}'_{\mathrm{T}} = -\tilde{E}'_1 + \tilde{E}'_2 = 2\gamma(V)V\tilde{p} \tag{51}$$

A similar operation is automatically performed on the momentum of tachyon 1. It is easier to find the individual momenta in K′ from Lorentz transformation combined with Equation (48):

$$\tilde{p}'_1 = \gamma(V)\left(\tilde{p} - \frac{V}{c^2}\tilde{E}\right) = \gamma(V)\frac{\tilde{E}}{c^2}(\tilde{v} - V);$$
$$\tilde{p}'_2 = \gamma(V)\left(-\tilde{p} - \frac{V}{c^2}\tilde{E}\right) = \gamma(V)\frac{\tilde{E}}{c^2}(-\tilde{v} - V) \tag{52}$$

The total momentum of the pair is determined as the sum of $\tilde{p}'_2$ and *negative* of $\tilde{p}'_1$:

$$\tilde{p}_T = -\tilde{p}'_1 + \tilde{p}'_2 = -2\gamma(V)\tilde{p} \tag{53}$$

The momentum of the mirror must then decrease by the same amount:

$$\Delta p_M = \tilde{p}'_T = 2\gamma(V)\tilde{p} \tag{54}$$

Now we can account for the energy needed to produce the pair of tachyons. The decrease in momentum of the mirror decreases its kinetic energy by the amount

$$\Delta E_M = -V\Delta p_M = -2\gamma(V)V\tilde{p} \tag{55}$$

This is exactly the result in Equation (51) – the energy of the two created tachyons. The conservation law holds (but keep in mind the price of it.)

Suppose now that there are two mirrors at a fixed distance L from one another along the x-axis (Fig. 8.6). The world line of the left mirror (which is placed at the origin of an inertial system K) is represented by the time axis ct. The world line of the right mirror is given by the vertical line $x = L$. The area on the plane (ct, x) between the verticals $x = 0$ and $x = L$ forms the world sheet of the whole segment $(0, L)$. Let a tachyon zip back and forth between the two mirrors with a speed $\tilde{v}$. Part of its history between the two successive reflections is given by the segment of the world line OP_1, and the whole process by the broken line $OP_1P_2P_3$ ...

Consider again the observer Peter in another reference frame K′, which is moving to the right along the x-axis of the original system K at a speed V. The coordinate axes of the new system are also shown in Figure 8.6. Peter observes the same process as we do. We notice that the axis x' intersects the broken world line of the tachyon at points A_0, A_1, A_2, ..., A_j, ... What does this mean?

Using the rules for determining coordinates of the events in the skewed coordinate system (Sect. 2.9), we can reconstruct the picture observed by Peter.

Recall that all the events simultaneous in Peter's reference frame lie on a line parallel to his x'-axis; if they occur at the zero moment by his system of clocks, they coincide with this axis. And vice versa, all the events "forming" the x'-axis, are simultaneous for Peter – they all occur at one moment $t' = 0$ of his time. This is also true for the intersection points A_0, A_1, A_2, ... in Figure 8.6. Each such point corresponds to an event – passing of the tachyon by this point. For Peter, all these events occur at dif-

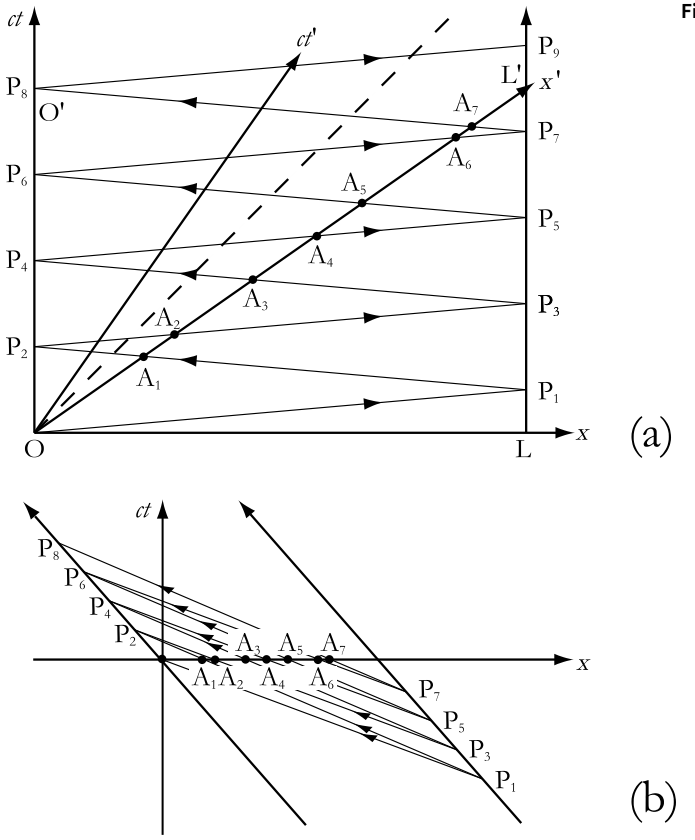


Fig. 8.6

ferent points in space, but at one moment of time. In other words, Peter sees the same tachyon in different places at once.

This is not the same thing as the observation of a transcendent tachyon discussed in the previous section! The transcendent tachyon moves infinitely fast, its world line is parallel to the corresponding spatial coordinate axis, and it is observed simultaneously at *all* points of its trajectory. The tachyon observed by Peter has a finite speed (unless $V = c^2/\tilde{v}$), its world line is not parallel to the x'-axis, and it is observed simultaneously only at the discrete set of points A_j, $j = 0, 1, 2, \ldots$ But whatever is observed in separate disconnected places at once is perceived as separate objects. In other words, Peter observes a few identical, but independent, tachyons between the mirrors! The number of these tachyons is equal to the number of intersection points between the x'-axis and the broken line $OP_1P_2P_3 \ldots$ within the world sheet of the cylinder. In turn, it is equal to the number of legs of the broken line inside the rectangle OO′L′L in Figure 8.6. We can find this number from Figure 8.6 as the integral part of the ratio OO'/LP_1 (from here on, italics indicate distances) plus 1. Now, if we do some algebra, we can express this number in terms of two speeds: the speed of tachyon and the speed of Peter.

It is seen from Figure 8.6 that the faster the tachyon moves, the shallower are the legs of its world line, and thereby more legs will fit into the rectangle OO′L′L. The faster Peter moves, the steeper is the x'-axis, and the higher the rectangle. Hence the number of intersections must be proportional to the product of the two speeds.

Let us now find this number rigorously. We have from Figure 8.6: $OO' = L\tan\theta$, where θ is the angle between the x- and x'-axes. Recall that, according to Equation (64) in Section 2.9, $\tan\theta = V/c$. Thus $OO' = LV/c$. Regarding LP_1, it represents just the time it takes the tachyon to make a one-way trip down the cylinder, multiplied by c: $LP_1 = cL/\tilde{v}$. Combining these expressions gives

$$N = \left[\frac{V\tilde{v}}{c^2}\right] + 1 \tag{56}$$

where the designation $[X]$ means the integral part of X.

The number of tachyons observed simultaneously by Peter does not depend on L, and is completely determined by the product $V\tilde{v}$. In order for this number to be >1, it is necessary that this product be not less than c^2. If it is less than c^2, then Peter observes, just as we do, only one tachyon at any moment. At $V\tilde{v} = c^2$, Peter can see one (transcendent) tachyon; at $V\tilde{v}$ slightly exceeding c^2 he sees already two tachyons, and he registers two tachyons in all cases where the product $V\tilde{v}$ changes within the range $c^2 \le V\tilde{v} < 2c^2$. For the range $2c^2 \le V\tilde{v} < 3c^2$ Peter can observe simultaneously three tachyons between the mirrors, and so on. If $\tilde{v} \to \infty$, the number of tachyons between the mirrors as observed by Peter can be arbitrarily large, while we have only one tachyon there. The number of tachyons is not invariant!

The reader can recall how emphatically we stressed in Chapter 1 that the total number of stable objects is one of the most important, absolute (Lorentz-invariant) characteristics of a system. Now we see this "sacred rule" outrageously violated by tachyons.

But this is not all. The mere picture of motion observed by Peter is also unusual. We know already that in each cycle, the tachyon approaching the right mirror can be observed by Peter as receding from this mirror, so that he can within each cycle see two tachyons moving away from the mirror at different speeds. Applying Lorentz transformation to their speeds in reference frame K, we find their speeds in K′:

$$\tilde{v}'_1 = \frac{\tilde{v} - V}{1 - \dfrac{\tilde{v}V}{c^2}}, \quad \tilde{v}'_2 = \frac{-\tilde{v} - V}{1 + \dfrac{\tilde{v}V}{c^2}} \tag{57}$$

Because of the difference in speeds, the two tachyons reach the left mirror at different times: one at the zero moment $t'_1 = 0$ and another at $t'_2 > 0$ (this event is represented by point P_2 in Figure 8.6).

The "multiplying" of the tachyon when an observer switches from system K to K′ occurs only under the condition $V\tilde{v} > c^2$. The same condition, as is seen from Equations (57), makes $\tilde{v}'_1$ negative, interchanging temporal coordinates of the events O and P_1. Equations (57) confirm that the velocities of the two tachyons in system K′ are different in magnitudes and both negative, that is, directed to the left.

Hence each cycle of the oscillatory motion of the tachyon in system K transforms into motion of two tachyons from the right to left mirror in system K′. The resulting picture of motion observed by Peter can be described with the following figurative model. Imagine two teams – male and female – of alien runners from another world, who call themselves "Tachyons." The members of the female team run at the speed $\tilde{v}'_1$ and members of the male team at the speed $\tilde{v}'_2$. The members of the two teams are interspersed and run in pairs between the mirrors, which move to the left at speed V. At certain moments of time t'_j, $j = 1, 2, \ldots, t'_{j+1} > t'_j$, a pair is born at the right mirror, and its male and female members rush to the left at their respective speeds. The female runner reaches the left mirror earlier, at which moment she catches up with the male member of the previous pair and they both disappear. At the moment the male runner of the pair reaches this mirror, the female runner of next pair catches up with him, and they also both disappear. When the male runner of this pair comes to the place, the female from the third pair reaches it too, and they again mutually annihilate, and so on. The Tachyons are born in pairs at the right mirror and annihilate at the left mirror, with the members of the neighboring pairs. And all this phantasmagoria of flickering phantoms are just different events of history of only one tachyon in system K, which are "projected" simultaneously on to Peter's system K′.

Next, we can imagine our tachyon evolving in time. Suppose, for instance, that a tachyon is an extended object (we will see soon that there are sound reasons for such an assumption!), and its size is increasing in time. Then we will see in K one tachyon bouncing between the mirrors and growing larger like an inflated balloon as it does so. Peter will see in his system something different and even more weird than before. He will see again many Tachyons at once racing in pairs from the right to the left mirror. But this time they are all of different size except for the moments of their birth and death. Each time when a pair is born at the right mirror, both of its partners are of the same size, but both are larger than the members of the previous pair and smaller than the members of the next pair. As they start towards the left mirror, the male Tachyon expands, which stands in total accord with tachyon history in K. But its female partner shrinks! At the left mirror she catches up with the male runner of the previous pair, who was born smaller than she, but since he was growing in the run while she was diminishing, they meet being equal in size and both annihilate. When her male partner reaches the left mirror, he is caught up with by the female runner of the next pair, who was born larger than him; but now they are both of the same size, and both annihilate each other.

We can easily understand why the participants of this carnival are generally all different in size. It is again a manifestation of relativity of time. What Peter sees in K′ are different events of the life of one object. While these events follow one after another in K, they appear all at once to Peter in K′. The Tachyons of different size that Peter sees simultaneously are merely different ages (and thereby different sizes) of the same tachyon in K.

But for Peter all the tachyons observed simultaneously appear to be separate independent entities, rather than only ghost images of the single real tachyon. What tests can we perform to find out which possibility is true?

First, we can measure one of the basic properties of the tachyon, say, its energy and momentum in our system K, and suggest that Peter does the same with his tachyons in his system K′. If each one of Peter's tachyons contributes to the total on an equal footing with its partners, then each can be considered, at least to some extent, as an independent particle. Otherwise, some of them are "ghosts" that do not really have mass or energy, or possess any other property of a particle. But we have seen that this type of test is rather ambiguous. It does show that all the tachyons carry energy and momentum and obey conservation laws like any well-behaved real object, but only if we "reinterpret" the energies of some of them by changing their sign.

Alternatively, we can try to *interrupt* the tachyon history in K at some moment, and ask Peter how this interruption affects his observations. Indeed, once we have admitted the interactions between tachyons and tardyons, it is natural to assume that we can at any moment influence the tachyon at our will. So, imagine that we "knock down" our tachyon, for instance, at the event A_4 by firing at it a tardyon that absorbs it when they collide. Immediately, its world line above this point on Figure 52.2a (that is, $A_4P_5A_5P_6$ ...) is obliterated, so that the events P_5, A_5, P_6, ... of the tachyon history are prevented from happening. Accordingly, Peter sees (Fig. 8.6b) that the right mirror stops to produce the next tachyon pairs after the moment P_3. On the other hand, this mirror is far away from the collision point and cannot be affected by this collision. Does this mean that these pairs would have all been ghosts, and only the previous ones were the real things? It does not. Suppose we decide to hit the original tachyon in K at a later moment, say at the event A_6, in which case Peter would not see any tachyons with numbers greater than 6. We see that the designation of a tachyon in K′ as a ghost or a real object depends on what we do to the tachyon in K and *when* we do it. So shooting the tachyon does not give us a clear-cut criterion.

But if the tachyons in system K′ are all real, how can their production by the mirror be affected by a distant event? Let us turn to Peter's report of his observations. What we see as the tachyon absorption is seen in K′ as its emission by the tardyon. Figure 8.6b, presented by Peter, shows that the mirror stops producing pairs after the event P_3. But the moments P_3, P_5, P_7 are *before* the event A_4 in Peter's reference frame! Hence shooting the tachyon affects in K′ its past history! How does the signal travel into past?

Peter may want to describe the observed phenomena in K′ without any reference to another reference frame. And relativity not only grants him the right to do so, but must also provide him with the means to do it (recall Sections 5.4 and 5.5). If the right mirror knows that it has to stop producing next tachyons *before* the moment of tachyon absorption, then there must be a physical effect responsible for it. However, unlike the situation with the relativistic train, now there is none to be seen except, maybe, for something yet unknown – what can that be?

In our attempts to explain things without referring to the system K, we are forced back to this system. Take a closer look at our thought experiment. The regular creation of tachyon pairs at one mirror and their coordinated annihilation with the members of subsequent pairs at the opposite mirror indicate a very special initial condition. This condition is periodic motion of just one tachyon in the rest frame of the mirrors. The coordinated motions of tachyons in system K′, even though they appear

to Peter to be independent, are intimately connected with each other by their common origin. This may seem to have something in common with the non-local quantum mechanical correlations of distant particles, discussed in Section 6.15. But the situation here is even more dramatic. In Section 6.15, the positron on Rulia appeared to know *instantaneously* what had happened to the electron on the Earth, and changed its state accordingly. Now, the right mirror appears to know about the distant tachyon absorption *in advance*, and changes its behavior (stops producing pairs) accordingly. It looks as if together with the introduction of tachyons, we must introduce the possibility of a new physical effect – the flow of information from the future into the past!

Can this be possible? And, if it can, how could it change the picture of the world? We will try to analyze this question in the next section.

8.8 To be, or not to be?

This section will not require heavy math. Its results can be illustrated by a couple of simple diagrams and accordingly allow a more casual style of writing. In what follows, the possible dramatic consequences of superluminal signaling are described in the form of a science fiction story.

