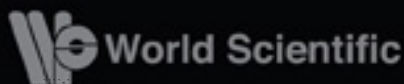


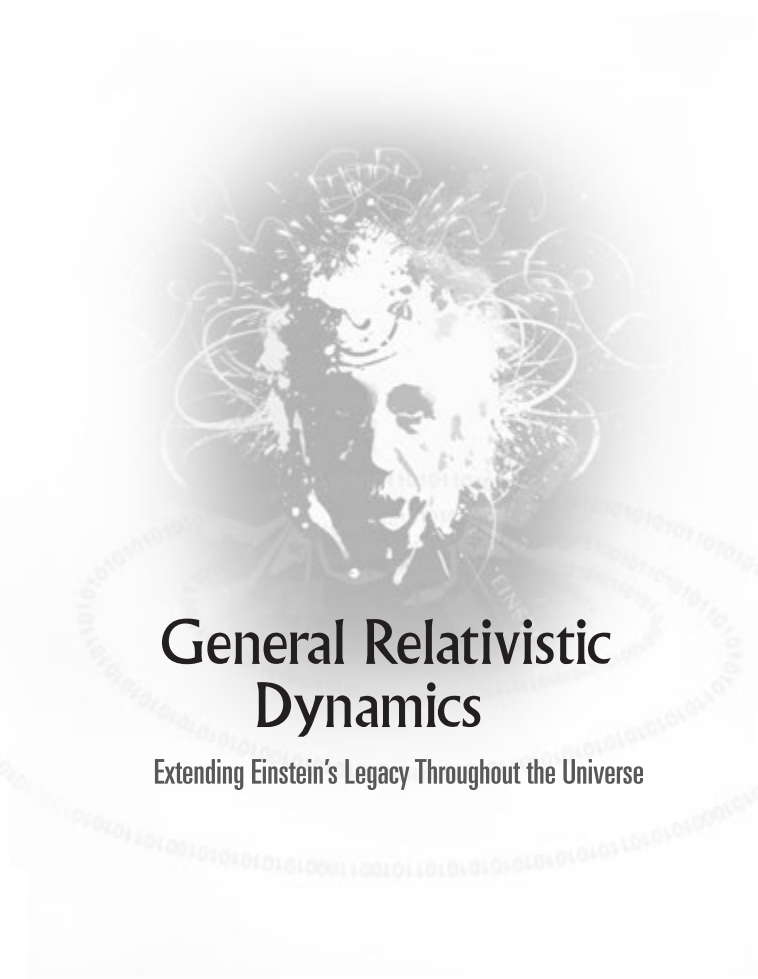


General Relativistic Dynamics

Extending Einstein's Legacy Throughout the Universe

Fred I. Cooperstock



A stylized portrait of Albert Einstein is centered in the upper half of the page. The portrait is rendered in a high-contrast, almost binary style, with his hair and facial features appearing as dark, textured shapes against a lighter background. Behind the portrait, there is a faint, glowing spiral pattern that resembles a galaxy or a gravitational well. The background of the entire page is a light gray with a subtle, repeating pattern of binary code (0s and 1s) arranged in a grid-like fashion.

General Relativistic Dynamics

Extending Einstein's Legacy Throughout the Universe

This page intentionally left blank



General Relativistic Dynamics

Extending Einstein's Legacy Throughout the Universe

Fred I. Cooperstock

University of Victoria, Canada

 **World Scientific**

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG

Copyright World Scientific Publishing Co.

Provided by IHS Markit under license with WILEY

No reproduction or networking permitted without license from IHS

Licensee=, User=,

Not for Resale,

Full document details: WSPC GENRL RELAT DYNA 2009



Documents

Full Document Hub

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

Cooperstock, F. (Fred)

General relativistic dynamics : extending Einstein's legacy throughout the universe /
Fred I Cooperstock.

p. cm.

Includes bibliographical references and index.

ISBN-13: 978-981-4271-16-5 (hardcover : alk. paper)

ISBN-10: 981-4271-16-0 (hardcover : alk. paper)

1. General relativity (Physics) 2. Gravity. 3. Gravitational fields. 4. Galaxies. I. Title.

QC173.6.C67 2009

530.11--dc22

2009012008

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Einstein illustration by *Travis Morgan*, morgantj@gmail.com

Galaxy illustration by *Robert Gendler*, www.robgendlerastropics.com

Copyright © 2009 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

Printed in Singapore.

In memory of Sima, Tevya and Aryeh

This page intentionally left blank

Contents

Chapter 1 Introduction	1
Chapter 2 Essentials of Special Relativity	11
2.1 Basic principles	11
2.2 The spacetime interval and the Lorentz transformation	13
2.3 Lorentz contraction and time dilation	15
2.4 Causality	17
2.5 Transformation of velocity and the aberration of light	19
2.6 Four-vectors and four-tensors	21
2.7 Special relativistic dynamics	23
2.8 Relativistic Doppler shift	26
Chapter 3 Bondi’s k-Calculus Approach to Special Relativity	31
3.1 Introduction	31
3.2 Velocity–Doppler factor connection	33
3.3 Composition law for velocities and Doppler factors . .	35
3.4 Derivation of the Lorentz transformation	38
3.5 The twin or clock paradox	40
Chapter 4 Essentials of General Relativity	47
4.1 The need for a new theory of gravity	47
4.2 The Principle of Equivalence	48
4.3 The metric tensor	49
4.4 Basic tensor calculus–introduction	50
4.5 Parallel transport, spacetime curvature and the Riemann tensor	54
4.6 Geodesics	56

4.7	Covariant conservation laws and the Einstein field equations	57
4.8	Einstein–Maxwell equations and motion of a charged body in general relativity	61
4.9	Summary of the steps from Newtonian gravity to Einstein’s general relativity	62
Chapter 5 Schwarzschild Solution and its Consequences		65
5.1	The metric	65
5.2	The measurement of distance and time in general relativity	66
5.3	The event horizon, black holes and singularities	68
5.4	The tests of general relativity	78
Chapter 6 Gravitational Waves		83
6.1	Introduction	83
6.2	Linearized field equations	83
6.3	The energy issue and the pseudotensor	85
6.4	The energy localization hypothesis	86
Chapter 7 The Normal Scales of Physics and the Planck Scale		93
7.1	The hierarchy of scales	93
7.2	The fundamental interactions of nature	95
7.3	The Planck scale and the issue of the quantization of gravity	96
7.4	Adding spin and charge to the Planck scale	97
7.5	Quantum limits, spectra, the value of α	100
Chapter 8 General Relativistic Cosmology		103
8.1	Sizes of astronomical elements	103
8.2	Early ideas about cosmology	104
8.3	Friedmann universes	108
8.4	The cosmological term	110
Chapter 9 Motion of the Stars in the Galaxy		115
9.1	Introduction	115
9.2	General relativistic effects on the stellar motions in galaxies	117

9.3 Modeling the observed galactic rotation curves 121

9.4 A velocity dispersion test for the presence of extra
matter 130

9.5 Summary comments on rotation velocities of galaxies . 131

Chapter 10 Clusters of Galaxies 135

10.1 Preliminary comments 135

10.2 Spherical dust collapse 136

10.3 Velocity of particles falling in vacuum toward a
spherical concentration of mass 138

10.4 The velocity of dust in collapse 142

10.5 Observing an idealized galactic cluster 147

10.6 Current evidence for dark matter 154

Chapter 11 Closed Timelike Curves and Time Machines161

11.1 The background 161

11.2 Creating closed timelike curves and Gödel’s spacetime 163

11.3 Re-examining the standard closed timelike curve
interpretation 165

11.4 The role of our experience in nature 171

11.5 Gott’s moving cosmic strings 172

Chapter 12 The Direction of Physics Research 179

Chapter 13 Summary with Concluding Commentary 187

Appendix A Critical Challenges and Our Replies 195

Appendix B Radial Velocity Derivation Details 213

Bibliography 217

Acknowledgements 225

Index 227

Copyrighted material

*... and the slavery of fear had made men afraid to think.
But such is the irresistible nature of truth, that all it asks,
and all it wants, is the liberty of appearing.*

Thomas Paine in “The Rights of Man” [1]

*The only justification for our concepts and systems of
concepts is that they serve to represent the complex of
our experiences; beyond this, they have no legitimacy.*

Albert Einstein in “The Meaning of Relativity” [2]

Chapter 1

Introduction

One often reads about the beauty and importance of general relativity (GR). It is viewed almost universally as our premier theory of gravity, superseding Newton's theory of gravity in a profound way. We share this view whole-heartedly. However, while there have been a very large number of papers written in the field, it is disconcerting to note how little there is in the way of a connection of this large accumulated effort with the great body of the world's physicists and astronomers. Many of these papers are of great mathematical complexity, decipherable only by the experts in the field. Even when decoded, many, if not most, have little relevance to the important issues of physics and astronomy. This is unfortunate as gravity is assuming a greater role of significance in science and the best theory of gravity should be readily accessible to the non-experts.

