



Illustrated Special Relativity

Through Its
Paradoxes



John dePillis & José Wudka
Illustrations and animations by John dePillis

Illustrated Special Relativity Through Its Paradoxes

**A Fusion of Linear Algebra,
Graphics, and Reality**

© 2013 by the Mathematical Association of America, Inc.

Library of Congress Catalog Card Number 2013956313

Electronic edition ISBN: 978-1-61444-517-3

Illustrated Special Relativity Through Its Paradoxes

A Fusion of Linear Algebra, Graphics, and Reality

John de Pillis

Professor of Mathematics

University of California, Riverside

and

José Wudka

Professor of Physics

University of California, Riverside

Illustrations and animations by John de Pillis



MAA

*Published and Distributed by
The Mathematical Association of America*

Council on Publications and Communications

Frank Farris, *Chair*

Committee on Books

Gerald Bryce, *Chair*

Spectrum Editorial Board

Gerald L. Alexanderson, *Co-Editor*

James J. Tattersall, *Co-Editor*

Robert E. Bradley

Susanna S. Epp

Richard K. Guy

Keith M. Kendig

Shawnee L. McMurran

Jeffrey L. Nunemacher

Jean J. Pedersen

Kenneth A. Ross

Marvin Schaefer

Franklin F. Sheehan

SPECTRUM SERIES

The Spectrum Series of the Mathematical Association of America was so named to reflect its purpose: to publish a broad range of books including biographies, accessible expositions of old or new mathematical ideas, reprints and revisions of excellent out-of-print books, popular works, and other monographs of high interest that will appeal to a broad range of readers, including students and teachers of mathematics, mathematical amateurs, and researchers.

777 Mathematical Conversation Starters, by John de Pillis

99 Points of Intersection: Examples—Pictures—Proofs, by Hans Walser.

Translated from the original German by Peter Hilton and Jean Pedersen

Aha Gotcha and Aha Insight, by Martin Gardner

All the Math That's Fit to Print, by Keith Devlin

Beautiful Mathematics, by Martin Erickson

Calculus and Its Origins, by David Perkins

Calculus Gems: Brief Lives and Memorable Mathematics, by George F. Simmons

Carl Friedrich Gauss: Titan of Science, by G. Waldo Dunnington, with additional material by Jeremy Gray and Fritz-Egbert Dohse

The Changing Space of Geometry, edited by Chris Pritchard

Circles: A Mathematical View, by Dan Pedoe

Complex Numbers and Geometry, by Liang-shin Hahn

- Cryptology*, by Albrecht Beutelspacher
- The Early Mathematics of Leonhard Euler*, by C. Edward Sandifer
- The Edge of the Universe: Celebrating 10 Years of Math Horizons*, edited by Deanna Haunsperger and Stephen Kennedy
- Euler and Modern Science*, edited by N. N. Bogolyubov, G. K. Mikhailov, and A. P. Yushkevich. Translated from Russian by Robert Burns.
- Euler at 300: An Appreciation*, edited by Robert E. Bradley, Lawrence A. D'Antonio, and C. Edward Sandifer
- Expeditions in Mathematics*, edited by Tatiana Shubin, David F. Hayes, and Gerald L. Alexanderson
- Five Hundred Mathematical Challenges*, by Edward J. Barbeau, Murray S. Klamkin, and William O. J. Moser
- The Genius of Euler: Reflections on his Life and Work*, edited by William Dunham
- The Golden Section*, by Hans Walser. Translated from the original German by Peter Hilton, with the assistance of Jean Pedersen.
- The Harmony of the World: 75 Years of Mathematics Magazine*, edited by Gerald L. Alexanderson with the assistance of Peter Ross
- A Historian Looks Back: The Calculus as Algebra and Selected Writings*, by Judith Grabiner
- History of Mathematics: Highways and Byways*, by Amy Dahan-Dalmédico and Jeanne Peiffer, translated by Sanford Segal
- How Euler Did It*, by C. Edward Sandifer
- Illustrated Special Relativity Through Its Paradoxes: A Fusion of Linear Algebra, Graphics, and Reality*, by John de Pillis and José Wudka
- In the Dark on the Sunny Side: A Memoir of an Out-of-Sight Mathematician*, by Larry Baggett
- Is Mathematics Inevitable? A Miscellany*, edited by Underwood Dudley
- I Want to Be a Mathematician*, by Paul R. Halmos
- Journey into Geometries*, by Marta Sved
- JULIA: a life in mathematics*, by Constance Reid
- The Lighter Side of Mathematics: Proceedings of the Eugène Strens Memorial Conference on Recreational Mathematics & Its History*, edited by Richard K. Guy and Robert E. Woodrow
- Lure of the Integers*, by Joe Roberts
- Magic Numbers of the Professor*, by Owen O'Shea and Underwood Dudley
- Magic Tricks, Card Shuffling, and Dynamic Computer Memories: The Mathematics of the Perfect Shuffle*, by S. Brent Morris
- Martin Gardner's Mathematical Games: The entire collection of his Scientific American columns*

- The Math Chat Book*, by Frank Morgan
- Mathematical Adventures for Students and Amateurs*, edited by David Hayes and Tatiana Shubin. With the assistance of Gerald L. Alexanderson and Peter Ross
- Mathematical Apocrypha*, by Steven G. Krantz
- Mathematical Apocrypha Redux*, by Steven G. Krantz
- Mathematical Carnival*, by Martin Gardner
- Mathematical Circles Vol I: In Mathematical Circles Quadrants I, II, III, IV*, by Howard W. Eves
- Mathematical Circles Vol II: Mathematical Circles Revisited and Mathematical Circles Squared*, by Howard W. Eves
- Mathematical Circles Vol III: Mathematical Circles Adieu and Return to Mathematical Circles*, by Howard W. Eves
- Mathematical Circus*, by Martin Gardner
- Mathematical Cranks*, by Underwood Dudley
- Mathematical Evolutions*, edited by Abe Shenitzer and John Stillwell
- Mathematical Fallacies, Flaws, and Flimflam*, by Edward J. Barbeau
- Mathematical Magic Show*, by Martin Gardner
- Mathematical Reminiscences*, by Howard Eves
- Mathematical Treks: From Surreal Numbers to Magic Circles*, by Ivars Peterson
- A Mathematician Comes of Age*, by Steven G. Krantz
- Mathematics: Queen and Servant of Science*, by E.T. Bell
- Mathematics in Historical Context*, by Jeff Suzuki
- Memorabilia Mathematica*, by Robert Edouard Moritz
- More Fallacies, Flaws, and Flimflam*, Edward J. Barbeau
- Musings of the Masters: An Anthology of Mathematical Reflections*, edited by Raymond G. Ayoub
- New Mathematical Diversions*, by Martin Gardner
- Non-Euclidean Geometry*, by H. S. M. Coxeter
- Numerical Methods That Work*, by Forman Acton
- Numerology or What Pythagoras Wrought*, by Underwood Dudley
- Out of the Mouths of Mathematicians*, by Rosemary Schmalz
- Penrose Tiles to Trapdoor Ciphers . . . and the Return of Dr. Matrix*, by Martin Gardner
- Polyominoes*, by George Martin
- Power Play*, by Edward J. Barbeau

Proof and Other Dilemmas: Mathematics and Philosophy, edited by Bonnie Gold and Roger Simons

The Random Walks of George Pólya, by Gerald L. Alexanderson

Remarkable Mathematicians, from Euler to von Neumann, by Ioan James

The Search for E.T. Bell, also known as John Taine, by Constance Reid

Shaping Space, edited by Marjorie Senechal and George Fleck

Sherlock Holmes in Babylon and Other Tales of Mathematical History, edited by Marlow Anderson, Victor Katz, and Robin Wilson

Sophie's Diary, Second Edition, by Dora Musielak

Student Research Projects in Calculus, by Marcus Cohen, Arthur Knoebel, Edward D. Gaughan, Douglas S. Kurtz, and David Pengelley

Symmetry, by Hans Walser. Translated from the original German by Peter Hilton, with the assistance of Jean Pedersen.

The Trisectors, by Underwood Dudley

Twenty Years Before the Blackboard, by Michael Stueben with Diane Sandford

Who Gave You the Epsilon? and Other Tales of Mathematical History, edited by Marlow Anderson, Victor Katz, and Robin Wilson

The Words of Mathematics, by Steven Schwartzman

MAA Service Center
P.O. Box 91112
Washington, DC 20090-1112
800-331-1622 FAX 301-206-9789

Contents

I. A FIRST PASS	1
Preface	2
0.1 Exposition and Paradoxes	2
0.2 Organization of this Book	5
1 Introduction to the Paradoxes	11
1.1 Aristotle vs. Galileo	11
1.2 Frames of Reference	12
1.3 Straight-Line Trajectories in 3-Space	13
1.4 Galilean Relativity	14
1.5 Special Relativity: A First Pass	16
1.6 A Symmetry Principle	18
1.7 Lorentzian Relativity	19
1.8 The Ubiquitous Shrinkage Constant	19
1.9 Paradox: The Accommodating Universe	22
1.10 Paradox: Time and Distance Asymmetry	26
1.11 Paradox: The Traveling Twin	29
1.12 Paradox: The Train in the Tunnel	33
1.13 Paradox: The Pea-Shooter	36
1.14 Paradox: The Bug and Rivet	40
1.15 <u>Exercises</u>	42
2 Clocks and Rods in Motion	43
2.1 The Perfect Clock	43
2.2 Synchronizing Clocks within a Single Frame	45
2.3 Moving Clocks Run Slow, Moving Rods Shrink	47
2.4 <u>Exercises</u>	50
3 The Algebra of Frames	54
3.1 Inertial Frames of Reference	54
3.2 Vector Space Structure of Frames	55
3.3 Several Parallel Moving Frames	56
3.4 Six Rules for Frames	58
3.5 <u>Exercises</u>	64
4 The Graphing of Frames	66
4.1 The Filmstrip Model of Spacetime	66
4.2 Constant Velocities in Spacetime	69
4.3 Worldlines are Parallel to the Home Frame Time Axis	71

4.4	Simultaneous and Static Events	72
4.5	Linearity of Line-of-Sight Functions	75
4.6	<u>Exercises</u>	80
II.	<u>GALILEAN TRANSFORMATIONS OF FRAMES</u>	83
5	Galilean Transformations	84
5.1	Key Ideas	84
5.2	Galilean Spacetime Diagrams	85
5.3	The Galilean Matrix	86
5.4	Pattern of the Galilean Matrix	89
5.5	Addition of Speeds via Matrices	90
5.6	Addition of Speeds via Areas	93
III.	<u>The SPEED OF LIGHT IS CONSTANT</u>	96
6	Constant c in Spacetime	97
6.1	Minkowski Spacetime Diagrams	97
6.2	Constant c and Simultaneity	98
6.3	How Constant c Destroys Simultaneity	100
6.4	<u>Exercise</u>	102
IV.	<u>LORENTZ TRANSFORMATIONS OF FRAMES</u>	103
7	Lorentz Transformations	104
7.1	The Lorentz Matrix	104
7.2	Pattern of the Lorentz Matrix	109
7.3	The Lorentz Sum of Speeds	109
7.4	Addition of Speeds via Matrices	111
7.5	Addition of Speeds via Areas	112
7.6	<u>Exercises</u>	116
8	The Hyperbola of Time-Stamped Origins	118
8.1	Invariance of Minkowski Length	118
8.2	The Time-Stamped Origins Theorem	120
8.3	Interpreting the Time-Stamped Origins Theorem	121
8.4	Tangent Lines of Simultaneity	122
8.5	<u>Exercises</u>	125
V.	<u>GRAPHIC RESOLUTION OF THE PARADOXES</u>	126
9	The Accommodating Universe Paradox	127
9.1	Preview	127

9.2	Setup for the Minkowski Diagram	127
9.3	Resolving the Accommodating Universe	129
9.4	<u>Exercises</u>	131
10	The Length-Time Comparison Paradoxes	132
10.1	An Overview of the Paradoxes	132
10.2	Resolving the Mutual Length-Time Paradoxes	137
10.3	Summary	138
10.4	<u>Exercises</u>	139
11	The Twin Paradox	141
11.1	An Overview of the Paradox	141
11.2	A Simplifying Assumption	141
11.3	Setup for the Minkowski Diagram	143
11.4	Resolving the Twin Paradox	143
11.5	General Relativity Confirmation	147
11.6	<u>Exercises</u>	151
12	The Train-Tunnel Paradox	153
12.1	An Overview of the Paradox	153
12.2	A Distance Lemma	155
12.3	The Train-Tunnel Minkowski Diagram	158
12.4	Explaining Mutual Contraction	159
12.5	Resolving the Train-Tunnel Paradox	160
12.6	<u>Exercises</u>	161
13	The Pea-Shooter Paradox	163
13.1	An Overview of the Paradox	163
13.2	The Fizeau Experiment: Adding Speeds	164
13.3	<u>Exercises</u>	168
14	The Bug-Rivet Paradox	171
14.1	The Minkowski Diagram	171
14.2	Coordinates in the Minkowski Diagram	174
14.3	The Slinky Connection	179
14.4	<u>Exercises</u>	181
VI.	ENERGY AND MASS	184
15	$E = mc^2$	185
15.1	How We Came to This Place	185
15.2	Speed-Dependent Mass: an Intuitive View	186
15.3	Equivalence of Mass and Energy	190
15.4	A Numerical Example	192

15.5 <u>Exercises</u>	194
VII. THE MATHEMATICS OF WAVES AND LIGHT	198
16 The Nature of Waves	199
16.1 Propagated Waves	199
16.2 Speed of Rope Wave is Constant	204
16.3 Shapes Traveling in One Dimension	204
16.4 The Wave Equation in One Dimension	209
16.5 Wave Propagation: The Skipping Stone Model	210
16.6 The Doppler Effect in Spacetime	213
16.7 <u>Exercises</u>	216
17 Measuring the Speed of Light	217
17.1 Early Thoughts on the Speed of Light	217
17.2 Rømer: The Speed of Light is Finite	218
17.3 Fizeau Measures the Speed of Light	221
17.4 de Sitter: c Independent of Source Speed	225
17.5 Michelson-Morley's Happy Failure	228
17.6 <u>Exercises</u>	234
VIII. MAXWELL'S EQUATIONS	238
18 Maxwell's Mathematical Toolkit	239
18.1 Preface	239
18.2 Language and Proportionality	240
18.3 1D Lengths & 2D Areas as 3D Vectors	241
18.4 Orientations of Lines and Surfaces	244
18.5 Vectors Modeling Reality	247
18.6 Inner and Cross Products	249
18.7 Riemann Sums and Integrals	251
18.8 Integrals of the Inner Product	255
18.9 <u>Exercises</u>	259
19 Electric and Magnetic Fields	262
19.1 Background	262
19.2 Electric Forces: Coulomb's Law	263
19.3 Electric Fields	266
19.4 Magnetic Fields	268
19.5 Magnetic Forces: Lorentz Forces	270
19.6 How Thomson Discovers the Electron	272
20 Electricity and Magnetism: Gauss' Laws	276
20.1 Flux of Vector Fields	276

20.2 Electric and Magnetic Flux	279
20.3 Gauss' Law for Electricity	280
20.4 Gauss' Law for Magnetism	282
20.5 <u>Exercises</u>	283
21 Towards Maxwell's Equations	289
21.1 Biot-Savart Law: <i>Magnetism from Electricity</i>	289
21.2 Quantitative Results for Biot-Savart	290
21.3 Ampère's Law	292
21.4 Maxwell Adds to Ampère's Law	294
21.5 Faraday's Law: <i>Electricity from Magnetism</i>	296
21.6 Lenz's Law: <i>The Positive Side of Negativity</i>	297
21.7 Maxwell's Four Equations	299
21.8 <u>Exercises</u>	301
22 Electromagnetism: A Qualitative View	304
22.1 Magnetic Waves from an Infinite Wire	304
22.2 Wave Propagation	306
22.3 The Geometry of Electromagnetism	310
23 Electromagnetism: A Quantitative View	313
23.1 Quantitative Preliminaries	313
23.2 A Quantitative View of Propagation	316
23.3 Theoretical Speed of Wave Propagation	323
23.4 Maxwell's Calculation of c	326
23.5 Mathematical Hits	328
23.6 <u>Exercises</u>	331
<u>IX. FINAL THOUGHTS</u>	334
24 Epilogue: Final Thoughts	335
24.1 A Coming of Age	335
24.2 Einstein's Annus Mirabilis	337
24.3 Comparing Relativities	339
24.4 Against Conventional Wisdom	343
24.5 Some Experimental Results	345
24.6 Bad Assumption, Good Result	348
24.7 A Limited Reality	349
24.8 PIES Reality	353
24.9 <u>Exercises</u>	355
<u>X. APPENDICES</u>	358
A Linear Algebra Overview	359

A.1 Mathematics as a Conduit to Reality	359
A.2 Vector Spaces	360
A.3 Functions	366
A.4 Linear Functions and Matrices	368
A.5 Eigenvectors and Eigenvalues	374
B Hyperbolic Functions	376
B.1 Overview	376
B.2 Even and Odd Functions	377
B.3 Invariant Areas of Transformed Hyperbolas	378
B.4 <u>Exercises</u>	383
C Deconstructing a Moving Train	385
C.1 Motion Alters Age	385
C.2 Minkowski Diagram for a Moving Train	385
C.3 <u>Exercises</u>	387
XI. SUPPLEMENTAL MATERIAL	388
D Dimensional Analysis	389
D.1 Unitless Quotients of Dimensions	389
D.2 Dimensions in Fractions	390
D.3 <u>Exercises</u>	393
E Rings of Functions and Square Matrices	396
E.1 Associative, Binary Operations	396
E.2 Rings over the Real Numbers	399
E.3 The Ring of Matrices	400
E.4 <u>Exercises</u>	401
F The Scientific Method	406
F.1 Reality of the Unseen	406
F.2 If-then Sentences	410
F.3 Property Lists	412
F.4 The Four-Step Scientific Method	413
F.5 Is X a Duck? Applying the Scientific Method.	414
F.6 Whence the Scientific Method?	417
F.7 The Logical Implication	419
F.8 Induction <i>vs.</i> Deduction	419
F.9 Necessary <i>vs.</i> Sufficient	420
F.10 Uncertainty, Popper, and Derrida.	422
F.11 Popper Logic	425
F.12 <u>Exercises</u>	427

G Logic of the Scientific Method	435
G.1 Implications Built from P, Q	435
G.2 Equivalence of Implications	438
G.3 Proof by Contradiction	440
G.4 <u>Exercises</u>	443
Bibliography	447
Index	451



His book is dedicated to my wife,
Susan de Pillis, whose support over the years
allows me to truly understand why authors write dedications.



I. A First Pass

Preface presents introductory remarks.

Chapter 1 offers an intuitive overview of some of the paradoxes in special relativity. Detailed graphical analysis follows in Chapters (9)-(14).

Chapter 2 concerns moving clocks (that always run slow) and moving objects (that always shrink in the direction of motion).



Preface

(0.1) Exposition and Paradoxes

The Nature of Exposition

(0.1.1)



There are two flavors of scientific exposition — the qualitative and the quantitative. We provide both. The qualitative narrative gives us a story, a description of *what* happens. The quantitative exposition also tells us what happens, but adds the *how* and *why* of these happenings.

If we only want to know *what* happens, then we are content with descriptions and assertions. We need no explanation. We are happy to learn that rods in motion shrink in the direction of motion, or that moving clocks always run slower than identical stationary clocks. But we need not know why.

But if we also want to know *why* things happen, then we must use observed measurements along with mathematical structures to represent physical objects and phenomena.¹

(0.1.2)

Why must we use mathematics in our quantitative analysis?

This is a question that confounded Albert Einstein, who wrote, ([13], pg. 28)

“How can it be that mathematics, being after all a product of human thought independent of experience, is so admirably adapted to the objects of reality?”

Einstein was not alone in wondering why mathematics proved necessary to explain Reality. Physicist Eugene Wigner [33] and mathematician R. W. Hamming [19], were also puzzled over the UNREASONABLE EFFECTIVENESS of mathematics.

¹It may be only apocryphal, but it has been said that when Stephen Hawking was writing his book, *A Brief History of Time* [20], he was advised by his editor that each formula he used would cut the sales by half.

(0.1.3) To see how mathematics, a chess game of the mind, proves its “unreasonable” effectiveness, note how often in this book we will take the following INTELLECTUAL ROUND TRIP:

- (i) Consider a natural phenomenon in the “real” world,
- (ii) Represent that phenomenon as a mathematical model,
- (iii) Run a mathematical analysis on that model, and, finally,
- (iv) Return the results of the mathematical analysis back to the original “real” world.