... Commander Fletcher was leaving the planet Rulia with heavy misgivings. His mission fell short of successful. He could not have persuaded the leadership of Rulia to stop developing the most powerful weapon of mass destruction – tachyonic beams. The superpower on planet Rulia was notoriously ambitious. After the first reports of the discovery of tachyons had been published, it saw in tachyons a weapon of unlimited potential, capable of instant destruction of the remotest recalcitrant planets.

Right in the midst of preparation of his spaceship Alvad for departure, Commander Fletcher received information that Alvad had been chosen as the first experimental target to be hit by a tachyon beam. Just the day before the arrival of Alvad, the dictator of Rulia, General Hiss, personally inspected the military facilities on the Rulian space station M. And right after the negotiations had finished with no agreement, he ordered preparations to fire the destructive beam from station M at the moment "P" of Rulian time, when Alvad would be well on its way back to Earth. General Hiss wanted to make sure that the far away target could be destroyed instantaneously. Therefore, he ordered the use of the high-power beam moving with a nearly infinite speed.

The moment Rulia was left behind, Commander Fletcher drew two space–time diagrams with the world lines of Rulia and Alvad. Knowing the shooting time and speed of the beam, and also the speed of Alvad, he calculated the arrival time of the beam. The calculation showed that the beam was to be expected in 12 h.

There was not much time left for discussions. Commander Fletcher ordered the activation of the protective shield and Alvad's tachyon firing facilities to be brought to the highest alert. His strict instruction was to avoid any hostile actions until the moment of attack. *If, and only if,* the Alvad is hit by the beam from Rulia should it respond by firing its own beam. The beam should be aimed at the Rulian space station

M and move with nearly an infinite speed relative to Alvad. The power of the beam must be sufficient to destroy all military facilities on M.

After about 12 h of anxious waiting, a violent jerk shook the ship. Nearly all of the protective shield was gone, vaporized in the twinkling of an eye. A few crew members were seriously hurt. Just a few seconds later, the gunmen of Alvad, following the initial orders, were about to fire their own tachyon gun. But at the very last moment, a new order came to postpone the firing. Nobody could understand what had happened.

The first mate ventured to interfere. "Commander," he said, "don't take my question as insubordination. But this is an emergency situation, so will you please tell me, what are we waiting for?"

Commander Fletcher was deep in thought. "Our shield is gone," he said. "We cannot afford to suffer another attack."

"Then why are we waiting? Are we going to give them a chance to strike again?"

"Quite the contrary," Commander Fletcher said as if answering his own thoughts rather than his mate's. "We want to minimize their ability to strike again."

"By not responding?" asked the mate.

"By responding at a proper moment."

"Sir, what moment can be more proper than the earliest one?"

"I will tell you when we check our calculations."

In half an hour, Commander Fletcher told his mate the exact time of the Alvad's response. It was in about 6 h from the present moment. For all this time the crew was again to follow strict orders – not to shoot.

Nothing serious happened during these 6 h. And when they had gone by, the powerful tachyon flux from Alvad zipped to Rulia with a nearly infinite speed and converted its military station M into atomic vapor.

Now the time has come for us to discuss this "Star Wars"-like episode. For a better understanding of the underlying physics, we consider the described situation from two viewpoints – that of Rulia (system R) and that of Alvad (system A).

A good way to see the basic features involved is to look at the two figures drawn by Commander Fletcher.

We start with Figure 8.7. On this figure, the vertical line O–ct represents the world line of Rulia, and O–ct' represents the world line of Alvad in Rulian coordinates (the Rulian time ct and spatial coordinate x in the direction to Alvad are drawn at right-angles to each other). The line O–x' represents the spatial coordinate axis of system A. The point S represents the starting moment of the whole story – the inspection visit of General Hiss on station M. The point P represents the tragic event in the history of Rulia – the shooting of the tachyon beam to destroy Alvad – and point Q represents the event of blasting off Alvad's protecting shield by this beam. Accordingly, the line PQ represents the world line of the beam. Point C gives the event – response of Alvad's tachyon gun. The interval QC gives the waiting time, calculated by Commander Fletcher and his Strategic Research Division team, and so bitterly disputed by the Commander's mate.

Why was this waiting interval needed? This becomes clear when we find the world line of the tachyon beam from the Alvad. You will remember that the speed of the

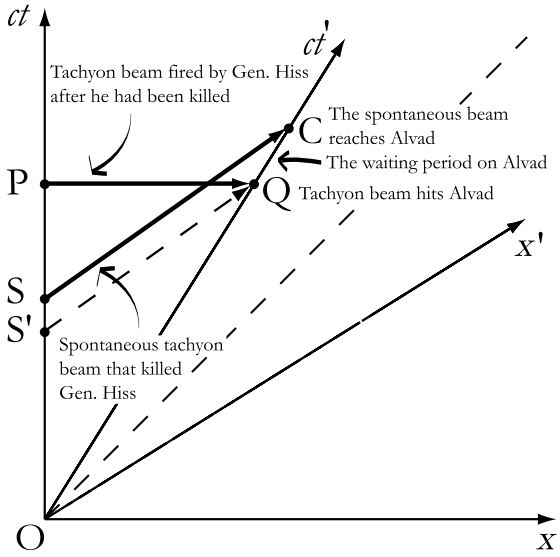


Fig. 8.7 The time loop in R.

beam was nearly infinite relative to Alvad. Therefore, the beam's world line must run parallel to the spatial axis x'. Its intersection S with O–*ct* would represent the arrival of the retaliating beam at Rulia's space station M. The arrival of the beam happens *earlier* in Rulian time than its departure from the Alvad. We are already familiar with this phenomenon. Moreover, it is seen from the diagram that the retaliating beam may have arrived at M and destroyed it even *before* the guns at M started shooting! As Commander Fletcher thought it over, he realized that this gives him a chance to do away with General Hiss along with the station and thus greatly reduce the chance of further strikes from Rulia. Commander Fletcher knew that General Hiss had personally inspected the station before the Alvad's arrival at Rulia. He knew exactly the inspection time from the message he had received before his departure from Rulia. He decided to hit station M with the beam right at that time. But he had seen from the diagrams that if he shoots immediately after having been attacked, the intersection point S′ would fall below S, and the beam would hit M slightly too early. Therefore, he decided, after painful hesitation, to postpone the shooting. Together with his research team, he calculated exactly the needed waiting period – it turned out to be about 6 h. His foresight was correct and his calculations were accurate. The postponed beam from Alvad turned the whole station M into an atomic cloud just at the moment S when General Hiss was there for the inspection.

Now we can get back to Figure 8.7.

Take the event S as a starting point of the loop. The line SP represents the period of Rulia's history from that event to the fatal shot of the Rulian tachyon gun (event P). The arrow PQ represents the world line of the tachyon beam from station M to Alvad. The arrow QC is the waiting period on Alvad. Finally, arrow CS represents the world line of the retaliating beam from Alvad to station M. This arrow closes the loop, but not the story. Think again about the events at the end point of this arrow:

which one is the cause and which one is the effect? From all we know, one would tend to conclude that the shot from the Alvad's gun was the cause, and the destruction of station M and death of General Hiss were the effect. But to the Rulians these two events (and also all the intermediate events on the world line CS) are reversed in time. They first observe the event S. They see it as the spontaneous disappearance of their station M in a gigantic explosion, accompanied by the outburst of a powerful tachyon beam in the direction in which the Alvad had departed after the negotiations. Then they (or their outpost observers) could watch how this beam caught up with the receding Alvad at a later moment and disappear in the Alvad's tachyonic gun, which was at that moment aiming precisely at what used to be station M.

The Rulian physicists have been arguing about the direction of the information flow in the beam. According to the reinterpretation principle, they were prone to consider the earlier event S as the cause, and the later event C as the effect. But there was much greater embarrassment for them and the entire population of Rulia – the question about the physical and legal status of General Hiss. It was a recorded physical and historical fact that station M was functioning normally and General Hiss was alive and active at the moment P long *after* the explosion on station M. It was at that later moment that he ordered the station to fire the tachyon beam at Alvad. And it was also a physical and historical fact established beyond doubt that General Hiss was killed by a spontaneous explosion of station M at the earlier moment S. Both facts together are, of course, logically incompatible, and the Rulians found it impossible to resolve this contradiction. Facing this unbearable situation drove them into a state of complete hysterical frenzy.

Now consider the same events from the viewpoint of the Alvad's crew. Look at the second diagram (Fig. 8.8) of the above events, which is now scaled to Alvad's coordinates (its temporal and spatial axes are drawn perpendicular to each other, and the corresponding "Rulian" axes are skewed).

It is convenient to start the account of the events as observed on the Alvad from event Q – the spontaneous explosion of the Alvad's protective shield. According to common sense, this event was the result of event P – the shot of the tachyon gun from station M. The corresponding interval PQ represents the world line of the tachyon beam fired from M at Alvad in that shot. But according to the reinterpretation principle, this world line would be better read as QP, because event Q happened before event P by the Alvad's clocks. The crew of the Alvad first observed the spontaneous explosion of their shield, accompanied by the outburst of the tachyon beam, which rushed towards Rulian station M and then disappeared into the barrel of its tachyon gun, which was at that moment aiming directly at Alvad. With this chronology, the spontaneous shield explosion appears to be the cause, and the emergence of the tachyon beam due station M and its subsequent absorption by its tachyon gun appear to be the effect.

The world line QC represents the waiting period between the explosion of the Alvad's shield and the retaliating shot of its tachyon gun. The shot at moment C, as already mentioned, was accurately timed to hit the Rulian station M with the infinitely fast beam precisely at the moment when General Hiss was there for the inspection. The world line CS represents the space–time trajectory of the corresponding beam.

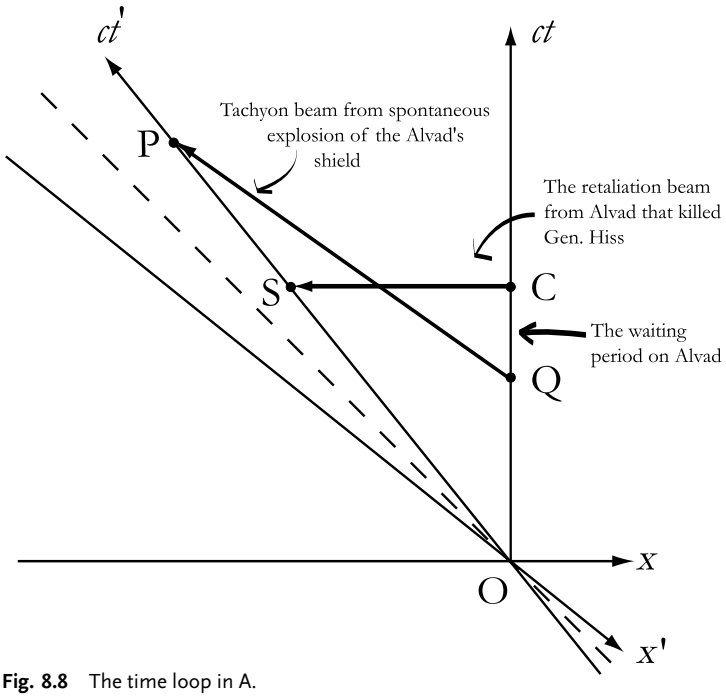


Fig. 8.8 The time loop in A.

The end point S of this line represents the destruction of station M together with General Hiss.

The total picture of all these events and their interpretation drove the crew on the Alvad no less crazy than the population on Rulia. The physicists on the Alvad bitterly disputed the nature of the beam QP. From the viewpoint of those on Rulia, the beam was, with absolute certainty, fired from station M. They witnessed and recorded all preceding developments and preparations leading to this tragic event. But the Alvad's records showed, with equal certainty, that the source of the beam was the explosion of their protective shield, accompanied by the outburst of tachyons towards Rulia, and the beam hit station M at a later time. If we indicate the directions along the beams by arrows pointing from the Alvad's past to its future, then both arrows QP and CS point from the Alvad to Rulia (recall that the speed of the beam CS was only *nearly* infinite, so that the event S occurred slightly later than event C). Now, if we judge only from the directions of the arrows, it is the Alvad that appears to be the aggressor!

Because both the crew of the Alvad and the informed inhabitants of Rulia knew that the opposite was true, they all were intent to agree that in the first arrow QP the actual flow of information must have been the opposite to the direction of time – that is, from P to Q. But this means that the information flew along this arrow from the future to the past! That may seem hard to swallow, but it becomes easier as we recall that for the space-like intervals the properties *past* and *future* are relative and depend on a reference frame.

But there was one thing much harder to comprehend: that of the physical and legal status of General Hiss, and accordingly, in view of his role as a leader of the superpower on the intergalactic scale, the political status and the current situation of all the observable Universe. On the one hand, there could be no doubt that General Hiss was alive and active at the moment P, when his order to start the war was coincident and correlated with the arrival of the beam QP from the Alvad. On the other hand, the earlier obliteration of station M together with General Hiss gave absolutely solid evidence of his death before the moment P. This logical and political ambiguity seemed to some unbearable. But Commander Fletcher thought that it was still better to have the Alvad returning to Earth with the embarrassed crew than to have instead an unambiguous atomic cloud with no crew at all.

The described results constitute a known logical paradox (so-called Tolman paradox [63]) that arises when we exchange the tachyons between two systems in such a way that the corresponding world lines form a loop. Whereas different observers may disagree about separate elements of the loop, they all agree that the loop as a whole is a self-contradictory entity. Looking again at either of Figures 8.7 and 8.8, we can summarize the contradiction in the following way: If General Hiss is *not* killed at the moment S, then he is alive at the moment P, and sends to Q a signal that triggers (through the chain of events QCS) his death at moment S. If he *is* killed at moment S, then he cannot send any signal at P, which cancels the chain PQCS, so he is *not* killed at moment S.

The same self-contradictory conclusion can, of course, be drawn regarding the Alvad's defensive actions: if the Alvad fired the retaliating beam that destroyed station M, then this station did not exist at a later moment P, there is no shot at Alvad and no destruction of its shield, and therefore Alvad never fired the retaliating beam.

To make things look even more impressive, although less dramatic, consider another version of the tachyon communication loop, that has been discussed in a paper by Benford, Book, and Newcomb [64]. In their paper, the tachyon beam carries instructive, rather than destructive, information.

Suppose tachyons had been discovered by Shakespeare's time, and Shakespeare types out the first copy of *Hamlet*, encodes it into modulated tachyon beam and transmits it into space. The modulated beam carrying the text of *Hamlet* catches up with an alien spaceship returning from Earth with an abducted earthling called Francis Bacon. The earthling intercepts and decodes the beam and thereby becomes the second man from Earth who knows *Hamlet*. Draw a space–time diagram of this communication and analyze it from the viewpoint of Bacon. If the product $V\tilde{v} > c^2$, then he receives the encoded message about *Hamlet* before Shakespeare sends it! It again appears that he receives a message from the future. However, what he really observes is quite different. All of a sudden, his receptor spontaneously emits a modulated beam of tachyons and types its decoded contents on a tape. The beam reaches the receding Earth at some later time and illuminates Shakespeare's transmitter, which is moving on its own in mystically perfect synchrony with modulations of the beam. When Bacon measures the energy of his tachyons, he finds its value to be negative. When he looks through the whole tape,

he is astounded to realize that instead of gibberish naturally expected from any spontaneous process, the tape contains a masterpiece that will fascinate generations to come.

This is a more sophisticated version of the unhappy incidence with the killed hare observed by Alice in Section 7.4. And both incidences contain, if not yet a logical paradox, a deep puzzle. On the one hand, Bacon, as Alice before, holds that if there is any causal connection between the observed events on his spaceship and the Earth, the spontaneous triggering of his transmitter must be the cause (it has started first), and the typing of Shakespeare's transmitter must be the effect (it started later, when the beam reached the Earth). Bacon reinterprets the designations of what should be the cause and what should be the effect in accordance with the principle of retarded causality. However, on the other hand, according to universal causality (any event has its cause), the triggering of his transmitter must have its own cause. Actually, however, it is totally spontaneous and has no preceding effects in the past that could cause it. It appears to be a non-causal event, just as the deadly outburst of tachyon beam from the hare, which so embarrassed Alice! In this respect, it violates more general principle of universal causality.