Our goal in this book is multi-fold: first, we wish to present the essentials of general relativity in a simple way so that any physicist who might have missed training in the field, once having digested the primary concepts and equations, will not have his or her eyes glaze over when confronted by some reference to or equation for general relativity. To do so, we will also cover the essentials of special relativity to provide a smooth transition to the general theory. For a detailed development of special relativity, the reader is directed to the truly classic work of L. D. Landau and E. M. Lifshitz [3] and the very intuitive approach of H. Bondi [4]. The special relativistic treatment in this book has been particularly influenced by these excellent thinkers.

The reader who is comfortable with the standard development of special relativity but who is unfamiliar with Bondi's intuitively appealing approach might wish to start with Chapter 3. In the standard approach, we discuss how the physical equivalence of all inertial reference frames plus the experimental result that the velocity of light is an invariant form the foundation of special relativity, Einstein's theory of space and time in the absence of gravity. These two cornerstones show us that the old transformations of the space-time coordinates with the Cartesian coordinates with which we are familiar, no longer hold and that time is no longer an absolute. The correct transformations using Cartesian coordinates preserve both the value and the form of the spacetime interval, the important measure linking time and space that will lead us into general relativity. Most importantly, we will show that the lengths of bodies and the intervals of time are seen to vary in remarkable ways when viewed in different frames of reference. We will get our first taste of the importance of extrema in physics, how the minimum possible time interval between the ticks of a clock is read in the rest frame of that clock (the "proper time interval") whereas the maximum possible length of a body is the length that is measured in the rest frame of that body (the "proper length"). Being extrema of opposite sense underlines the reciprocal nature of time and space.

In the development of special relativity, we have our first contact with the vectors and tensors of four-dimensional spacetime. These mathematical structures are used with the important Principle of Least Action, which forms the basis of the fundamental equations of physics and relativistic dynamics.

In Chapter 3, Bondi's approach to special relativity builds upon the Doppler factor between two observers in relative motion. Some of the basic results of special relativity are re-derived using Bondi's simple and appealing framework. It enables us to resolve the so-called twin or clock paradox, the asymmetric aging of twins who reunite after a voyage of separation. We will see that there is nothing mysterious that occurs at the turnaround point in the voyage of the accelerating twin. Different paths in spacetime track different spans of time in analogy with the different paths in space which track different spans of distance.

Having covered the essentials of special relativity, we proceed into

the next major step in Chapter 4, the development of general relativity, Einstein's theory of gravity. (The reader who is comfortable with the basics of general relativity might wish to proceed to Chapter 5 or 6, depending upon his/her familiarity with the subject.) We first discuss the Principle of Equivalence, the local equivalence between accelerated reference frames and gravitational fields that was the guiding light for Einstein in his quest for a relativistic theory of gravity. We also focus on the importance of the approximate aspect of the Equivalence Principle, how it is spacetime curvature rather than accelerated reference frames that constitutes true gravity, a point well-articulated by J. L. Synge. That being said, the usefulness of the Equivalence Principle remains, as illustrated in the lead-in to the spacetime metric tensor as describing a gravitational field.

We then proceed with the basic mathematics, tensor calculus, that is required for technical work in general relativity. (Depending upon the degree to which the reader may wish to follow technical aspects, he or she may wish to skim over the sections that follow in Chapter 4 and then concentrate more carefully beginning with Chapter 5.) With the basic aspects of general coordinate transformations covered, we proceed to explore the nature of curved spaces, introducing the important Riemann tensor, which characterizes gravity in an invariant manner. We then focus upon a key GR departure from Newtonian gravity, the removal of gravity from the category of forces. A freely gravitating body is seen to move with zero intrinsic acceleration, following the extremal paths, the geodesics of the spacetime that the body occupies. Motion under gravity as a force is replaced with free motion following the special paths in curved spacetime in analogy with airline pilots who follow the geodesics, the great circles on the globe, to minimize distance between two points.

The energy and momentum conservation laws of special relativity are first generalized to arbitrary coordinate systems using the new generalized derivative, the "covariant derivative". Guided by the Principle of Equivalence, we are led to the conservation laws for general relativity. Consistency with these laws and the demand for melding with Newtonian gravity under appropriate conditions brings us to the Einstein tensor. This incorporates gravity on the left hand side of the Einstein field equations to equal the source, the energy-

momentum tensor, on the right hand side. Having established the Einstein field equations with the basic background structures and concepts, we are prepared to study the relativistic world of gravity.

We first study the simple important and very interesting Schwarzschild solution of the Einstein equations in Chapter 5, the gravitational field in vacuum under the condition of spherical symmetry. To do so, we require the concepts of proper distance and proper time in general metrics, including gravity. The study of the Schwarzschild spacetime introduces interesting issues, event horizons, black holes and singularities, issues that have ignited the imagination of the general public for decades. As compared to special relativity, there is a paucity of experimental corroboration for the theory of general relativity. We outline the various issues regarding these tests.

As there are waves of electromagnetic nature emanating from the acceleration of charges, Einstein's general relativity predicts that there must be waves of gravitational nature arising from the acceleration of masses. In Chapter 6, we discuss these waves briefly. Gravity waves have never yet been observed directly but their existence is inferred from the motion of certain sources such as the binary pulsar, PSR1913+16. However, issues concerning energy and momentum for such waves are not clear-cut. Energy localization has been an enduring controversy in general relativity. We will bring forward our hypothesis that energy, including the contribution from gravity, is most logically localized in the regions of the energy-momentum tensor. This has the unsettling implication that gravitational waves, assuming the reality of their existence to which we certainly subscribe, are not carriers of energy in vacuum.

Part of the natural appeal of physics is that it encompasses all scales of dimension, from the very tiniest size, a size so small as to challenge the imagination, to the universe itself which may in fact be infinite. In Chapter 7, we wind our way through the hierarchy of scales in nature, connecting them to the fundamental forces. We know that the macroscopic physics of our everyday experience breaks down completely when we reach the quantum scale of atomic physics at dimension 10^{-10} m. At this level and smaller, the world as we know it is gone. New forces come into play, the "strong force" binding the nucleus and the "weak force" responsible for the decay of particles such as the neutron. Even the familiar electromagnetic force from

macroscopic physics must be “quantized”, taking on a new character.

It has been almost universally assumed that gravity must also be quantized at a certain stage. We have argued that this is not necessarily the case since gravity is fundamentally different: all particles and fields other than gravity exist *within* spacetime whereas gravity *is* spacetime, i.e. its curvature. In our view, this aspect sets gravity apart as the enveloper of all the rest of physics and removes the necessity for its quantization. However, it is well to ask whether gravity might play a non-classical role at a scale that arises from equating the Compton wavelength of a particle where a quantum duality sets in, to the dimension of contraction at which a body exhibits an event horizon, bringing into play the full nonlinearity of general relativity. This occurs at the Planck scale, of dimension 10^{-35} m. While we can look at this number with its long chain of zeros, we cannot begin to visualize it as an actual length in any normal sense.

In his earliest work in atomic physics, Bohr had set into motion the quantization of the hydrogen atom with *ad hoc* rules that were remarkably successful for their time. We were inspired by the work of Bohr to attempt an *ad hoc* quantization at the Planck scale, adding spin and charge to the Planck mass.

Interestingly, at the extreme of Planck quantum states, a new level of the dimensionless fine structure constant α arises, namely $1/128$ as opposed to the approximate value of $1/137$ of atomic physics. The $1/128$ value has a serendipitous aspect as this is almost precisely the α value governing high energy radiation in Z-boson production and decay. We know of no particular reason for these numbers to match. Perhaps it is a sheer coincidence. However, R. P. Feynman used to impress upon us that it is the confluence of numbers that can foreshadow important truths in physics.

The preceding focused upon the smallest of scales. Proceeding to the very largest series of scales beginning in Chapter 8, we first take a brief excursion into cosmology, the very largest scale in nature. We provide some perspective by building the image of the vast dimensions that we will encounter by the use of scaling, reducing the size of the Sun to that of the head of a pin. From there, we can better picture the vastness of empty space between the stars and then the immensity of a galaxy. We will discuss some of the early ideas about the cosmos and how the modern picture developed.

There has been much recent interest in the idea that the universe is currently in a state of *accelerated* expansion. We will discuss this aspect briefly in conjunction with the cosmological term in the Einstein equations. We will argue that this term should be viewed most logically as another form of matter, albeit exotic, and not a “geometrical” adjunct to the theory.