Without this round trip from Reality to mathematics and back to Reality, we would not know, for example, that

- (0.1.4)**
- Moving clocks must run slow (*Figure (1.10.2)*, pg. 26) and moving rods must shrink in the direction of motion (*Figure (1.10.4)*, pg. 27).
 - The front end of a moving train is always younger than its back end (*Appendix (C)*, pg. 385).
 - The assumption of perfectly rigid bodies implies that time can be reversed in the sense that a future event would affect the present ((*14.2.4*), pg. 177).

NOTE: A stark example of this intellectual round trip between Reality and mathematics is the set of Maxwell’s equations (*Chapters (18)-(23)*).

The successes of (0.1.4), drawn from graphs and formulas, show how mathematics can model Reality and take us to truths about the physical world that lie well beyond the apprehension of our five humble input ports, or senses. But beware!



- (0.1.5)** **The Occasional Illusion:** Although we will be applauding the success of mathematical modeling throughout this book, we must also be mindful that we often grasp Reality with a slippery hand. For example, in (24.6), we show how Descartes applied a false hypothesis to obtain a true result — Snell’s Law. Fermat, who disputed Descartes’ premise, applied an alternate (true) hypothesis to arrive at the same result ((*24.6.2*), pg. 348).

As indicated above, this book offers both qualitative and quantitative expositions since the reader may benefit from one style or the other — or both.

The Nature of Paradox

- (0.1.6) **Paradoxes are phenomena** or arguments that challenge our intuition to such a degree that we are tempted to declare the statements to be outright contradictions to what we “certainly” know is true.

Part of the fun is to acknowledge these first impressions of disbelief, and then to resolve them through the force of logic and mathematics, and thereby clarify the contours of Reality. These are characteristics of non-intuitive assertions or paradoxes which we now define:

- (0.1.7)
- DEFINITION:** A PARADOX is a statement or proposition that is *seemingly contradictory or absurd, but, in reality, expresses a truth without contradiction.*

- (0.1.8) **How are paradoxes explained?** As events will reveal, paradoxes often arise from the ambiguity of spoken language and reluctance to abandon “obvious” intuitive truths.

For example, our intuition that tells us that the earth is flat, that the sun moves around the earth, and that atoms, if they exist at all, must be relatively solid — they can not consist mostly of empty space.

More generally, we might say that paradoxes arise when we fail to apply CRITICAL THINKING² which includes the *questioning and evaluating of basic assumptions.*

² *Critical thinking* is not the same thing as criticizing. The declaration, “I don’t like those shoes you are wearing!” is not necessarily a product of critical thinking which must be based on *clear identification of assumptions* and their logical consequences.

(0.2) Organization of this Book

- Preface: *This chapter*

SR³ QUALITATIVE MATERIAL (*what* the paradoxes are)

- **Chapter (1) Introduction to the Paradoxes:** This is an introductory overview of six paradoxes in special relativity without full explanations or proof.
 - **Chapter (2) Clocks and Rods in Motion:** It is stated, without extensive proof, that moving clocks always run slow and moving objects always shrink in the direction of motion.
-

SR QUANTITATIVE MATERIAL (*questions of why and how*)

- **Chapter (3) The Algebra of Frames:** The inertial frame, \mathcal{F} , the basic setting for special relativity, is a choice of an origin, a coordinate system and clocks at every point in space, which together allow a unique specification of the location and time of each event. Without loss of generality, we can restrict the three space coordinates to one space coordinate x (1.3.1). Examples of inertial frames experiencing no forces with a clock at each point would be a railroad car or a ruler moving at constant speed. Distance x and time t at that point x defines an ordered pair, $[x, t]$, the set of which is a *two-dimensional vector space* (Section (3.2)). Finally, we present the six rules for frames (Section (3.4)).
- **Chapter (4) The Graphing of Frames:** Since an inertial frame is mathematically modeled as a two-dimensional vector space (Ch. (3)), each frame can also be represented as a two-dimensional *spacetime graph*. Each ordered pair $[x, t]$ can then be pictured as a point on the spacetime graph that has coordinate axes X for distance and T for time.

The tools of linear algebra now come into play: If we have two frames, \mathcal{F}^A and \mathcal{F}^B , say, moving parallel to each other with a constant relative speed (*imagine two railroad cars on parallel tracks*), then to peek into one frame from another along a virtual *line-of-sight*, is to establish a relationship or pairing between the observer point $[x, t]_A$ in frame \mathcal{F}^A and the observed point $[x', t']_B$ in frame \mathcal{F}^B . The set of all pairings, $[x, t]_A \rightarrow [x', t']_B$, defines a

³SR is an abbreviation for *Special Relativity*.

mathematical *function* that also turns out to be, in mathematical terminology, a *linear transformation* ((A.4.1b), pg. 368).

- **Chapter (5) Galilean Transformations:** The standard tools of linear algebra allow us to create a unique 2×2 matrix that totally reflects the properties of the corresponding linear transformation $[x, t]_A \rightarrow [x, t]_B$ described in Chapter (4). What does this matrix look like? Since Galileo assumed that there was only one universal clock by which all observers in all frames can simultaneously synchronize their watches (*an intuitive idea behind railroad timetables, and one that we see in spy movies all the time*), it follows, in a non-obvious way, that the speed of light must be infinite. Assuming infinite speed of light, we arrive at a unique 2×2 matrix called the Galilean transformation matrix that allows us to view between the pair of inertial frames along line-of-sight corridors ((5.4.1a), pg. 90).
- **Chapter (6) Constant c in Spacetime:** By 1905, c , the speed of light in a vacuum, was determined not to be infinite, but to have a value of 186,000 miles per second. Einstein's fundamental assumption is that this single value of c must be the same for all observers regardless of the speed of either the light-source or the observer (*No matter how fast you chase after a photon of light, its speed is still the constant c*). With this additional assumption, spacetime graphs are called *Minkowski diagrams*. With the aid of Minkowski diagrams, we learn that two events occurring simultaneously in one frame can never occur simultaneously in another frame that is moving relative to the first frame. That the constant speed of light destroys simultaneity (*Figure (6.2.3)*) is anything but clear from just the reality of our five senses.
- **Chapter (7) Lorentz Transformations:** Since the speed of light c is *not* infinite and is measured to be constant by *all* observers regardless of motion, we are able to use the geometry of the Minkowski diagram to derive the linear transformation $[x, t]_A \rightarrow [x, t]_B$ that pairs an observer at $[x, t]_A$ in \mathcal{F}^A viewing an observer at $[x', t']_B$ in \mathcal{F}^B through a virtual line-of-sight corridor that runs between the two (parallel) frames. This pairing, the *Lorentz linear transformation*, has a corresponding 2×2 matrix representation called the *Lorentz matrix* ((7.2.2), pg. 109). With this non-Galilean matrix in hand, we find that speeds no longer add in the “usual” way that accords with our experience. We show this to be true analytically through a matrix analysis (*Section (7.4)*) and then geometrically using areas (*Section (7.5)*)

- **Chapter (8) *The Hyperbola of Time-Stamped Origins:*** Consider infinitely many parallel frames, each moving with a different relative speed. For fixed time t_0 in each frame, plot the respective origins in the Minkowski X - T diagram. The resulting curve is a well-defined hyperbola. Different fixed times t_0 produce different hyperbolas. (*This hyperbola reduces to a horizontal line in Gaussian/Newtonian physics.*) Each moving frame \mathcal{F} gives rise to a tangent line (*of the t_0 hyperbola*) which then corresponds to all the t_0 simultaneous events in frame \mathcal{F} .
- **Chapter (9) *The Accommodating Universe:*** The theory of special relativity holds that particles and signals can never travel faster than the speed of light. However, conditions are described under which an intergalactic traveler (skateboarder) apparently does travel faster than light. The analysis reveals how this “contradiction” is a consequence of the (mis)use of language.
- **Chapter (10) *The Length-Time Comparison Paradoxes:*** Let two inertial frames be in parallel constant motion relative to each other. Then observers in each frame see the clocks in the other (*moving*) frame run slower than the clocks of their own (*stationary*) frame. Similarly, identical rods in the other (*moving*) frame are shorter than the rods in their own (*stationary*) frame. This chapter answers the questions: How can clocks in each frame simultaneously run slower than clocks in the other frame? How can identical rods in each frame simultaneously be shorter than rods in the other frame?
- **Chapter (11) *The Twin Paradox:*** On the day of their birth, one twin sets off in space in a very fast rocket while the other twin stays on Earth. After 25 Earth years, the traveling twin’s space ship turns around and takes another 25 years to return to Earth. The paradox is that the Earth-bound twin has aged 50 years while the space ship twin, upon return, has aged only 30 years. On one hand, all moving clocks run slow, so naturally, the moving space ship clock slows down and reads only 30 years instead of 50 years on its return to Earth. On the other hand, it is the Earth clock that is moving as observed from the space ship. If so, then why doesn’t the (*moving*) Earth clock run slower than the (*stationary*) space ship clock?
- **Chapter (12) *The Train-Tunnel Paradox:*** Viewed from the tunnel, a moving train inside the tunnel avoids the shock of simultaneous lightning bolts that strike the tunnel entrance and the exit. Viewed from the train, the (*moving*) tunnel becomes shorter than

the train. Do the (*simultaneous?*) lightning bolts now strike the “too-long” train?

- **Chapter (13) *The Pea-shooter Paradox:*** Our intuition tells us that if a train is going at $u = 50$ mph and a ball on the train is thrown forward at $v = 20$ mph, then an observer standing outside the train will see the ball traveling at $u+v = 20+50 = 70$ mph. Not so. The actual (Lorentzian, or relativistic) sum of speeds, is always $u \oplus v = (u+v)/(1+uv/c^2)$ which, for small speeds, is imperceptibly less than the familiar combined speed $u + v$. There are two ways to derive $u \oplus v$: analytically through matrix calculation (7.4) and graphically through addition of areas (7.5).
- **Chapter (14) *The Bug-Rivet Paradox*** is one of the least explored paradoxes of special relativity. The assumption of absolute rigidity — where objects can never be stretched or compressed — implies that *the future affects the present!* Using a principle demonstrated by the Slinky toy (14.3), special relativity tells us that a speeding rivet (which, from a stationary frame, is seen to be shortened), reverts to its at-rest length after its speed is reduced to zero. In our example, (14.1), a certain collision is designed to bring the rivet speed to zero. However, perfect rigidity implies that the rivet will “sense” the impending collision and will start to reach its at-rest length even *before* the collision occurs.

ENERGY and MASS

- **Chapter (15) $E=mc^2$** This celebrated equation is almost an after-thought once the connection is made between the pea-shooter paradox, Chapter 13, and conservation of momentum.

THE MATHEMATICS of WAVES and LIGHT

- **Chapter (16) *The Nature of Waves:*** We develop the mathematical model of propagated waves since light exhibits the properties of a propagated wave under certain circumstances.
- **Chapter (17) *Measuring the Speed of Light:*** We offer details of some of the earliest and ingenious methods for measuring c , the finite speed of light. Even in the time of Galileo and Newton, it could not be proven that the speed of light was finite. In this chapter, you will see how you can use your home microwave oven to measure c in a 20-second experiment (*Exercise (17.6.6d)*, pg. 237).

MAXWELL'S EQUATIONS (*a unification of theories of electricity, magnetism, and optics*)

- **Chapter (18) Maxwell's Mathematical Toolkit:** At the very outset of his seminal paper in special relativity [12], Einstein acknowledges the work of Clerk Maxwell and his unifying work in which he developed just four equations ((21.7.3), pg. 300) that unified the theories of electricity, optics, and magnetism. In this chapter, we see how observable (one-dimensional) lengths and (two-dimensional) areas are interpreted mathematically as three-dimensional vectors, or arrows (See *Vectors Model Reality* (18.5), pg. 247).
- **Chapter (19) Electric and Magnetic Fields:** This is an overview of electricity and magnetism basics including Coulomb's Law (19.2.4b)
- **Chapter (20) Electricity and Magnetism. Gauss' Laws:** After developing the notion of flux, we state Gauss' Law for both electricity (20.3.1) and magnetism (20.4.1).
- **Chapter (21) Towards Maxwell's Equations:** After discussion of the Biot-Savart Law (21.1), the Ampère Maxwell Law (21.3), along with the laws of Faraday (21.5) and Lentz (21.6), we finally present the elegant four equations that Maxwell fashioned (21.7) to embrace all the laws previously stated.
- **Chapter (22) Electromagnetism. A Qualitative View:** We offer a general description of the geometry of electromagnetic wave propagation.
- **Chapter (23) Electromagnetism. A Quantitative View:** An analytic explanation of topics in the previous chapter are presented. We see how Maxwell calculated c , the speed of light. (*Actually, he calculated the speed of electromagnetic waves, not knowing at the time that light was one example of electromagnetism*). He did this by merely computing $c = \sqrt{\epsilon_0\mu_0}$, the square root of the product of ϵ_0 , the permittivity (19.2.5a) from Coulomb's law, and μ_0 , the permeability (21.2.2) from the Biot-Savart Law. This is an amazing mathematical prediction.

FINAL THOUGHTS

- **Chapter (24) Epilogue. Final Thoughts:** We find an overview of Einstein's Annus Mirabilis of 1905 (24.2), a comparison of the

Galileo/Newton relativities with that of Einstein/Lorentz (24.3), a note on the reliability of Conventional Wisdom, and a description of a feud between Descartes and Fermat, each of whom derived Snell’s Law (24.6). The difference is that Descartes falsely assumed that light would gain speed as it entered water, while Fermat correctly asserted that light would slow down. Yet they both arrived at the same conclusion.

APPENDIX

- **Appendix (A) Linear Algebra Overview:** A brief exposition of vector spaces, functions, and linear transformations along with their matrix representations, eigenvalues and eigenvectors.
 - **Appendix (B) Hyperbolic Functions:** Graphical properties of the hyperbola $f(x) = 1/x^2$ become the mathematical foundation for Section (7.5), pg. 112, *Addition of Speeds via Areas* in the special case that the speed of light $c = 1$. Properties of hyperbolas and their areas are central to developing a graphical resolution of the pea-shooter paradox.
 - **Appendix (C) Deconstructing a Moving train:** A railroad car at constant speed is a physical object. When modeled mathematically, it becomes an inertial frame (*a ruler with a clock at each point* ((1.5.1), pg. 16)). What is not obvious in the “real” physical reality, is that observers on a platform will see that the front of the train is younger than the rear of the train. The Minkowski diagram ((C.2.1), pg. 386) justifies this conclusion.
-

(0.2.1)

Acknowledgments. Lectures from this book were presented in the winter of 2005 at the University of Padua, Italy, with the generous support of Professoressa *Maria Morandi Cecchi*, Dipartimento Matematica Pura ed Applicata. The authors benefited from conversations with Professors *Jack Denny* of Arizona State University, *Al Kelley* of the University of California, Santa Cruz, Professors *Fred Metcalf*, *Larry Harper*, *David Ellerman*, and *John Baez* of the University of California, Riverside, Professor *John Mallinckrodt*, California State Polytechnic University, who commented on the paradoxes, *Dr. Joan Kleypas* of the National Center for Atmospheric Research in Boulder, Colorado, and *Richard Reeves*,. Special thanks are owed to *Jerry Alexanderson* and particularly to *Don Albers* of the Mathematical Association of America who was consistently supportive of this book with its color illustrations, and the animations.

(1) Introduction to the Paradoxes

(1.1) Aristotle vs. Galileo

(1.1.1) ARISTOTLE (384–322 B.C.) believed that a stone fell toward the earth because the stone and the earth were both in the “Earth” category among the four basic elements, Earth, Air, Fire, and Water. According to this reasoning, smoke, which consists of air and fire, wants to be closer to the sky (air) and further from an unlike element (earth). Hence, earthly objects move naturally toward the earth while “airy” objects move naturally upward. The assumed fifth element, the heavenly substance he called the QUINTESSENCE along with an assumed PRIME MOVER accounted for the “perfect” circular and uniform motion of the heavens. Finally, Aristotle asserted that motion of any object required a continuing force on that object as long as motion persisted.

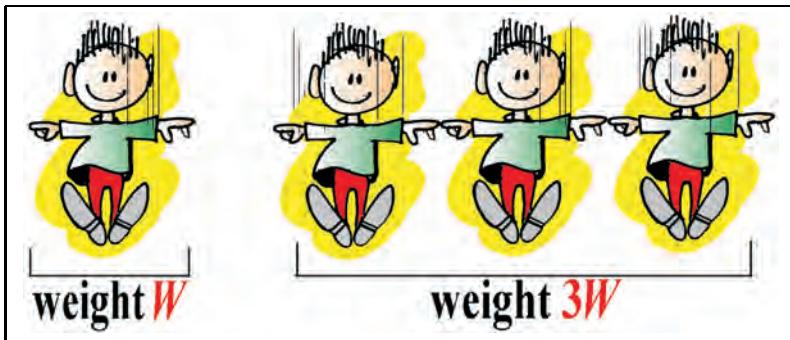
As if to simplify the discussion, ST. THOMAS AQUINAS (1225–1274) later declared that something is in motion when it is both *what it is already* and *something else that it is not yet*. (*Apparently, new explanations do not always lead to clarity.*)

In a complete departure from the convoluted explanations of Aristotle and Aquinas, GALILEO GALILEI (1564–1642) claimed that a force was *not* necessary to maintain constant, straight-line motion. This property is in evidence even today where it can be observed that an object in space will drift along forever without external forces.

Galileo’s thesis was that a force was necessary only to produce a *change* in speed, that is, an acceleration or deceleration. For example, it is the force of friction that brings a rolling ball to a stop on a flat surface.

Aristotle and Galileo disagreed on another point. Where Aristotle said a heavy body would fall faster than a lighter one (*Doesn’t a rock fall faster than a feather?*), Galileo claimed that two bodies, when dropped from the same height, always fall at the same rate, when acted upon by the force of gravity alone. (*The effect of friction due to air resistance must be negligible, which is not the case with a falling feather.*)

(1.1.2)

FIGURE: Galileo's Thought Experiment for Falling Bodies

A body of weight W (left panel) falls at the same rate as the heavier body of weight $3W$ (right panel) since it can be considered as three independent bodies, each of weight W .

If Aristotle were correct in claiming that heavier bodies fall faster than lighter ones then once the three individuals of Figure (1.1.2) touch fingers, forming a collective weight $3W$, they would start to fall faster. And once their fingers separated, the speed of each individual, being only W units, would slow down.

(1.2) Frames of Reference

A Frame of Reference is a way of specifying the position of each event with respect to a given point, and the time at which the event occurred.

For example, while directing a tourist from a train station to her hotel room, we need to tell her how many meters east she should walk, then how many north, and how many stories up the building she should go; we can also tell her shell be there a 4 PM. So we imagine three lines, one east-west, one north-south and one up-down, and a clock at every point; That is, from our location (the “origin”), we can specify any point and any time—that is, any event—by providing 4 numbers.

In sum, the chosen set of lines and clocks represent a reference frame. Using this frame we can tell the where and when of any event. The corresponding 4 numbers are the coordinates of that event with respect to that reference frame.

Here is a more formal definition:

(1.2.1)

DEFINITION: A FRAME OF REFERENCE, \mathcal{F} , is a choice of an origin, a coordinate system and clocks at every point in space, which together allow a unique specification of the location and time of each event.

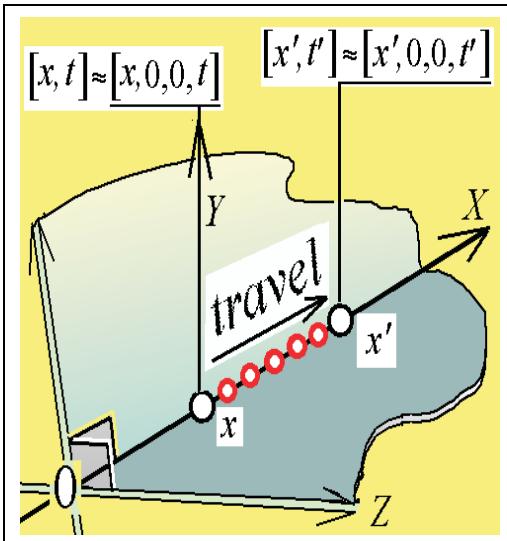
(1.3) Straight-Line Trajectories in 3-Space

(1.3.1)

Representing 3D Motion as 1D Motion: If the motion of a particle is only along a straight line in a space with X , Y , and Z -axes, then we can choose our X -axis, which is also a straight line, to exactly coincide with that straight line path in 3-space. As illustrated by Figure (1.3.2), the particle traveling along this X -axis has spacetime coordinates of the form $[x, 0, 0, t]$ —the Y and Z coordinates are fixed at 0 while the x and t values vary with time. Hence, the coordinates $[x, t]$ contain the same information about the particle's motion as does $[x, 0, 0, t]$.