The authors [64] summarized the results of the one-way communication in the following way:

... no amount of reinterpretation will make Bacon the author of *Hamlet*. It is Shakespeare, not Bacon, who exercises control over the content of the message. For any tachyon trajectory the time ordering of the end points is relative ... But the direction of *information transfer* is necessarily a relativistic invariant."

This conclusion sounds as the most plausible explanation of the observed effects in one-way communication. What else could account for the fact that Bacon's transmitter "spontaneously" produces a masterpiece that is produced independently and yet letter by letter the same, after a proper prehistory, billions of miles away?

But the generalization that allows information to flow from the future to the past along space-like trajectories is not by itself sufficient for closed loops. Suppose we want to know what happens if Bacon now signals the obtained information back to Earth. This would be similar to Commander Fletcher firing back to Rulia, and does not even require special timing, since now there is no need to hit the Earth at some special moment of time. Bacon's only concern would be that the signal comes back to Earth long *before* Shakespeare's time. As we know, this can always be done if the distance between the Earth and the spaceship and the ratio $V\tilde{v}/c^2$ are large enough. So imagine Bacon signaling back the complete text of *Hamlet* he had obtained from Shakespeare, and his message arriving at England about half a century before Shakespeare's birth. The message is intercepted by someone named, say, Snakespeare. This person might have observed the transmission as a spontaneous outburst of a modulated beam from his transmitter to a distant spaceship, but the information flow is again from the future to the past. He records the information letter by letter, and publishes it. Very soon he and his work become famous. When Shakespeare is born, *Hamlet* is already taught in colleges, and one of the first plays he watches in the Royal Theatre is the play *Hamlet* by the great Snakespeare. Naturally, Shakespeare does not produce *Hamlet*.

Here we arrive again at the logical paradox. Suppose that the whole story started from Shakespeare. If Shakespeare wrote *Hamlet*, then he triggers the above chain, and as a result he does not write *Hamlet*. If he does not write it, the chain is canceled, and he writes it.

The type of logical loops associated with the modulated tachyon beams carrying instructive information can be not only puzzling, but also as dramatic as the case with Rulia. The instructive information can be used for destructive purposes. Imagine someone in today's America transmitting information about the technology needed to produce the atomic bomb, in a modulated tachyon beam addressed to a dictator of the past!

Depending on whether General Hiss, the dictator of Superpower Rulia, is dead or alive is the fate of the Universe. To be, or not to be? This is the kind of dilemma that can be brought about in the world along with the concept of tachyons.

Either tachyons do not exist, and the above dilemma disappears together with them, or they do exist but in such a way that resolves the paradox. So, what is the Nature's verdict on tachyons? To be, or not to be? So far, the answer to this question is suspended in limbo.

8.9
They *are* non-local!

What remedies can we offer for the mind-boggling situations brought about in the world by the hypothesis of tachyons? The most radical and simple one is that tachyons just do not exist. But what if they do? Then there must be some limitations on the emission–absorption processes that would preclude the appearance of tachyons carrying information into the past. And, since the past and the future are interchangeable for tachyons, we also have to ban signaling into the future. As Feinberg put it: "Tachyons cannot be used for sending reliable signals either forward or backward in time in the sense that one cannot fully control the outcome of the experiments with emission or absorption of tachyons." This stands in accord with the fact that tachyon emission in many proceses with tachyon exchange in the previous sections did indeed appear as a spontaneous process totally unexpected and out of control by an observer in a given reference frame.

In the following sections we shall consider a few unusual properties of tachyons, which will help resolve the above "existential" paradox.

The first such property is that tachyons are intrinsically non-localizable in a sense that a space region occupied by a tachyon cannot in principle be reduced to a point. Rather, a tachyon tends to occupy the whole space. The reason for this is the tachyon's imaginary rest mass $\tilde{m}_0 = i m_0$. We already know from previous sections that this results in the energy–momentum spectrum of a free tachyon, which is fundamentally different from that of a tardyon:

$$\tilde{p}^2 c^2 = \tilde{E}^2 + m_0^2 c^4 \tag{58}$$

That is, while the momentum of a tardyon is smaller than its energy, the momentum of a tachyon is greater than its energy. In particular, the tachyon retains a non-zero momentum even when it has the zero energy [see Eq. (27)]. The product $m_0 c$ is the lowest possible limit of the tachyon's momentum. We can write this as a mathematical restriction:

$$|\tilde{p}| \geq m_0 c \tag{59}$$

Now, recall that according to quantum mechanics, any particle can be described as a superposition of its de Broglie waves. The envelope of the resulting bundle of waves (the wave packet) determines the spatial region occupied by the particle, and the propagation rate of the bundle (the group velocity of the packet) determines the velocity of the particle. We want to apply these rules to a tachyon to find out how it is localized in space. Before doing any math, let us try to understand qualitatively how the restriction (59) may affect the result.

When waves of many different wavelengths cross the same region of space, they interfere constructively in some places and destructively in others, depending on correlations between their phases. The places where the waves cancel each other out are essentially empty (with regard to the particle in question). The place of constructive interference is where we see a big splash, and accordingly there is a high probability of finding a particle there. Figuratively, we can imagine that whatever may constitute the particle (its mass, charge, etc.) is "smeared out" over the region of high probability until an attempt at another observation is made or an interaction with another object changes the original state.

Let us estimate qualitatively the role of different wavelengths in this process. Short waves are absolutely necessary to shape out the fine details of the packet. Manipulating with a set of different short waves, properly adjusting their phases and amplitudes, one can produce a huge splash within a small range – construct a physical state in which a particle occupies practically one point in space. According to de Broglie relationships (Sect. 6.13), short waves correspond to large momentums and energies. Therefore, physicists use high-energy particles (naturally occurring in cosmic rays or produced in accelerators) to probe the properties of matter on a small scale. Long waves, on the other hand, are necessary to shape out the form of the packet on a large scale. In particular, they are needed to cancel the reiterating splashes produced by short waves far away from the main splash. Without long waves, these smaller splashes remain in existence, and therefore there is also a probability of finding the particle far away from the main splash. The particle remains spread out in space rather than being perfectly localized at a point. Without long waves in a particle's spectrum, Nature lacks the tools to localize the particle completely. Long waves correspond to small momentums. Therefore, without sufficiently small momentums, the particle cannot be ideally localized to a point size.

Let us apply this result to tachyons. According to Equation (59), there is a cutoff in the long-wave range (small momentums) of a tachyon spectrum. This means that tachyons cannot, in principle, be point-like particles [65].

The simplest non-local subliminal object of finite size is a sphere with an effective radius a_0. The spherical symmetry is complete – you cannot invent anything more symmetrical (or simpler!) than a sphere. Corresponding models had been extended and widely used also for the description of tachyons, but we will show in the following sections that a model of a tachyon with complete internal symmetry cannot be true.

8.10 Cerenkov radiation by a tachyon and Wimmel's paradox

In the early 1970s, two researchers, Wimmel and Jones, at approximately the same time, considered the same problem that Sommerfeld had once considered – that of Cerenkov radiation by a charged superluminal particle. But they picked up where Sommerfeld left off. They pointed out explicitly in their papers [66, 67] a few difficulties associated with Cerenkov radiation by a tachyon (let us abbreviate this to CRT).

The first difficulty had been known to Sommerfeld himself: the field of a point charge moving faster than light is infinite on the shock front (Mach cone). The physical reason for this can be explained without heavy math in the following way.

As we found in the previous section, in order to construct a point-like object from waves, waves of all lengths are equally needed to extinguish each other properly everywhere except for only one point of space where we want the object to be. The contribution of the arbitrarily short waves (and thereby arbitrarily high frequencies) is vital here because we want the total cancellation of all waves arbitrarily close to the desired point, and their reinforcement at precisely this point. You cannot carry out such a fine adjustment without short waves!

It is the contribution of the short waves (high frequencies) that makes the total radiation field on the Mach cone and total radiation power infinite. This leads to infinitely large values for both radiation power and the drag force on the charge.

Later, with the advance of quantum mechanics, some researchers had tried to get around this difficulty by taking account of the quantum nature of the emission process, according to which the emission occurs as a quantum jump with the release of the whole lump of electromagnetic energy $\varepsilon = \hbar\omega$ in the form of a photon of frequency ω (see Sect. 7.1 and 7.4). According to this mechanism, a particle cannot emit a photon with more energy than the particle had by the moment of emission. This is forbidden by the conservation of energy. Thus, all frequencies higher than $\omega > \varepsilon/h$ must be forcibly exempt ("cut off") from the Cerenkov radiation spectrum for a point-like tachyon with energy ε, and the corresponding procedure was called the cutoff. The absence of the high frequencies makes all the observable quantities for the point charge finite. But the resulting expression for the radiation power is not Lorentz-invariant, and therefore it cannot give the correct description of the process.

The only natural way to get rid of the high frequencies without the artificial cutoff is to consider an extended, rather than a point-like, charge distribution as a source of field. This makes all the observable variables finite, because for the extended charge the relative contribution of the short waves (high frequencies) sharply decreases of its own accord.

To construct the extended charge, the constituent waves must produce a finite splash within a finite region of space. The requirement for the waves to cancel each other is dropped within the region occupied by the object. The wavelengths smaller than the size of the object are no longer necessary, except maybe for a moderate amount needed to delineate a possible sharp boundary of the object. As a result, the spectrum of the constituent waves (called form-factor) falls off at high frequencies, and their net contribution to the radiation power and the drag force becomes finite.

This fact had been known to Sommerfeld, but his treatment using it was not successful either, because he used the model of a non-deformable (that is, non-relativistic) sphere (see Sect. 6.8). Jones refined Sommerfeld's treatment by taking account of the fact that the shape of a tachyon must undergo a relativistic length contraction (as we will see later, it takes the form of length extension at sufficiently high speeds.) He used a model of a relativistic sphere of radius a (in its own reference frame). He also assumed the total charge q to be uniformly distributed over the volume of the sphere. This led him to a finite (and very simple!) expression for the drag force:

$$\frac{\mathrm{d}\tilde{E}}{\mathrm{d}t} = -f\,\tilde{v}, \quad f = -\frac{9}{8}\frac{q^2}{a^2} \tag{60}$$

This expression retains its form in all reference frames moving along the initial direction of $\tilde{\boldsymbol{v}}$. Consider, for instance, another reference frame moving along this direction to the right with speed V. We have in the new frame (using $\tilde{v} = \mathrm{d}\tilde{E}/\mathrm{d}\tilde{p}$)

$$\left.\begin{aligned} \mathrm{d}\tilde{E}' &= \gamma(V)\,(\mathrm{d}\tilde{E} - V\,\mathrm{d}\tilde{p}) = \gamma(V)\left(1 - \frac{V}{\tilde{v}}\right)\mathrm{d}\tilde{E} \\ \mathrm{d}t' &= \gamma(V)\left(\mathrm{d}t - \frac{V}{c^2}\mathrm{d}x\right) = \gamma(V)\left(1 - \frac{V\tilde{v}}{c^2}\right)\mathrm{d}t \end{aligned}\right\} \tag{61}$$

The radiation power in the new system is

$$\frac{\mathrm{d}\tilde{E}'}{\mathrm{d}t'} = \frac{1 - \dfrac{V}{\tilde{v}}}{1 - \dfrac{V\tilde{v}}{c^2}}\frac{\mathrm{d}\tilde{E}}{\mathrm{d}t} = \frac{1}{\tilde{v}}\,\frac{\tilde{v} - V}{1 - \dfrac{V\tilde{v}}{c^2}}\,\frac{\mathrm{d}\tilde{E}}{\mathrm{d}t} \tag{62}$$

But the factor in the middle of the right-hand side of Equation (62) is just the tachyon speed $\tilde{v}'$ in the new system (Sect. 3.1). Therefore,

$$\frac{\mathrm{d}\tilde{E}'}{\mathrm{d}t'} = \frac{\tilde{v}'}{\tilde{v}}\,\frac{\mathrm{d}\tilde{E}}{\mathrm{d}t} \tag{63}$$

Finally, using Equation (60) for $\mathrm{d}\tilde{E}/\mathrm{d}t$, we obtain

$$\frac{\mathrm{d}\tilde{E}'}{\mathrm{d}t'} = -f\,\tilde{v}' \tag{64}$$

This expression has the same form as Equation (60) – it is Lorentz-invariant, with the drag force being just invariant (having the same numerical value in another reference frame).

Since the force changes the speed of the tachyon, its world line must be curved. The resulting world line of the tachyon turns out to be the hyperbola (Fig. 8.9):

$$x^2 - c^2 t^2 = b^{-2}, \quad b = \text{constant} \tag{65}$$

This expression is also invariant because its left-hand side has the form of the interval. Everything seems to be OK.

And yet the results are not 100% satisfactory. Here are the comments of Jones himself: "Both expressions are Lorentz invariant only under a restricted class of Lorentz transformations involving only x and t. Those Lorentz frames in which tachyon trajectories are rectilinear (that is, moving along the x-axis), constitute a class of preferred frames, but they are singled out by the initial conditions and not by the law of motion hence no violation of Lorentz invariance is implied."

But this non-violation is restricted by very special initial conditions for tachyon velocity and acceleration. In particular, Jones has to assume that "... the apparatus that produces a tachyon would do so with its initial 3-velocity and 3-acceleration parallel in the rest frame of the apparatus." A natural question then arises: what physical agent causes such special conditions?

Wimmel saw a real paradox in the fact that the equations for Cerenkov radiation from a tachyon with internal spherical symmetry yielded a Lorentz-invariant solution only for the case when the tachyon's acceleration is parallel to its velocity, and for reference frames moving along this common direction. He concluded his analy-

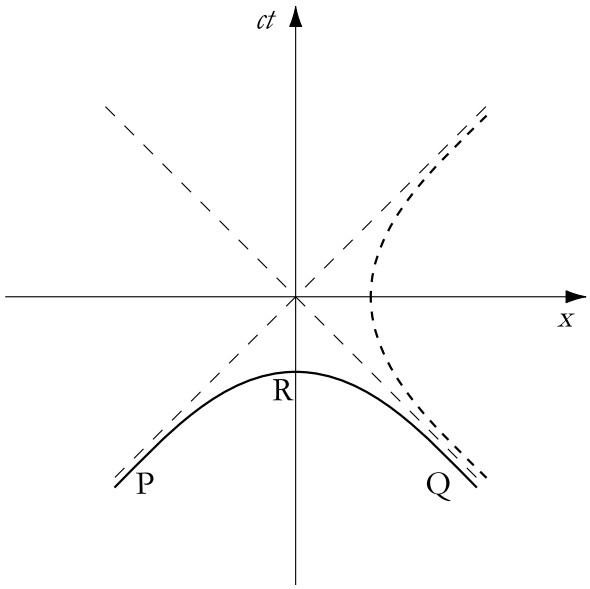


Fig. 8.9 The Cerenkov tachyon world line.

sis of this problem by a general negative statement: "No unique equation $d\tilde{E}/dt$ for Cerenkov energy loss can exist."