As a primary goal, in Chapters 9 and 10, we will show why general relativity must be brought into the greater sweep of dynamical problems in the universe. These entail the motions of stars in the galaxies and the motions of galaxies within clusters. Until now, it was believed that Newtonian gravity was the appropriate theory for these scales and that general relativity came into significant play only in situations of ultra-strong (or at least very strong) gravity or for the dynamics of the universe as a whole. This is a bizarre view: overall, the gravity of the universe is weak. Since general relativity is seen as necessary to describe this largest scale, why would one expect GR to be unnecessary for the second and third largest scales in nature, those of the clusters of galaxies and of the galaxies themselves? To this point, when one encountered the expression “general relativistic dynamics”, the reference was to those very special situations in nature such as the case of a closely orbiting pair of neutron stars where very strong gravitational fields and very high velocities prevailed. These cases are certainly very interesting but they are of very limited range. It is worth repeating: the new reality to be faced is that general relativity reaches into the dynamics of essentially all of the key basic building blocks in nature, the arrays of the billions of stars in the galaxies and the clusters of the galaxies themselves. In the vast majority of these cases, the gravity is not very strong and the velocities are not very high by standard relativity measure.

The importance in following this new path is immediate: without general relativity, one is left with serious issues first brought to bear from the work of F. Zwicky and V. Rubin in having to account for the higher-than-expected velocities of stars within galaxies and of galaxies within clusters, velocities having to be rationalized on the basis of Newtonian gravity. This had led to the belief that the normal baryonic matter that we see is but a small fraction of the total matter in the universe, that there is an immense quantity of so-called “dark matter” that is required to drive these anomalous velocities. This

matter was seen to reveal its existence solely by gravitation. Clearly if the gravitational laws that underpin this belief are removed, the paradigm shifts. Many would argue that the galactic motions are no longer the main reasons for belief in dark matter, that the primary reason is now the need for the extra matter to quickly form the galaxies and their clusters shortly after matter decoupled from radiation in the early universe. While early universe studies are the glamorous focus of current interest, some perspective is useful here: progress in early universe study has been impressive, but firm pronouncements as to what is required to explain COBE and WMAPs strike us as unjustified. Alternative scenarios should be, and surely will be explored in the years to come.

In Chapter 9, we will describe the paradigm shift due to general relativity. We will describe how S. Tieu and the author accounted for the high velocities of stars in spiral galaxies by the application of general relativity and without the requirement for the vast stores of exotic dark matter as is the case when Newtonian gravity is taken as the underlying theory of gravity. This raises the phenomenon of *general* relativistic velocity to a level of fundamental importance as an observational tool for situations of weak gravity and velocities far below that of light, domains previously reserved for Newtonian gravity. An essential point is this: where general relativity gives a result for a given physical system different from that provided by Newtonian gravity, we choose the general relativistic result. There is nothing controversial about this choice—the great majority of the world’s physicists would do likewise as general relativity is regarded as our premier theory of gravity. The issue is the generally prevailing belief that the galactic systems that we studied should have given the same descriptions with the applications of both Newtonian gravity and general relativity. While our approach and departure from expectations have received a great deal of attention world-wide in the media and by many physicists and astronomers in hundreds of communications, it has also been opposed by an interesting variety of critics. We have answered their critiques in detailed papers and in Appendix A, we present a simpler account of both the nature of the criticisms and our replies to them.

In Chapter 10, we will then proceed to the next size scale to present an account of how we can rationalize the relatively large ve-

locities of galaxies within clusters without the aid of dark matter, again using general relativity. We will focus on our study of the spherical collapse under gravity of a pressureless ball of fluid as an idealized model of a freely gravitating cluster of galaxies. We will present the contrast between the local and asymptotic measures of the velocities of the particles, the former that would be perceived by observers in the vicinity of the galaxies and the latter that astronomers would actually perceive over the vast distances separating us from the populations of galaxies under study. It is well understood that even in classical physics, velocity is an observer-dependent quantity. In special relativity, that dependence assumes a more complicated form and it is even more complicated in general relativity. While these aspects of velocity have been understood, they have been under-appreciated in the case of general relativity. Our approach in taking these aspects into careful account leads to an alternative to dark matter as an explanation for the high galactic velocities seen in clusters of galaxies.

A key theme of this book is one of repositioning general relativistic as opposed to Newtonian dynamics as an essential tool for the complete description of the motions of bodies under gravity. Our work with Tieu on galactic dynamics focuses on the hitherto unjustified neglect of the application of general relativity and the interesting effects its incorporation produces. By contrast, in Chapter 11 we will also discuss our work on the *misuse* of general relativistic dynamics in the notion of “closed timelike curves”, often interpreted as time machines, and how they can be seen most logically as mere mathematical rather than physical constructs.

Finally, in Chapter 12 we will discuss the overall direction of current theoretical physics research, its fixation on the need for unification of the fundamental forces with gravity. We will argue that there are reasons for regarding gravity as fundamentally different from the other forces in nature, in fact that it is not a force at all. Rather than the standard forces that serve as mechanisms coupling objects to each other as ingredients *within* spacetime, gravity is a property of spacetime itself, its curvature. This is the essence of gravity. While most relativists, if pressed, would acknowledge this to be the case, the focus has been lost (or never fully appreciated) by many. One aim of this book is to restore the realization of gravity’s essence.

This book will have accomplished its mission if readers will appreciate the wonders that Einstein has brought to our recognition of relativity in a now much broader framework as one of the cornerstones of modern physics.

.....

This page intentionally left blank

Chapter 2

Essentials of Special Relativity

2.1 Basic principles

Special relativity is Einstein's theory of space and time in the absence of gravity. It has at its base, Newton's principle of relativity which states that physics is the same in all inertial reference frames. An inertial reference frame is one in which a body moves with a constant velocity in the absence of an unbalanced force.

As well, special relativity incorporates the experimental fact that there exists a maximum velocity for the propagation of a signal in nature, namely c , the speed of light in vacuum. Simply put: there is a speed limit for interactions in nature.^a Given this experimental fact and given that all inertial frames are physically equivalent, it follows that the speed of light in vacuum (the speed limit) must be the same, i.e. c , in all such frames. This is certainly counter-intuitive: if A throws a ball to B at speed V , B will measure the ball speed as V if B is at rest relative to A but B will measure a slower speed if B is running away from A and a faster speed if B is running

^aThe existence of a maximum speed for the propagation of interactions leads to the conclusion that there can be no truly rigid bodies in nature. For example, if one were to believe that a steel rod is truly rigid, a push leading to motion at one end would require an *instantaneous* pressure wave to flow to every other part so that each part would move in step. However such an infinite speed pressure wave does not exist and hence there is some buckling of the rod, however small [3]. Note the discussion of how close one can come towards perfect rigidity in [5].

towards A. If photons, the quanta of light, behaved like ordinary macroscopic balls, the velocity measurements for them would follow in the same manner. However, what works for balls does not work for light. We must constantly remind ourselves, as Feynman had done with such gusto, that in physics, it is the experiment that is the ultimate arbiter. While the injection of common sense reasoning is very important for our understanding of physical reality, it is the experiment that can over-ride our most cherished presuppositions, and when this occurs, we must yield.^b

The constancy of the speed of light leads to the necessity that the notion of absolute time, a prime example of a cherished presupposition of classical Newtonian physics, must be abandoned. Intervals of time as well as intervals of space, the measured lengths of bodies, are seen to be relative to one's frame of reference. This in turn leads to the necessity of focusing upon "events" in space and time, not just where some event E has occurred, but where *and when* that event has occurred relative to the frame of reference that is being used to label it mathematically. It is essential because, as we shall see, the temporal relationship between events is no longer an absolute in relativity but rather it is dependent upon the reference frame of the observer. It should be noted that in normal parlance, the word "event" usually refers to some happening of significance. In physics, events need not have any particular significance and are generally defined simply as points in a spacetime plot, with time plotted on one axis and space plotted on axes perpendicular to the time axis.

A particularly shocking consequence of the invariance of the speed of light in vacuum is the necessity to abandon the seemingly common-sense notion that the simultaneity of events is absolute. To illustrate this, consider a railway car moving along the track at speed V . In the middle of the car, a match is struck and receptors fixed at both ends of the car record the arrival of the first photons from the flash. Clearly they record the arrivals at the same time as they are totally equivalent receptors. Within the frame of the car, there is no physical significance to their being at the back or the front of the car; the distance from the match is the same for both and the speed of the photons is the same in both directions.

^bThis applies with particular force in the bizarre domain of quantum phenomena.

However, observers standing along the track have a different picture of the events. They see the receptor at the back of the car moving towards the source of the light and the receptor at the front moving away from the source. While the photons are traveling to the back end, relative to the reckoning of the outside observers, the back end receptor has a *decreased* distance for the photons to reach it while the front end receptor has an *increased* distance for the photons to reach it. Since the speed of the photons as viewed by the outside observers along the track is the same value c in both directions, just as was the case for the rest frame of the car, clearly they will say that the event of photons reaching the back end of the car *preceded* their arrival at the front end. Simultaneity relative to the rest frame of the car does not translate into simultaneity in the frame of the outside observers along the track.