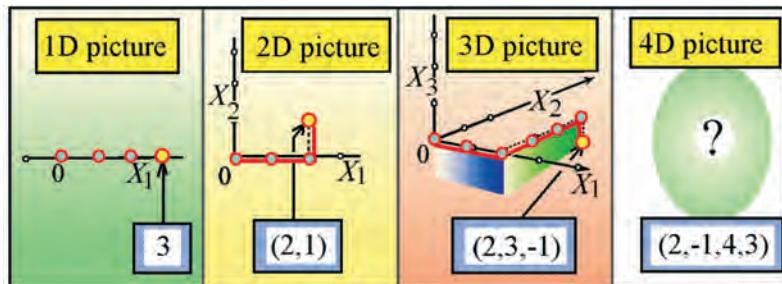
(1.3.2)

FIGURE: 1D Representation of 3D Straight-Line Trajectory.



Without loss of generality, we choose the X - Y - Z axes so that the X -axis falls on the straight-line trajectory of an object in 3-space. Two positions (x coordinates) are shown of an object traveling on the X -axis at times t_0 and t_1 . The two positions are economically represented, without loss of information, by the ordered pairs, $[x, t]$ and $[x', t']$ — no y or z values are needed.

(1.3.3)

FIGURE: When Pictures Fail

Although time is involved, Figure (1.3.2) does not show a separate time axis, T which would necessarily add an extra dimension to the 3-dimensional graph. Such a complete spacetime diagram, or frame of reference \mathcal{F} (1.2.1) with coordinates $[x, y, z, t]$, would require four space dimensions to plot which, in our 3-dimensional universe, we cannot do.

But if the trail of a uniformly moving particle in 3-space can be represented as having only one space dimension (and one time dimension), i.e., we need only two fixed coordinates, as in (1.3.2), then we can represent all the information of this object's motion in a 2-dimensional spacetime graph.

(1.4) Galilean Relativity

(1.4.1)

It was Galileo who first formulated a principle of relativity where speeds are measured relative to a frame of reference (1.2.1). This principle can be observed, for example, when a passenger on an airplane (*frame of reference*) traveling at 600 mph relative to the ground (*another frame of reference*) throws a ball vertically in the air. The ball travels on a vertical up-and-down path *relative to the passenger's frame of reference*. The



ball does not, as intuition and the illustration might suggest, snap toward the back of the airplane at 600 mph.

Galileo considered two ships (frames of reference (1.2.1)), each of which has an experimenter below deck in a cabin without windows. Each ship is moving with constant speed, which could be zero, relative to land. Galileo then asserted that there is no experiment that the experimenters can devise that would reveal whether or not their respective ships are in motion. Moreover, all their experiments within the ship (e.g., balls rolling down tilted planes, collisions of billiard balls) would produce identical results.

- (1.4.2) Using the model of two ships, each of which is a separate frame of reference, Galileo stated his relativity principle as follows:

Shut yourself up with some friend in the main cabin below decks on some large ship, and have with you some flies, butterflies, and some other small flying animals...hang up a bottle that empties drop by drop into a wide vessel beneath it ... have the ship proceed with any speed you like, so long as the motion is uniform and not fluctuating this way and that ... The droplets will fall ... into the vessel beneath without dropping toward the stern, although, while the drops are in the air the ship runs many spans ... the butterflies and the flies will continue their flights indifferently toward every side, nor will it ever happen that they are concentrated toward the stern, as if tired out from keeping up with the course of the ship.

The point of this quote is that passengers on the ship can not determine whether the ship is moving (with constant speed) relative to the shore unless they view the shore from their cabin. Within the cabin frame of reference, the ship speed is zero, while in a frame anchored to the shore, the ship may have a non-zero speed. Speed or velocity is therefore relative to the frame from which it is measured.

This quote can be found in a 1953 translation by Stillman Drake [9] and in the more recent 2004 book by Roger Penrose [27], *The Road to Reality*.

(1.5) Special Relativity: A First Pass

Briefly put, SPECIAL RELATIVITY is the study of physical frames of reference (1.2.1)—coordinate systems with clocks at each point—with the additional condition that the frames move at constant speed relative to each other. Inertial frames experience no forces or acceleration and objects or observers within each frame are fixed at their respective positions. The PRINCIPLE OF RELATIVITY asserts that all laws of physics, including those of electromagnetism and quantum mechanics, are equally valid in each inertial frame. In sum,

(1.5.1)

DEFINITION: An INERTIAL FRAME OF REFERENCE is a frame of reference [(1.2.1), pg. 13] which is non-accelerating.

In an *inertial* frame, whose objects experience no forces, a freely-moving ball on a horizontal friction-free table would remain stationary.

The theory of special relativity is based on the following assumptions:

(1.5.2)

TWO BASIC PREMISES (POSTULATES)

- (1.5.2a) The laws of physics for mechanics (for objects), electromagnetism, and quantum theory, are the same and equally valid for all inertial frames of reference (points of view) (1.2.1).
- (1.5.2b) The speed of light in a vacuum, denoted c , is the same for all observers, regardless of the speed of its source, or the speed of the observers.

(1.5.3)



Assumptions, postulates, axioms, or premises, like those of (1.5.2), are, by definition, fundamental properties, observations, or “starting points” that are given without proof. Scientists and mathematicians hold to the importance of clearly identifying these basic and unproven assumptions. This is so that we can better understand their consequences — theorems and properties that necessarily follow. Chapter F on the scientific method and chapter G on mathematical logic expand on this principle.

(1.5.4)

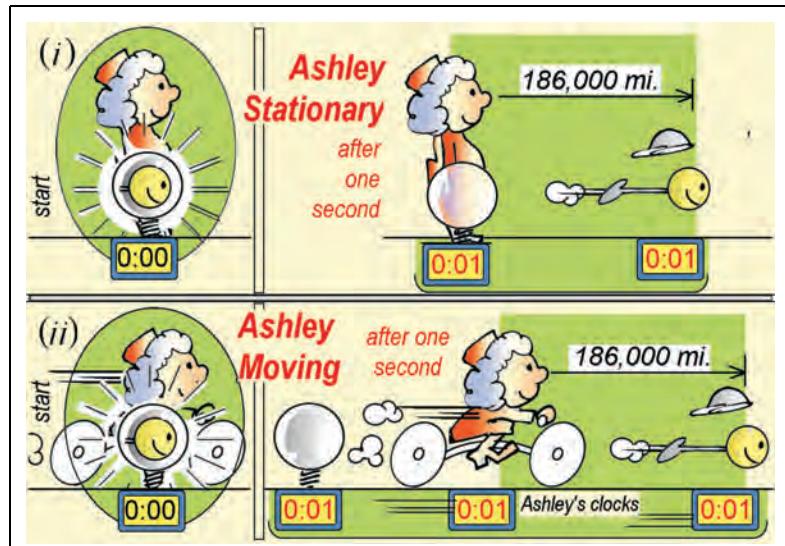


Assumption (1.5.2b) is, admittedly, anti-intuitive. Although you can run after a *material* object, like a train or a ball in flight, to make its speed appear slower (relative to you), you can not do this with *electromagnetic* light “particles” or photons.

If you release photons from a light source, then you will always measure their speed to be **186,282 miles per second**, or **2.99792458×10^8 meters per second**.¹ regardless of your motion or the motion of the light source.

(1.5.5)

FIGURE: Chasing a Photon from a Light Beam



At time **0:00** seconds, Ashley is synchronized with the emitted photon at the light source. One second later, Ashley’s clock reads **0:01**. Ashley will always find herself separated from the emitted photon by a distance of

¹ We approximate these values with $c \approx 186,000 \text{ miles/hr} \approx 3 \times 10^8 \text{ meters/sec}$.

186,000 miles, whether she is

- (i) stationary relative to the light source or
 - (ii) moving relative to the light source.
-

(1.5.6)

The Importance of Maxwell's Work. In listing the two postulates of special relativity (1.5.2), one might think they are independent of each other. In fact, if the laws of electromagnetism, i.e., Maxwell's equations, are valid in all inertial frames, as is implied by (1.5.2a), then we must accept the consequences of these equations, one of which is (1.5.2b), namely, c , the speed of all electromagnetic waves in a vacuum (*and the speed of light in particular*) is the same in all frames from which it is measured. (See Section (23.4), pg. 326, to see how Maxwell calculated c using just two laboratory measurements.)

(1.6) A Symmetry Principle

If frame \mathcal{F}^A measures the relative speed of frame \mathcal{F}^B to be rightward, at, say, 30 miles per hour, then is it possible for frame \mathcal{F}^B to measure the relative speed of frame \mathcal{F}^A to be leftward at 20 miles per hour?

Such a lack of symmetry is not possible if all inertial frames are equal in the sense that no frame is any more special or unique than any other frame. Thus, if \mathcal{F}^A sees frame \mathcal{F}^B moving rightward at 30 miles per hour, then frame \mathcal{F}^B must see frame \mathcal{F}^A moving leftward at 30 miles per hour, or at -30 miles per hour.

We state this symmetry principle more formally:..

(1.6.1)

THE SYMMETRY PRINCIPLE FOR RELATIVE SPEEDS. *Inertial frames \mathcal{F}^A and \mathcal{F}^B are moving parallel to each other with constant speed. If observers in frame of reference \mathcal{F}^A measure the speed of frame \mathcal{F}^B to be v in the positive direction, then conversely, observers in \mathcal{F}^B will measure the speed of frame \mathcal{F}^A to be $-v$, i.e., observers in \mathcal{F}^B measure the speed of \mathcal{F}^A to have the same magnitude but the negative (opposite) direction.*

(1.7) Lorentzian Relativity

- (1.7.1) || **Note:** At this point, our exposition is purely descriptive—we do not explain how these results arise. Proof of why moving clocks run slow, and why moving rods shrink, will come later in Theorem (2.3.3).

- (1.7.2) Lorentzian Relativity _____

The following results, (1.7.2a) and (1.7.2b), are inevitable consequences of the two assumptions noted in (1.5.2), along with the symmetry principle for speeds (1.6.1). Suppose a clock and an object are fixed in frames that are observed to be moving relative to an observer in another frame. Then

- (1.7.2a) The moving clock is always *observed* to run slower than stationary clocks. (*This effect is too small to be observed until the clock reaches speeds close to the speed of light.*) Within its own frame, a stationary observer does not see an identical clock running any slower.
- (1.7.2b) The length of the moving object physically contracts in the direction of its motion as measured by the stationary observer. Within its own frame, a stationary observer does not measure an identical rod to be any shorter.

A further consequence of these two results concerns the unique rôle of the speed of light in a vacuum, namely,

- (1.7.2c) The speed of light in a vacuum is an upper limit for the speed of all matter and all signals in the Universe—no material object or wave can achieve a speed greater than the speed of light in a vacuum.

(1.8) The Ubiquitous Shrinkage Constant

- (1.8.1) In (1.7.2a) and (1.7.2b), the clock and the object are fixed in frames that are observed to move at a specific speed v . The observed “shrinkage” in length and time is quantified by the speed-dependent con-

stant, σ_v which we define next in (1.8.2).

(1.8.2)

DEFINITION: *The speed-dependent Lorentz contraction factor is defined to be*

$$\sigma_v \stackrel{\text{def}}{=} \sqrt{1 - (v/c)^2} < 1 \quad \text{whenever speed } v < c.$$

(This value of σ_v will be justified in (7.1.3b), pg. 106.)

Specifically, suppose clocks are SYNCHRONIZED so that the clocks at the origin $x = 0$ in each frame read time $t = 0$ (*A more formal definition will be found at (3.4.5a), pg. 62*). Then the moving clock will always tick at a rate $\sigma_v \leq 1$ times the rate of the stationary clock. Similarly, if d_0 is the observed rod length in a stationary frame, then observers in the stationary frame observe the length of the moving rod to be $\sigma_v d_A = d_0$.

(1.8.3)

NOTE: *One proof of the fact that moving rods shrink, and moving clocks slow down* is found in Theorem (2.3.3), pg. 49. This theorem uses properties of Figure (2.3.2) which are based only on the Pythagorean Theorem.

The properties of length contraction and time dilation are summarized in the following table.

(1.8.4)

TABLE: *At-Rest and Speed- v Lengths and Times*

	Observed at rest	Observed at speed v
Clock time	(t_0) hrs.	$(\sigma_v \cdot t_0)$ hrs.
Rod length	(d) ft.	$(\sigma_v \cdot d)$ ft.
$\sigma_v = \sqrt{1 - (v/c)^2}$		

NOTE: *Objects do not shrink when measured in their own frames and clocks do not slow down when time is measured in their own frames.*



(1.8.5) **The online animation² of Lorentz frames** shows that in relativistic physics, a ruler contracts to 80% of its at-rest length when it travels at relative speed $0.6c$ (*Note that $\sigma_v = 0.8$ when $v = 0.6c$ in (1.8.2)*). The online animation of **Galilean frames** shows that in classical Galilean physics, a ruler traveling at any speed enjoys no length contraction.

The following examples are products of the contraction formula (1.8.2).

(1.8.6) **EXAMPLE: Any Degree of Contraction is Achievable.** From (1.8.2), we see that the number $\sigma_v = \sqrt{1 - (v/c)^2}$ can be as close to zero as we like as long as speed, v , is close enough to c , the speed of light in a vacuum. (*More formally, we write " $v \rightarrow c$ implies $\sigma_v \rightarrow 0$ ".*) From Table (1.8.4), any rod of length d can have an arbitrarily small observed length, $(\sigma_v \cdot d)$, if its speed, v , is close enough to c .

As a special case of (1.8.6), we have

(1.8.7) **EXAMPLE: The Ever-Shrinking Tunnel.** A tunnel or corridor has an at-rest length of 100 miles. Therefore, from (1.7.2b) or (1.8.6), a traveler moving through the tunnel who sees the tunnel flying by at speed v can make that tunnel shrink in the direction of travel to any small non-zero length, namely, $(\sigma_v \times 100)$ miles. From the traveler's point of view, the moving tunnel's length will be multiplied (shrunk) by the Lorentz contraction factor, $0 < \sigma_v = \sqrt{1 - (v/c)^2} \leq 1$ which can be made as small as we like if v is taken close enough to c .

However, from the point of view of an outsider who is *not* moving relative to the tunnel, the tunnel's length remains at the fixed length of 100 miles.

When we compare similar observations and measurements of an observer *outside* a tunnel with those of a moving skateboarder *inside* the tunnel (*see Figure (1.9.4c)*), we generate the Accommodating Universe Paradox in Section (1.9).

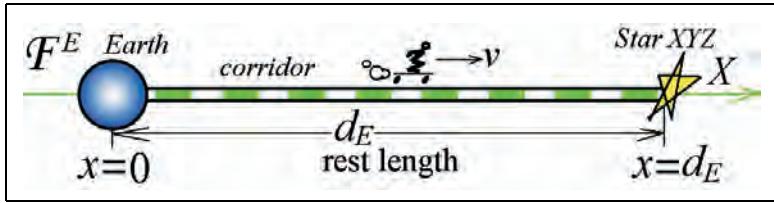
²<http://www.special-relativity-illustrated.com>

(1.9) Paradox: The Accommodating Universe

(1.9.1) **Inertial frame \mathcal{F}^E** — contains the “dumbbell” configuration of

- (i) the Earth at the point $x = 0$,
- (ii) Star XYZ at point $x = d_E$ and
- (iii) a tunnel or corridor with at-rest length d_E in frame \mathcal{F}^E that connects Earth and Star XYZ .

Inertial frame \mathcal{F}^S — whose X -axis is parallel to that of \mathcal{F}^E and contains a skateboard and a skateboarder fixed at point $x = 0$, the origin of \mathcal{F}^S . The origin of frame \mathcal{F}^S is fixed to the skateboard and travels through the tunnel from Earth to Star XYZ at speed v relative to frame \mathcal{F}^E .

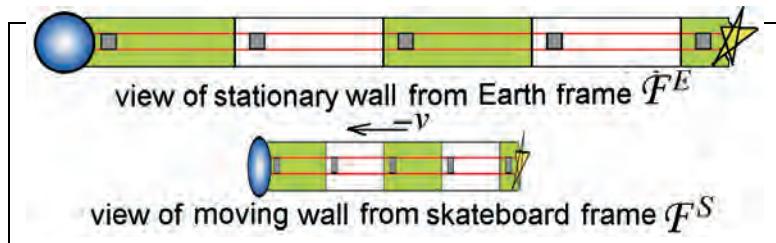


From (1.6.1), the skateboarder in frame \mathcal{F}^S , will see the corridor traveling leftward at speed $-v$. From properties of the ubiquitous shrinking factor (1.8.4), the skateboarder will measure the corridor to have a length of only $\sigma_v d_E$. This means *as speed v approaches the speed of light, the length of the corridor becomes arbitrarily small from the skateboarder's point of view.*³

What does this mean?

(1.9.2) **The Universe is ACCOMMODATING** in the sense that the faster the skateboarder races toward the star, the shorter will be the length $\sigma_v d_E$ of the corridor from the skateboarder's point of view in \mathcal{F}^S .

³The same argument implies that any two points in the entire universe become arbitrarily close to each other from a traveler's point of view if the speed is close enough to c .



Hence, the skateboarder, traveling at speed v along a shortened corridor will reach Star XYZ in an arbitrarily short time.

To set the stage, we will need one more definition.

(1.9.3)

DEFINITION: A LIGHT YEAR is defined as the distance in a vacuum that light travels after a one-year journey. Similarly, a LIGHT SECOND, LIGHT MINUTE, etc. is the distance traveled by light in a second, or a minute, etc.

Statement of the Paradox

Suppose the length of the corridor between Earth and Star XYZ is $d_E = 150$ light years (1.9.3) as seen from Earth. This means a photon requires 150 years to travel from Earth to Star XYZ . Yet, as mentioned in (1.9.2) above, for sufficiently high speeds, the skateboarder can travel through the corridor from Earth to Star XYZ in less than 150 years.

Question: How can the skateboarder travel faster than light?

(1.9.4)

Analysis: In both the Earth's frame of reference \mathcal{F}^E and the skateboarder's frame \mathcal{F}^S (1.9.1), the time of travel of any object traveling at speed w over distance $dist$ is

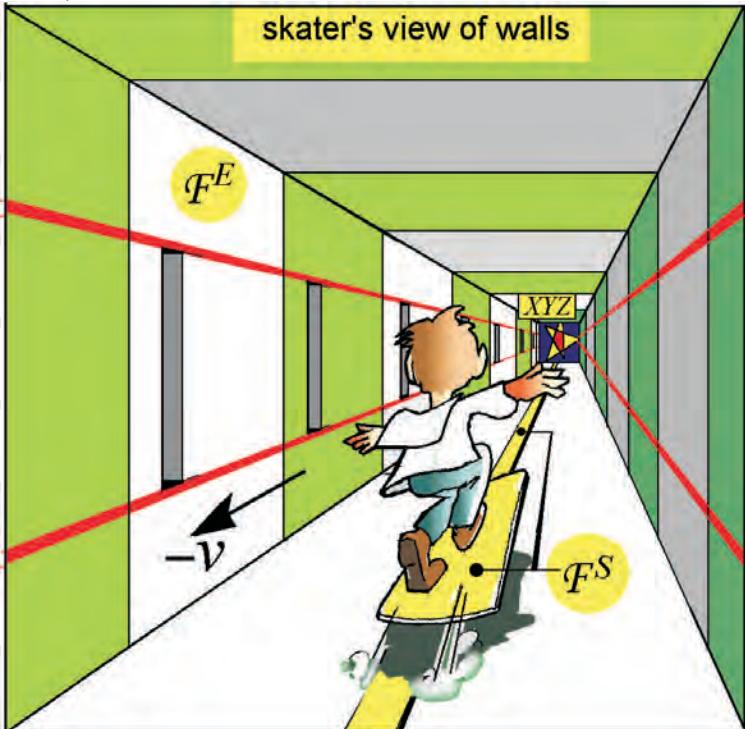
$$(1.9.4a) \quad time = \frac{dist}{w}.$$

Using (1.9.4a) in frame \mathcal{F}^E ,

(1.9.4b)

- Skateboarder time in \mathcal{F}^E required to cover distance d_E at speed v is $\boxed{d_E/v}$ seconds.
- Photon time in \mathcal{F}^E required to cover distance d_E at speed c is $\boxed{d_E/c}$ seconds.