Now, if we take a look at the the world line in Equation (65), we will also see something weird. The world line is shown in Figure 8.9 (the reader should not confuse the tachyon world line in this figure with Figure 8.3, showing energy-momentum curves in momentum space). Take a closer look at the branch PRQ. It is, of course, space-like and symmetrical with respect to axis *ct*. As time progresses from $-\infty$ to $-t_0$, the tachyon moves from the left to the origin. As it approaches the origin, its speed increases without bounds. The moment it reaches the origin, its speed becomes infinite, and the tachyon becomes for this instant a transcendent tachyon. After this moment it proceeds to move to the right with decreasing speed, but with negative energy and backward in time! As Wimmel puts it: "The classical Cerenkov energy loss is finite only if a tachyon is not a point particle, and leads to tachyon world lines that bend back in time when they go through the state zero energy." We already know what it means. The motion along the branch RQ should be reinterpreted (and would actually appear to an observer) as the motion of the tachyon with positive energy from right to left forward in time. Therefore, the observer sees two symmetrical tachyons moving along the *x*-axis towards each other with equal and increasing speeds. They meet at the origin and disappear (mutually annihilate), leaving after them an expanding front of electromagnetic (and gravitational!) radiation diverging from the *x*-axis. We know that in order for the two particles to annihilate each other in empty space, they must be particle and antiparticle with respect to each other. Therefore, we have, as other authors [60, 61] had done, extended the reinterpretation principle by adding the requirement that apart from the operations in Equations (46) and (47), we must change a tachyon to its antitachyon. If we denote the tachyon as τ, and the antitachyon as $\bar{\tau}$, then, symbolically,

$$\tau \Rightarrow \bar{\tau} \tag{66}$$

But this leads to another paradox: whenever you launch a tachyon, its fate is known in advance! It is bound to annihilate with its own anti-self at a point where its energy becomes zero. This annihilation, prescribed by Equation (65), is unavoidable and must be considered as an essential part of CRT.

This, in turn, leads to other difficulties concerning causality. Indeed, the impending annihilation implies a mystical connection between the participants, like a preliminary arrangement between them about the exact moment of time, initial energies, and directions of their emission in order for them always to meet at the right place and time (at $x = 0$ and $t = -t_0$). In other words, the emission of a tachyon in a certain direction is only possible if there is a source of antitachyons straight ahead, and this source fires in the oncoming direction at just the right moment to ensure annihilation of our tachyon just when it becomes transcendent. This stands in flat contradiction with all we know about the world in general, and about annihilation as a totally random event.

The bizarre conclusion that we arrived at can be considered as yet more evidence against the existence of tachyons: since no tachyon can be born and emitted in what-

ever direction without a corresponding source of antitachyon lurking in wait in that direction, and since asking for such sources to wink into existence any time at our will would be too much, tachyons just cannot be born.

Such a conclusion, however, would be premature. It turns out that all three paradoxes described above – the existential paradox with closed loops in spacetime, Wimmel's paradox with the absence of a unique function of energy for radiation loss in CRT, and the latest paradox with annihilation – can be removed if we take into account an additional feature of CRT connected with an important property of the tachyon itself – its general asymmetry [68, 69].

In the remaining part of the section we will show that CRT does not generally possess an axial symmetry about its velocity vector $\tilde{\boldsymbol{v}}$. As a result, the spectrum depends, in addition to the frequency ω, on two additional variables: the azimuth angle φ and the angle of misalignment χ, whose physical meaning will be made clear below.

The absence, in the general case, of the axial symmetry of CRT follows directly from transformation properties of the 4-vector $(\omega/c, \boldsymbol{k})$, where $\boldsymbol{k}$ is the wave vector of radiation of frequency ω. Owing to these properties (or because of the Doppler effect, which is the same), the radiation axially symmetrical with respect to $\tilde{\boldsymbol{v}}$ in some inertial reference frame K_0 is asymmetric in another reference frame K moving with velocity $\boldsymbol{V}$ *perpendicular to* $\tilde{\boldsymbol{v}}$. This effect is manifest with maximum clarity when the tachyon speed in K_0 is $\tilde{v} = \infty$ and accordingly CRT is a diverging cylindrical wave. So imagine a reference frame K moving relative to K_0 at $\boldsymbol{V} \perp \tilde{\boldsymbol{v}}_0$. In this reference frame, the tachyon speed is again $\tilde{v} = \tilde{v}_0 = \infty$, and the front of the diverging radiation is also cylindrical (recall Sect. 2.6!). But now, apart from the *shape* of the wave front, we are also interested in the *spectrum* of the emitted radiation. And the spectrum is sensitive to relative motion (recall the discussion of the Doppler effect in Section 5.3)! Because of the Doppler frequency shift, the frequency ratio of radiation emitted in the two opposite directions – one parallel to $\boldsymbol{V}$ and one antiparallel to it – is $(c - V)/(c + V)$, and becomes arbitrarily small as $V \to c$.

In the general case ($\tilde{v}_0 \neq \infty$), the transition $K_0 \Rightarrow K$ is accompanied by changes in both the magnitude and direction of the tachyon velocity. This complicates the situation, but the qualitative picture of the phenomenon remains the same. Below we consider in some detail the basic features of the CRT symmetry breaking, its possible observable manifestations, and the way it removes the above difficulties in the hypothesis of tachyons.

Suppose we managed to launch a tachyon whose CR (Cerenkov radiation) is axially symmetrical with respect to its velocity $\tilde{\boldsymbol{v}}_0$. Consider an observer moving *along* the direction of $\tilde{\boldsymbol{v}}_0$. For this observer, the tachyon is moving with different speed, but along the same direction, and its CR will remain axisymmetrical. We did not specify the speed of the observer, therefore the result is true for any speed in this direction. In other words, for all inertial reference frames, moving relative to each other with different speeds but all along the original direction of $\tilde{\boldsymbol{v}}_0$, the axial symmetry of CRT is conserved. All these reference frames can be obtained from one another by continuous Lorentz-transformation with relative velocity $\boldsymbol{V} \uparrow\uparrow \tilde{\boldsymbol{v}}_0$ or $\boldsymbol{V} \uparrow\downarrow \tilde{\boldsymbol{v}}_0$. We will call the set of these reference frames the "privileged" frames with respect to the given tachyon, and will denote them, and also all the physical quantities of the tachyon mea-

sured in these frames, with the subscript 0. In all reference frames other than "privileged", the CRT (for the given tachyon!) will not possess axial symmetry.

Similar loss of symmetry is known for the charged tardyons moving sufficiently fast in a homogeneous transparent medium (Sect. 6.8). The Mach cone of their Cerenkov radiation is symmetrical with respect to their velocity $\boldsymbol{v}$ in the laboratory reference frame – the rest frame of the medium where the particle moves. If we board another reference frame moving relative to this medium with some velocity parallel to $\boldsymbol{v}$, the observed radiation remains axisymmetrical. However, if we transfer to a reference frame moving perpendicular to $\boldsymbol{v}$, this symmetry will be broken. In this case we can attribute the asymmetry to the influence of the medium. In the new reference frame the medium as a whole moves relative to the observer in a direction different from the direction of the moving particle. This additional direction is singled out as an optical axis of the medium, because the moving medium is optically anisotropic (see Sect. 2.10). We say that the isotropy of CRT is lost because the particle is moving in an anisotropic medium at an angle to its optical axis.

In the case with tachyons, the situation is fundamentally different. The motion occurs in a vacuum, there is no medium to which we could attribute the breaking of symmetry. Such a situation can only be possible if *the tachyon itself has an intrinsic vector* $\boldsymbol{e}$, which singles out a "privileged" direction in space. Such a direction can, of course, be different for another tachyon because of the isotropy of space. But for each tachyon this direction (which can be considered as its individual characteristic) is, in the absence of external forces, a conserving property. We can summarize this in the following statement: the tachyon must be a vector particle – it must carry an intrinsic vector in the sense defined above.

The physical nature of the tachyon vector $\boldsymbol{e}$, most probably, cannot be found from general principles alone. However, it is clear from the above that the transformation properties of $\boldsymbol{e}$ must be the same as those of a segment of length or electric dipole moment. It would therefore be natural to assume that a tachyon can be characterized by an intrinsic dipole moment $\boldsymbol{e} = \boldsymbol{d} = q\boldsymbol{l}$, where q is the charge and l is the dipole length.

There is a ban on the dipole moment of stable particles in view of T-invariance [70]; however, the T-invariance itself is not universal, and a tachyon is not a stable particle: in any event, it must emit gravitational Cerenkov radiation [71].

We want to stress, however, that the hypothesis of the dipolar nature of the vector $\boldsymbol{e}$, although quite natural, is not at all necessary. The vector $\boldsymbol{e}$ may be a dipole, but may as well be something else. We will discuss this question in more detail in the next section, and for now will focus only on the properties of CRT and their consequences.

Now we want to find out the basic features of CRT for an extended charge. Following the approach used by Wimmel and Jones, we will carry the search started by them a little further. We will consider here the same spherical charge q of radius a_0 and $\boldsymbol{d} = 0$ (in this case the intrinsic vector $\boldsymbol{e}$ must have a non-dipole nature), but will go beyond the domain of "privileged" reference frames (the reader can look for more detail in [68].) Suppose first that the tachyon is moving along its intrinsic direction $\boldsymbol{e}$. The "spherical" tachyon will observed as an ellipsoid of rotation about the direction $\boldsymbol{e}$

with transverse semi-axis a_0 and longitudinal semi-axis l undergoing Lorentz contraction at $\tilde{v} \to c$:

$$l = \frac{a_0}{\hat{\gamma}(\tilde{v}_0)}, \quad \hat{\gamma}(\tilde{v}_0) \equiv -i\gamma(\tilde{v}_0) = \left(\frac{\tilde{v}_0^2}{c^2} - 1\right)^{-\frac{1}{2}} = \cot\theta_0 \tag{67}$$

where θ_0 is the Cerenkov angle in system K_0. Because the tachyon "lives" beyond the light barrier, its length contraction described by Equation (67) is pretty weird: the tachyon appears to us a sphere at $\tilde{v}_0 = \sqrt{2}c$, as a flattened (contracted) sphere at $c < \tilde{v}_0 < \sqrt{2}c$ and as an elongated sphere at $\tilde{v}_0 > \sqrt{2}c$. In the last case it actually experiences length dilation, because its longitudinal diameter extends!

Consider now the same tachyon from a reference frame moving relative to K_0 with a speed V in a direction perpendicular to $\boldsymbol{e}$. The change of reference frame does not change the direction of the intrinsic vector $\boldsymbol{e}$. But the tachyon velocity in the new system will no longer be parallel to $\boldsymbol{e}$ – it will acquire a transverse component, $\boldsymbol{V}$ – and will accordingly make with $\boldsymbol{e}$ a certain angle χ. We also expect that in the new reference frame the CRT is not axisymmetrical, that is, different powers and frequencies may be radiated away along different generatrices of the Cerenkov cone (Fig. 8.10). These factors combined must result in an expression for the radiated power in an arbitrary reference frame, which depends on the misalignment angle χ. We want to find this expression.

Carrying out this idea, we direct the z_0-and z-axes of the two systems along the relative velocity $\boldsymbol{V} \perp \boldsymbol{e}$; while the x_0 and x-axes are parallel to $\boldsymbol{e}$. We will then have the fol-

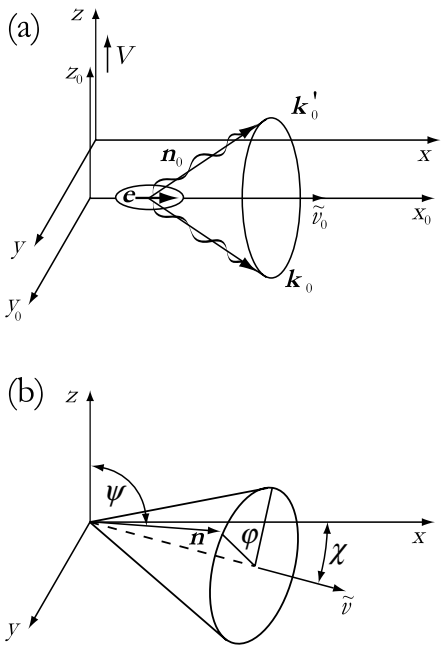


Fig. 8.10 (a) CRT in K_0; (b) CRT in K.

lowing expressions for the components and the magnitude of the tachyon velocity $\tilde{\boldsymbol{v}}$ in K:

$$\tilde{v}_z = -V\,, \quad \tilde{v}_x = \gamma^{-1}(V)\,\tilde{v}_0 \tag{68}$$

$$\tilde{v}^2 = \tilde{v}_0^2 - \frac{V^2\,\tilde{v}_0^2}{c^2} + V^2 \tag{69}$$

According to Equation (68), the vectors $\tilde{\boldsymbol{v}}$ and $\boldsymbol{e}$ make an angle χ determined by

$$\sin\chi = \frac{V}{\tilde{v}} \tag{70}$$

This angle is an additional tachyon variable (degree of freedom) determining the orientation of the vector $\boldsymbol{e}$ with respect to its velocity. The magnitude of χ could in principle be estimated by measuring the observed asymmetry of CRT.

It follows from the invariance of the spherical fronts whose envelope forms the shock wave that in system K one would also observe the Mach cone. The Cerenkov radiation will make an angle θ with the vector $\tilde{\boldsymbol{v}}$ such that $\cos\theta = c/\tilde{v}$. Hence we find, applying Equations (68)–(70):

$$\sin\theta_0 = \frac{\sin\theta}{\cos\chi}\,, \quad \beta \equiv \frac{V}{c} = \frac{\sin\chi}{\cos\theta}\,, \quad 0 \le \sin\chi < \cos\theta \tag{71}$$

Apply now the Lorentz transformation directly to Equation (60) for total loss! We are considering a reference frame moving perpendicularly to the x-axis. In this case the differentials $\mathrm{d}\tilde{E}$ and $\mathrm{d}t$ transform similarly, and we obtain

$$\frac{\mathrm{d}\tilde{E}_0}{\mathrm{d}t_0} = \frac{\mathrm{d}\tilde{E}}{\mathrm{d}t} = -f_0\,\tilde{v}_0 \tag{72}$$

Because we are now in a reference frame where the tachyon's speed is given by Equation (70), we have to express the original speed in K_0 in terms of $\tilde{v}$. After that we express $\tilde{v}$ in terms of the tachyon energy and rest mass. Using Equations (70) and (71), after some algebra we obtain

$$\frac{\mathrm{d}\tilde{E}}{\mathrm{d}t} = -f_0\,c\left(\frac{\tilde{E}^2 + m_0^2\,c^4}{\tilde{E}^2 - m_0^2\,c^4\tan^2\chi}\right)^{1/2}\,, \quad 0 < \tan\chi < \frac{\tilde{E}}{m_0\,c^2} \tag{73}$$

This gives the solution of Wimmel's paradox! We solve it by extending the pool of variables describing CRT. It is true that there is no universal expression for the radiation power of CRT of the form $\mathrm{d}\tilde{E}/\mathrm{d}t = -F(\tilde{E})$. This general negative statement is correct if we expect the function F to depend on a single variable E. But if we introduce the second independent variable χ, which is, as shown above, absolutely essential for

a complete description of the tachyon, we obtain a simple expression, Equation (73) for the radiation power, which is true in an arbitrary reference frame.

I want to stress that Equation (73) is obtained for a model of a "spherical" tachyon. And even for this model, the simplest possible, there emerges, seemingly out of nothing, an internal vector $\boldsymbol{e}$. This indicates that not even a spherical tachyon can possess *complete* spherical symmetry. This must be even more so for more complicated models. Therefore, the above analysis can be considered as a general proof of the vector nature of tachyons, whereas identifying the tachyon vector $\boldsymbol{e}$ with electrical dipole $\boldsymbol{d}$ is, as already mentioned, only a more or less justified assumption. A tachyon *without* the dipole moment or any other vector property familiar to us still must possess an intrinsic vector $\boldsymbol{e}$!

Where does this vector come from? We will look for the answer in the next section.

8.11 How symmetry breaks

As in the previous sections, we will start with the simplest possible model – a tachyon with complete internal symmetry (the "marble"). But before moving further, we want to define the concept of "internal symmetry" for a superluminal object.