Had the outside observers calculated according to the classical compounding of velocities, they would have viewed the photon velocities towards the rear and front ends as $V - c$ and $V + c$ respectively rather than $-c$ and c and they would have deduced simultaneity in their frame as well. Counter to our intuition and experience with material objects, the speed V of the light source has no effect on how the outside observers gauge the velocity of the light. The key is the invariance of the speed of light in relativity in abolishing the absoluteness of simultaneity.

2.2 The spacetime interval and the Lorentz transformation

Consider two events: a photon emitted at (x_1, y_1, z_1) at time t_1 and absorbed at (x_2, y_2, z_2) at time t_2 relative to a reference frame K. Since the photon speed is c , the distance traversed can be expressed as $c(t_2 - t_1)$ or as $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$. For any two events, we define the square of the spacetime interval s_{12} between them as

$$s_{12}^2 = c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2. \quad (2.1)$$

Its value is seen to be 0 for the events on the path of the photon given the two ways of expressing the same distance traversed by the photon. Note that this interval squared has been defined with

differing signs, a positive time interval squared and a negative space interval squared. This is crucial in what follows and it incorporates the essential difference between time and space.

We could equally well consider the photon events from the point of view of a frame K^* and we would find (2.1) again with the symbols starred, apart from c which would be redundant since $c = c^*$ by experiment. Thus we have

$$s_{12}^2 = s_{12}^{*2} = 0 \quad (2.2)$$

for the spacetime interval squared between photon events. This can also be expressed for infinitesimally separated photon events as

$$ds^2 = ds^{*2} = 0 \quad (2.3)$$

where

$$ds^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2 \quad (2.4)$$

is now the expression for the infinitesimal spacetime interval squared.

It can be shown [3] that the invariance of the spacetime interval for events on the path of a light ray leads to the invariance of the spacetime interval for events that are *not* on the path of a light ray, i.e. when the interval as expressed in either (2.1) or (2.4), is not zero. *The invariance of the spacetime interval between events is an essential fact in relativity.*

Let us label an event as $E : (t, x, y, z)$ relative to observer frame K . Let an observer K^* , moving at speed V along the x direction carrying axes (x^*, y^*, z^*) , label this same event. The K^* observer will assign spacetime coordinates to this event as $E : (t^*, x^*, y^*, z^*)$. It is straightforward to show that the familiar classical transformation between the coordinate systems $x = x^* + Vt^*, y = y^*, z = z^*, t = t^*$ incorporating the absoluteness of time, will not correctly relate the coordinates of the event in the two frames because this transformation will not preserve spacetime intervals in the form of (2.1) or (2.4) between distinct events. The transformation leads to an unwanted cross-term in $dx^* dt^*$ for infinitesimally separated events. It is simple to verify that the correct transformation that preserves the form is the Lorentz transformation

$$x = \frac{x^* + Vt^*}{\sqrt{1 - V^2/c^2}}, y = y^*, z = z^*, ct = \frac{ct^* + Vx^*/c}{\sqrt{1 - V^2/c^2}}. \quad (2.5)$$

In Section 3.4, we will derive this transformation using measurements with beams of light.

The V/c term in the new time transformation and the square root factors display the smallness of the effect of relativity when the relative velocity between the frames is much less than c and the criticality of the effect when V is close to c . The transformation also contrasts the stark change that relativity imposes upon the relationship between t and t^* as compared with the Newtonian absolute time relation $t = t^*$.

With frame K^* moving with velocity V relative to K , it follows that K moves with velocity $-V$ relative to K^* . Therefore the inverse transformation is effected by replacing every un-starred coordinate with its starred counterpart while letting V go into $-V$. The result is

$$x^* = \frac{x - Vt}{\sqrt{1 - V^2/c^2}}, y = y^*, z = z^*, ct^* = \frac{ct - Vx/c}{\sqrt{1 - V^2/c^2}}. \quad (2.6)$$

2.3 Lorentz contraction and time dilation

From the Lorentz transformation, it is easy to show that the length of an object as measured in its rest frame, the “proper length”, is maximal: an observer moving relative to the object measures a shorter length, the phenomenon called “Lorentz contraction”. Consider a meter stick at rest in the K frame along the x axis, its left end at x_L and its right end at x_R at all times t . The K frame observers measure its length l as $l = x_R - x_L = 1$ meter. Using the first of (2.5), we will deduce the length l^* of the stick as observed in the frame K^* . We have, for the relationship between the coordinates,

$$x_L = \frac{x_L^* + Vt_L^*}{\sqrt{1 - V^2/c^2}}, \quad x_R = \frac{x_R^* + Vt_R^*}{\sqrt{1 - V^2/c^2}}. \quad (2.7)$$

By choosing different combinations for the values of t_L^*, t_R^* , we can realize any length that we wish for the stick length $l^* = x_R^* - x_L^*$ as viewed in K^* . This being the case, a logically useful *definition* of length is called for. This is the value of $x_R^* - x_L^*$ when taken at a *common* time t^* for the left and right end x^* values. Then, subtracting the first from the second of (2.6), the time terms cancel

and we have

$$x_R - x_L = \frac{x_R^* - x_L^*}{\sqrt{1 - V^2/c^2}} \quad (2.8)$$

Thus $l^* = l\sqrt{1 - V^2/c^2}$ and since $\sqrt{1 - V^2/c^2} < 1$ for non-zero V , we see that $l^* < l$, i.e. the stick appears shorter than 1 meter.

One might ask if this kind of derivation could have been applied just as well to the inverse transformation, the first of (2.6), to deduce the opposite, that it is the frame in which the stick is seen to be moving that shows a bigger length. This is not the case. Recall that the left end of the stick is at $x = x_L$ for arbitrarily chosen t and similarly for the right end. Thus, attempts to use this equation merely reveals that any length can be attributed to the stick in the frame K^* by choosing values of t appropriately.

A key aspect is the lack of symmetry in the comparison. The stick is at rest in one special frame and this sets it apart from all other frames in which it is seen to be moving. However, another kind of comparison does have symmetry. If observers in K measure the lengths of sticks that are at rest in K^* , the K observers will say that these sticks are shorter and if observers in K^* measure the lengths of sticks that are at rest in K , the K^* observers will say that these sticks are shorter. This is a comparison with perfect symmetry: each stick has its proper length in its rest frame and each stick is being compared symmetrically with length as judged in the adjacent moving frame. This reflects the total physical equivalence of the two frames of reference and the objects being compared.

We now consider how lengths of intervals in time compare. A time interval read in the rest frame of a clock, what we define as the “proper time”, is minimal: observers in a frame moving relative to the clock would deduce that the time interval is longer, the phenomenon generally referred to as “time dilation”. To see this, we now consider a clock at rest in the K frame and compare an interval of time that is read in the clock’s rest frame K to the corresponding interval read in the frame K^* in which the clock is seen to be moving. Let the clock’s position in K be x_C which is the case for all times t . Let the first click demarking the start of the interval be at time t_1 and the final click demarking the end of the interval be at time t_2 . The length of the interval Δt is equal to $t_2 - t_1$. From the second of

(2.6), we have

$$ct_1^* = \frac{ct_1 - Vx_C/c}{\sqrt{1 - V^2/c^2}}, \quad ct_2^* = \frac{ct_2 - Vx_C/c}{\sqrt{1 - V^2/c^2}} \quad (2.9)$$

for the two ticks of the clock demarking the time interval. Subtracting the first equation from the second in (2.9), we have $t_2^* - t_1^* = \Delta t^* = \Delta t / \sqrt{1 - V^2/c^2}$. Since the square root factor is less than 1, we see that $\Delta t^* > \Delta t$. Again we see the reciprocal nature of space and time: lengths shrink and time intervals expand. Proper lengths are maximal and proper time intervals are minimal. Note that since ds^2 in (2.4) is an invariant and since $dx = dy = dz = 0$ in the rest frame of a clock, the proper time interval can also be characterized by ds/c .

2.4 Causality

Due to the presence of both positive and negative signs in ds^2 , the spacetime interval squared, the values of ds^2 can be positive, negative or zero. Spacetime intervals that are zero in value are called “null” or “lightlike” intervals. Intervals whose squares are positive are called “timelike” and intervals whose squares are negative are called “space-like”. For discussions involving intervals and the important issue of causality, it is useful to define the “light cone”. This is the geometric figure that the totality of possible light rays intersecting the base event traces out in a spacetime diagram. It is simplest to display this with two spatial dimensions in the x, y plane with time running in the axis perpendicular to this plane. The photons that are confined in space to the x, y plane and that intersect the spacetime point $(t, x, y, z) = (0, 0, 0, 0)$ trace out trajectories in this picture that are straight lines through the $(0, 0, 0, 0)$ origin point and are at constant slope with respect to the t axis. Clearly the assembly of all such photons traces out two cones, the one above the (x, y) plane displaying the photon events in the future of the base event and the cone below displaying the past photon events.