(1.9.4c)



- Skateboarder time in \mathcal{F}^S required to cover distance $\sigma_v d_E$ at speed v is $\boxed{\sigma_v d_E/v}$ seconds.
- Photon time in \mathcal{F}^S required to cover distance $\sigma_v d_E$ at speed c is $\boxed{\sigma_v d_E/c}$ seconds.

(1.9.5)

Skateboarder Travels Faster Than Light: A Specious Argument. Consider the ratio of the times required, for the skateboarder

and for a photon, to cover the distance between Earth and Star XYZ .

(1.9.5a)

from (1.9.4c), skateboarder time = $\sigma_v d_E/v$

from (1.9.4b), photon time = d_E/c

$$\text{which implies } \frac{\text{skateboarder time}}{\text{photon time}} = \frac{\sigma_v d_E/v}{d_E/c} = \frac{\sigma_v}{(v/c)} .$$

To fix ideas, set the skateboarder's time equal to 1% of the photon's time. Then calculation of the ratio $\frac{\sigma_v}{(v/c)}$ of (1.9.5a) yields the result

$$(1.9.5b) \quad \frac{\text{skateboarder time}}{\text{photon time}} = \frac{\sigma_v}{(v/c)} = 0.01 \text{ implies } (v/c) \approx 0.99995.$$

(1.9.6)

Something is Horribly Wrong. We have just shown in (1.9.5b) that if the skateboarder travel time is 99% less than the photon travel time (*for the same Earth \rightarrow Star XYZ journey*), then the skateboarder speed v is also *less than* the photon speed c . How can both statements be true at the same time?

Why Argument (1.9.5) is Specious:
An exercise in critical thinking.

(1.9.7)

CRITICAL THINKING: An example. The error of (1.9.6) lies not in the calculations, but in the interpretation of their meaning. Recall that we used the following assumed “truths.”

- (1.9.7a) Moving lengths observed to have speed v , shrink by a factor of σ_v (*from (1.7.2b)*).
- (1.9.7b) Neither matter nor information (electromagnetic waves) can move faster than the speed of light (*from (1.7.2c)*).
- (1.9.7c) Clocks in Earth frame \mathcal{F}^E and the skateboarder's frame \mathcal{F}^S run at the same rate. (*In (1.9.5a), we chose the clock of \mathcal{F}^S to measure the skateboarder's journey time and the clock of \mathcal{F}^E to measure the photon's journey time.*)

In the list of (1.9.7), we started by accepting that (1.9.7a) and (1.9.7b) are true—moving lengths do shrink and nothing travels faster than light. Given these assumptions, we then concluded that (1.9.7c) must be false—when we compare travel times of two objects (*skateboarder*

and photon) we can not use clocks from two different frames. So, if magnitudes of lengths depend on the reference frame, so will the length of time intervals.

Chapter (9) offers a full analysis of (1.9.4b), (1.9.4c), and (1.9.5b).

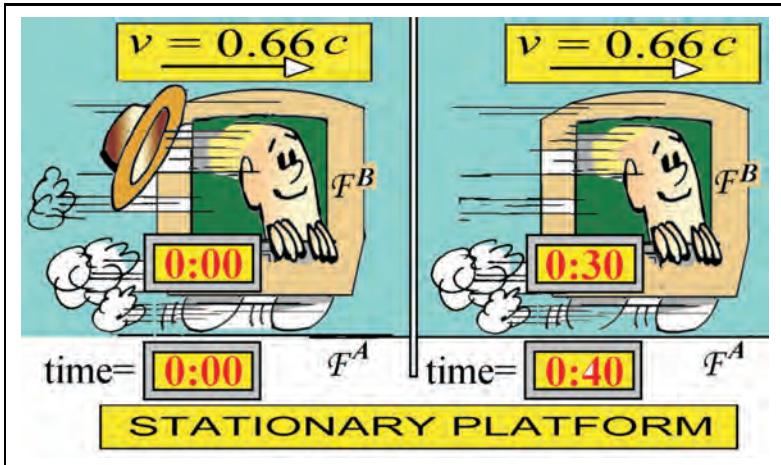
(1.10) Paradox: Time and Distance Asymmetry

Moving Clocks Run Slow

The paradox of Time and Distance Asymmetry comes in two parts:

- (1.10.1) **I. The Clocks on a Train.** In both panels of Figure (1.10.2), the car (frame \mathcal{F}^B) is traveling to the right at speed $0.66c$, or 0.66 times the speed of light, as seen from the platform (frame \mathcal{F}^A). Each panel shows two identical clocks, one in the frame of the moving train and the other in the frame of the stationary platform.

(1.10.2) **FIGURE: Motion Makes Clocks Run Slow**



- LEFT PANEL OF (1.10.2): A snapshot at time $t = 0$ shows the clock of the moving car and a stationary platform clock simultaneously synchronized to read $0:00$ milliseconds.
- RIGHT PANEL OF (1.10.2): The same clock of the car in the moving frame, \mathcal{F}^B reads only $0:30$ milliseconds while, at the same moment,

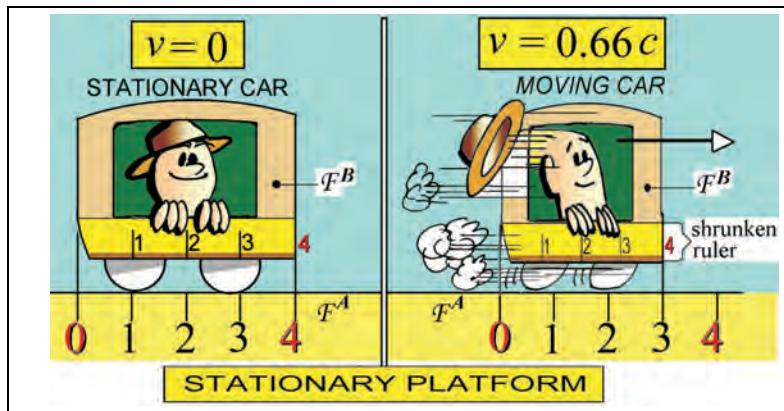
another stationary platform clock in frame \mathcal{F}^A registers 0:40 milliseconds.

The Ubiquitous Contraction Factor: From Definition (1.8.2), when $v = 0.66c$, the Lorentz contraction factor is $\sigma_v = 0.75$ (within two-place decimal accuracy). Thus, the moving car records 0:30 = $\sigma_v \cdot \boxed{0:40} = 0.75 \cdot \boxed{0:40}$ at the moment that time 0:40 is read by the clock on the stationary platform.

Moving Rulers Shrink in the Direction of Motion

- (1.10.3) **A Quantitative View of (1.7.2b) A Ruler on a Moving car.** In both panels of Figure (1.10.4), the car travels rightward at $0.66c$, or 0.66 times the speed of light. Each panel shows two identical rulers, one on the moving train and the other on the stationary platform.

(1.10.4) **FIGURE: Motion Shrinks Distance**



- LEFT PANEL OF (1.10.4): *The car, at rest, is 4 units wide as measured by identical rulers, one on the car and the other on the fixed platform.*
- RIGHT PANEL OF (1.10.4): *The car is now in motion with a speed of $0.66c$ as it passes a second point on the platform. This snapshot shows the platform ruler measures the moving car width as only 3 units.*

The Ubiquitous Contraction Factor: From (1.8.2), the Lorentz contraction factor $\sigma_v = 0.75$ when $v = 0.66 c$. Hence, the ruler on the moving car contracts in the direction of motion from 4.0 units, as measured in its stationary state, to $3.0 = \sigma_v \cdot 4.0 = 0.75 \cdot 4.0$ units as measured while the car is moving at constant speed $0.66 c$.

Statement of the Paradox

It can be argued that the assertion that motion alone will cause clocks to slow (*Figure (1.10.2)*) and rulers to shrink (*Figure (1.10.4)*) is a paradox in itself.

But we wish to go further: We agree that all rulers and clocks are identical when observed at rest. Table (1.8.4) tells us that, in all cases, at observed speed v moving clocks run slow by a factor of σ_v and moving objects shrink (in the direction of motion) by the same factor σ_v . The (apparent) contradiction of asymmetry is this:

(1.10.4a) *From the railway platform* point of view (*Figure (1.10.2)*), the moving clock on the railway car runs slower than the stationary platform clock by a factor of σ_v . Also, (*Figure (1.10.4)*), the moving ruler (rod) attached to the railway car is shorter than the identical ruler on the platform by a factor of σ_v .

(1.10.4b) *From the railway car* point of view, the moving clock is the one on the platform. Hence, it is now the platform clock that runs slower than the railway car clock by a factor of σ_v . Similarly, the moving ruler is now the one on the platform. Hence, it is now the platform ruler that is shorter than the railway car ruler by the same factor σ_v .

.....

- (1.10.5) **How can (1.10.4a) and (1.10.4b) both be true?** That is, how can each observer—one in the railway car and one on the platform—see each other's clock run slower than his own? And how can each observer see the other's (identical) ruler as smaller than his own?

Comments on How the Paradox is Resolved

Spoken language is the culprit with respect to this paradox. We use

the verb “see” as in (1.10.5), when *an* observer in one frame measures the time in another frame—we imagine only one person doing the seeing.

However, in Figures (2.4.2) and (2.4.3), pg. 51, it will be shown that to measure lengths of moving rods and rates of moving clocks, a team of *two* observers in the stationary frame is always required. With this configuration of observers, the paradox of mutual shrinkage and mutual slowing of moving clocks is resolved.

Statement of the Twin Paradox

(1.11.1)



A Qualitative View. A pair of twins carry identical clocks. One twin travels into space at a speed close to that of light. After the return to Earth, the traveling twin is biologically many years younger than the Earthbound twin.

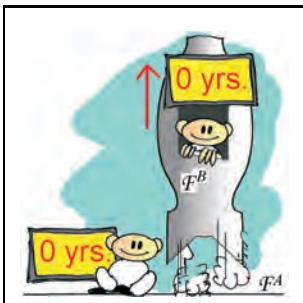
Since observers in each frame see the moving clocks of the other frame running slow (see (1.7.2a) and Figure (1.10.2)), the paradox, or apparent contradiction, is this:

- If each twin sees the moving clock of the other twin running more slowly than his/her own clock, how can one of the twins, the traveling one, end up being younger than the Earthbound twin?

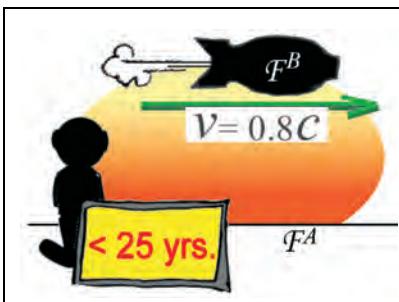
In this case, travel into the future—by the traveling twin—is possible.

(1.11.2)

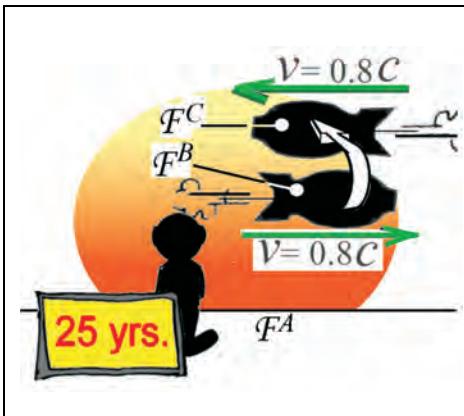
A Quantitative View. The following four panels, (1.11.2a)-(1.11.2d), illustrate the twin paradox from the Earthbound twin’s point of view when the speed of the rocket ship is $0.8c$ (80% the speed of light). After 50 years, as measured by the earthbound clock, the traveling twin returns, having aged only 30 years biologically.

(1.11.2a) **FIGURE: The Take-Off:**

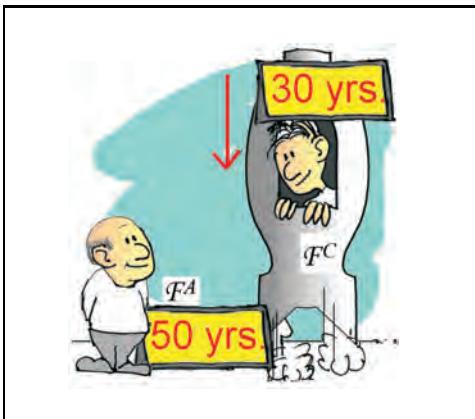
At the moment of takeoff, two identical clocks, the Earth clock in frame \mathcal{F}^A and the spaceship clock in frame \mathcal{F}^B are synchronized to read $t = 0$ years. The time unit of each clock is in “years.”

(1.11.2b) **FIGURE: In Flight**

For the next twenty five years, as timed by the Earth clock in \mathcal{F}^A , the spaceship in frame \mathcal{F}^B races away from Earth, traveling at a constant speed of $0.8 c$ ($8/10$ the speed of light).

(1.11.2c) **FIGURE: Transfer to the Returning Spaceship**

At the moment the Earth clock reads 25.0 years, the traveling twin instantly switches from frame \mathcal{F}^B , the outbound ship, to frame \mathcal{F}^C , an incoming spaceship going in the opposite direction. The incoming ship is traveling at a constant speed of $0.8 c$ ($8/10$ the speed of light).

(1.11.2d) **FIGURE: Return to Earth: The Reunion**

According to the Earth clock in \mathcal{F}^A , the return journey of frame \mathcal{F}^C containing the traveling twin, takes another 25 years. Therefore, the full round trip has taken 50 years. On the other hand, the spaceship clock, which is identical to the Earth clock, will have ticked off only 30 years!

Informal Resolution of the Paradox

- (1.11.3) *The Earthbound Twin Point of View.* The Earthbound twin in frame \mathcal{F}^A always sees the clock of frame \mathcal{F}^B (the spaceship) running slow since it is moving relative to the Earth frame \mathcal{F}^A with speed $0.8c$. Specifically, the Lorentz contraction factor, $\sigma_v = 0.6$ (1.8.2) guarantees that the 50 years on the stationary Earth clock correspond to a total of $\sigma_v \times 50 = 0.6 \times 50 = 30$ years on the moving spaceship clock—15 years are registered on the spaceship clock in each direction. There is no paradox here.
- (1.11.4) *The Spaceship Twin Point of View.* The spaceship twin in \mathcal{F}^B also sees the clock on Earth in \mathcal{F}^A running slow since it is moving at relative speed $0.8c$. The same Lorentz factor, $\sigma_v = 0.6$, guarantees that the 30 years of the spaceship clock (see (1.11.3)), correspond to $0.6 \times 30 = 18$ years on the (relatively moving) Earth clock—9 years are registered on the slower Earth clock in each direction. (See Table (1.11.5).)

Now here is the surprising part. After the “instantaneous” change of frames (*from outgoing \mathcal{F}^B to incoming \mathcal{F}^C*), the travel-

ing twin in frame of reference, \mathcal{F}^C , now sees the Earth clock reading 41 years—a full 32 years later than the Earth clock reading of 9 years he had just seen from frame \mathcal{F}^B . (See Table (1.11.6).)

This “surprise” is due to the fact that during the moment of spaceship transfer, the traveling twin is no longer in an *inertial* frame since he suffers an extreme acceleration (change of speed) which affects the way clocks tick during that moment. See (11.5), pg. 147 for details.

Here is a tabulated summary of answers (1.11.3) and (1.11.4) of the twin paradox which is illustrated in panels (1.11.2a)–(1.11.2d):

(1.11.5)

TABLE: *Two Phases of the Spaceship Clock Journey as Seen from the Stationary Earth Clock: $\sigma_v = 0.6$*

Leg	View from Stationary Earth
(1) Outbound	<i>The stationary Earth clock registers 25 years. The moving spaceship clock runs slow and registers only $0.6 \times 25 = 15$ years.</i>
(2) Homebound	<i>The stationary Earth clock registers 25 more years (for a total of $25 + 25 = 50$ years). The moving spaceship clock runs slow and registers $0.6 \times 25 = 15$ more years (for a total of $15 + 15 = 30$ years).</i>

(1.11.6) **TABLE:** *Three Phases of the Moving Earth Clock as Seen from the Stationary Spaceship: $\sigma_v = 0.6$*

Leg	View from Stationary Spaceship
(1) Outbound	<i>The stationary spaceship clock registers 15 years, as was shown in (1.11.5). The moving Earth clock runs slow and registers only $0.6 \times 15 = 9$ years.</i>
(2) Instantaneous turn-around	<i>After passing from outgoing frame \mathcal{F}^B (where the Earth clock reads 9 years) to the inbound frame \mathcal{F}^C (where the Earth clock is seen to read 41 years), the Earth clock is seen by the traveling twin, now in \mathcal{F}^C, to have gained 32 years.</i>
(3) Homebound	<i>The stationary spaceship clock registers 15 more years (for a total of $15 + 15 = 30$ years). The moving Earth clock runs slow and registers $0.6 \times 15 = 9$ more years (for a total of $9 + 32 + 9 = 50$ years).</i>

But *why* does the spaceship twin see a time-jump on the moving Earthbound clock as described in (1.11.6)(2)? Once spacetime diagrams, known as Minkowski diagrams (6.1.1), pg. 97, are understood, the facts of this phenomenon are rendered accessible through Figure (11.3.1), pg. 143.

(1.12) Paradox: The Train in the Tunnel

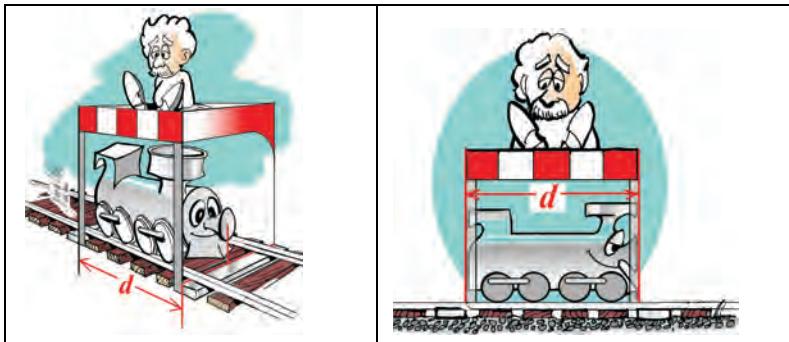
Statement of the Paradox

(1.12.1) **Train-in-Tunnel Paradox: A Qualitative View.** When at rest,

a train and a tunnel have the same length, d . This is illustrated in Figure (1.12.1a).

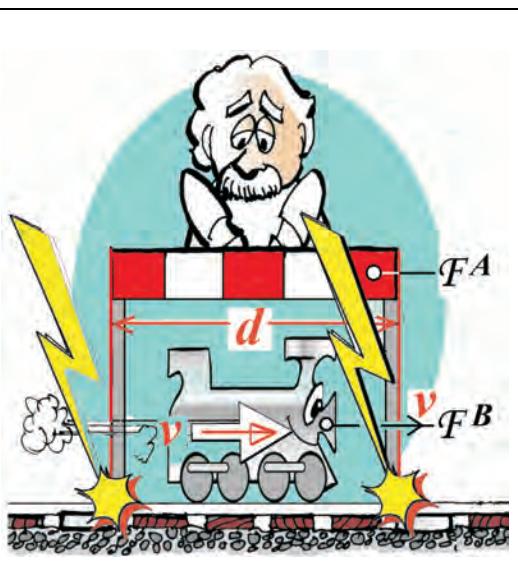
The following panels illustrate a pair of lightning strikes at each end of the tunnel and describe two *different* perceptions—one from Bernie who sits on the tunnel in frame, \mathcal{F}^A , and the other from the train's point of view in frame \mathcal{F}^B .