The shape of such an object is observed in regular reference frames, where it can be directly measured, and therefore has direct physical meaning. Its "internal symmetry," on the other hand, can only be revealed in its rest frame, that is, in a superluminal frame of reference, and therefore cannot be directly observed. Nevertheless it can be "reconstructed" by the measurements in an ordinary reference frame – in direct analogy with determining the proper mass of the tachyon from its measured energy $\tilde{E}$ and momentum $\tilde{p}$ – as the invariant $m_0 = c^{-2}(\tilde{p}^2 c^2 - \tilde{E}^2)^{1/2}$. And just as the quantity m_0 is a fundamental characteristic of a tachyon, even though there is no reference frame where we could measure it directly, we can speak about a proper or intrinsic shape of a tachyon as an invariant characteristic that can be determined from measurements in regular reference frames.

Let in one such frame K_0 the tachyon be an ellipsoid of revolution about its velocity vector $\tilde{\boldsymbol{v}}$, and the length of its corresponding longitudinal axis is $a_z(\tilde{v})$. Then its proper length, taking account of the Lorentz factor [see Eq. (12)], is determined as the invariant

$$a_0^* = -i\,a_0 = a_z(\tilde{v})\,\gamma(\tilde{v}), \quad \gamma(\tilde{v}) \equiv \left(1 - \frac{\tilde{v}^2}{c^2}\right)^{-1/2} \tag{74}$$

If we impose a condition $a_\perp = a_0$ on the transverse semi-axis, then the tachyon, similar to a tardyon under the same condition, might be expected to possess complete internal symmetry in the sense defined above: it would require only one characteristic, a_0, to describe its shape.

However, we will see that even for this model – the simplest possible – only one radius a_0 is not enough for the complete description of its shape, and therefore the

symmetry of this "spherical" tachyon cannot be equivalent to that of a spherical tardyon. This directly follows from the fact that the longitudinal "proper" radius of the superluminal sphere [Eq. (13)] is imaginary, whereas the transverse radius is real. But in order to clarify the nature of the symmetry breaking connected with this circumstance, a rigorous proof is needed.

With this goal in mind, let us assume the opposite: suppose that complete symmetry is also possible at $\tilde{v} > c$. Then in any inertial frame at any velocity $\tilde{\boldsymbol{v}}$, the corresponding shape of the object must be an ellipsoid of revolution about $\tilde{\boldsymbol{v}}$, with the ratio of semi-axes $a_\perp / a_z = i\gamma(\tilde{v})$, according to the fact that an object with a complete symmetry cannot be characterized by any direction other than $\tilde{\boldsymbol{v}}$. Taking the z-axis of the inertial frame K_0 along the direction of $\tilde{\boldsymbol{v}}$, we have for an arbitrary moment of time t the equation of the surface of the ellipsoid

$$\frac{x^2}{a_0^2} + \frac{y^2}{a_0^2} + \frac{(z - \tilde{v}\,t)^2}{a_z^2(\tilde{v})} = 1 \tag{75}$$

It follows, in view of Equation (74), that

$$x^2 + y^2 + \gamma^2(v)(z - v\,t)^2 = a_0^2\,, \quad v < c \tag{76a}$$

and

$$x^2 + y^2 - \gamma^2(\tilde{v})(z - \tilde{v}\,t)^2 = a_0^2\,, \quad \tilde{v} > c \tag{76b}$$

Equation (76 b) differs from Equation (76 a) only in the sign of the middle term. This distinction causes a fundamental difference between the behavior of tachyons and tardyons. Namely, Equation (76 a) is Lorentz-invariant, whereas Equation (76 b) is not. We will show this with a concrete example – by transition to another inertial frame K moving relative to K_0 with a speed V along the positive x-direction.

The invariance of Equation (76 a) follows directly from the group properties of Lorentz transformations and the spherical symmetry of the tardyon in its rest frame. We can see it in two steps – first by returning to the rest frame of the tardyon, where it is obviously spherical, and then by looking at it from the new system K, where it will be Lorentz-contracted along the new direction of its velocity. In any reference frame, its contraction is uniquely determined by its velocity in this frame, so it does not require any new variables for its description.

For a tachyon, on the other hand, the two-step transition $K_0 \Rightarrow K$ via its rest frame is impossible in the set of Lorentz transformations with $V < c$. We will therefore consider direct transformation from K_0 to K. Setting the axes $\tilde{x}$, $\tilde{z}$ of the new system to be parallel to x and z, respectively, and performing the Lorentz transformation in Equation (76):

$$x = \gamma(V)(\tilde{x} + V\tilde{t})\,, \quad y = \tilde{y}\,, \quad z = \tilde{z}\,, \quad t = \gamma(V)\left(\tilde{t} + \frac{V}{c^2}\tilde{x}\right) \tag{77}$$

we obtain

$$\gamma^2(V)\,(\tilde{x}+V\tilde{t})^2+\tilde{y}^2\pm\gamma^2\,(\tilde{v})\left[\tilde{z}-\tilde{v}\left(\tilde{t}+\frac{V}{c^2}\,\tilde{x}\right)\right]^2=a_0^2 \tag{78}$$

where the + and – signs refer to cases $v < c$ and $\tilde{v} > c$, respectively. The new speed of the tachyon in K is

$$\tilde{v}'=\left(\tilde{v}^2-\frac{V^2\,\tilde{v}^2}{c^2}+V^2\right)^{1/2} \tag{79}$$

It is easy to see that

$$\gamma\,(\tilde{v}')=\gamma\,(V)\,\gamma\,(\tilde{v}) \tag{80}$$

and the direction of the new velocity $\tilde{\boldsymbol{v}}'$ makes an angle χ with the $\tilde{z}$-axis, for which

$$\sin\chi=\frac{V}{\tilde{v}'} \tag{81}$$

Let us now carry out a purely spatial rotation through the angle χ in the system K so that the new axis z' is directed along the velocity $\tilde{\boldsymbol{v}}'$:

$$\left.\begin{aligned}\tilde{x}&=x'\cos\chi-z'\sin\chi\,,\quad \tilde{y}=y'\\ \tilde{z}&=x'\sin\chi-z'\cos\chi\,,\quad \tilde{t}=t\end{aligned}\right\} \tag{82}$$

Putting Equations (82) into Equation (78) and using Equations (79)–(81), we find

$$\left[\frac{\tilde{v}}{\tilde{v}'}x'-\gamma\,(V)\,\frac{V}{\tilde{v}'}(z'-\tilde{v}'\,t')\right]^2+y'\pm\left[\frac{V}{\tilde{v}'\,\gamma\,(\tilde{v})}x'+\frac{\tilde{v}}{\tilde{v}'}\gamma\,(\tilde{v}')\,(z'-\tilde{v}'\,t')\right]^2=a_0^2 \tag{83}$$

For $v < c$ it follows after some algebra that

$$x'^2+y'^2+\gamma^2\,(\tilde{v}')\,(z'-\tilde{v}'\,t')^2=a_0^2 \tag{84}$$

which is, as it should be, the equation for an ellipsoid of rotation about the direction of $\tilde{\boldsymbol{v}}'$ in the new system K.

The situation is absolutely different for $\tilde{v} > c$. In this case [the "–" sign in Equation (83)!] the mixed terms containing the product $x'\,(z'-\tilde{v}'\,t')$ do not cancel, but add up! As a result, the shape of the tachyon in K is characterized by the ellipsoid

$$Px'^2+y'^2+Q\,(z'-\tilde{v}'\,t')^2-Rx'\,(z'-\tilde{v}'\,t')=a_0^2 \tag{85}$$

where

$$P \equiv \frac{\gamma^2(\tilde{v}') - \tan^2\chi}{\gamma^2(\tilde{v}') + \tan^2\chi} > 1\,, \quad Q \equiv -\gamma^2(\tilde{v}')P\,, \quad R \equiv \frac{4\tan\chi}{1+\gamma^{-2}(\tilde{v}')\tan^2\chi} \tag{86}$$

This is not an ellipsoid of rotation! And neither of its axes coincides with the direction of $\tilde{\boldsymbol{v}}'$! The lengths of the three principal axes are

$$a_1 = \frac{2\sqrt{2}\,a_0}{\sqrt{P+Q+\sqrt{(P-Q)^2+R^2}}}\,,\ a_2 = 2\,a_0\,,\ a_3 = \frac{2\sqrt{2}\,a_0}{\sqrt{P+Q-\sqrt{(P-Q)^2+R^2}}} \tag{87}$$

As we see, they are all different. In particular, at $V \to c$ we have from Equation (79) $\tilde{v}' \to c$, and Equations (87) give

$$a_1 \to 0\,, \quad a_2 = 2\,a_0\,, \quad a_3 \to 2\,a_0\left(2\,\frac{\tilde{v}^2}{c^2} - 1\right)^{1/2} \tag{88}$$

For the angle δ betwen the direction $\tilde{\boldsymbol{v}}'$ and the axis a_3 we have

$$\tan 2\delta = \frac{R}{P-Q} = \frac{4\tan\chi}{\left(1+\gamma^2(\tilde{v}')\right)\left(1-\gamma^{-2}(\tilde{v}')\tan^2\chi\right)} \tag{89}$$

Thus, the mere transition to a new frame of reference $K_0 \Rightarrow K$ destroys the axial symmetry of the tachyon – just as it did for its radiation field. Therefore, the shape and attitude of a superluminal ellipsoid is always determined by at least two independent variables – the velocity vector $\tilde{\boldsymbol{v}}'$ and the angle χ between $\tilde{\boldsymbol{v}}'$ and the z-axis (if you remember, the z-axis of the system K_0 is along the "privileged" direction in space where the tachyon is axisymmetrical). The direction $\boldsymbol{z}$ is "memorized" by the tachyon at the transition $K_0 \Rightarrow K$, which indicates the existence of a constant intrinsic vector $\boldsymbol{e}$.

The state of the tachyon can be also determined by the unit vector ξ along the semiaxis a_3 (z'-axis in K) (Fig. 8.11). Vector ξ depends on $\boldsymbol{e}$ by the relation (89) and (as long as the physical meaning of $\boldsymbol{e}$ is not clarified) has the advantage of direct geometrical clarity.

Because of the Lorentz invariance of physical laws, the break of the axial symmetry of the tachyon in transitions to other reference frames means that also in a single reference frame a tachyon can be arbitrarily oriented to the direction of its velocity. This could be manifest in some observable effects depending on the tachyon shape. For instance, one could observe Cerenkov radiation from two identical tachyons, with a symmetrical spectrum for one and an asymmetric one for another (see the previous section). Such an observation would be experimental evidence that the tachyons are actually in different states of "polarization."

We therefore come to conclusion that the superluminal objects described by Equation (76 b) are non-scalar objects.

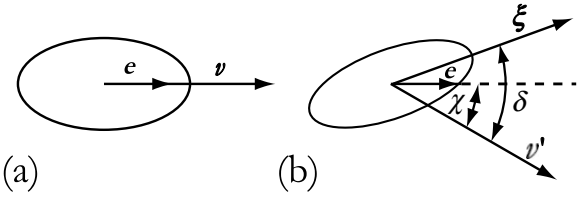


Fig. 8.11 (a) A tachyon in the "symmetry system" K_0. (b) The same tachyon in system K moving relative to K_0 with velocity $\mathbf{V} \perp \mathbf{e}$. The symmetry is broken. The state of the tachyon is determined by its speed v' and also the angle χ between the vectors $\mathbf{e}$ and $\mathbf{v}'$ (or angle δ between $\boldsymbol{\xi}$ and $\mathbf{v}'$).

It would be instructive to observe the origin of this result by comparing the tachyons with non-scalar tardyons. For such a tardyon, the change of its orientation with respect to its velocity is caused by the absence of total internal symmetry (non-sphericity, a dipole moment, spin, etc.) Consider a particular case when the asymmetry is caused by the existence of the electrical dipole moment $\boldsymbol{d}$ (Fig. 8.12). The tardyon is a sphere in its rest frame, so *without* the dipole moment it would possess complete symmetry. Let the charges forming the dipole moment $\boldsymbol{d}$ be distributed as shown in Figure 8.12a – they occupy two semi-spheres (in their rest frame) separated by the equatorial plane PQ. In a system K_0 moving parallel to $\boldsymbol{d}$, the tardyon is an ellipsoid of revolution about is velocity. The charges here are again separated by the equatorial plane. If we now transfer to another system K moving relative to K_0 with velocity $\boldsymbol{V} \perp \boldsymbol{d}$, the vector $\boldsymbol{d}$ will not be affected by the transition, whereas the direction of the tachyon's velocity will change by the angle χ determined by Equation (81). The tardyon remains the ellipsoid of revolution about the new direction of its velocity (Fig. 8.12b). But the plane PQ separating the + and – charges will now make with $\boldsymbol{d}$ an angle (recall Sect. 2.10!)

$$\phi = Arc\tan \frac{c^2}{Vv\gamma(V)} \tag{90}$$

The new positions of the separation plane and the equatorial plane of the ellipsoid are shown in Figure 8.12b. The position of the separation plane in K does not coincide with the equatorial plane! Even though the *geometric shape* of the tardyon retains its axial symmetry with respect to its velocity, its *internal physical structure* (electric charge distribution) and therefore its physical state no longer have this symmetry.

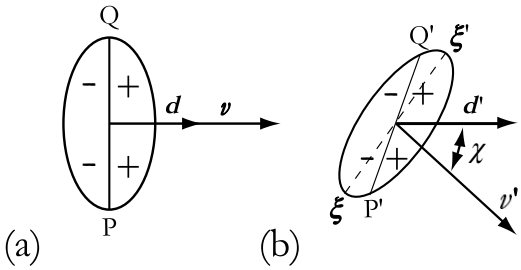


Fig. 8.12 (a) A tardyon with the dipole moment $\boldsymbol{d}$ (system K_0). (b) The same tardyon in system K moving relative to K_0 with velocity $\mathbf{V} \perp \mathbf{e}$. The line $P'Q'$ is determined by the intersection of the plane of the figure with the plane separating the + and – charges. The line $\xi\xi'$ is a similar "trace" of the equatorial plane of the particle.

This result becomes self-evident if, instead of continuous charge distribution, we consider two spatially separated point charges.

It is important to realize that for a tardyon to lose its scalar nature, it has to be imparted with some intrinsic asymmetry of the kind considered above. For a tachyon, in contrast, the conclusion regarding its vector character follows without any assumptions about its internal physical properties and is manifest in the fact that, unlike tardyons, its cannot retain the axial symmetry of its geometric shape. In this case, the geometric shape becomes a physical factor.

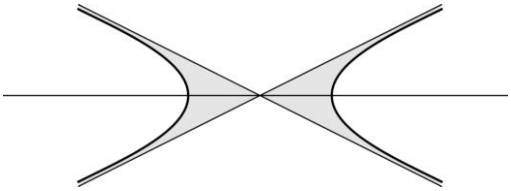
We can still ask: what specifically is the source of the vector nature of tachyons?

Such a source has already been mentioned: original anisotropy – its imaginary proper length along the direction of its velocity. At the core of this, somewhat formal, explanation lies very clear physics: the source of the vector nature of a tachyon is its superluminal motion. The superluminal velocity singles out a space-like axis in Minkowski's world, which is purely spatial in some inertial frame. An axially symmetrical mass distribution around this direction turns out to be sufficient for "memorizing" it under Lorentz transformations. In a regular (subluminal) scalar object the memory of the initial direction of its velocity is totally erased by transition to its rest frame. In a tachyon, this memory can only be destroyed with the tachyon itself. It is recorded in the physical structure that singles out a direction in space along which the tachyon is axisymmetrical. We cannot destroy it because we cannot transfer to a tachyon's rest frame. This leads us to the concept of an intrinsic vector $\boldsymbol{e}$, which can be parallel to the tachyon's velocity in some reference frames, and in others will make a non-zero angle with it. Thus, even though the vector $\boldsymbol{e}$ is caused by the superluminal velocity of the tachyon, it is not uniquely determined by velocity alone.