Events that are connected to a base event by a timelike interval, when displayed in a spacetime diagram are seen to be connected by a line that lies within the light cone^c and hence can be causally re-

^cFor example, the history of a body at rest at the spatial origin $x = y = z = 0$

lated with an absolute sense of which event came “before” and which event came “after” the base event. By contrast, events separated by a spacelike interval to the base event are connected by a line outside of the light cone and hence can never be causally related to the base event. (No signal can exceed the speed of light in vacuum.) For such events, it is easy to show that one event can occur before, after or even simultaneous with the base event, depending upon the frame of reference chosen to label them. This is not a cause for consternation as such outside-the-light cone events can never be causally related to the base event. Such events can never occur at the same spatial position as the base event and hence are said to be “absolutely separated” from the base event.

This is to be compared to timelike separated events which can be seen to have occurred at the same spatial position by the right choice of reference frame but which can never be seen to occur at the same time. As well, for two events A and B that are timelike separated, if A occurs before B in one frame, it will be seen to be so in any other frame. For such events, there is a definite ordering in time. Let us say that event B occurred after event A in some reference frame. Then it will be so in all frames. This is so because to realize a reference frame in which there is a reverse order in time, one would have to proceed through a succession of transformations passing through the frame in which the events are simultaneous. But this is impossible, given that the events are timelike separated.

Having a definite ordering in time is consoling as the causal ordering connection that could link such events in a given frame must retain that ordering in all frames to avoid a logical inconsistency. In graphic terms for example, if event A were a mosquito biting one’s arm and a later event B would be the slap of the mosquito into paste, it would be a physical absurdity were the events to be seen in reverse order in a different frame. While a film run backwards would show it as such, it would not be a realization of unfolding events in nature. The laughter that such reverse-run films provoke is a testament to their non-physicality.

With time intervals being non-absolute in relativity, changing from frame to frame as do the space intervals, it is useful to regard

traces out the t axis as its spacetime trajectory, the axis of the light cone, which is, of course, within the light cone.

the set of spacetime coordinates (ct, x, y, z) as the components of a four-dimensional vector $x^i = (x^0, x^1, x^2, x^3)$ with $i = 0, 1, 2, 3$, whose “length” squared is defined as $(x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 = (ct)^2 - x^2 - y^2 - z^2$. This is a dimension step up from elementary three-dimensional mathematics with the added twist of a different sign for the square of the new fourth coordinate $x^0 = ct$ relative to that of the three space coordinates. It is necessary to incorporate this difference in keeping with the invariance of the spacetime interval under a transformation between coordinate systems. The difference in sign is the characteristic that makes time essentially different from space.

One often reads that relativity places space and time on the same footing with time being just another coordinate like the space coordinates. However, this is misleading: one should say that space and time, while being of equal significance in their alterability, are in a sense placed on a *reciprocal* rather than an equal footing. As well, it is essential to emphasize that our experience in nature is that while we can fix our spatial position, we cannot stop the flow of time.

Actually, x^i is more precisely referred to as a Lorentz “contravariant” four-vector. This name distinguishes it from a slightly different four-vector (with a subscript rather than superscript index) x_i , the Lorentz “covariant” four-vector. Its components are related to x^i as $x_0 = x^0, x_\alpha = -x^\alpha, \alpha = 1, 2, 3$. This is very useful as it enables us to express the invariant $c^2t^2 - x^2 - y^2 - z^2$ very succinctly as $x_i x^i$ with the repeated index denoting a summation over 0, 1, 2, 3. Four-dimensional scalar products must always be formed with repeated indices where one is a subscript (covariant) and the other a superscript (contravariant).

2.5 Transformation of velocity and the aberration of light

We now consider how the motion of an object appears from the vantage point of the frames K and K^* . Let $v_x = dx/dt$ be the x-component of the velocity of the object and let v_y, v_z be the y and z components, all as viewed in K . Similarly, let $v_x^* = dx^*/dt^*$ be the x^* velocity component seen in K^* and v_y^*, v_z^* be the y^*, z^* components as seen in K^* . To compare the velocities seen in the two frames,

we take differentials of (2.5), keeping V , the given relative velocity between K and K^* , constant:

$$\begin{aligned} dx &= \frac{dx^* + V dt^*}{\sqrt{1 - V^2/c^2}}, \\ dy &= dy^*, \quad dz = dz^*, \\ cdt &= \frac{cdt^* + V dx^*/c}{\sqrt{1 - V^2/c^2}}. \end{aligned} \quad (2.10)$$

Dividing dx, dy, dz by dt in (2.10), we find the relationship between the velocity components in the two frames as

$$\begin{aligned} v_x &= \frac{v_x^* + V}{1 + V v_x^*/c^2}, \\ v_y &= \frac{v_y^* \sqrt{1 - V^2/c^2}}{1 + V v_x^*/c^2}, \\ v_z &= \frac{v_z^* \sqrt{1 - V^2/c^2}}{1 + V v_x^*/c^2}. \end{aligned} \quad (2.11)$$

This rather complicated transformation reverts to the simple Newtonian connection

$$v_x = v_x^* + V, \quad v_y = v_y^*, \quad v_z = v_z^* \quad (2.12)$$

in the Newtonian limit of velocities much smaller than c .

Using (2.11), we can relate the directions of the velocities of objects relative to the coordinate axes. With no loss in generality, we consider the motion of the object to be in the x, y plane with θ being the angle between the velocity vector (v_x, v_y) and the x -axis as viewed in K and θ^* being the corresponding angle as viewed in K^* . Thus

$$\begin{aligned} v_x &= v \cos \theta, & v_y &= v \sin \theta, \\ v_x^* &= v^* \cos \theta^*, & v_y^* &= v^* \sin \theta^* \end{aligned} \quad (2.13)$$

where

$$v = \sqrt{v_x^2 + v_y^2}, \quad v^* = \sqrt{v_x^{*2} + v_y^{*2}}. \quad (2.14)$$

Thus, from (2.13) and (2.11),

$$\tan \theta = \frac{v^* \sqrt{1 - V^2/c^2} \sin \theta^*}{v^* \cos \theta^* + V}. \quad (2.15)$$

By setting $v^* = c$, the object becomes a photon and the formula in (2.15) becomes the relativistic expression for the aberration of light, the change in the direction of the propagation of light in changing to a new moving reference frame. In the Newtonian limit ($V \ll c$), the aberration formula reduces to the familiar expression

$$\Delta\theta = \theta^* - \theta = (V/c) \sin \theta^*. \quad (2.16)$$

2.6 Four-vectors and four-tensors

Just as x^i transforms to x^{i*} as in (2.6), so too it is useful to define any set A^i of four functions of the coordinates x^i as the components of a general Lorentz four-vector. This is provided that these functions transform to new functions A^{i*} in the same manner as x^i (i.e. as in (2.6) when one changes to a new coordinate frame x^{i*}). Lorentz four-vectors (and four-tensors) are the basic mathematical objects in special relativity. The essence of what makes a vector or a tensor lies in the nature of the transformation of the object (the set of functions) when expressed in a different reference frame. For vectors, once we designate four arbitrary functions A^i of the coordinates x^i as the components of a Lorentz four-vector in the x^i coordinate frame, the new form of this four-vector A^{i*} in any other Lorentz frame x^{i*} is completely determined by the transformation (2.6) with A^0, A^1, A^2, A^3 replacing ct, x, y, z (as well as their starred counterparts) in (2.5).

A vector can also be described as a tensor of first rank. A set of 16 functions B^{ik} with i and k each taking on the values 0, 1, 2, 3 are said to form the components of a second rank Lorentz four-tensor provided this set transforms as does A^i for each of the two indices in B^{ik} ; similarly for third and higher rank tensors. The rank of a tensor is given by the number of indices attached to it. Thus, a vector is also a tensor of first rank. A scalar is a tensor of zero rank (no indices).

Since velocity^d $v^\alpha = dx^\alpha/dt$, $\alpha = 1, 2, 3$ is of great importance in physics, we require the four-vector generalization, the four-velocity u^i . One might first be tempted to define this as dx^i/dt . But recalling the relativity of time intervals, while dx^i is a four-vector, dt is not a

^dWe adopt the convention of [3] that Greek indices range over the space indices.

scalar and hence the product of dx^i with $1/dt$ is not a four-vector. However ds is a scalar, and therefore four-velocity is defined as^e

$$u^i = dx^i/ds. \quad (2.17)$$

Also of importance is four-acceleration w^i which in special relativity is naturally defined as

$$w^i = du^i/ds. \quad (2.18)$$

However, we will see in what follows that this procedure for acceleration is inadequate in general relativity.