(1.12.1a) **FIGURE:** Tunnel and Train (length d) at Rest



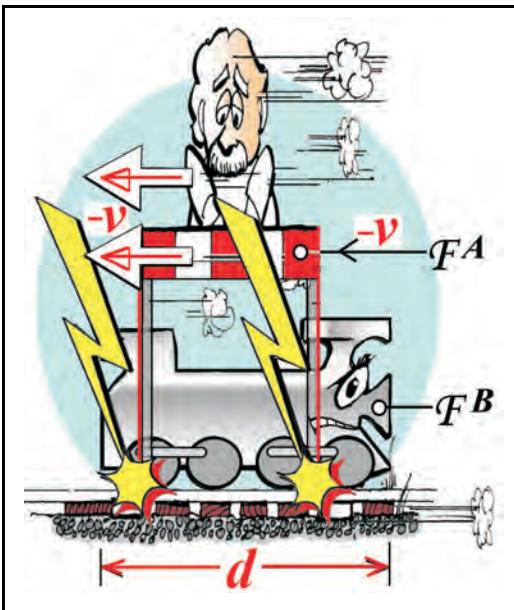
(1.12.1b) **FIGURE:** Lightning Avoids the Short Train:

From Bernie's point of view in frame \mathcal{F}^A , the tunnel is stationary while the train in frame \mathcal{F}^B is in motion. According to (1.10.4), pg. 27, the train shrinks in the direction of motion. (*Note how the wheels have become elliptical.*) Hence, the two lightning bolts that are simultaneous in Bernie's frame, \mathcal{F}^A , never hit the shorter train.



(1.12.1c) **FIGURE:** Does Lightning Hit the Too-Long Train?

Relative to the stationary frame \mathcal{F}^B (the train), the frame, \mathcal{F}^A (the tunnel), is observed to be moving. The moving tunnel \mathcal{F}^A is seen from the train's frame \mathcal{F}^B to shrink in the direction of motion (1.7.2b). If the (shrunken) tunnel is always shorter than the train, then, shouldn't the train be electrocuted by the “simultaneous” lightning strikes at the ends of the tunnel?



(1.12.2)

Summary: From the viewpoint of the tunnel's frame, \mathcal{F}^A , the shortened train in frame \mathcal{F}^B avoids the lightning (see Figure (1.12.1b)). Yet, in a change of viewpoint from which the train is stationary, we see the moving, shortened tunnel where “simultaneous” lightning strikes hit the too-long train (see Figure (1.12.1c)).

Question: How does the train escape electrocution in one frame but not in another?

Informal Resolution of the Paradox

The apparent inconsistency between the figures is explained by the fact that the double lightning strike in Figure (1.12.1b) is *simultaneous* whereas, in Figure (1.12.1c), the lightning strikes are not simultaneous. (*In (A.1.1), we first noted this loss of simultaneity.*) In Figure (1.12.1c), the right bolt of lightning strikes ahead of the “too-long train” and, a little later—only after the tail of the train has entered the tunnel—does the left bolt of lightning decide to strike. In both frames of reference, the train avoids electrocution.

How do we know this to be true? These sequences are read from the spacetime Minkowski diagrams in Section (C.2).

(1.13) Paradox: The Pea-Shooter

- (1.13.1) **Relation to $E=mc^2$.** The relativistic addition of speeds, illustrated in this example, lies at the heart of the celebrated mass-energy equation, $E = mc^2$ (*A full exposition appears in Chapter (15)*.)

A Brief Preview: In our everyday experience, the speed of an object—like a bullet—is equal to the speed v of the gun plus the speed w of the bullet relative to the gun. The total speed is the ordinary sum $v + w$ which is very accurate *provided* the speeds are much smaller than c .

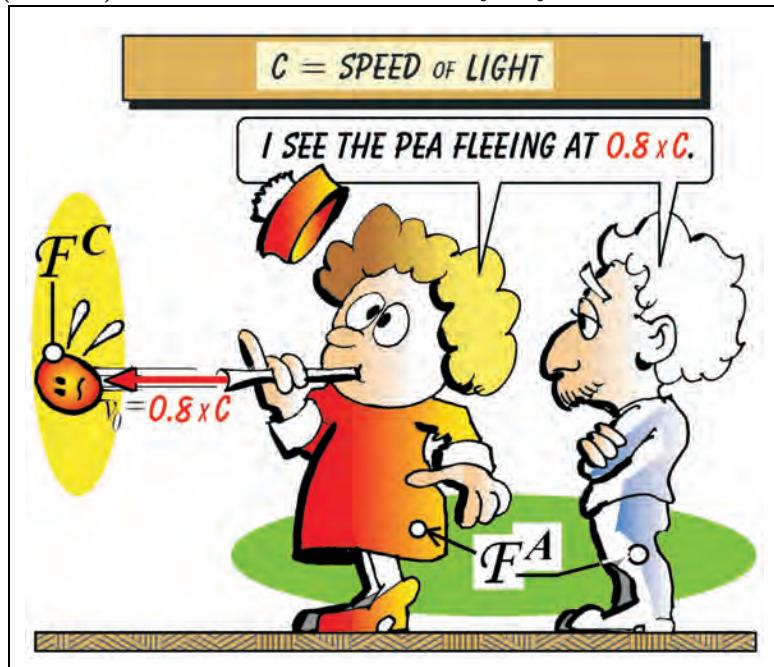
However, to calculate accurately with very high speeds v and w , we need a more general relativistic “sum,” $v \oplus w$, which will be defined in (1.13.3). Hence, if mass m has speed v with consequent momentum $[m]v$, then an addition of speed w produces a new momentum $[m](v \oplus w)$ instead of the classical momentum $[m](v + w)$. This implies that mass m must increase with increasing speed (*Section (15.2)*). Finally, the increase in mass is found to be due to (equivalent to) the energy required to produce the increase in speed (*Section (15.3)*)).

Statement of the Paradox

- (1.13.2) **The Pea-Shooter Paradox: A Quantitative View**

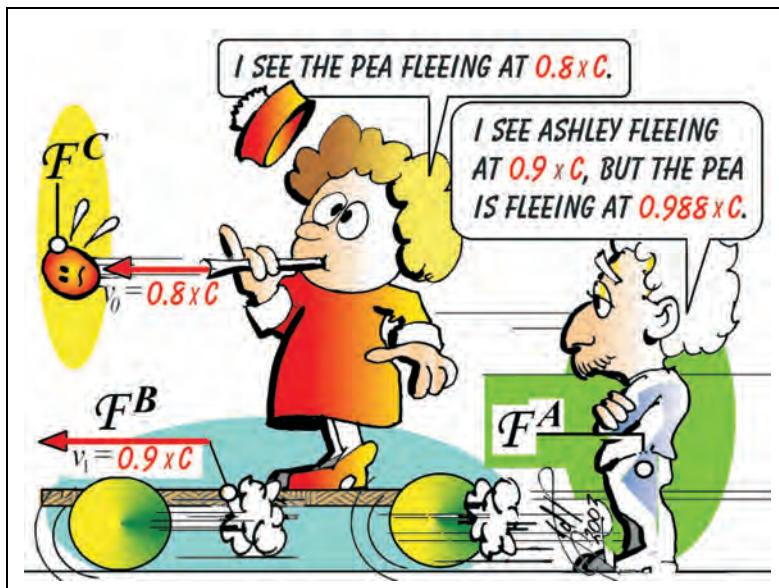
Bernie and Ashley are stationary relative to each other—they are both in the same frame of reference \mathcal{F}^A . From their common point of view (frame), the pea, in moving frame \mathcal{F}^C , is seen to have a leftward speed of $v_0 = 0.8c$ or 0.8 times the speed of light.

(1.13.2a) FIGURE: A Shared Frame of Reference



(1.13.2b) FIGURE: From Bernie's Frame of Reference

Bernie, in frame F^A , sees Ashley's wagon in frame F^B speeding leftward at $0.9c$. Ashley sees the pea in frame F^C moving with speed $0.8c$, yet Bernie sees the same pea (propelled by the wagon) moving leftward with speed $0.988c$, a speed that is *less than* $1.7c = 0.8c$ (*pea speed seen by Ashley*) + $0.9c$ (*wagon speed seen by Bernie*).



Informal Resolution of the Paradox

Consistent with our personal experience, speeds are additive. That is, when a train passes in front of us at 30 mph while a passenger walks toward the front of the train at 4 mph, then we expect to see the passenger moving at

$$(1.13.2c) \quad 30 \text{ mph} + 4 \text{ mph} = 34 \text{ mph}$$

relative to us.

To discuss the “true addition of speeds,” we use \oplus , a different plus sign, to write the assertion that

$$(1.13.2d) \quad 0.8c \text{ (mph)} \oplus 0.9c \text{ (mph)} = 0.988c \text{ (mph)}$$

which is *not* the value $0.8c + 0.9c = 1.7c$ we would expect by applying

the usual arithmetic sum used in (1.13.2c).

(1.13.3)

DEFINITION: The true addition of speeds, $v \oplus w$. Given c is the speed of light in a vacuum. Then for speeds v and w , the sum, $v \oplus w$, is called the LORENTZ SUM of speeds v and w and is defined by

$$(1.13.3a) \quad v \oplus w \stackrel{\text{def}}{=} \frac{(v + w)}{1 + (vw/c^2)}.$$

The defining equation (1.13.3a) shows that for low speeds, such as $v = 30$ mph and $w = 4$ mph, the quantity (1.13.2c) is a close approximation to the true sum, $v \oplus w$ of (1.13.3). In fact, whenever we have speeds v and w whose product $v \cdot w$ is tiny relative to c^2 , we have from (1.13.3)

$$(1.13.4) \quad v \oplus w = \frac{(v + w)}{1 + \underbrace{(vw/c^2)}_{\text{negligible}}} \approx v + w.$$

Justification for this unlikely “sum” of speeds (1.13.3a) will be developed in (7.3.1a).

(1.13.5)

NOTE: Letting $c = 1$. It is often more convenient to take $c = 1$ from the outset so that instead of dealing with speeds of the form v/c and w/c , we can work with “normalized” speeds, $-1 < v, w < 1$.

(1.13.5a) **EXAMPLE:** One way to force c to equal one is to introduce a new measure of length that equals the distance light travels in one second. Since $c = 2.997925 \times 10^8$ meters/second (or 186,282.4 miles/hour), we can define

$$(1.13.5b) \quad \text{One KABOOGLE} \stackrel{\text{def}}{=} 2.997925 \times 10^8 \text{ meters.}$$

Then $c = 1$ kaboogle/second.

(1.14) Paradox: The Bug and Rivet

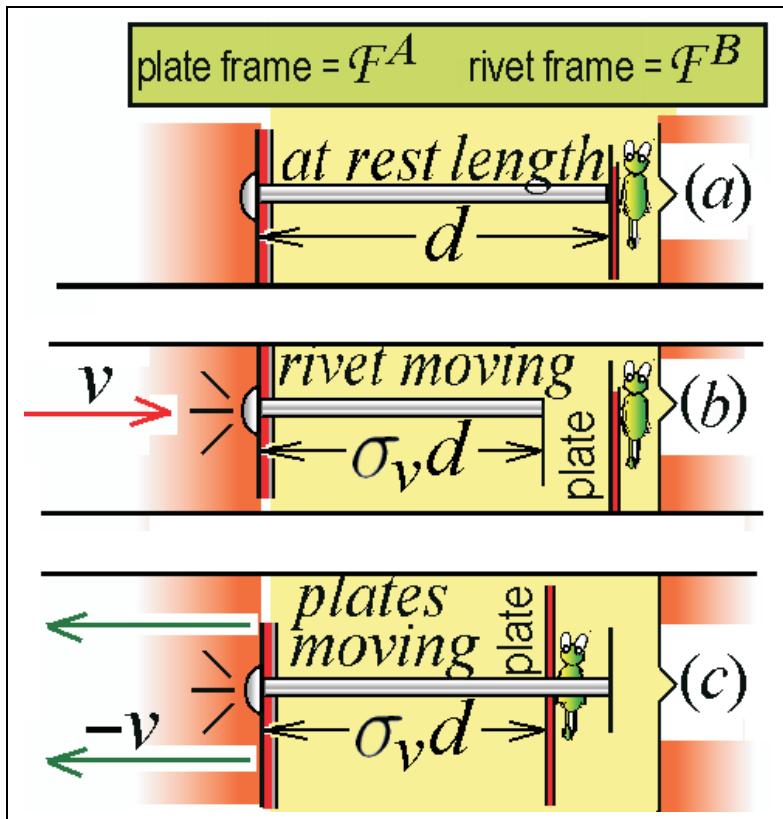
(1.14.1)



The Bug-Rivet Paradox. *A Qualitative View:* Two vertical parallel metal plates in inertial frame \mathcal{F}^A have left-hand surfaces that are separated by a distance d . The plates have aligned holes large enough to allow a rivet shank to pass through, but not large enough for the rivet head.

(1.14.2)

FIGURE: Setup for the Bug-Rivet Paradox:



- Fig. (1.14.2), Band (a): The at-rest length of the rivet shank d equals the at-rest distance between the left surfaces of the parallel plates. We ignore all momentum issues and assume the mass of the rivet is minuscule compared to that of the plates.

- **Fig. (1.14.2), Band (b):** From the (stationary) plates' point of view in \mathcal{F}^A , the rivet's frame, frame \mathcal{F}^B , moves at constant speed v . The length of its moving shank, as measured from frame \mathcal{F}^A , is therefore less than d . Once the rivet head collides with the surface of the left metal plate, the rivet head stops abruptly but, as seen from frame \mathcal{F}^A , the shank length eventually increases to d (see the X - T spacetime diagram in Figure (1.14.2)).
- **Fig. (1.14.2), Band (c):** From the (stationary) rivet's point of view in \mathcal{F}^B , i.e., we are riding on the rivet, the plates in frame \mathcal{F}^A are in leftward motion, traveling at constant speed $-v$. The distance between the two moving plates, as measured from the stationary rivet frame, \mathcal{F}^B , is therefore less than d . Once the (moving) left-hand plate collides with the (stationary) rivet head, the left plate stops.

The Paradox Question: *Is the bug squashed when the rivet head is stopped by the left surface of the left-hand plate?*

If configuration (b) of Figure (1.14.2) prevails, then the answer is no—the rivet length is less than d at the moment of collision and the rivet point never reaches the bug.

If configuration (c) of Figure (1.14.2) prevails, then the answer is yes—the distance between the two plates is less than d , the length of the rivet shank. Hence, the rivet point will reach the bug that is hiding behind the right-hand plate. Subsequently, the distance between the two plates increases to its at-rest value.

- (1.14.3) **The resolution of the paradox** lies in the fact that the length of the rivet must expand beyond its at-rest length before it finally returns to its at-rest length. (*Proof and justification will be found in Chapter (14) which is illustrated by a property of the Slinky® toy.*) The situation is similar to a stretched spring relaxing and returning to its unstretched length.

Otherwise, if the rivet is assumed to be absolutely rigid, then the following pair of false statements must be valid:

(1.14.4)

**Absolute rigidity implies:**

- *The moment the head stops (as it collides with the plate), the tip also stops—simultaneously. This is equivalent to saying the information, “I’ve hit the plate” is transmitted to the end of the rivet at a speed greater than c . See (1.7.2c).*
- *The future affects the present in the sense that in one frame, the rivet tip “knows” before the event that the rivet head will collide with the plate. See (14.4.4)*

(1.15) Exercises**(1.15.1) Exercise**

Using the fact that c , the speed of light, is 186,000 miles per second, show that the time parameters of Figure (1.10.2) require a platform length of 744 miles.

(1.15.2) Exercise

The Skateboarder’s Necessary Speed, v . At what speed, v , must the skateboarder of (1.9.4c) travel to reach StarXYZ in
 (a) one hour? (b) one day? (c) one year? (d) $1\frac{1}{2}$ years?

(1.15.3) Exercise

Slow Speeds 1. Compute the value of $(vw)/c^2$ for the speeds $v = 30$ miles per hour and $w = 40$ miles per hour. (See Appendix D for the conversion of “miles per hour” to “miles per second.”)

(1.15.4) Exercise

Slow Speeds 2. With the values of v and w of Exercise (1.15.3), how close is the “intuitive” speed $v + w$ to the true speed of $v \oplus w$ given by (1.8.2)?



(2) Clocks and Rods in Motion

(2.1) The Perfect Clock

We start with an informal definition of a clock:

(2.1.1)

DEFINITION: A CLOCK is a device that ticks, pings, or presents some discernable signal or readout at regular intervals.

(2.1.1a) **Starting the Clock.** The clock is started when an initial photon passes the clock and triggers the clock's first tick or ping as required in Definition (2.1.1). Although 0.00 is a natural starting time, the initial starting time can be set arbitrarily.

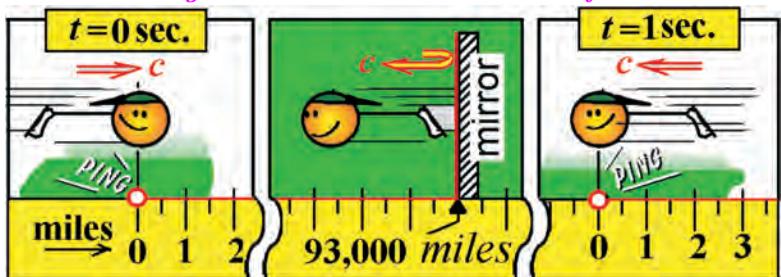
(2.1.1b) **Setting the Tick Rate.** A second passing photon defines the length of the time interval that will be repeated for all subsequent ticks or pings—the clock “remembers” the time interval between the first and second photons and it waits for this interval to pass before it issues each subsequent tick.

(2.1.2)

Defining the One-Second Ticking Rate. Figure (2.1.3) shows a clock with properties of (2.1.1a) and (2.1.1b) whose ticking rate is being set to one-second intervals.

(2.1.3)

FIGURE: Setting the One-Second Tick Rate of a Clock.



- LEFT PANEL OF (2.1.3): The photon passes over the clock at point 0 and simultaneously sets off the clock's first tick. The initial time may be set at 0.00 seconds (2.1.1a).

- MIDDLE PANEL OF (2.1.3): *At the 93,000 mile mark, the photon strikes the mirror and starts its return journey to the clock at the origin, 0.*
- THIRD PANEL OF (2.1.3): *On completion of the 186,000 mile round trip, the photon passes the clock, setting the second tick which the clock records as 1.00, or 1-second. From this point on, the clock ticks at exactly the rate of these first two ticks. The time interval between ticks is permanently set (2.1.1b) and is defined (as follows) to be one second.*

(2.1.4)

DEFINITION: ONE SECOND is that time interval required for a photon of light to travel 186,000 miles in a vacuum. Any clock that remains stationary, and is constructed according to Figure (2.1.3), is said to read the PROPER TIME of its own frame of reference.

(2.1.5)

How time can be identified with distance. More accurately, the time of a photon traveling in a vacuum can be correlated to a precise *distance* once we rephrase Definition (2.1.4) as follows:

| Whenever light (in a vacuum) travels a distance of 186,000 miles, then exactly one second of time will have passed.

(2.1.6)



In our relativistic setting, where c is constant for all observers, we constructed a clock (2.1.2), (2.1.3) whereby *time* is a secondary quantity derived from two primary quantities, *distance* and c , the *speed of light in a vacuum*. For example, in (2.1.5), *time* is defined as a certain *distance* covered by a *speeding* photon. By contrast, in the Newtonian, nonrelativistic setting, $\text{speed} = \text{distance}/\text{time}$ is a secondary quantity derived from the two primary quantities, *distance* and *time*.

(2.1.7)

Why is this clock of (2.1.2), (2.1.3) “perfect?” Assumption (1.5.2b) tells us that the speed of light does not change, even if its source and observer are moving relative to each other. Moreover, since all frames of reference share the same yardstick to mark distances along the direction of the photon’s motion, the same 1-second interval of proper (2.1.4) time will be recorded within all inertial frames of reference.

As we shall see in Section (2.2) if we follow the construction described in Figure (2.1.3), then we can produce as many identical clocks as we like that will simultaneously read the same proper time within a single frame, \mathcal{F} .

(2.2) Synchronizing Clocks within a Single Frame

Informally, we may define two clocks in a single frame \mathcal{F} (1.2.1) as being *synchronized* if they show the same time simultaneously.

Moreover,

(2.2.1)

DEFINITION: Any two clocks that are constructed so as to tick at exactly the same rate when observed at rest are called IDENTICAL clocks.

(2.2.2) **What We May Not Do: Invalid Synchronizing Methods.**

- We cannot shout, mail, or wire the readings of one clock to another remote clock since any such message takes time to deliver — any time lag renders the reported time as stale and invalid.
- Also, it does no good to remove the distance problem by bringing all the clocks to one place in order to synchronize them because eventually, we have to move them back to their original positions. As we have noted (1.7.2a), motion itself will slow the clock down and so, will distort its readings.

(2.2.3)

What We May Do: A Valid Synchronizing Method. We know from (2.1.5) and Exercise (2.4.1) that if a photon travels a total distance of d miles, then $t = d/186,000$ seconds will have passed.