There may or may not be additional anwers to the question about the physical nature of vector $\boldsymbol{e}$. From the above analysis we can only suggest that the vector $\boldsymbol{e}$ is connected somehow with the form factor of the tachyon, and the latter depends on the initial conditions at the moment of its birth.

There is, however, an additional twist to the whole story about the vector nature of tachyons. Take another look at Equations (76). If in Equation (76 b), while keeping the condition $\tilde{v} > c$, we change "by hand" the sign at $\gamma^2(\tilde{v})$, we will obtain the equation for a one-folded *hyperboloid of revolution* about $\tilde{\boldsymbol{v}}$ rather than an ellipsoid. But the change of sign will do much more than just reshape the ellipsoid (object of finite size) into a hyperboloid (an object of infinite size). It will also restore its universal axial symmetry! The superluminal objects of such shape would remain axisymmetrical in all reference frames. This is also true for tachyons shaped as two-folded hyperboloids, which were introduced by Recami and co-workers [72]. The region occupied by such a tachyon is bounded by the two folds of the hyperboloid, and by the two-folded conical surface (Fig. 8.13). The whole structure, while being always axially symmetrical with respect to the direction of tachyon's velocity, is not rigid, because the solid angle at the apex of the cone depends on the tachyon speed. However, the most interesting thing about it is that this angle is equal to the angle of the Mach cone produced by an object with this speed. This may tempt one to identify the object itself with the cone of its Cerenkov radiation. But this would be wrong! The cone

Fig. 8.13



of Cerenkov radiation trails the tachyon. Here we indeed have one trailing cone, but in addition there is one in front of the apex! Apart from the orientation of this leading cone, the geometry of both cones is the same as that of Cerenkov radiation, but their physical nature is not. It is just the tachyon's substance itself so exotically distributed over the region shown in Figure 8.13. Moreover, the authors have shown that such a distribution will not emit Cerenkov radiation at all!

When I first learned about this idea, my reaction to it was very skeptical, for four reasons. First, the authors achieved universal symmetry at a high price – they had to assume that the square of the four-dimensional interval ds^2 changes sign under the superluminal Lorentz transformations in Equation (9). In other words, they saved the invariance of tachyon symmetry at the cost of the invariance of the interval. Second, in view of the paradoxes in the above sections, the asymmetry of tachyons and their radiation should be considered as a virtue, because it helps resolve the paradoxes. The restoration of the tachyon symmetry, and their non-radiation, apart from the price paid, would bring the paradoxes back! Third, an infinite size of a particle is hardly a plausible physical property. If there are any interactions between them and the rest of the world, they will effectively restrict the tachyon size in the longitudinal direction by non-hyperbolic surfaces (Recami and co-workers themselves suggested such a restriction by two planes [72]). Fourth, the additional restricted surfaces will again cause the symmetry break of the whole object at the $K_0 \Rightarrow K$ transition. Hence we can say that the symmetry breaking comes back as an unavoidable property of any localized superluminal object.

By the time of writing this book, however, my skepticism had nearly disappeared – for reasons which will become clear in the next sections.

8.12
Paradoxes revised

The fundamental asymmetry of the tachyon, discussed above, must have profound consequences. In particular, it dramatically affects its motion and thereby its ability for superluminal signaling. In this section, we will revise all the above paradoxes associated with this ability, and show that the asymmetry precludes tachyons from transporting any information and thereby restores the "world order."

The asymmetry of CRT must cause the appearance of the drag force transverse to velocity $\tilde{\boldsymbol{v}}$, and accordingly there must appear a transverse component of the acceleration. This will curve the *spatial* path of the tachyon! This kind of curvature has a crys-

tal clear analogy in non-relativistic mechanics: go back to the Introduction and consider a falling item dropped by a passenger in a subway car. The item will fall with acceleration, but along a straight path relative to the car. This is similar to an accelerated motion of the tachyon in an inertial reference frame K_0, where the tachyon acceleration is parallel to its velocity. The car in our example plays the role of such a "privileged" reference frame K_0. However, the same trajectory of the falling item is curved (parabolic) in the reference frame of the platform (system K) – because the item's velocity relative to platform is not vertical, and acceleration due to gravity has the component transverse to it! This is similar to the curved trajectory of the tachyon seen from system K. The analogy is incomplete because the asymmetry in the case of a falling item has an external source, the gravitational field of the Earth, whereas the source of the tachyon's asymmetry is the tachyon itself. However, the effect of the asymmetry on motion is essentially the same in both cases.

Our problem now will be to find the components of the tachyon acceleration in an arbitrary reference frame K. To this end, we have first to determine the components of force $\boldsymbol{f}$. We can do it by writing the Lorentz transformation for 4-force $G = \gamma(V)\{\boldsymbol{f}\tilde{\boldsymbol{v}}/c^2, \boldsymbol{f}_0/c\}$ (Sect. 4.2). In the original system K_0 we have only two non-zero components: the time component $G_{c\tilde{t}_0} = \gamma(\tilde{v}_0)(f_0\tilde{v}_0/c^2)$, and the z-component $G_{z_0} = \gamma(\tilde{v}_0)(f_0/c)$, In system K, moving perpendicularly to the direction $\tilde{\boldsymbol{v}}_0$, with the axes $\tilde{x}$, $\tilde{y}$, $\tilde{z}$, parallel to those of K_0, there will be already three non-zero components: G_{ct}, $G_{\tilde{x}}$, $G_{\tilde{z}}$. We are interested in the last two (spatial) components. They are related to corresponding components in K_0 by Lorentz transformations:

$$G_{\tilde{x}} = \gamma(V)(G_{x_0} - G_{c\tilde{t}_0}), \quad G_{\tilde{z}} = G_z \tag{91}$$

The z-component does not transform because it is transverse to the direction of the relative motion of the two systems. Now the diligent reader is invited to perform a simple algebraic exercise: express all the Gs in Equation (91) in terms of real physical force $\boldsymbol{f}$ as indicated above, use $\gamma(\tilde{v}) = \gamma(V)\gamma(\tilde{v}_0)$, and show that

$$f_{\tilde{x}} = -\frac{V\tilde{v}_0}{c^2}f_0, \quad f_{\tilde{z}} = \gamma^{-1}(V)f_0 \tag{92}$$

The components we have found are related to the K-axes that are parallel to corresponding axes of the original system K_0. But what we actually want to find is the force components related to the direction of the tachyon's *velocity in K*, which is tilted to the $\tilde{z}$-axis (Fig. 8.14). The angle χ between these directions is given by Equation (81). Therefore, we now have to carry out a purely spatial rotation within the system K – to rotate it about the y-axis through an angle χ. Call the new axes x, y, z. The z-axis is directed along $\tilde{\boldsymbol{v}}$, and the x-axis is perpendicular to $\tilde{\boldsymbol{v}}$.

As can be found by inspection of Figure 8.14, the components of $\boldsymbol{f}$ along x and z are related to the components of Equation (92) by

$$\left.\begin{aligned} f_x &= f_{\tilde{x}}\cos\chi + f_{\tilde{z}}\sin\chi \\ f_z &= -f_{\tilde{x}}\sin\chi + f_{\tilde{z}}\cos\chi \end{aligned}\right\} \tag{93}$$

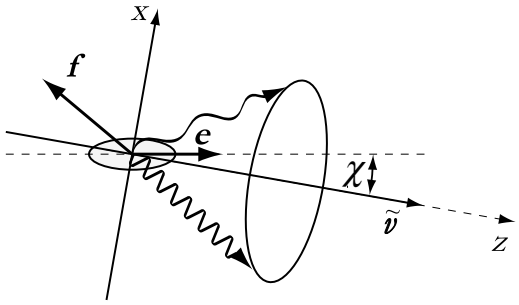


Fig. 8.14 The radiation drag force on the tachyon.

Putting here Equation (92) gives

$$\left.\begin{aligned} f_x &= \left(-\frac{V\tilde{v}_0}{c^2}\cos\chi + \gamma^{-1}(V)\sin\chi\right) f_0 \\ f_z &= \left(\frac{V\tilde{v}_0}{c^2}\sin\chi + \gamma^{-1}(V)\cos\chi\right) f_0 \end{aligned}\right\} \tag{94}$$

Our next step is to express $\tilde{v}_0$ in terms of $\tilde{v}$:

$$\tilde{v}_0^2 = (\tilde{v}^2 - V^2)\gamma^2(V) = \tilde{v}^2\left(1 - \frac{V^2}{\tilde{v}^2}\right)\gamma^2(V) = \tilde{v}^2\gamma^2(V)\cos^2\chi$$

Putting this into Equation (94) and expressing also all the speeds in terms of the angles according to Equations (70) and (71), we will finally arrive at fairly simple equations for the force components:

$$f_x = -\frac{\tan\chi\tan^2\theta}{\sqrt{1-\tan^2\chi\tan^2\theta}}f_0\,; \quad f_z = \frac{f_0}{\sqrt{1-\tan^2\chi\tan^2\theta}} \tag{95}$$

Now, let us read these equations. Look at Figure 8.15 and recall that the value of f_0 is negative – it is a drag force! Accordingly, the *transverse* component of the radiation force on the tachyon in K is in the *positive* x-direction. The tachyon's behavior in directions *perpendicular* to its velocity is the same as that of a tardyon, and so is its response to a transverse force. Therefore, it must accelerate in the direction of the corresponding force component – in the positive x-direction – just as a tardyon would do. Now, if you look at Figure 8.15, you will see that such acceleration would tend to *decrease* the angle χ – it will bend the tachyon's path so as to make its velocity parallel to vector $\boldsymbol{e}$ again.

Concerning longitudinal acceleration, for a tachyon it must be *opposite* to the applied force – recall our analysis in Section 8.5. This means that the tachyon will accelerate under the negative (braking) force!

We can confirm these results quantitatively if we use the results in Section 4.2. You remember, we obtained there Equation (27) in Section 4.2 – one of the basic relations in the relativistic dynamics of a point mass. Now the time has come to use this equation. Applying it to our case, we have for the x-component of the acceleration

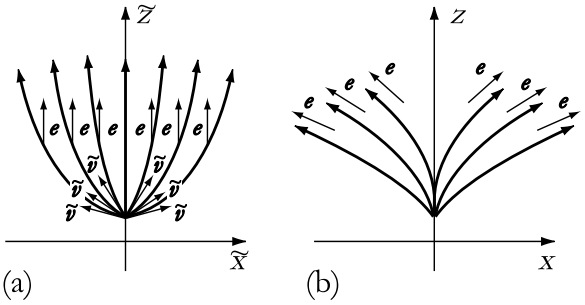


Fig. 8.15

$$a_x = \frac{f_x}{m_0 \hat{\gamma}\,(\tilde{v})} \tag{96}$$

and for the z-component

$$a_z = \frac{f_z - \dfrac{\tilde{v} f_z}{c^2}\,\tilde{v}}{m_0 \hat{\gamma}\,(\tilde{v})} = \frac{f_z}{m_0 \hat{\gamma}^3\,(\tilde{v})} \tag{97}$$

Using Equations (95) and the identity $\hat{\gamma}^{-1}(\tilde{v}) = \tan\theta$ [Eq. (67)], we obtain

$$a_x = -\frac{\tan\chi \tan^3\theta}{\sqrt{1 - \tan^2\chi \tan^2\theta}}\, a\,; \quad a_z = -\frac{\tan^3\theta}{\sqrt{1 - \tan^2\chi \tan^2\theta}}\, a \tag{98}$$

where $a = -f_0/m_0$.

Let us introduce the angle α between the vectors $\boldsymbol{a}$ and $\tilde{\boldsymbol{v}}$. We see from the last equation that both a_x and a_z are positive, in accordance with our quantitative conclusion. Therefore, $\tan\alpha = \tan\chi$, $\alpha = \chi$. It follows that $\boldsymbol{a} \uparrow\uparrow \boldsymbol{e}$, and the radiation reaction tends to decrease χ, "pressing" vector $\tilde{\boldsymbol{v}}$ to $\boldsymbol{e}$. Since at the same time it increases the magnitude $\tilde{v}$, the tachyon moves ever faster and its trajectory becomes more and more parallel to $\boldsymbol{e}$ (Fig. 8.15 a). Therefore, in the limit $\tilde{v} \to \infty$ the speed of tachyon obtained under the given initial condition becomes parallel to $\boldsymbol{e}$ in any reference frame. (This does not mean that at $\tilde{v} \to \infty$ CRT becomes axially symmetrical again. We have already mentioned that at $\tilde{v} \to \infty$ the condition $\chi = 0$ is not by itself sufficient for selecting the subset K_0.)

We can also show another remarkable property of the tachyon trajectories. Let us start again from system K_0. We already mentioned that the equation $\boldsymbol{f} = d\boldsymbol{P}/dt$ leads in this system to the trajectory

$$z_0^2 - c^2 t_0^2 = -(c^2/a)^2 \tag{99}$$

We can "project" the shape of this trajectory into system K by expressing coordinates $x_0 = 0$; z_0, t_0 in terms of the coordinates $\tilde{x}$, $\tilde{z}$, $\tilde{t}$ of system K:

$$x_0 = \gamma(V)(\tilde{x} + V\tilde{t}) = 0\,; \quad z_0 = \tilde{z}\,; \quad t_0 = \gamma(V)\left(\tilde{t} + \frac{V}{c^2}\tilde{x}\right) \tag{100}$$

As a result we will obtain

$$\tilde{z}^2 - \beta^2\tilde{x}^2 = -(c^2/a)^2 \tag{101}$$

In this equation the value of $\beta = V/c$ is a constant parameter. On the other hand, according to Equations (71), this constant is the combination of the variables χ and θ. We thus can conclude that for each tachyon the quantity

$$\beta c = V = (c/\cos\theta)\sin\chi = \tilde{v}\sin\chi \tag{102}$$

is the integral of motion along its whole trajectory! As the tachyon is moving, the variables χ and θ change, but in such a coordinated manner that their combination in Equation (102) remains constant. Also, this constant shows the result that we already obtained in a different way from the analysis of the tachyon acceleration: as the tachyon's speed increases, the angle χ decreases, so that the trajectory bends towards the direction $\boldsymbol{e}$.

Figure 8.15 shows two different families of tachyon trajectories with different β corresponding to different initial values χ, θ, that is, pertaining to different initial $\tilde{\boldsymbol{v}}$ with common $\boldsymbol{e}$ ("self-channeling" of the tachyon beam (Fig. 8.15 a), and the opposite case ("$\boldsymbol{e}$-divergence" of the beam, Fig. 8.15 b).

The effect of the curving of the trajectory has been experimentally observed for a Cerenkov dipole in a medium. However, in that case the dipole's velocity can deflect *away* from the dipole's axis. This may appear as the experimental disproof of our theory. But it is not. Recall that at $v < c$ the radiation force decreases the speed v! Therefore, assuming that the same rule, Equation (102), applies to a tardyon in a medium, we see that it prescribes for this case an increase of χ with time, in accordance with the experiment!

However, the main difference from the dipole in the medium is that the bend of the *tachyon's* trajectory is a general phenomenon not necessarily connected with the existence of the dipole moment. It is a fundamental property determining the way in which a tachyon moves.

With this general result, we are now prepared to come back to the paradoxes we discussed above. It turns out that the paradoxes disappear the moment we take account of the properties of tachyon motion. In order to transport the signal or information in a certain direction, one must exercise full control over the characteristics of motion of the beam used for the transport. However, the properties of tachyon's motion tell us that we cannot control these characteristics, especially the direction of motion. Suppose you want to hit a target or send a message in a certain direction. You shoot a tachyon bullet, or you beam a signal in this direction, but the moment the carrier of the signal is out of your transmitter, it curves away from the direction to your target. It will go where the intrinsic vector $\boldsymbol{e}$ tells it to go, not where the gun had been directed. Since we have no control over the intrinsic properties of a tachyon, we cannot predict its motion.