Another important four-vector is the energy-momentum four-vector p^i ,

$$p^i = mcu^i. \quad (2.19)$$

where m is the mass. The spatial components $p^\alpha = (p_x, p_y, p_z)$ constitute the relativistic components of the linear momentum

$$p^\alpha = mv^\alpha / \sqrt{1 - v^2/c^2}. \quad (2.20)$$

and the time component p^0 is E/c where E is the relativistic energy

$$E = mc^2 / \sqrt{1 - v^2/c^2}. \quad (2.21)$$

From this, we have the base level of energy of a body, the energy as measured in the body's rest frame ($v = 0$), to be the very important and familiar mass-energy identification

$$E = mc^2. \quad (2.22)$$

Straightforward calculations with (2.20) and (2.21) reveal the important connections between relativistic energy, momentum and velocity

$$E^2 = p^2 c^2 + m^2 c^4 \quad (2.23)$$

and

$$p = Ev/c^2. \quad (2.24)$$

where $p = \sqrt{p_x^2 + p_y^2 + p_z^2}$, the magnitude of the three-momentum.

^eNote that u^i thus defined is dimensionless. It could have been defined with ds replaced by ds/c to recover the usual dimensionality for velocity. However, it is customary not to do so and it becomes irrelevant in the increasingly common usage of coordinates for which c is taken to be 1, as we will adopt in later chapters.

2.7 Special relativistic dynamics

Having the basic elements in place, we turn to the question of how bodies move within the theory of special relativity. There is a basic principle that deals with this and much more. It is called the “Action Principle” or, more formally, the “Principle of Least Action”. Many have come to regard this principle as the most basic in all of physics. In the broadest terms, it states that for any physical system, there exists an invariant integral called the “action” such that when the system evolves from state A to state B, it does so in such a manner as to minimize the action. The integral is evaluated over the region in which the system evolves. In the case of a body, it is over its path (a curve) in spacetime between the initial and final spacetime points. If we are considering the evolution of a field such as an electromagnetic field, it evolves over its domain, three-dimensional space between an initial and a final time. Accordingly, the action integral for this case is over spacetime, covering all of three-dimensional space and over time between the limits of the initial and final states. The invariant action integral is expressed in generality for arbitrary assumed evolutions and the physically correct evolution is the one that minimizes this integral. Just as in simple calculus, the minimum is found by setting the first derivative to zero but now, the derivative must be evaluated over a path rather than at a point. Performing these calculations requires a knowledge of the calculus of variations which is beyond the scope of this book. Readers who wish to follow this mathematics should consult [3].

In their most elegant form, the great equations of physics, the Maxwell equations of electromagnetism, the field equations of general relativity, even quantum mechanics can be realized through the Action Principle. Fermat’s Principle in optics, describing the path of a light ray through a medium of varying refractive index as one that minimizes the time, is a realization of the Action Principle. It is interesting to contemplate the fact that the most basic phenomena of nature are tied to extrema. The challenge is to discover the appropriate Lagrangians to create the action integrals and for this purpose, there is no magic wand available. However, there is a widespread sentiment that all *bona fide* physical theory emerges from an action integral, that the Action Principle is paramount. Feynman [6] provides a highly recommended lively account of the Action Principle

in his characteristic style.

Richard P. Feynman (1918–1988) was one of the most distinguished and accomplished theoretical physicists of the 20th Century. His contributions to quantum electrodynamics earned him the Nobel Prize along with J. Schwinger and S. I. Tomonaga. His three-volume “Lectures on Physics” [6] continue to be a great resource for beginning students, graduate students and researchers in physics. One of the most colorful and inspiring lecturers, his interests and contributions spanned a variety of fields beyond physics.

For a free body in special relativity, the invariant available to us is the spacetime interval and hence the action integral S is simply

$$S = K \int ds, \quad K = \text{constant} \quad (2.25)$$

taken between the limit spacetime points $A = (t_1, x_1, y_1, z_1)$ and $B = (t_2, x_2, y_2, z_2)$ representing the spacetime points from which the body begins its motion to where it ends its motion.

The action integral is a solely mathematical construct. The actual physical path is revealed in setting the variation of the action to zero, in searching out from the infinite number of imaginable unphysical types of evolutions, the unique physical path that constitutes the minimum, i.e.

$$\delta S = 0. \quad (2.26)$$

The δ is the variational calculus derivative representing the first derivative of this integral taken not at a single point as in ordinary calculus but over the entire path between A and B.

Using the calculus of variations, the result is

$$\frac{du^i}{ds} = 0 \quad (2.27)$$

which tells us that the four-velocity u^i of the body is a constant. This comes as no surprise but it is useful to follow through this simplest application of the very important Action Principle (see [3]).

The constancy of u^i over the path implies, through (2.19) that the four-momentum p^i is constant, and hence the energy and linear momentum are constant.

While the choice of action integral for a free body is quite straightforward, the choice for more complicated systems is correspondingly more complicated. To be noted is that when the action is expressed as an integral over time, the integrand is the Lagrangian, L , the function that we first encounter in classical mechanics,

$$S = \int L dt. \quad (2.28)$$

It can be shown [7] that the dynamics embodied in the key physics equation (2.26) is equivalent to the Lagrangian satisfying Lagrange's equations

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}^\alpha} \right) = \frac{\partial L}{\partial x^\alpha}. \quad (2.29)$$

However, while the Lagrangian that we first encounter in classical mechanics is the difference between the kinetic (T) and potential (U) energies,

$$L = T - U = \frac{1}{2}mv^2 - U \quad (2.30)$$

the Lagrangian in relativistic physics applies to fields as well as particles and assumes a variety of complicated forms. As the simplest example, for a free body, the potential energy is a constant which can be taken to be zero for convenience. The comparison of (2.25) and (2.28) using

$$ds = c dt \sqrt{1 - \frac{v^2}{c^2}} \quad (2.31)$$

[which follows from (2.4)] yields the relativistic Lagrangian for a free body

$$L = Kc \sqrt{1 - \frac{v^2}{c^2}}. \quad (2.32)$$

For small velocities, the relativistic Lagrangian (2.32) must approach the classical mechanics Lagrangian (2.30) (with $U = 0$). We make this comparison by expanding the square root in (2.32) and retain up to the lowest non-vanishing order term in the velocity,

$$L \approx -Kv^2/2c. \quad (2.33)$$

Comparing the v^2 terms in (2.33) and (2.30) gives $K = -mc$ and hence the relativistic Lagrangian for a free body is^f

$$L = -mc^2 \sqrt{1 - \frac{v^2}{c^2}}. \quad (2.34)$$

With the relativistic Lagrangian (2.34), we apply the standard formalism to derive the relativistic momentum and energy:

$$p^\alpha = \partial L / \partial v^\alpha \quad (2.35)$$

or

$$p^\alpha = \frac{mv^\alpha}{\sqrt{1 - v^2/c^2}}. \quad (2.36)$$

The energy is^g

$$E = p^\alpha v^\alpha - L = mc^2 / \sqrt{1 - v^2/c^2}. \quad (2.37)$$

We see that the energy and momentum derived in this way match the expressions that we had found in (2.21) and (2.20) by using (2.19). However, the latter have the added virtue of showing us that $(E/c, p^\alpha)$ constitute a Lorentz four-vector. As a result, given the energy and momentum in one Lorentz frame x^i , by the Lorentz transformation, we know how to find these quantities in any other Lorentz frame x^{i*} with the transformation format of (2.6):

$$\begin{aligned} p_x^* &= \frac{p_x - VE/c^2}{\sqrt{1 - V^2/c^2}}, \\ p_y &= p_y^*, \quad p_z = p_z^*, \\ E^* &= \frac{E - Vp_x/c}{\sqrt{1 - V^2/c^2}}. \end{aligned} \quad (2.38)$$

2.8 Relativistic Doppler shift

The power of Lorentz covariance is well-illustrated in the relativistic Doppler formula, displaying the difference in the frequency of light

^fAn additive constant could be retained but it has no physical significance and is most conveniently set to zero.

^gHere α is summed over 1, 2, 3.

perceived by an observer in motion relative to that observed in the rest frame of the source of light. The derivation requires the introduction of a new four-vector called the “wave four-vector” with symbol k^i . The time component k^0 is taken to be the angular frequency^h ω divided by c and the three-vector k^α is defined by

$$k^\alpha = \frac{\omega}{c} n^\alpha \quad (2.39)$$

where n^α is a unit vector in the direction of wave propagation. We recall that while we can designate any set of four components k^i to constitute a Lorentz four-vector, the above formulated definition has value only if its components in a new frame x^{i*} have the same significance in terms of frequency and wave propagation direction as they had in the original frame x^i . That they do indeed have this significance is seen from the fact that the inner product

$$k^i x_i = \omega t - k^\alpha x^\alpha \quad (2.40)$$

is the phase of the wave which is a scalar. By the “quotient rule” (see [8]), since x_i is an arbitrary Lorentz four-vector and the phase of the wave is a scalar (see [9]), the wave four-vector as defined above for all frames is truly a *bona fide* Lorentz four-vector.