Each clock is constructed so that it starts to tick (at one-second intervals, say) as soon as it is hit by a photon. Before a clock starts ticking, we set an initial time according to its distance from a common photon emitter at the *Origin Clock* positioned at $x = 0$ with a setting of $t = 0$. Once a single burst of simultaneously emitted photons is

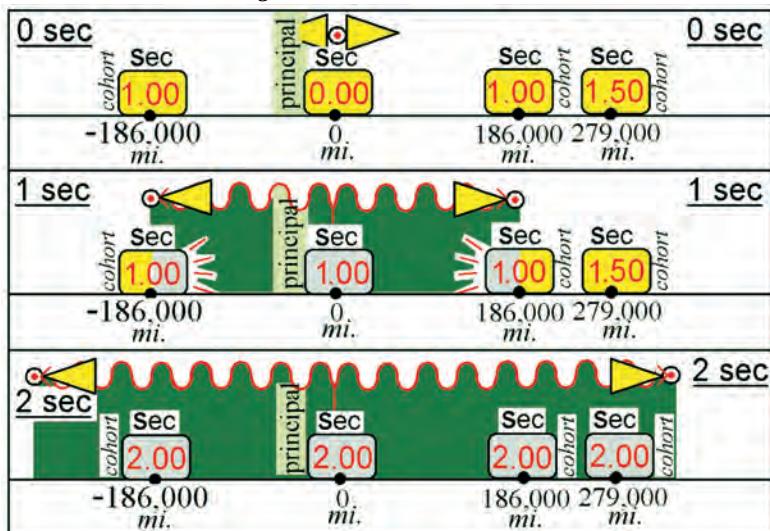
sent into space, the Origin Clock starts ticking and the other non-Origin or clocks, COHORT CLOCKS, will start ticking at the instant of collision with one of these photons.

To fix ideas, Figure (2.2.4) shows how to synchronize three cohort clocks at respective positions $x = -186,000 \text{ miles}$, $x = 186,000 \text{ miles}$, and $x = 279,000 \text{ miles}$. The general definition will be given in (2.2.5).

(2.2.4)



FIGURE: Synchronizing Three Cohort Clocks with the Origin Clock



0 sec The Origin Clock will eject photons in the plus and minus X directions at exactly 0.00 seconds. As per (2.1.1a), the three cohort clocks will start to tick only when their start mechanisms collide with a photon.

A photon from the Origin Clock at $x = 0$ will take exactly one second to reach each of the two cohort clocks that are positioned at $x = -186,000 \text{ miles}$ and $x = 186,000 \text{ miles}$. A photon will require $1\frac{1}{2}$ seconds to reach the cohort clock situated at $x = 279,000 \text{ miles}$. As allowed by (2.1.1b), before they start ticking we may preset the cohort clocks with initial readings 1.00 seconds and 1.50 seconds, respectively.

1 sec It is one second later according to the Origin Clock. At this instant, photons collide with the start mechanism of the two cohort clocks where the initial readings had been set to $\boxed{1.00}$ sec. The cohort clock at position $x = 279,000 \text{ mi}$ will have to wait for its collision photon $1/2$ second from now, that is, when all (three) running clocks will be reading $\boxed{1.50}$ seconds.

2 sec All three cohort clocks have been started by their photon triggers. The three clocks are now synchronized with the Origin Clock and with each other.

We generalize the algorithm of Figure (2.2.4) which is valid for three cohort clocks, and define the synchronization of N stationary clocks within a frame of reference.

(2.2.5)

DEFINITION: Given a functioning Origin Clock at $x = 0$ and N cohort clocks, at points $x = d_1, x = d_2, \dots, x = d_k, \dots, x = d_N$. Then if clock k is initially set at time d_k/c (2.1.1a) and is started by a photon emitted from the Origin Clock in accordance with (2.1.1b), then the N cohort clocks and the Origin Clock are said to be mutually SYNCHRONIZED within their shared frame of reference.

(2.2.6)



The age of points. Since every point in a frame has its own clock, and all the clocks are synchronized to simultaneously read the same time, it will be useful to think of a point in a frame as having the **same age** as every other point in that frame. In terms of Definition (2.1.4), the *proper time* at any point is equivalent to the *age* of that point.

(2.3) Moving Clocks Run Slow, Moving Rods Shrink

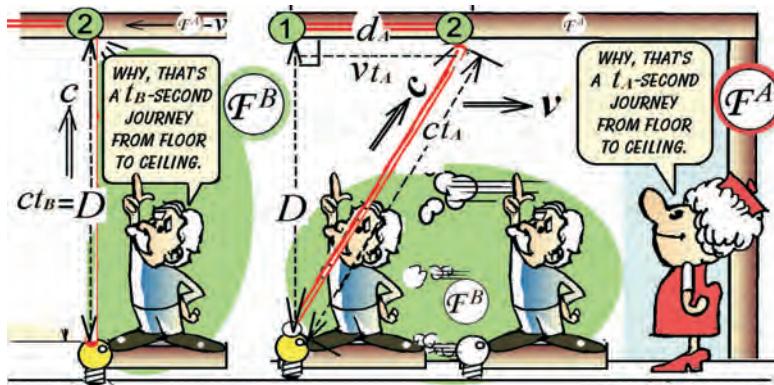
(2.3.1)

NOTE: In the following thought experiment, Ashley's Clock in \mathcal{F}^A and Bernie's clock in \mathcal{F}^B are **synchronized** in the sense that their

clocks simultaneously read $t = 0$ at the precise moment the photon starts its journey from the light source on the floor to the ceiling. This ensures that the respective clock readings, t_A and t_B , coincide with the changes in time, Δt_A and Δt_B , since the release of the photon.

(2.3.2)

FIGURE: *The Vertical Photon Path in \mathcal{F}^B is a Diagonal Path in \mathcal{F}^A*



Ashley's frame \mathcal{F}^A contains a horizontal bar with the marks ① and ②. Bernie's frame \mathcal{F}^B contains a light bulb that is affixed to the floor of a platform that can move left or right. D is the distance from the floor of \mathcal{F}^B to the horizontal bar of \mathcal{F}^A .

- LEFT PANEL OF (2.3.2): In Bernie's frame \mathcal{F}^B , the light bulb is vertically beneath the moving mark ① at time $t = 0$, the precise moment it releases a photon that travels vertically toward the leftward-moving ceiling. At time t_B in \mathcal{F}^B , after traveling D units vertically, the photon hits the leftward-moving ceiling at the mark ②. Since $\text{distance} = \text{speed} \times \text{time}$,

(2.3.2a)

$$D = c t_B = \text{vertical distance from floor to ceiling}$$

where c is the speed of light and t_B is the time (as measured in \mathcal{F}^B) required for the photon to travel D units vertically from floor to ceiling.

- RIGHT PANEL OF (2.3.2): Ashley's frame \mathcal{F}^A is fixed and Bernie's rolling platform (including the light bulb) has rightward speed v . Since $\text{distance} = \text{speed} \times \text{time}$,

(2.3.2b)

$$d_A = v \cdot t_A = \text{the at-rest distance in } \mathcal{F}^A \text{ between ① and ②}$$

where t_A is the time in \mathcal{F}^A for the light bulb to travel rightward at speed v from a point vertically below ① to a point vertically below ②.

The path of the photon from floor to ceiling (that was a vertical line in \mathcal{F}^B) is seen by Ashley to be a diagonal line in \mathcal{F}^A . Since distance = speed \times time,

$$(2.3.2c) \quad ct_A = \text{diagonal distance}$$

where c is the speed of the photon and t_A is the \mathcal{F}^A time required for the photon to travel the diagonal path from floor to ceiling.

Based on the right triangle lengths of Figure (F.12.5b), we show why moving clocks run slow and why moving rods shrink.

(2.3.3)

THEOREM: Moving Clocks Run Slow, Moving Rods Shrink. Inertial frame \mathcal{F}^B has speed v relative to frame \mathcal{F}^A . Then clocks in the moving frame \mathcal{F}^B run at a rate that is $\sigma_v \leq 1$ (1.8.2) times the rate of clocks in \mathcal{F}^A . That is, if clocks of the two frames are synchronized (3.4.5a) at their origins (at time $t_A = t_B = 0$), then when the moving clock time t_B runs slower than all stationary clocks in \mathcal{F}^A . Specifically,

$$(2.3.3a) \quad t_B = \sigma_v t_A$$

where the shrinking constant $\sigma_v = \sqrt{1 - v/c^2}$ is given by (1.8.2).

If a rod in the moving frame \mathcal{F}^B has at-rest length d_A as measured in the direction of motion, then it has length

$$(2.3.3b) \quad d_B = \sigma_v d_A$$

as measured from the stationary frame \mathcal{F}^A .

PROOF: Moving Clocks Slow Down. Combining paths of the right triangle, we see that the three lengths of Figure (2.3.2) listed in (2.3.2a), (2.3.2b), and (2.3.2c), correspond to sides of a right triangle. The Pythagorean Theorem implies

$$(2.3.3c) \quad (ct_A)^2 = (D)^2 + (vt_A)^2$$

$$= (ct_B)^2 + (vt_A)^2 \quad \text{from (2.3.2a).}$$

Solving for t_B , we obtain

$$(2.3.3d) \quad t_B = \sqrt{1 - (v/c)^2} t_A = \sigma_v t_A$$

which proves (2.3.3a).

Moving Rods Shrink: We refer to Figure (2.3.2) and change our viewpoint so that frame \mathcal{F}^B is the stationary frame and frame \mathcal{F}^A is moving leftward relative to \mathcal{F}^B . In this case, points ① and ② are moving leftward past the (fixed) light bulb in \mathcal{F}^B .

At time $t = 0$ in \mathcal{F}^B , point ① is directly overhead the stationary light bulb. At time t_B in \mathcal{F}^B , the leftward-traveling point ② passes vertically over the light bulb. Since distance = speed \times time,

$$(2.3.3e) \quad \begin{aligned} d_B &= v \cdot t_B & d_B \text{ and } t_B \text{ are measured in frame } \mathcal{F}^B \\ &= v \cdot \sigma_v t_A & \text{from (2.3.3d)} \\ &= \sigma_v \cdot d_A & \text{from (2.3.2b)} \end{aligned}$$

which proves (2.3.3b). The proof is done. ■

(2.4)

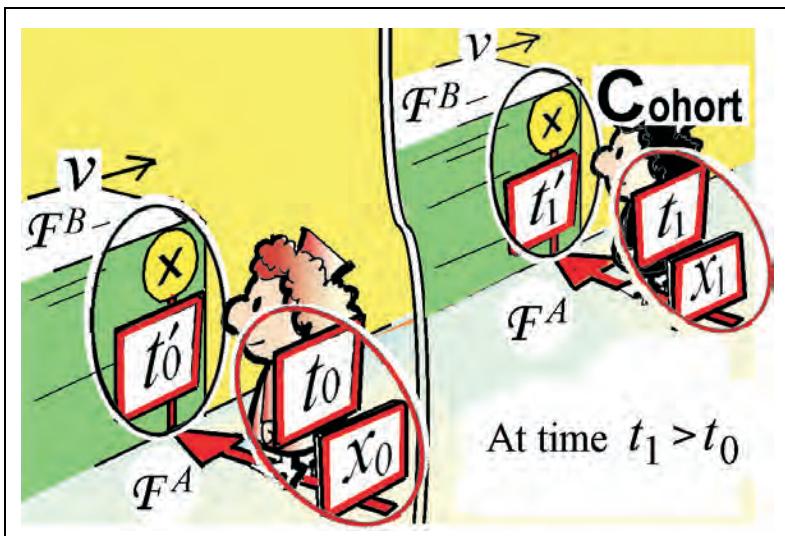
Exercises

(2.4.1) Exercise

Show that any time interval t' can be defined in the manner of (2.1.2) and Figure (2.1.3). Apply your construction to the definition of $t' = 1/4$ second, and $t' = 1/1,000$ second.

Hint: Calibrate the clock by setting the reflecting mirror d' miles from the point 0 instead of the 93,000 miles that was used in Figure (2.1.3). The distance required for the round trip of the photon is now $2d'$ miles, so the time for the round trip, the time between the very first pair of clicks, will be $t' = 2d'/c$ which, depending on the value of d' , can be any time interval desired. The units of speed c and distance d' determine the units of time t' , i.e., seconds, minutes, hours, etc. (See Appendix D.)

(2.4.2) Exercise Speed of a Moving Point, and the Rate of a Moving Clock:
The Two-Observer Method



Step 1: Frame \mathcal{F}^B is moving rightward (*in the positive direction*) with speed v relative to the platform frame \mathcal{F}^A . A single mark, x and a clock are placed at the right-hand end of the train. At time t'_0 in \mathcal{F}^A , the x mark on the train in \mathcal{F}^B passes in front of Ashley at point x_0 on the platform (*left panel*).

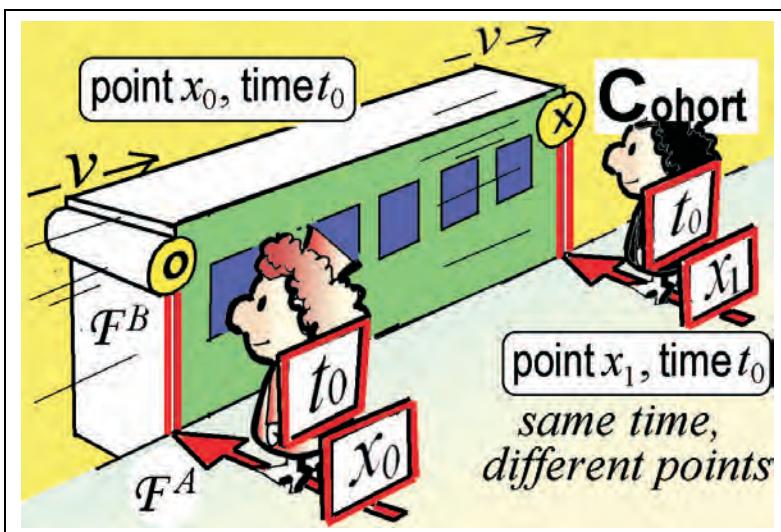
Step 2: At a later time, $t_1 > t_0$ in \mathcal{F}^A , the mark x passes a cohort observer standing at point x_1 on the platform in \mathcal{F}^A (*right panel*).

Questions: How can we use x and t coordinates in the two panels of (2.4.2) to measure

- (a) *the speed of the moving point* x in \mathcal{F}^B as measured from the platform frame \mathcal{F}^A ? (Ans. $v = \Delta x / \Delta t = (x_1 - x_0) / (t_1 - t_0)$)
- (b) *the ratio $\Delta t_B / \Delta t_A$ of the time changes* Δt_B and Δt_A of the respective moving frames \mathcal{F}^B and \mathcal{F}^A ? (Ans. $\Delta t_B / \Delta t_A = (t'_1 - t'_0) / (t_1 - t_0)$).

(2.4.3)
Exercise

Measuring Moving Lengths: The Two-Observer Method In the following exercise, we place two marks on a train or rod—an **x** on the front endpoint and an **o** on the back endpoint.



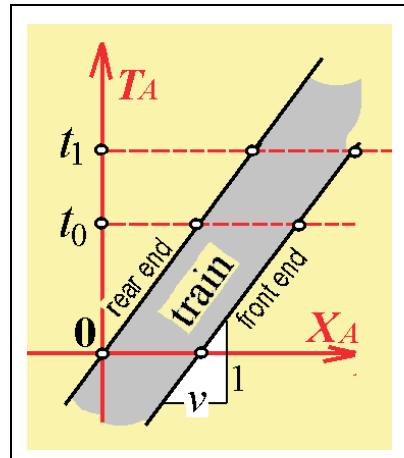
Ashley, and infinitely many cohort observers are standing with synchronized clocks at each point x on the platform. At time t_0 , there are only two unique observers, Ashley and a cohort (*shown in the diagram*), each of whom simultaneously sees an endpoint of the train, **o** and **x**, at their respective platform positions, x_0 and x_1 .

Question: (c) How can we use the x and t coordinates of (2.4.3) to calculate the length of the moving train in \mathcal{F}^B as measured in stationary frame \mathcal{F}^A . (*Ans.* Moving train length = $(x_1 - x_0)$.)

(2.4.4)
Exercise**Graphing the Measurements for Finding Speed, Clock Rates, and Moving Lengths: *The Two-Observer Method***

Complete the figure to the right by locating the quantities x_0 , x_1 , t_0 , t_1 , t'_0 , t'_1 that appear in the figures and the Question sections of (2.4.2) (a), (b), and (2.4.3) (c).

Each single point of the grey area of the figure is represented by a unique coordinate pair $[x, t]$ in \mathcal{F}^A . The train travels rightward with speed v .



(3) The Algebra of Frames

(3.1) Inertial Frames of Reference

We shall see how a physical idea—*the inertial frame of reference*—is modeled as a mathematical concept—a set of *ordered* $[x, t]$ pairs or *vectors*. The following is a repeat of Definitions (1.2.1), pg. 13 and (1.5.1), pg. 16.

DEFINITION: (1.2.1), (1.5.1) A FRAME OF REFERENCE, \mathcal{F} , is a space or environment in which each point x can be located by a coordinate system and the time t_x at point x is determined by a clock at that point. An INERTIAL FRAME OF REFERENCE is a frame of reference which is non-accelerating, that is, free bodies move at constant speed in a straight line or are at rest.

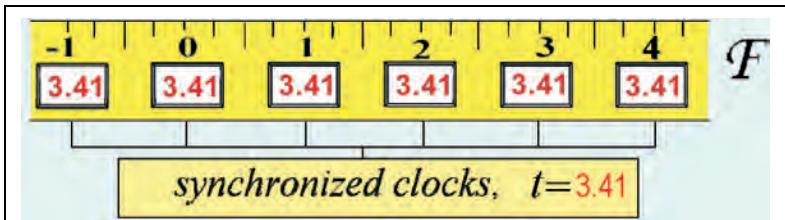
For example, in three-space each point is described by its space coordinates $[x, y, z]$ and its time coordinate t . The quartet $[x, y, z, t]$ therefore describes the state of each point in the frame of reference \mathcal{F} .

(3.1.1)

NOTE: Following the simplification of Section (1.3), we choose our X -axis so as to replace the three space coordinates with just the x coordinate.

(3.1.2)

FIGURE: *The Ruler-Clock Structure of a Frame* (Click [here](#) to see a synchronization of two Galilean frames. Click [here](#) to see a Lorentzian synchronization of two frames.)



A point x on the ruler at time t on the clock is represented by the ordered pair $[x, t]$. In this figure, clocks are shown to be synchronized so at all

points, all clocks simultaneously read the same time, namely, 3.41. In this case, all frame coordinates take the form $[x, 3.41]$.

(3.2) Vector Space Structure of Frames

- (3.2.1)** As seen in Figure (3.1.2), each point x on the ruler of the frame of reference corresponds to an ordered pair $[x, t]$ where x is the distance mark for x on the ruler and t is the time at the point x .

Do these ordered pairs have the mathematical structure of a vector space as required by Definition (A.2.1)? The following proposition says yes.

- (3.2.2)**

PROPOSITION: *The set of all ordered pairs $\{[x, t]\}$ in frame of reference \mathcal{F} forms a vector space as required by (A.2.1). Specifically, for all $[x_1, t_1], [x_2, t_2], [x, t]$ in \mathcal{F}*

$$(3.2.2a) \quad [x_1, t_1] + [x_2, t_2] = [x_1 + x_2, t_1 + t_2].$$

$$(3.2.2b) \quad \text{For all scalars } \alpha, \alpha[x, t] = [\alpha x, \alpha t].$$

PROOF: (PLAUSIBILITY ARGUMENT.)

Justifying (3.2.2a): Suppose a particle or observer from an arbitrary point on the ruler moves

- (3.2.2c)** rightward 8 feet, in 4 seconds, *then moves*
leftward 6 feet, in 3 seconds.

In natural language: The particle covered a net total of $(8 - 6)$ ft. = 2 ft. to the right in $(4 + 3)$ seconds = 7 seconds. In other words, a net total of 2 ft. were covered in a total of 7 seconds. To pass from natural to mathematical language, rewrite (3.2.2c) as

- (3.2.2d)**

$$\begin{array}{r} & \text{ft. sec.} \\ & [+8, +4] \\ + & [-6, +3] \\ \hline = & [+2, +7] \end{array}$$

where $+$ in (3.2.2d) corresponds to the phrase “*then moves*” in (3.2.2c).