This conclusion about the uselessness of tachyons for communication purposes becomes even more fundamental if we take account of the quantum nature of radiation. According to the classical theory of CRT, the Cerenkov emission occurs for all tachyon energies down to $\tilde{E} = 0$. At this point the radiation can only be possible if the final tachyon energy can become negative. As we know, at this moment one could actually observe the annihilation of the tachyon and antitachyon with positive energies (reinterpretation principle). As we pointed out, such an outcome admits a mysterious connection between the partners that appear to have conspired in advance about all essential characteristics of their motion and their initial positions, which are needed for them to meet. But this unphysical conclusion follows only because we use a classical approach beyond the limits of its applicability. The classical treatment of CRT is valid under the condition

$$\hbar\omega \ll \tilde{E} \tag{103}$$

When the tachyon is accelerated by its own radiation to almost infinite speed, its energy goes to zero, and the condition in Equation (103) breaks practically for the whole frequency range. The radiation now has to be treated quantum mechanically. In this case we have discrete acts of the photon emission instead of continuous radiation. If the tachyon after the emission has to be reinterpreted as an antitachyon (which happens when $\hbar\omega > \tilde{E}$), we have the act of annihilation as random as the emission of the photon itself. The direction of the emission is now also random, no longer given by the Cerenkov classical rule in Equation (31), Section 6.8, and accordingly there must be a random tachyon recoil. As a result, the "true" tachyon trajectory must be more and more like a trajectory of a particle performing the Brownian motion and in addition "blown away" with a superluminal speed (Fig. 8.16). The points where the trajectory is broken correspond to random emissions of a photon.

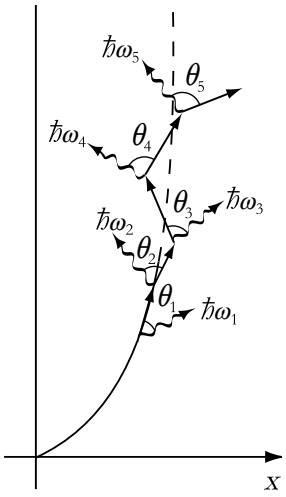


Fig. 8.16

The smooth curve of the type shown in Figure 8.15 is only an averaged result of these random "breakups."

We may therefore conclude that at least the kind of tachyons discussed here cannot be used for a signal or energy transfer in any given direction, for two reasons: first, because of the uncontrollable spread of individual vectors $\boldsymbol{e}$ taking tachyons away from this direction, and second, because of degradation and Brownian wanderings of tachyons due to photo-decay. This also refers to tachyons that may be stable with respect to photo-decay, owing to unavoidable gravitational Cerenkov radiation [71].

There is no room left for "prearranged" annihilations, causality violations, or other paradoxical processes, except for random spontaneous effects, in which the statistical aspects of tachyon behavior are exactly the same as those of any regular particle. All our thought experiments with Alice, Peter, and Tom, Commander Fletcher and General Hiss, Shakespeare, Snakspeare, and Francis Bacon, where the superluminal signals and bullets were moving in straight lines, cannot be performed with the kind of superluminal particles allowed by the theory. Let alone the fact that such particles have not yet been discovered!

8.13 Laboratory-made tachyons

In this section we consider a weird object, which has the elements moving faster than light in vacuum and obeys the kinematic of tachyons. It can be easily manufactured in the laboratory, and even at home if you have a simple pulse laser, a pair of lenses, and a conical mirror. The embryo of such an object has been considered in Section 6.14. It is a system of two crossing plane waves in vacuum. Now we will take two more steps to transform this system into an *electromagnetic model of a tachyon* (EMT).

Step 1. Instead of monochromatic waves, use the narrow wave packets propagating in the directions $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ (Fig. 8.17 a). The intersection line between the packets slides along the bisector of the angle BOB′ with the speed $\tilde{v} = c/\cos\theta$.

Step 2. Instead of only two flat wave packets, make a system with an axial symmetry by rotating Figure 8.17 a about the line OO′. As a result, we will obtain the conical surface A, the envelope of all the flat fronts, and the apex O, where the energy density is maximum (Fig. 8.17 b).

There is a simple way of manufacturing a system of such waves with axial symmetry. Consider a conical mirror and let a flat packet AB be incident on the mirror along its symmetry axis (Fig. 8.17 c). As a source of the packet with a small width one can use a picosecond pulse laser (at a pulse period $\sim 10^{-12}$ s, the pulse width is about 3×10^{-2} cm). Upon reflection in the mirror the flat packet will be converted into a conical wave with the maximum intensity at the apex. The region of high energy density around the apex has a size of the characteristic wavelength of the packet. The maximum intensity depends on the ratio z/λ, where z is the distance traveled by the apex from the moment of its formation. For a typical wavelength of about half a micron and $z = 10$ cm, the energy density at the apex can be a few million times that in

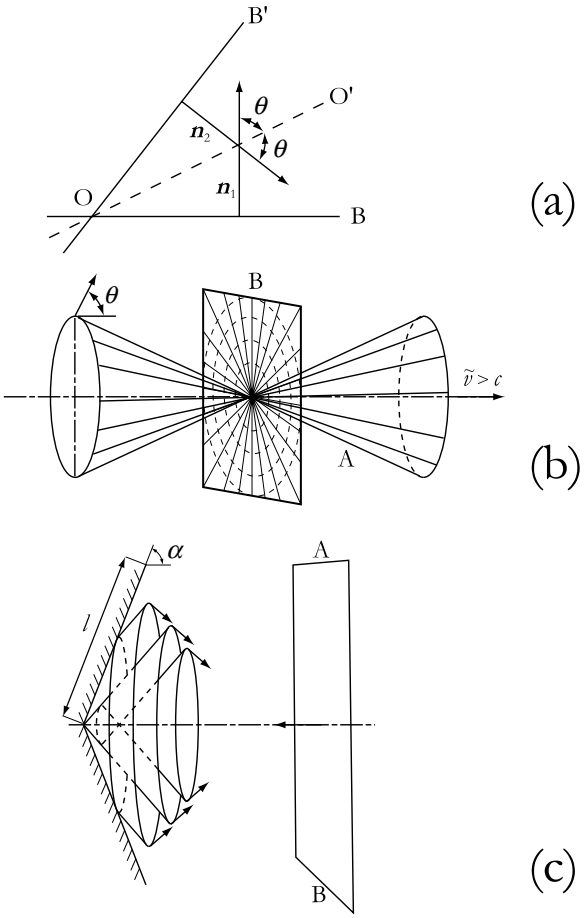


Fig. 8.17 Diagram illustrating the formation of a radiation cone with the properties of a tachyon.

the initial packet. This allows us to speak about the high-intensity region as a lump of electromagnetic energy and associate it with a moving particle. The speed of the lump is, of course, the same as that of the apex:

$$\tilde{v} = \frac{c}{\cos\theta} \tag{104}$$

Since the speed is superluminal, the particle we associate it with must be the tachyon. This is a very special tachyon consisting entirely of the electromagnetic field, and it therefore deserves a special name. Let us call it the electromagnetic tachyon (EMT).

As the open angle of the mirror changes from π to $\pi/2$, the angle θ changes from zero to $\pi/2$, and the speed in Equation (104) changes accordingly from c (the reflection from a flat mirror, no cone is formed) to ∞ (the cone is degenerated into a cylin-

der). In mirrors with an open angle of less than $\pi/2$ the "tachyon" formed moves in the opposite direction – from the base to the apex of the mirror.

In a vacuum, the EMT formed exists within a lifetime determined by the transverse size of the incident packets and the angle of the mirror. The "extra" plane B shown in the figure does not form if we use the conical mirror as described above.

The drawback of this model is that the electric (and magnetic) field on the cone is different from that of a real charged tachyon considered in Section 8.10. With the above technique it cannot be made polarized along the generatrices of the conical surface. This drawback can be eliminated if we use a more sophisticated technique, which can hardly be carried out under domestic conditions. The idea is to obtain real Cerenkov radiation from a bunch of electrons in a medium, and then shake the cone off the bunch by letting the system go out into vacuum. The expanding cone can then be reflected from a properly positioned conical mirror (Fig. 8.18). The object produced would be much closer to a "true" tachyon – its Mach cone has been "appropriated" from a real superluminal (in the medium) charged particle! The electric field lines will be "combed" along generatrices of the cone, and magnetic field lines will form circular loops as shown in Figure 8.18. Such field structure is characteristic of the Cerenkov radiation by any charged particle.

Denote with E the total electromagnetic energy of the reflected packet. What is its total momentum? When the cone has just formed, it has one fold with its open side ahead and its apex behind. We can consider the total energy as the sum of the elements ΔE contained in narrow strips along the generatrices of the cone. Each strip moves perpendicular to itself and has the momentum $\Delta P = \Delta E/c$, with a longitudinal component $(\Delta E/c)\cos\theta$. When we sum all the elements, the transverse components will all cancel each other out, because owing to the axial symmetry to each transverse

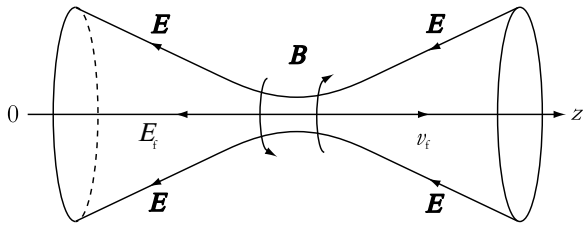


Fig. 8.18

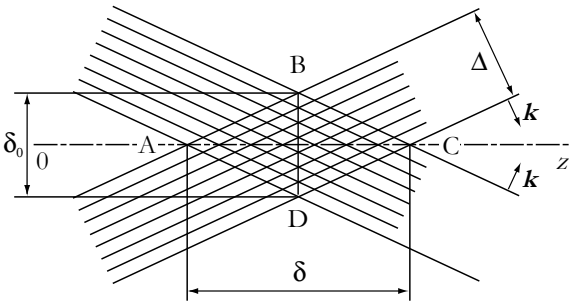


Fig. 8.19

component there is another one with the opposite sign. Only the longitudinal components contribute to the whole, resulting in the net momentum

$$P = \frac{E}{c}\cos\theta \tag{105}$$

As a result, the conical electromagnetic field formed after the reflection will have a non-zero rest mass:

$$m_0 = \frac{1}{c^2}\sqrt{E^2 - p^2\,c^2} = \frac{E}{c^2}\sin\theta \tag{106}$$

Any object with non-zero rest mass has the rest frame where this mass can be measured directly. The reader will not confuse this rest frame with the rest frame of the mirror! In the rest frame of the cone the whole energy distribution stands still – it has zero momentum. Applying the general Equation (106) to this frame, we will see that for $P = 0$ there must be $\theta = \pi/2$, and the cone degenerates into a cylindrical wave. The speed of this system relative to the mirror coincides with the speed of the "center of mass" of the cone:

$$v = \frac{\mathrm{d}E}{\mathrm{d}P} = \frac{P}{E}\,c^2 \tag{107}$$

and, as is seen from Figure 8.17 (and Equation 105), is equal to the z-component of velocity of the conical front:

$$v = c\cos\theta \tag{108}$$

Equations (105)–(108) describe a regular subliminal object.

Now we will show that the same object has all the basic properties of a tachyon. First, as we already know, the apex of the cone outruns its center of mass – it moves with the superluminal speed given by Equation (104). We have formally identified this maximum with a particle, and now we can see that the radiation direction in the trailing cone is exactly the same as it would be for the Cerenkov radiation from this particle.

Now, we know that the tachyon must be essentially a non-local particle. Its transverse (with respect to velocity $\tilde{\boldsymbol{v}}$) size $r_\perp$ is constant, and longitudinal size undergoes Lorentz contraction

$$r = i\,\frac{r_0}{\gamma(\tilde{v})} = r_0\sqrt{\frac{\tilde{v}^2}{c^2} - 1} \tag{109}$$

The lump at the apex possesses the same properties! Indeed, for a quasi-monochromatic radiation the width of the incident packet is proportional to the number N of the characteristic wavelengths that fit into it: $\Delta = N\lambda = 2\pi N/k$. The region of the

strong field formed at the apex is bounded by the surface obtained by rotation of the rhomb ABCD about the z-axis (Fig. 8.19). The transverse size of the rhomb is

$$\delta_\perp = \Delta/\sin\theta = 2\,\pi\,\frac{N}{k\sin\theta} \tag{110}$$

We can show the invariance of this expression by noting that $k\sin\theta = k_\perp$ is the transverse component of the wave vector, which remains invariant under the Lorentz transformations along the z-direction.

More generally, we can consider the electromagnetic field of the cone as consisting of the photons with momenta $p = \hbar k$. The individual photon momentum p is E/cN_f, where N_f is the net number of photons in the field. Thus, $k = p/\hbar = E/c\hbar\, N_f$ and

$$\delta_\perp = 2\,\pi\,\frac{N\,N_f\hbar c}{E\sin\theta} \tag{111}$$

However, according to Equation (106), $E\sin\theta = m_0 c^2$ is the invariant rest mass of the cone, and therefore

$$\delta_\perp = 2\,\pi\,N\,N_f\,\frac{\hbar}{m_0\,c} = \mathrm{inv} \tag{112}$$

Thus, the length $\delta_\perp$ can be interpreted as an invariant transverse size of the tachyon associated with the cone. For the longitudinal size one has, in view of Equation (110),

$$\delta = \Delta/\cos\theta = \delta_0\tan\theta = \delta_0\left(\frac{\tilde{v}^2}{c^2} - 1\right)^{1/2} \tag{113}$$

which is identical with Equation (109). We see that the geometrical properties of the lump are exactly the same as those of the tachyon.

Next we want to show a really weird thing. We know that the superluminal motion of the lump is actually the motion of geometrical region of high energy density, and its identification with the tachyon is rather formal. Actually, the whole cone with both its sheets behaves as a normal object with an energy E, momentum $P < E/c$ given by Equation (105) and a rest mass given by Equation (106). And yet we will take the small advantage of having a tiny region around the apex moving with a superluminal speed, and introduce, also formally, but according to the already known general definition, the energy $\tilde{E}$ and momentum $\tilde{P}$ for the object moving with this speed:

$$\tilde{E} = i\,m_0\,c^2\,\gamma(\tilde{v})\,, \quad \tilde{P} = i\,m_0\,\tilde{v}\,\gamma(\tilde{v}) \tag{114}$$

Here the role of m_0 is played by the rest mass of the cone, and the role of $\tilde{v}$ is played by the speed of its apex. By introducing the new characteristics in Equation (114) specific for a tachyon, we assign to this tachyon the mass m_0. Thus the cone and the associated tachyon are both characterized by the same rest mass, and their speeds,

according to Equations (104) and (108) satisfy the condition $v\tilde{v} = c^2$. Recall that if a tardyon has a counterpart with the same mass on the other side of the light barrier, whose speed satisfies this condition, the two particles are mutually dual. Their 4-momenta must be mutually perpendicular. Let us check if the energy–momentum in Equation (114) is perpendicular to the energy–momentum of the cone. Using the relations $v = Pc^2/E$, and $\tilde{v} = \tilde{P}c^2/\tilde{E}$, we have

$$0 = v\tilde{v} - c^2 = \frac{P\tilde{P}}{E\tilde{E}}\,c^4 - c^2 \tag{115}$$

It follows that

$$E\tilde{E} - P\tilde{P}c^2 = 0 \tag{116}$$

Yes! The energy–momentum vectors are equal in magnitude (m_0) and mutually perpendicular. And, owing to the invariance of a scalar product, this property holds in *all* inertial reference frames.

Thus, the introduction of apparently irrelevant characteristics [Eq. (114)] unexpectedly provides us with an additional consistent description of the same object. Having emerged as a tardyon, it now appears to have more and more properties of a tachyon.