Having this established, the work is essentially done; we know how Lorentz four-vectors transform, in the manner of (ct, x, y, z) in (2.5), i.e.

$$\begin{aligned} k^0 &= \frac{k^{0*} + (V/c)k^{1*}}{\sqrt{1 - V^2/c^2}}, \\ k^1 &= \frac{k^{1*} + (V/c)k^{0*}}{\sqrt{1 - V^2/c^2}}, \\ k^2 &= k^{2*}, \quad k^3 = k^{3*}. \end{aligned} \quad (2.41)$$

Consider the source of light to be at rest in x^i with the observer whose frame is x^{i*} moving with velocity V relative to the source and in the x direction as usual. Let α be the angle between the direction of wave propagation and the x^* axis as viewed in the x^{i*} frame. Then

$$k^{1*} = (\omega^*/c) \cos \alpha. \quad (2.42)$$

^h $\omega = 2\pi\nu$ where ν is the actual frequency.

Substituting (2.42) into (2.41) and using (2.39), we find

$$\omega^* = \frac{\omega \sqrt{1 - V^2/c^2}}{1 + (V/c) \cos \alpha}. \quad (2.43)$$

This is the Doppler shift formula in all accuracy, even for relativistic velocities. Note that with $V > 0$ and α between 0 and $\pi/2$, the observed frequency ω^* is less than ω , the red-shift. Also, for $V \ll c$ and α not close to $\pi/2$, (2.43) yields the familiar classical expression for the fractional change in frequency

$$\frac{\omega^* - \omega}{\omega} = \frac{\Delta\omega}{\omega} = -(V/c) \cos \alpha. \quad (2.44)$$

Returning to (2.43), consider the special case where $\alpha = \pi/2$, i.e. the observer sees the light arriving perpendicular to hisⁱ direction of motion.

In that case, $\cos \alpha = 0$ and

$$\omega^* = \omega \sqrt{1 - V^2/c^2}. \quad (2.45)$$

Thus, the observer sees a red-shift with the same value (as required by symmetry) whether he moves to the right ($V > 0$) or to the left ($V < 0$). It is very small (of order V^2/c^2) compared to the usual shifts for non-relativistic velocities. This is a purely relativistic effect, stemming from the non-absoluteness of time, without even an infinitesimal residue in pre-relativity physics. It is often referred to as the “transverse Doppler shift”. By contrast, in classical physics, a frequency shift can only occur if there is a component of the light propagation velocity in the direction of motion of the observer.

To make sense of sources moving at very high velocities, astronomers must use the relativistic equations above to properly interpret their observations.

We have covered the essential aspects of special relativity in this chapter. For further details, the reader is directed to the classic treatise of Landau and Lifshitz [3].

ⁱFor this and subsequent references to a person in the abstract, “his” means “his/her” and “he” means “he/she”.

Lev D. Landau (1908–1968) is generally regarded as one of the most distinguished theoretical physicists of the 20th Century. His major works spanned several areas of physics and he was awarded the Nobel Prize for his important development of the theory of superfluidity. His ten-volume Course on Theoretical Physics with E. M. Lifshitz remains a key research tool for physicists world-wide. Tragically, Landau suffered major injuries in a traffic accident from which he never fully recovered.

Before proceeding to general relativity, we revisit special relativity from the ingenious approach of Bondi.

This page intentionally left blank

Chapter 3

Bondi's k -Calculus Approach to Special Relativity

3.1 Introduction

Bondi [4] has developed a unique highly intuitive approach to special relativity that displays some of the essential characteristics with spacetime diagrams.

Sir Hermann Bondi (1919–2005) was a distinguished mathematician and cosmologist. With T. Gold and F. Hoyle, he developed the Steady State theory of the universe, widely believed but eventually discarded upon the discovery of the cosmic microwave background radiation. As well as his development of the k -calculus approach to special relativity, Bondi wrote several important papers on a variety of subjects in general relativity.

It makes extensive use of the relativistic Doppler factor k that relates inertial reference frames in relative motion and the fundamental new aspect of relativity vis-a-vis Newtonian physics, the invariance of the speed of light. We will see how clearly this approach resolves the so-called twin paradox, a key stumbling block for many who are first introduced to relativity. The word “calculus” to describe the Bondi method is misleading as his method relies upon elementary

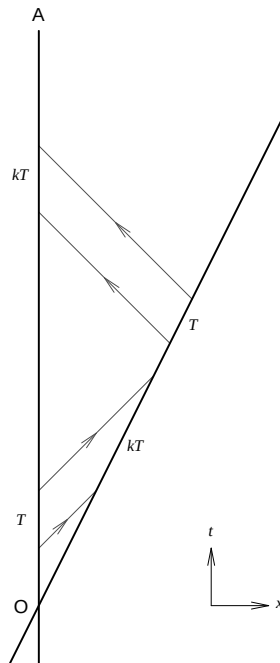


Figure 3.1: Light shone from Ava to Beta and vice-versa with intervals displaying their complete equivalence.

algebra and could be taught to junior high school students.

We consider two inertial observer twins Ava (A)^a and Beta (B)^b in relative motion. In Figure 3.1, the horizontal axis is x and the vertical axis is t . The diagram is displayed from the perspective of Ava who is shown to be at rest in (x, t) and Beta is seen to move in the positive x direction.

The spacetime diagrams that follow have the closest connections to simple plane geometry from the vantage point of the rest observer.

Some time after Ava meets Beta (the point at which their lines cross in the diagram and at which point they both set their clocks to 0), Ava shines light toward Beta for a period T by her (Ava's) clock. Beta receives the light for a proportional amount of time kT

^aAva—Greek—"An eagle".

^bBeta—Czech—"Dedicated to God".

where the k constant symbol is the Doppler factor connecting the observers. The lines with arrows indicate light rays.

Later, Beta sends light toward Ava for the same period T and since Ava and Beta are totally equivalent physically, Ava observes the light for the same period as Beta had observed the earlier light from Ava, namely kT . Note that the geometrical appearance aspects such as the slopes of outgoing light rays, are necessarily biased in favor of the rest observer. Only relative to the rest observer are the light ray slopes always at 45 deg.

3.2 Velocity–Doppler factor connection

Ava wishes to determine Beta's speed relative to herself. For this, a different process is required, as shown in Figure 3.2. Immediately upon meeting Beta at O, Ava shines light to Beta for a period T and as in Figure 3.1, Beta receives the light for a period kT . Ava and Beta agree that as long as Beta receives light from Ava, she (namely Beta) sends light back to Ava.

Since this emission period from Beta is kT , the reception of Beta's light by Ava is the Doppler factor k times the emission period, i.e. $k(kT) = k^2T$.

From here it is a simple procedure to deduce the relative velocity between Beta and Ava. All that is required is to determine the distance between Ava and Beta at a convenient point and how much time has elapsed since their meeting at O to achieve that separation. The point P that is chosen is where Beta has received the last photon from Ava and has sent her last photon back to Ava. Ava reckons that she has sent her last photon reaching P at time T and has received the last photon back from Beta at time k^2T . Thus, the time for the photon to reach P and return is $k^2T - T$. Multiplying this by c , the photon speed, we get twice the OP separation. Therefore the OP separation is $D = c(k^2T - T)/2$. Ava also reckons that the last photon reached P at the half-way point in time from the emission at time T and the reception at time k^2T , namely at time $T^* = T + (k^2T - T)/2 = (k^2 + 1)T/2$. Finally, the relative velocity v is D/T^* or

$$v = c \frac{k^2 - 1}{k^2 + 1}. \quad (3.1)$$

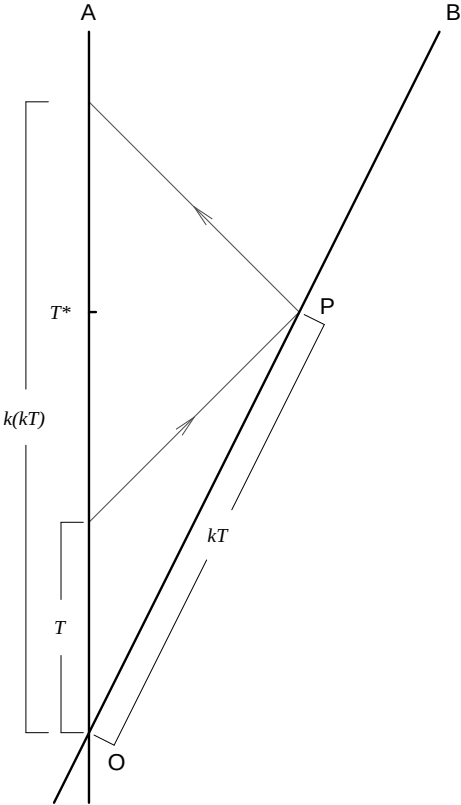


Figure 3.2: Determination of the speed of Beta relative to Ava.