The generalization of (3.2.2d) is that for all distances x_1, x_2 , and all times, t_1, t_2 ,

$$(3.2.2e) \quad [x_1, t_1] + [x_2, t_2] \stackrel{\text{def}}{=} [x_1 + x_2, t_1 + t_2]$$

which justifies (3.2.2a) since this agrees with the definition of vector addition (A.2.4b).

Justifying Definition (3.2.2b): From (3.2.2e), we have

$$(3.2.2f) \quad \underbrace{[x, t] + [x, t]}_{2[x, t]} = [2x, 2t],$$

which gives us $a[x, t] = [ax, at]$ when $a = 2$. Using continuity, (3.2.2f) can be extended to (A.2.4c), scalar multiplication for *all* numbers $a \neq 2$. (See Exercise (3.5.1).)

This ends the proof. ■

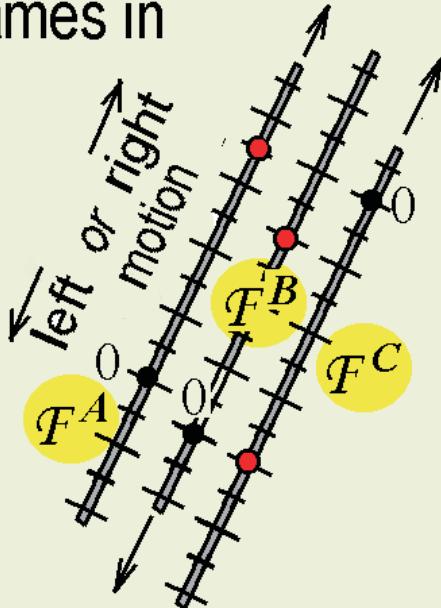
(3.3) Several Parallel Moving Frames

- (3.3.1) **The Train-as-Frame Metaphor:** We expand on the analogy drawn in (1.2.1), pg. 13 which presents a train as a model of an inertial frame. We now consider the case of several inertial frames on parallel tracks, and the way communication proceeds between pairs of frames. In fact, an observer in one train (*inertial frame* \mathcal{F}^A) can communicate with another frame on a parallel track only via a line-of-sight corridor that is perpendicular to both trains. That is, a passenger in one train (*frame*) can only look perpendicularly out his/her own window into the window of a parallel train (*frame*).

.....

(3.3.1a) A Railroad Model for Three Frames

Mathematical model of
three frames in
parallel
motion



Inertial frames, \mathcal{F}^A , \mathcal{F}^B , and \mathcal{F}^C are moving in left-right motion relative to each other on parallel paths. One can think of three infinite-length trains (frames) traveling on separate parallel tracks. Each moving train can be considered to be its own X-axis and the seated passengers inside the train are stationary observers, each with his/her own clock and position x .

(3.3.2)



The distance between the parallel tracks is assumed to be negligible so that there is no time delay for information traveling between the moving frames (trains).

(3.4) Six Rules for Frames

There are six rules, (3.4.1)-(3.4.6), that describe the structure of frames and how communication proceeds between them.

(3.4.1)

All Frames are Physically Identical When at Rest. All parallel frames have rulers that are identical when measured at rest and clocks that are identical in that they run at the same rate when measured at rest. As shown in Figure (3.1.2), each single frame of reference consists of a ruler (*X-axis*) with identical synchronized clocks placed at each point, x .

(3.4.2)

Like Being on Flypaper, Every Observer Remains Fixed Within His/Her Own Proper Frame (or train or X-axis). Once an observer decides to move along an *X-axis*, like a passenger in a railroad car who decides to stroll down the center aisle, that observer is instantly assigned his/her *own* parallel *X-axis*, or HOME FRAME of reference, \mathcal{F} , relative to which he/she is, once again, stationary.

The newly-occupied stationary or rest frame is not necessarily unique. This is seen in Figure (3.3.1a) where each of the moving red dots acquires its own frame, \mathcal{F}^A , \mathcal{F}^B , or \mathcal{F}^C , inside of which the red dot is fixed. The entire frame (*with its fixed red dot*) is seen by an outsider to be moving. That is, observers (*red dots*) within their own proper “railroad cars” are always stationary.¹

¹We will restrict our attention to frames that have constant speed relative to each other.

(3.4.3) **Origin Observers and Clocks.**

DEFINITION: For each frame of reference, $\mathcal{F}^A, \mathcal{F}^B, \mathcal{F}^C, \dots$, the one stationary observer fixed at the origin, $x = 0$, is called the ORIGIN OBSERVER of that frame. The clock at $x = 0$ is the ORIGIN CLOCK. All other observers, fixed at some $x \neq 0$, are called COHORTS of the origin observer in the frame.

(3.4.4) **Line-of-Sight Functions Between Frames.**

In (3.3.1a), pg. 57, we represented frames as a set of trains or rulers moving on parallel tracks. There is a clock at each point of the train or ruler.

In frame (*ruler*) \mathcal{F}^A , Ashley is at position $x = 2$ and time $t = 3.41$, or at $[2, 3.41]_A$ and in the parallel frame (*ruler*) \mathcal{F}^B , Bernie is at position $x = 0$ (*the origin*) and time $t = 3.00$, or at $[0, 3.00]_B$.

As per (3.3.1), at the \mathcal{F}^A end of the line-of-sight corridor (*the vertical red line in the diagram*), Ashley sees Bernie's coordinates $[0, 3.00]_B$ while Bernie, at the opposite \mathcal{F}^B end of the line-of-sight corridor, sees Ashley's coordinates $[2, 3.41]_A$. The *linking* of their coordinates is mathematically expressed by a function

$$G : \mathcal{F}^A \rightarrow \mathcal{F}^B, \quad (\text{A.3.2}),$$

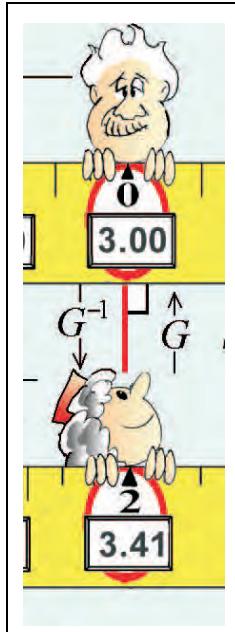
where

$$G([2, 3.41]_A) = [0, 3.00]_B$$

or, in terms of the inverse function $G^{-1} : \mathcal{F}^B \rightarrow \mathcal{F}^A$,

$$G^{-1}([0, 3.00]_B) = [2, 3.41]_A.$$

The physical line-of-sight pairing between all coordinates of \mathcal{F}^A and \mathcal{F}^B can be expressed as input-output pairs of the function G as follows:



(3.4.4a)

DEFINITION: *Frames (rulers with clocks) \mathcal{F}^A and \mathcal{F}^B are traveling parallel to each other. If coordinates $[x, t]_A$ of frame \mathcal{F}^A and $[x', t']_B$ of frame \mathcal{F}^B lie at opposite ends of an imaginary line that is perpendicular to both frames (rulers), then this perpendicular line is called a LINE-OF-SIGHT CORRIDOR between the frames. The linked coordinates at the corridor end-points define the LINE-OF-SIGHT FUNCTION*

$$G : [x, t]_A \rightarrow [x', t']_B \quad \text{or} \quad G([x, t]_A) = [x', t']_B$$

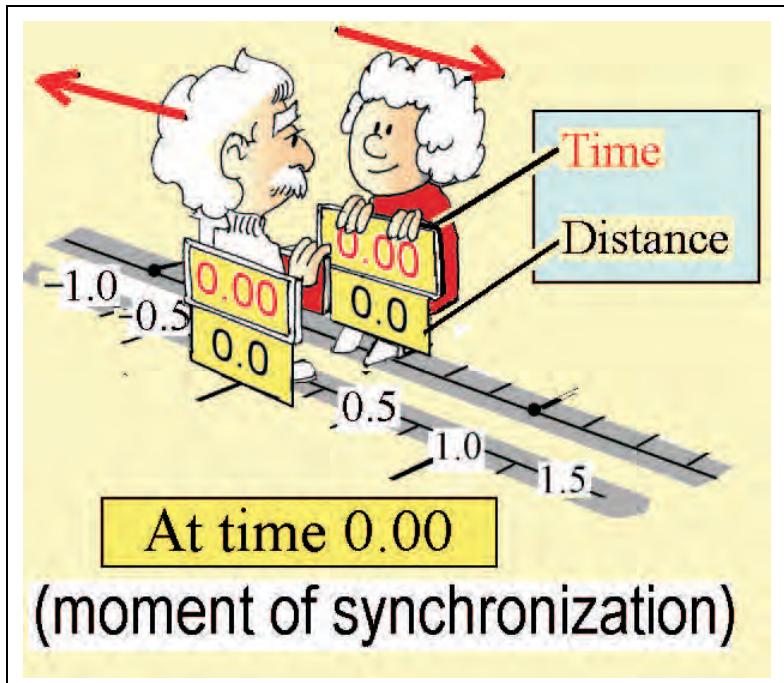
(See Definition (A.3.2) and Figure (3.4.4).)

|| In sum, Figure (3.4.4) shows Ashley in frame \mathcal{F}^A and Bernie in (parallel) frame \mathcal{F}^B viewing each other's coordinates through functions $G : \mathcal{F}^A \rightarrow \mathcal{F}^B$ and $G^{-1} : \mathcal{F}^B \rightarrow \mathcal{F}^A$.

.....

NOTE: The Structure of G . In Chapters (5) and (7), we shall see that line-of-sight functions (3.4.4a) will necessarily be *linear (representable as 2×2 matrices)* as defined by Definition (A.4.1). Functional correspondence, (3.4.4) must be constrained by physical requirements such as *synchronization* between pairs of frames, which is to say, that at time $t = 0$ in each frame, origin observers (at $x = 0$) are at opposite ends of a perpendicular line-of-sight corridor. Synchronization is formally defined in (3.4.5) following.

(3.4.5)

All Pairs of Frames are Synchronized at $[0, 0]$.

Synchronization between a pair of frames, \mathcal{F}^A and \mathcal{F}^B , means that at $x = 0$ and time $t = 0$ in frame \mathcal{F}^A , the origin observer (3.4.3) looks through the line-of sight corridor at $[0, 0]_A$ in \mathcal{F}^A and sees the coordinates $[0, 0]_B$ of the origin observer in \mathcal{F}^B and vice versa.

We formally define synchronization in terms of the line-of-sight function $G: \mathcal{F}^A \rightarrow \mathcal{F}^B$ (3.4.4a):

(3.4.5a)

DEFINITION: *Origin observers (in respective frames \mathcal{F}^A and \mathcal{F}^B) are mutually SYNCHRONIZED AT $[0, 0]$ if the line-of-sight function, $G: \mathcal{F}^A \rightarrow \mathcal{F}^B$, has the property that*

$$G(\vec{0}_A) = \vec{0}_B,$$

that is,

$$G([0, 0]_A) = [0, 0]_B,$$

where $\vec{0}_A = [0, 0]_A$ and $\vec{0}_B = [0, 0]_B$ are the respective origins $x = 0$ at time $t = 0$ of frames of reference, \mathcal{F}^A and \mathcal{F}^B .

(3.4.5b) **QUESTION:** *Are there preferred points or observers?* Although we claim there is no preferred frame, there always is a special, or preferred point within each frame—its origin $x = 0$. In fact, as noted in (3.4.5a), all (*parallel*) frames must have their respective origins $x = 0$ aligned along a perpendicular line-of-sight corridor at local time $t = 0$. In this sense, the origin and the origin observers (3.4.3) within each frame are special or “preferred.”

(3.4.6) Symmetry Between Frames.

The following notation G_v for the line-of-sight function enhances that of (3.4.4a) where v is the relative speed between frames.

(3.4.6a) **NOTATION:** *Suppose frame \mathcal{F}^B has speed v as seen from frame \mathcal{F}^A , and $[x, t]_A$ in frame \mathcal{F}^A and $[x', t']_B$ in frame \mathcal{F}^B are physically at opposite ends of a line-of-sight corridor. We then write*

$$(3.4.6b) \quad G_v : [x, t]_A \rightarrow [x', t']_B,$$

where symbol G_v is an enhanced form of the symbol G (3.4.4a).

Postulate (3.4.6c) following uses the terminology of (3.4.6b) to recast the symmetry postulate (1.6.1) in terms of line-of-sight functions as illustrated in Figures (3.4.4) and (3.4.5). (*A postulate or axiom is an assertion that is accepted without proof (1.5.3)*).

(3.4.6c) POSTULATE: (*Symmetry Principle and Lines of Sight.*) When frame \mathcal{F}^B is moving with constant speed, v , relative to frame \mathcal{F}^A , then frame \mathcal{F}^A will be observed to have speed $-v$ relative to frame \mathcal{F}^B . Symbolically,

$$(3.4.6d) \quad G_v : \mathcal{F}^A \rightarrow \mathcal{F}^B \quad \text{if and only if} \quad G_{-v} : \mathcal{F}^B \rightarrow \mathcal{F}^A.$$

(3.4.6e)

PROPOSITION: **Analytic Statement of Symmetry Principle.** The line-of-sight functions G_v and G_{-v} are inverses of each other, which is to say

$$(3.4.6f) \quad G_v^{-1} = G_{-v}.$$

PROOF: In terms of individual coordinates, we rewrite (3.4.6d) as

$$[x, t]_A \xrightarrow{G_v} [x', t']_B \quad \text{and} \quad [x', t']_B \xrightarrow{G_{-v}} [x, t]_A$$

if and only if $[x, t]_A \xrightarrow{G_v} [x', t']_B \xrightarrow{G_{-v}} [x, t]_A$ abbreviation of previous line

if and only if $[x, t]_A \xrightarrow{G_{-v} \circ G_v} [x, t]_A$ Definition (A.3.6)
of composition, \circ

if and only if $G_{-v} \circ G_v = I_2$. Definition (A.3.4)
Factors G_{-v} and G_v are
inverses of each other,

which is to say

$$(3.4.6g) \quad G_{-v} = \text{Inverse}(G_v) = G_v^{-1} \quad \text{and}$$

$$(3.4.6h) \quad G_v = \text{Inverse}(G_{-v}) = (G_{-v})^{-1}.$$

It is Equation (3.4.6g) above that directly gives us (3.4.6f). (Equation (3.4.6h) indirectly yields (3.4.6g). See Exercise (3.5.4).) This proves the proposition. ■



(3.5) Exercises

(3.5.1)

Exercise

Justifying Scalar-Vector Multiplication. Using a continuity argument, here is how vector addition alone (A.2.4b), pg. 361 implies the scalar multiplication property (A.2.4c).

(3.5.1a) Show that for all integers, n , $n[x, t] = [nx, nt]$.

(3.5.1b) Show that for all integers n , $(1/n)[x, t] = [(1/n)x, (1/n)t]$.

(3.5.1c) Show that for all rational numbers (fractions), $q = m/n$, $q[x, t] = [qx, qt]$.

(3.5.1d) Assuming the scalar product function

$$(*) \quad \Phi(q) = [qx, qt]$$

is continuous in q , show that for all real numbers a we have

$$(**) \quad a[x, t] = [ax, at].$$

(Continuity means $\lim_i \Phi = \Phi(\lim_i)$, or, when $\lim_i q_i = a$, $\lim_i \Phi(q_i) = \Phi(\lim_i q_i) = \Phi(a)$.)

Hint: Let $\lim_i q_i = a$ where $\{q_i\}$ is a sequence of rational numbers. Then from the continuity of Φ ,

$$(***) \lim_i \Phi(q_i) = \Phi(\lim_i q_i) = \Phi(a).$$

From (*), the rightmost term of (***)) is $\Phi(a) = [ax, at]$ which also equals its leftmost term, namely,

$$\lim_i \Phi(q_i) = \lim_i [q_i x, q_i t] \quad \text{from (*)}$$

$$= \lim_i q_i [x, t] = a[x, t] \quad \text{from (3.5.1c).}$$

(3.5.2)
Exercise

The Frame for Slowing Clocks. Show how the example of Figure (1.10.2) can be expressed in terms of a spacetime diagram. That is, in each of the two panels, describe the placement of identical rulers in frame \mathcal{F}^{car} of the car and in frame $\mathcal{F}^{platform}$ of the platform.

(3.5.3)
Exercise

The Frame for Shrinking Lengths. Show how the example of Figure (1.10.4) can be expressed in terms of a spacetime diagram. That is, in each of the two panels, describe the placement of identical rulers in frame \mathcal{F}^{car} of the car and in frame $\mathcal{F}^{platform}$ of the platform.

(3.5.4)
Exercise

Inverse of G_v . Show that (3.4.6h) implies (3.4.6g).

Hint: Take the multiplicative inverse of both sides of (3.4.6h) and use the fact that for any invertible function f we have $(f^{-1})^{-1} = f$.



(4) The Graphing of Frames

(4.1) The Filmstrip Model of Spacetime

Our objective is to provide an intuitive way of graphing particles in motion.

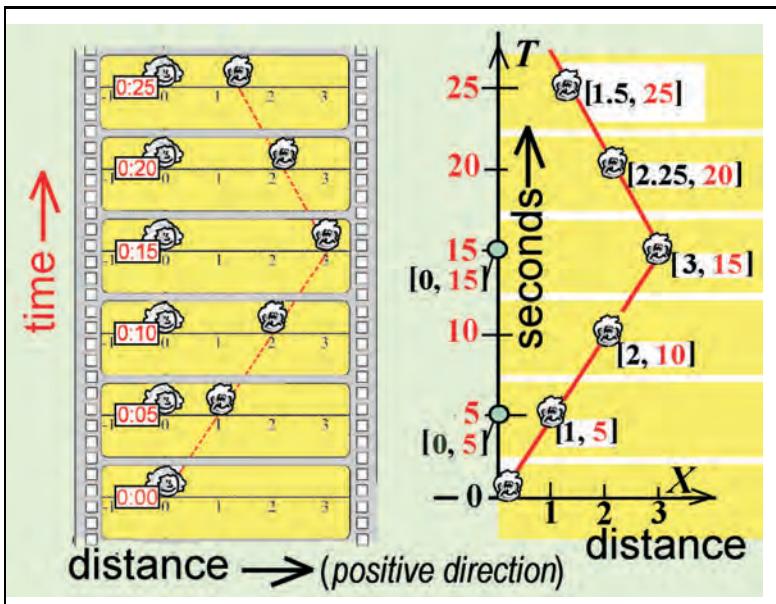
In 3-dimensional space, we may represent a particle by the quartet $[x, y, z, t]$. When the particle moves, it will change its location with time, that is, the space coordinates become functions of the time coordinate t , that is, $x = x(t)$, $y = y(t)$, and $z = z(t)$. The set of these coordinates gives us a picture of the *motion* of that point as well as its position.

We shall concentrate on motion in 1-dimensional space only so that points are represented by coordinates of form $[x, t]$. (*We have seen in Section (1.3) that this “special case” is, in fact, sufficiently general for our purposes.*)

We already have a familiar model for representing objects that move in only one dimension — it is the filmstrip, which records snapshots of an object’s position at regular time intervals.

We now expand upon this idea.

(4.1.1)

FIGURE: Bernie's Motion from Ashley's Point of View:

Five-second intervals of time correspond to each vertically advancing filmstrip frame in the left half of the figure. As we follow successive filmstrip frames in the upward direction, we view the left-to-right (positive) changes in Bernie's position along the X -axis.

(4.1.1a) The Setup for the Filmstrip of Figure (4.1.1). As indicated in Figure (4.1.1) the positive direction of distance is to the right and the negative direction is to the left—regardless of the direction in which Bernie and Ashley are facing.

Although the film strip is two-dimensional, Bernie's motion is in one-dimensional (*horizontal*) in space only. The second (*vertical*) axis tells us the time that Bernie finds himself at any position on the horizontal axis.

(4.1.1b) Extracting the X - T Graph (*Mathematics Models Reality*): In Figure (4.1.1), lift Bernie's path, like a decal, from the filmstrip (left-hand panel) to create the two-dimensional graph (right-hand panel). In this natural way, Bernie's 1-dimensional, left-to-right or right-to-left position x , say, corresponds to the horizontal X -axis reading and the time t corresponds to the vertical T -axis reading.

The right-hand side of Figure (4.1.1) motivates the following definition:

(4.1.2)

DEFINITION: Given an object in a frame of reference (1.2.1) that can move in a single direction. An object's position at point x and time t is represented by ordered pairs of the form

$$(4.1.2a) \quad [x, t] \quad \text{or} \quad \begin{bmatrix} x \\ t \end{bmatrix}$$

in a graph with distance axis X and time axis T . This graph is called the SPACETIME GRAPH (or SPACETIME DIAGRAM) where each ordered pair (4.1.2a) is called an EVENT in spacetime.