We can push the analogy a little further. Let us write down another scalar product – for the energy–momentum of the tachyon and that of a photon in the Cerenkov cone (Fig. 8.20):

$$\frac{\tilde{E}}{c}\omega - \tilde{\boldsymbol{P}}\boldsymbol{k} = \frac{\tilde{E}}{c}\omega - \tilde{P}k\cos\theta = im_0\,\gamma(\tilde{v})\,\omega\left(1 - \frac{\tilde{v}}{c}\cos\theta\right) \tag{117}$$

Using again the invariance of the scalar products, we can write for two arbitrary reference frames

$$\gamma(\tilde{v})\,\omega\left(1 - \frac{\tilde{v}}{c}\cos\theta\right) = \gamma(\tilde{v}')\,\omega'\left(1 - \frac{\tilde{v}'}{c}\cos\theta'\right) \tag{118}$$

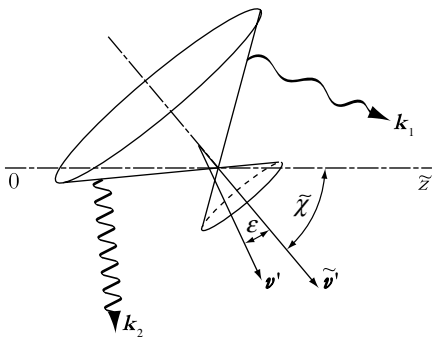


Fig. 8.20

But we know [look at Equation (104)!] that at least in one reference frame the quantity $1 - \frac{\tilde{v}}{c}\cos\theta = 0$. Therefore, the whole scalar product must be zero, and once this takes place in one reference frame, in must hold in all of them! Thus, we can write

$$\tilde{v}\cos\theta = \tilde{v}'\cos\theta' = c \tag{119}$$

This means that the picture of the superluminal lump at the apex of the circular cone connected with it is Lorentz-invariant, in complete analogy with CRT studied in the previous section.

At the same time in the systems K moving perpendicularly to the z-axis, the axial symmetry of the *electromagnetic radiation* in the cone breaks down. Indeed, as we perform the transition to such a system, the velocities of the cone and its apex are turned through different angles χ and $\tilde{\chi}$, for which $\sin\chi = V/v$ and $\sin\tilde{\chi} = V/\tilde{v}$. The misalignment between the new directions $\boldsymbol{v}$ and $\tilde{\boldsymbol{v}}$ is caused by the Doppler effect which produces the asymmetry of the spectrum: while the cone of radiation around the direction of velocity retains its circular form, the radiation frequency and intensity begin to depend on the azimuth angle φ (different "generatrices" of the cone acquire different "weight", see Figure 8.20). This leads to different trajectories of the cone apex and its center of mass.

The resulting asymmetry is exactly the same as that of CRT in Section 8.10, because in both cases we have the same cone of radiation.

Thus, all considered kinematic relations admit the "tachyonic" interpretation of the cone. This allows us to speak about the duality of its description. Depending on whether we choose the center of mass velocity $\boldsymbol{v}$ or the apex velocity $\tilde{\boldsymbol{v}}$ as a characteristic property of the electromagnetic cone, we can accordingly describe it as either a regular tardyon with the 4-momentum $(E/c, \boldsymbol{P})$, or as a dual tachyon with the 4-momentum $(\tilde{E}/c, \tilde{\boldsymbol{P}})$. Here again it would be relevant to recall the purely theoretical possibility discussed above – for a subliminal system with a rest mass m_0 to have a "ghost" partner in another world beyond the light barrier – the dual superluminal object of the same mass. Well, now we can demonstrate a simple home-made or laboratory-made model where *both* dual partners are not only damn real, but are both enrolled in one!

There still remain questions about the analogy drawn between the cone and conventional tachyon. Here is one: it is more or less O.K. when we choose to consider the superluminal lump as a tachyon, and the trailing cone as its Mach cone of Cerenkov radiation. But how should we interpret the *leading* cone that we have? In our system, the energy "radiated" through the trailing cone, is "supplied" from the leading one. This is not what one would expect to observe if "true" tachyons were discovered, and the difference may appear to preclude us from identifying fully our model with a "true" tachyon. But such a conclusion is not necessarily correct. We know already that a picture of a superluminal particle with trailing Mach cone is not the only model of a tachyon. Recall another alternative – the model proposed by Recami and co-workers [72]. In their model a tachyon is a "rigid" two-folded structure as shown in Figure 8.13. The trailing fold is geometrically similar to the Mach cone for the par-

ticle with the given speed and size (determined by its rest mass). The leading cone looks like a mirror reflection of the trailing one. The whole system thus looks like a doubled Mach cone (with the hyperboloidal cavities in each), and one might ask how we should interpret the extra fold? The answer to this is that the sheets are not Cerenkov radiation. They are similar to a Mach cone in shape, but physically different (see the previous section).

Similarly, our electromagnetic cone, even though it is made of Cerenkov radiation once kidnapped from a real charge, does not generate from the apex in its current afterlife. Rather, the apex itself should now be considered as a special point in the formed radiation structure.

We can therefore suggest this structure as a realization of the theoretical model proposed by Recami and co-workers. The existence and physical properties of the structure give their model additional credibility. But we must bear in mind that our structure is not totally identical with their model either. Whereas their model is always axially symmetrical, the real cone here still suffers the symmetry break – not in its geometrical shape, but in its physical properties described above. The reason for this is that it consists of waves with non-zero frequencies. Another distinction is that EMT is finite in space and time, whereas the original model [72] is infinite. However, after all, the EMT is *a real object*, and it might drop some hints at how we could refine our theoretical models of tachyon!

The most important feature of EMT is that we cannot use it for superluminal signaling! Recall the simple experiment described in Section 6.14. We have the same situation here. A mere attempt to pass the lump through an aperture to specify the trajectory along which we want to transport the signal destroys the whole structure. Nor can we "shoot down" our tachyon by affecting in any way the instantaneous lump at the apex. Suppose, for instance, that we want to block its way by putting a small absorbing screen at point z (Fig. 8.21). The screen would indeed absorb all the energy of the lump, but it will not prevent the tachyon from propagating farther down the same path. The energy at point $z + \Delta z$ further down the way comes from the section Δl of the leading cone, not from the point z. Thus, we cannot transport the tachyon along any chosen path by opening it, and we cannot prevent it from moving along this path by closing it!

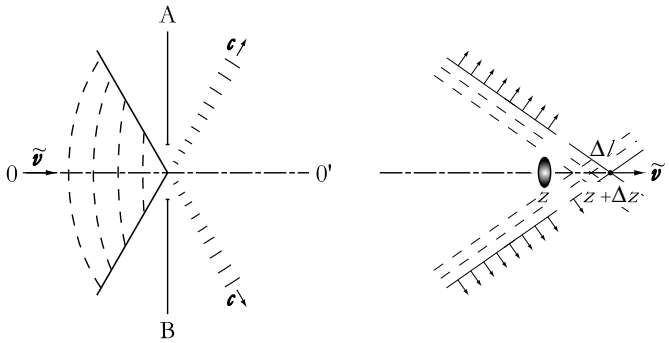


Fig. 8.21

The lump does not carry signal or energy along its trajectory. It is not a self-identical entity keeping memory of its previous self. It is a continuously decaying and renewing entity having no memory of its past or any anticipation of its future.

The EMT is not precisely the original exotic concept of a tachyon, getting us (and the whole Universe) into trouble, puzzles, and paradoxes. It offers none of the embarrassing situations with signal transfer from the future, world line loops, and tachyon shooting between mirrors, and in this respect it is as good as non-existing. And yet, here it is, a real self-dual tachyon–tardyon object, the double-faced Janus! And both of its faces are fully and consistently described by the special theory of relativity.

References

1 C. W. Misner, K. S. Thorne, J. A. Wheeler: *Gravitation*, Freeman, San Francisco, 1973.

2 N. Herbert: *Faster Than Light*, New American Library, New York, 1988.

3 J. Marangos, *Nature*, **2000**, *406*, 243–244.

4 L. G. Wang, A. Kuzmich, A. Dogarin, *Nature*, **2000**, *406*, 277–279.

5 C. Liu, Z. Dutton, C. H. Behroozi, L. V. Hau, *Nature*, **2001**, *409*, 490–493.

6 D. F. Phillips, A. Fleischhauer, A. Mair, R. L. Walsworth, *Phys. Rev. Lett.*, in press.

7 R. E. Slusher: Self-induced transparency, in *Progress in Optics*, ed. E. Wolf, Amsterdam, 1974, Band XII.

8 I. Poluektov, Yu. Popov, V. S. Roitberg, *Sov. Phys., Usp.*, **1974**, *114*(2), 97.

9 N. G. Basov, R. V. Ambartsumyan, V. S. Zuev, P. G. Krykov, V. S. Letokhov, *Sov. Phys., JETP*, **1966**, *23*(1), 16–22.

10 P. G. Kroukov, V. S. Letokhov, *Sov. Phys., Usp.*, **1969**, *99*(2), 169–227.

11 R. W. Chiao, *Phys. Rev. A*, **1993**, *48*, 34–37.

12 K. Frasier, *Skeptical Inquirer*, **1998**, *22*(2), 37.

13 V. A. Fock: *The Theory of Space, Time, Gravitation*, Pergamon Press, Oxford, 1958.

14 E. Taylor, J. A. Wheeler: *Spacetime Physics*, Freeman, San Francisco, 1966.

15 P. G. Bergmann: *Introduction to the Theory of Relativity*, Prentice-Hall, Englewood Cliffs, NJ, 1942, Ch. 4, 5.

16 L. Landau, I. Lifshits: *The Field Theory*, Butterworth-Heinemann, 1980.

17 J. D. Fix: *Astronomy. Journey to the Cosmic Frontier*, WCB/McGraw-Hill, New York, 1997, 148–149.

18 H. P. Robertson, *Rev. Mod. Phys.*, **1949**, *21*, 374.

19 E. M. Dewan, *Am. J. Phys.*, **1963**, *31*, 383–386.

20 G. Cavalleri, G. Spinelli, *Nuovo Cimento B*, **1970**, *66*(1), 11–20.

21 Ø. Crøn, *Am. J. Phys.*, **1977**, *45*(1), 65–70.

22 H. Nicolic, *Am. J. Phys.*, **1999**, *67*(11), 1007–1012.

23 J. D. Jackson: *Classical Electrodynamics*, 2nd edition, John Wiley & Sons, New York, 1975, 554–556.

24 P. Tipler: *Elementary Modern Physics*, Worth Publishers, New York, 1992.

25 K. Krane: *Modern Physics*, John Wiley & Sons, New York, 1992.

26 P. G. Hewitt: *Conceptual Physics*, Harper Collins College Publishers, New York, 1992.

27 J. C. Hafele, *Am. J. Phys.*, **1972**, *40*, 81–85.

28 J. C. Hafele, R. E. Keating, *Science*, **1972**, 166–170.

29 Hafele, Keating.

30 Ya. B. Zeldovich, I. D. Novikov: *The Theory of Gravitation and Stellar Evolution*, Nauka, Moscow, 1971, Sects 3 and 4.

31 Sagnac exp.

32 Sagnac exp.

33 K. Thorne: *Black Holes and Time Warps, Einstein's Outrages Legacy*, Norton, New York, 1994.

34 Y. Sofue, M. Fudsimoto, *Astrophys. J.*, **1983**, *265*(2), 722–729.

35 A. Hewish, J. S. Bell, D. J. Pilkington, F. P. Scott, A. R. Collins, *Nature*, **1968**, *217*, 709.

Special Relativity and Motions Faster than Light. Moses Fayngold

ISBN: 3-527-40344-2

36 V. L. Ginzburg, *Sov. Phys., Usp.*, **1968**,.

37 V. L. Ginzburg,V. V. Zheleznyakov, V. V. Zaitsev, *Sov. Phys., Usp., 12*(3), 378–398.

38 A. Sommerfeld, *Gotting. Nachr.*, **1904**, *99*, 201.

39 P. Cerenkov, I. Tamm, I. Frank, *Sov. Phys., Usp.*, **1958**, *68*, 376–415.

40 J. V. Jelley: *Cerenkov Radiation and Its Applications,* Pergamon Press, Oxford, 1958.

41 B. Bolotovsky,V. Ginzburg, *Sov. Phys., Usp.*, **1972**, *106*(4), 577–592.

42 L. S. Rodberg,V. F. Weisskopf, *Science*, **1957**, *125*, 627.

43 I. Shapiro.

44 M. Rees.

45 M. Fayngold: On a Superluminal Spot of the Type of a Moving Focus, in *The Einstein Collection*, Nauka, Moscow, 1976, 276.

46 L. Landau, E. Lifshits: *Electrodynamics of Continuous Media*, Nauka, Moscow, 1982, 80.

47 D. Bohm: *Quantum Mechanics,* Prentice-Hall, New York, 1952.

48 A. Aspect, P. Grangier, G. Roger, *Phys. Rev. Lett.*, **1982**, *49*, 91–94.

49 G. F. Clauser, A. Shimony, *Rep. Prog. Phys.*, **1978**, *41*, 1881–1927.

50 G. Greenstein, A. G. Zajonc: *The Quantum Challenge*, Jones and Bartlett, Sudbury, MA, 1997.

51 M. P. Silverman: *More Than One Mystery. Explorations in Quantum Interference*, Springer, New York, 1994.

52 A. Einstein, B. Podolsky, N. Rosen, *Phys. Rev.*, **1935**, *47*, 777–780.

53 L. V. Hau et al., *Nature*, **1999**, *397*, 594.

54 S. E. Harris, *Phys. Today*, **1997**, *50*(7), 36–42.

55 M. Schubert,l B. Wilhelmi, *Einführung in die Nichtlineare Optik,* Teil II, Teubner, Leipzig, 1978.

56 L. Carroll, *Alice's Adventures in Wonderland,* Easton Press, Norwalk, CT, 1977.

57 L. Carroll, *Through the Looking Glass,*.

58 R. Mignani, E. Recami, *Nuovo Cimento A*, **1973**, *14*, 169; *16*, 208.

59 Camill Flammarion.

60 O. M. Bilaniuk,V. K. Deshpande, E. C. G. Sudarshan, *Am. J. Phys.*, **1962**, *30*(10), 718–723.

61 O. M. Bilaniuk, E. C. G. Sudarshan, *Phys. Today*, **1969**, *22*, 43.

62 J. Feinberg, *Phys. Rev.*, **1967**, *159*(5), 1089–1105.

63 R. C. Tolman: *Relativity, Thermodynamics, Cosmology,* Oxford, 1934, 205.

64 G. A. Benford, D. L. Book,W. A. Newcomb, *Phys. Rev. D*, **1970**, *2*(2), 263.

65 R. Fox, C. G. Cuper, S. G. Lipson, *Proc. R. Soc. London, ser. A*, **1970**, *316*, 515–524.

66 H. K. Wimmel, *Natur*e Phys. Sci., **1972**, *236*(66), 79–80.

67 F. C. Jones, *Phys. Rev. D*, **1972**, *6*(10), 2727–2735.

68 M. Fayngold, *Sov. Phys. Teor. Mater. Fiz.*, **1981**, *46*(6), 395–406.

69 M. Fayngold, *Sov. Phys., Ukr. Fiz. Zh.*, **1982**, *27*(3), 440–442.

70 A. M. Perelomov, *Sov. Phys., Dokl.*, **1962**, *146*, 75.

71 A. Peres, *Phys. Lett. A*, **1970**, *31*(7), 361–362.

72 A. O. Barut, G. D. Maccarone, E. Recami, *Preprints of Nuclear Physics Institute,* Catania, Italy, INFN/AE, June **1982**.

73 M. Fayngold, *Sov. Phys., Radiofiz.*, **1979**, *22*(5), 531–541.

74 M. Fayngold, *Sov. Phys., Tech. Phys.*, **1980**, *25*(5), 552–556.

Index

Special Relativity and Motions Faster than Light. Moses Fayngold

ISBN: 3-527-40344-2