Solving for k yields the relativistic Doppler factor k in terms of relative velocity^c

$$k = \sqrt{\frac{1 + v/c}{1 - v/c}}. \quad (3.2)$$

The following properties are noted:

1) $k = 1$ for $v = 0$ which is logical since there is no Doppler shift when there is no relative velocity.

2) $k > 1$ if $v > 0$ which is logical since a relative recession entails an increase in period, decrease in frequency, increase in wavelength, or red-shift.

3) Similarly, $k < 1$ for $v < 0$. In this case there is a relative approach and hence a blue shift.

4) If $v \rightarrow -v$ then $k \rightarrow 1/k$.

3.3 Composition law for velocities and Doppler factors

We now determine the relativistic composition law for velocities. Consider a third observer Cayla (C),^d introduced as in Figure 3.3. Let Ava emit light to Beta for a period T . Beta receives the light for a proportional time period $k_{AB}T$ where k_{AB} is the Doppler factor between Ava and Beta. For as long as Beta receives the light, she transmits to Cayla who receives it for a period $k_{BC}(k_{AB}T)$ where k_{BC} is the Doppler factor between Beta and Cayla. At this point we invoke the key feature of special relativity, the invariance of the speed of light: we can view the direct transmission of light from Ava to Cayla for a period T being received by Cayla for a time k_{AC} with the same photon lines that were already used for the previous two-step process. This is because the light speed does not get boosted in the two-step process as compared to the direct transmission. As a result, we can equate the reception times by Cayla of the one-step and two-step process. Canceling the common factor T , we have

$$k_{AC} = k_{AB}k_{BC}. \quad (3.3)$$

^cThe positive root solution is chosen to maintain the same direction of flow of time for the two observers.

^dCayla–“pure”.

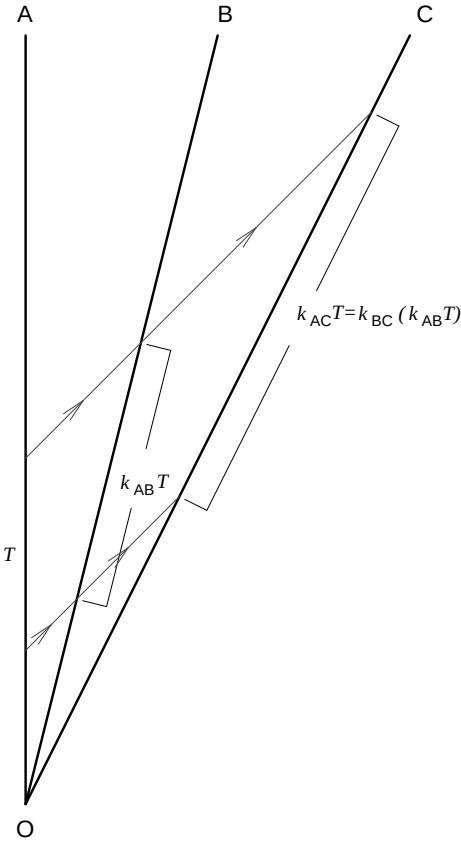


Figure 3.3: Determination of the compounding velocities in relativity.

Similarly, if we were to introduce a fourth observer Della (D)^e, we would have

$$k_{AD} = k_{AB}k_{BC}k_{CD} \quad (3.4)$$

and so on for an arbitrary sequence of observers. Let us now work in units where $c = 1$ (we will restore the symbol later). Returning to the three observers in Figure 3.3, from (3.1) and (3.3) we have

$$v_{AC} = \frac{k_{AC}^2 - 1}{k_{AC}^2 + 1} = \frac{k_{AB}^2 k_{BC}^2 - 1}{k_{AB}^2 k_{BC}^2 + 1}. \quad (3.5)$$

Now using (3.2) repeatedly in (3.5) for the k factors in terms of relative velocities, after simplification we find (with the explicit c now restored)

$$v_{AC} = \frac{v_{AB} + v_{BC}}{1 + v_{AB}v_{BC}/c^2}. \quad (3.6)$$

This is the familiar relativistic law for the composition of velocities. If we let c approach infinity, we retrieve the usual Newtonian velocity composition law

$$v_{AC} = v_{AB} + v_{BC}. \quad (3.7)$$

If the relative velocities of the observers are small compared to c , the effect of using (3.6) instead of (3.7) is small. However, for “relativistic” velocities, the effect is dramatic. For example, with $v_{AB} = v_{BC} = 3c/4$, the correct composition law (3.6) yields $v_{AC} = 24c/25$, a velocity less than c as must be the case and considerably different from $1.5c$ that would result from (3.7).

Of particular interest is the elegant simplicity of the composition law for Doppler factors as a simple multiplicative sequence in (3.4) in contrast to the awkward composition law for velocities in relativity (3.6). The simplicity of the former is understandable as it reflects the invariance of the speed of light and it is the Doppler factor that characterizes the connection between observers vis-a-vis light propagation. Note, however, that the Newtonian composition law, (3.7) (and its familiar extension to more observers, even non-collinearly) for velocities does have the simplicity that the Doppler factor composition displays. This is a reflection of the absoluteness of time in Newtonian physics. The Newtonian composition proceeds pictorially

^eDella-English-“A woman from the island of Delos”.

with vectors pasted end-to-end, having each observer's clock in step with all the rest.

In the next section, we display the non-absoluteness of time as an algebraic formula.

3.4 Derivation of the Lorentz transformation

The Lorentz transformation is the recipe for the labeling of event coordinates by one inertial observer relative to another. To derive this recipe, we return to our two twin observers Ava and Beta and consider an event E which Ava labels (t, x) and Beta labels (t^*, x^*) . Ava and Beta synchronize their clocks when they meet at O and Ava decides to send a photon so that it arrives in coincidence with the event E. At that point, it is reflected back to Ava. Since for Ava, the event is at position x , the photon must have been sent from Ava at time $t - x/c$ to arrive at time t at E. It will arrive back to Ava a time x/c later, i.e. at time $t + x/c$.

Since the speed of the photon is also c for Beta, the photon intersects with Beta at time $t^* - x^*/c$ for the inbound path and at time $t^* + x^*/c$ for the outbound path. From our earlier discussion regarding the relationship between time intervals, we see from Figure 3.4 that

$$t^* - x^*/c = k(t - x/c) \quad (3.8)$$

where k is the Doppler factor between Ava and Beta.

Similarly we can focus on an emission time interval of $t^* + x^*/c$ from Beta to Ava with a reception time interval $t + x/c$ by Ava's reckoning. Therefore we have

$$t + x/c = k(t^* + x^*/c) \quad (3.9)$$

Eliminating x^* between (3.8) and (3.9), we find

$$2t^* = t(k + 1/k) - (x/c)(k - 1/k) \quad (3.10)$$

$$2x^* = x(k + 1/k) - (ct)(k - 1/k) \quad (3.11)$$

Substituting the expression for k from (3.2) into (3.10) and (3.11), we find as in (2.5), the important Lorentz transformation

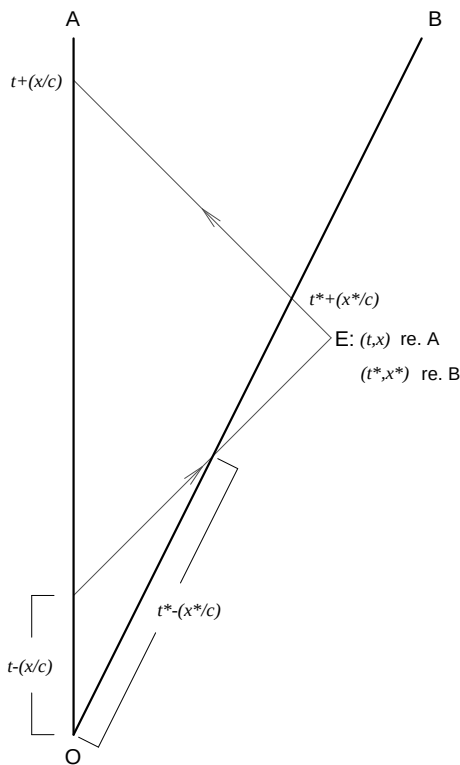


Figure 3.4: Derivation of the Lorentz transformation by the Bondi method.