The path of all coordinates of a moving object over time—the set all its events—is called the WORLDLINE of that object in spacetime.

For an object to move in three directions, ordered pairs (4.1.2a) are replaced with $[x_1, x_2, x_3, t]$.

(4.1.3)

NOTE: The space and time axes are not necessarily perpendicular to each other.

(4.1.4)

Question: Why do we use the term “event” to describe the point $[x, t]$ in a spacetime graph?

Answer: An object in spacetime, which is generally in motion, will have coordinates of the form $[x, t]$ which are generally changing in time. (*The coordinates of an object will change since t is always changing, even if the object is fixed at some point x .*) Hence, when we refer to a particular coordinate $[x, t]$ in spacetime, we are describing not only the location of an object, but the time it held that position. Specifying where and when something is corresponds to the intuitive definition of an *event*.

(4.2) Constant Velocities in Spacetime

Some of the most interesting effects in special relativity occur when we consider objects moving with constant velocity relative to one another (1.5) which we now formally define.

(4.2.1)

DEFINITION: CONSTANT VELOCITY is that motion along a straight line in space, where equal distances are covered in equal times. Therefore, any motion that veers off a straight line in space, or that changes speed along a straight line, has NON-CONSTANT VELOCITY.

(4.2.2)

NOTE: The notation \vec{v} that uses the overhead arrow, indicates speed v and direction. When we are only interested in the speed, a number for which no directions are involved, then we use the unornamented, symbol, v .

The following theorem equates constant speed (a physical quantity) with worldlines that are straight (a geometric quantity).

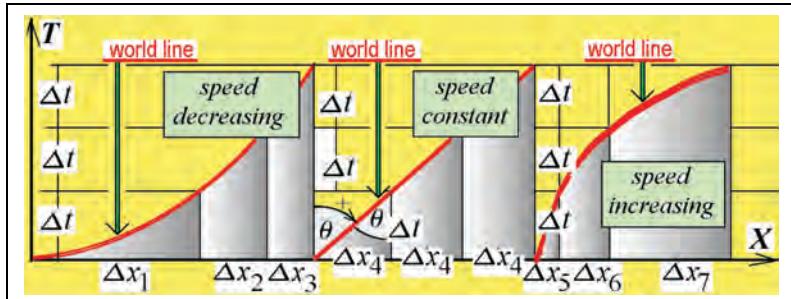
(4.2.3)

THEOREM: (*Constant Speed \Leftrightarrow Straight Line Trajectory*) A particle moving on the X -axis has constant speed if and only if its worldline (trajectory) in the spacetime graph (4.1.2) is a straight line.

PROOF: (*informal*) Figure (4.2.4) shows that even though motion of a particle proceeds along a straight 1-dimensional track—the X -axis—the 2-dimensional graph of this motion in spacetime can be either curved or straight. The straight path corresponds to constant speed. Here are details.

(4.2.4)

FIGURE: *Straight-Line Motion Along the X-Axis*



The Δt intervals on the vertical T -axis of Figure (4.2.4) are all equal. Since average speed over the Δx_i th interval is $v_i = \Delta x_i / \Delta t$, we have

- LEFT-HAND WORLDLINE IN (4.2.4): Since $\Delta x_1 > \Delta x_2 > \Delta x_3$,

$$\frac{\Delta x_1}{\Delta t} > \frac{\Delta x_2}{\Delta t} > \frac{\Delta x_3}{\Delta t}, \quad \text{which means } v_1 > v_2 > v_3, \quad \text{or } v \text{ decreases as time increases.}$$

- MIDDLE WORLDLINE IN (4.2.4): Since all distances equal Δx_4 , the average speed over each of the three distance intervals is

$$\frac{\Delta x_4}{\Delta t} = v_4 \quad \text{which means speed } v \text{ is constant for all time } t.$$

- RIGHT-HAND WORLDLINE IN (4.2.4): Since $\Delta x_5 < \Delta x_6 < \Delta x_7$,

$$\frac{\Delta x_5}{\Delta t} < \frac{\Delta x_6}{\Delta t} < \frac{\Delta x_7}{\Delta t}, \quad \text{which means } v_5 < v_6 < v_7, \quad \text{or } v \text{ increases as time increases.}$$

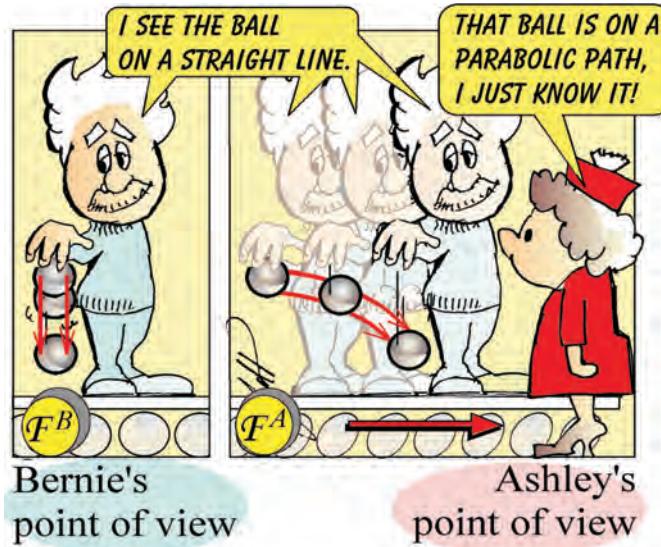
This ends the proof. ■

(4.2.5)

A curved trajectory might be observed when the moving frame is not inertial (1.5.1). If the spacetime graph is not straight, the velocity is not constant which implies the object is accelerating—a force is acting on it. In the following figure, (4.2.6), the force in question is that of *gravity*. Hence, the trajectory of the falling ball is seen to have two shapes, depending on the frame of the observer.

(4.2.6)

FIGURE: One Trajectory, Two Perceptions.



Bernie, from his frame, \mathcal{F}^B , sees the accelerating ball following a straight, vertical path. But Ashley, from her frame, \mathcal{F}^A , sees the ball moving to her right which causes her to observe the same ball on a downward parabolic path.

(4.3) Worldlines are Parallel to the Home Frame Time Axis

- (4.3.1) From (3.4.2), if each observer is at rest at some fixed point x_0 in frame \mathcal{F}^A say, then the worldline of that observer in the X_A - T_A spacetime graph is a straight line that is parallel to the time axis T_A . This follows since stationary observers have coordinates $[x_0, t]$ where distance x_0 is constant and time t varies.

As a special case, if the (fixed) observer is the Origin Observer (3.4.3) at $x = 0$, then his/her world line coincides with the T_A -axis itself.

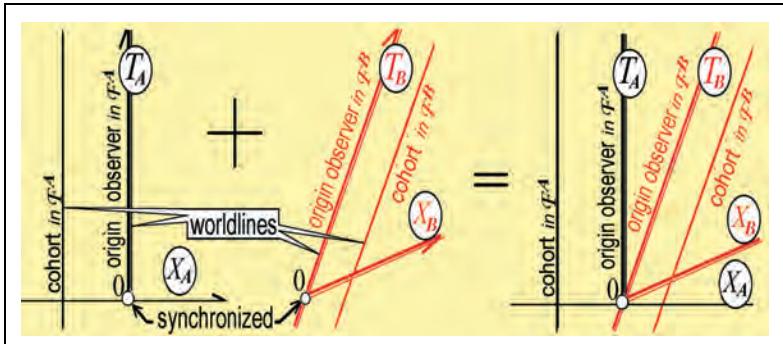
Suppose an observer fixed in a home frame \mathcal{F}^A sees an observer fixed in a moving frame \mathcal{F}^B that has constant speed relative to \mathcal{F}^A . Then the spacetime graph showing both frames will show the two time axes, T_A and T_B , at some angle relative to each other. (*Due to synchronization (3.4.5) T_A and T_B intersect at common time $t = 0$.*) Also, from Figure (4.2.4), we know that in \mathcal{F}^A the B -observer will trace a straight line at some angle; if the B -observer is at the origin in \mathcal{F}^A , then this straight line corresponds to the T_B -axis as seen in \mathcal{F}^A .

Hence, if a spacetime graph shows the relative motion of n inertial frames of reference, then there are at most n slopes that can be assigned among all the worldlines that are drawn in the graph.

.....

The following spacetime graph illustrates the only two slopes that are possible for all worldlines of moving objects that belong to either one of two synchronized inertial frames of reference, \mathcal{F}^A , or \mathcal{F}^B .

(4.3.2)

FIGURE: Two Frames Imply Only Two Slopes for All Worldlines

These graphs depict the worldlines of objects or observers in two synchronized inertial frames of reference, \mathcal{F}^A and \mathcal{F}^B , that are in constant motion relative to each other. In each separate spacetime diagram, the origin observer and all her (fixed) cohorts have worldlines that are parallel straight lines—all worldlines in each frame have the same slope. The worldline of the origin observer at $x = 0$ for any t , coincides with her own time axes, T_A or T_B . When the spacetime diagrams of the frames overlap (the rightmost graph of Fig. (4.3.2)), the origins must overlap due to synchronization (3.4.5a). In sum, in the overlapping spacetime diagrams, there can only be two slopes for the worldlines of all the observers in frames \mathcal{F}^A and \mathcal{F}^B . (Do not be distressed to see the X -axes, X_A and X_B , are not both at the same angle. As we shall see in (6.2.3), pg. 99, these axes represent respective lines of simultaneity—at time $t = 0$ —and will necessarily be drawn at different angles.)

(4.4) Simultaneous and Static Events

There is a straightforward, graphical way to characterize *simultaneous events* and *static particles* in spacetime diagrams. First, we present a formal definition of these terms:

(4.4.1)

DEFINITION: (See Figure (4.4.4).)

SIMULTANEOUS EVENTS in a specific frame \mathcal{F} are represented as a pair of events, $[x, t_0]$ and $[x', t_0]$, in a spacetime graph with the same time coordinate t_0 . A LINE OF SIMULTANEOUS EVENTS is the set of all events $[x, t_0]$ that take place at time t_0 .

(4.4.1a) Geometrically, we say that a line that is parallel to the distance axis, X , is a line of simultaneous events.

A STATIONARY, or STATIC SET OF EVENTS at point x_0 is the set of points, $[x_0, t]$, with a fixed space coordinate, x_0 , and a range of the time coordinate, t —the observer is fixed and time varies.

(4.4.1b) Geometrically, a stationary or static event is represented as a worldline that is parallel to the time axis, T .

(4.4.2)

NOTE: We have already seen in (4.1.3) that the X and T axes need not be mutually perpendicular. This non-perpendicularity is illustrated in Figure (4.4.4).

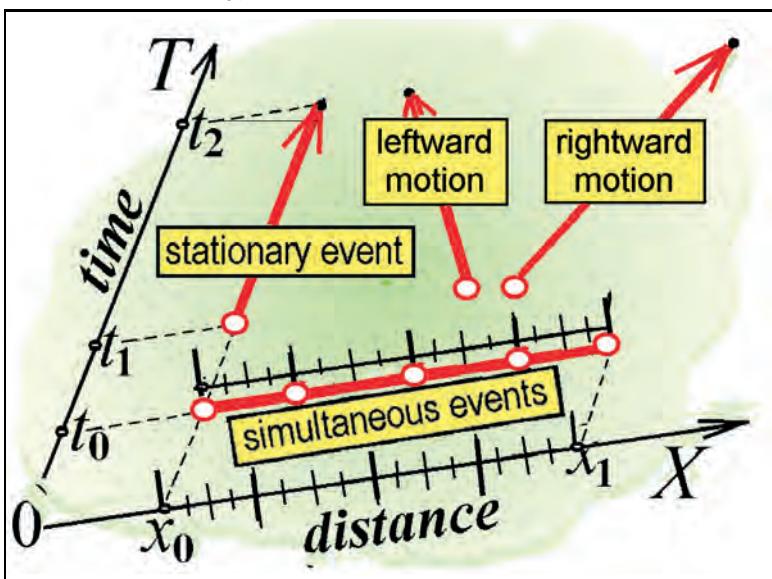
(4.4.3)

NOTE: One can interpret simultaneous events in frame \mathcal{F} that occur at time t_0 —points of the form $[x, t_0]$ —in the spacetime graph as a subset of the X -axis of \mathcal{F} at time t_0 (*the X-axis at “age” t_0 in \mathcal{F}*).

Hence, lines of simultaneity at time t_0 , are always **parallel** to the “usual” X axis at time 0.

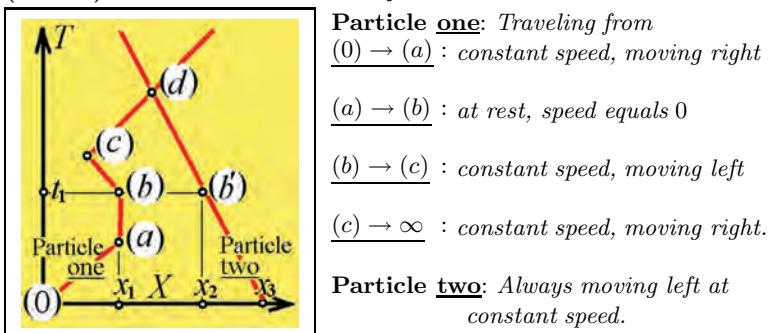
In sum, A stationary event has a worldline (trajectory) parallel to the T -axis. The worldlines of leftward (or rightward) traveling objects are characterized by their positive (or negative) angles with respect to the T -axis. Simultaneous events lie on a line that is parallel to the X -axis.

(4.4.4)

FIGURE: Worldlines of Simultaneous, Stationary, and Other Events .

(4.4.5)

EXAMPLE: Worldlines of two particles are illustrated in the following figure.

(4.4.5a) FIGURE: Worldlines for Two Particles

(In Figure (4.4.5a), the axes happen to be perpendicular but, as Figure (4.4.4) shows, this is not necessary.) Particle one and Particle two move in one-dimensional space. Except for the points, (a), (b), and (c), where Particle one experiences a force that causes an idealized instantaneous change in direction, there are no forces involved. At point (d), both particles theoretically seem to collide since they share

the same $[x, t]$ coordinates. However, we always assume particles that are in motion relative to each other occupy different parallel frames (3.4.2) that are a negligible distance apart.



(4.5) Linearity of Line-of-Sight Functions

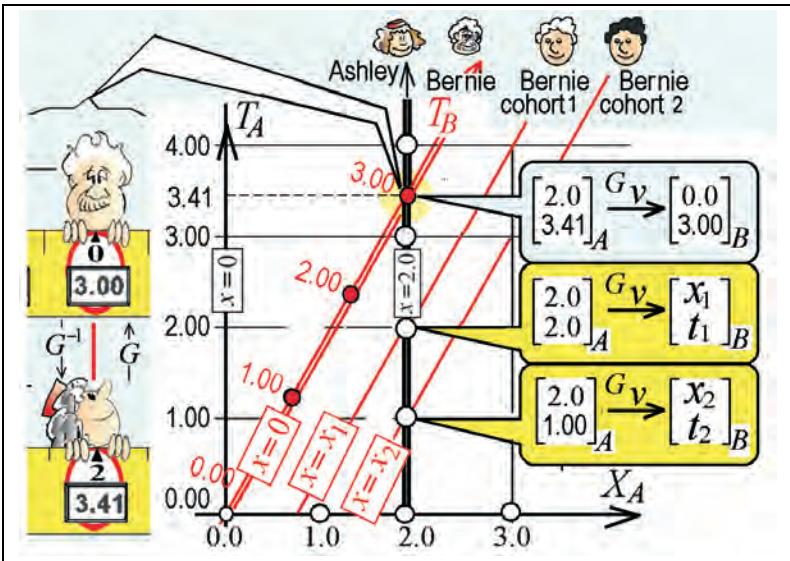
Figure (4.5.1) will provide a graphical context for the line-of-sight function G_v , one instance of which is illustrated in Figure (4.3.2). Moreover, note that

- **To choose an arbitrary observer** at an arbitrary time in \mathcal{F}^A , is to choose a point on a black, vertical worldline. Similarly, choosing an arbitrary observer at an arbitrary time in \mathcal{F}^B is to choose a point on a red, slanted worldline.
- **Each** (black, vertical) worldline of \mathcal{F}^A intersects one and only one (slanted, red) worldline of \mathcal{F}^B .
- **The two points at the intersection** of worldlines—point $[x, t]_A$ on the (vertical, black) worldline in \mathcal{F}^A and point $[x', t']_B$ on the (slanted, red) worldline in \mathcal{F}^B —define the endpoints of the line-of-sight corridor between \mathcal{F}^A and \mathcal{F}^B . That is, these points represent the same event in different coordinate systems so that the line-of-sight function G_v linking these two points is expressed by the equation $G_v([x, t]_A) = [x', t']_B$.

For example, Figure (4.5.1) shows the event defined by the intersection of Ashley's and Bernie's worldlines corresponds to the linked points $[2.0, 3.41]_A$ in Ashley's frame \mathcal{F}^A and $[0.0, 3.0]_B$ in Bernie's frame \mathcal{F}^B .

(4.5.1)

FIGURE: A Point Where a Line-of-Sight Corridor Links Ashley's Vertical Worldline in \mathcal{F}^A to a Slanted Worldline in \mathcal{F}^B .



The three colinear points, $[2.0, 3.41]_A$, $[2.0, 2.0]_A$, and $[2.0, 1.0]_A$ in \mathcal{F}^A lie on Ashley's worldline, which is a straight (vertical) line, all points of which are linked by the line-of-sight function G_v to the three colinear points $[0, 3.0]_B$, $[x_1, t_1]_B$, and $[x_2, t_2]_B$ which lie on a straight line (not a worldline) in \mathcal{F}^B . Intuitively, the straight worldline of Ashley in \mathcal{F}^A casts a “shadow” of itself onto an identical straight line in \mathcal{F}^B . The following holds true:

(4.5.2)

CONCLUSION: Any line-of-sight function $G_v : \mathcal{F}^A \rightarrow \mathcal{F}^B$ sends straight lines of the X_A - T_A spacetime graph to straight lines of the X_B - T_B spacetime graph.

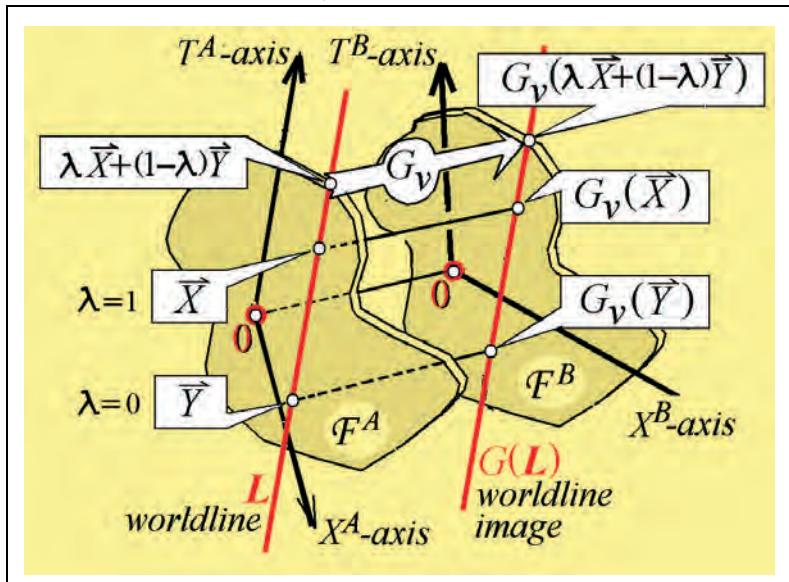
In fact, G_v may be interpreted as a function that leaves each point of any straight fixed, where only the coordinates (names) are changed from \mathcal{F}^A to \mathcal{F}^B coordinates.

The following figure illustrates a line-of-sight “name-changing” function $G_v : \mathcal{F}^A \rightarrow \mathcal{F}^B$ that sends points (vectors) of a straight line L in \mathcal{F}^A to points (vectors) $G_v(L)$, which also form a straight line in \mathcal{F}^B .

(4.5.3)

FIGURE: Line-Preserving Functions

$$G_v : \mathcal{F}^A \rightarrow \mathcal{F}^B, G_v^{-1} : \mathcal{F}^B \rightarrow \mathcal{F}^A$$



In any vector space, a straight line \$L\$ containing points \$\vec{X} \neq \vec{Y}\$ must be the collection of points of the form

$$\lambda\vec{X} + (1 - \lambda)\vec{Y} \text{ for } -\infty < \lambda < \infty$$

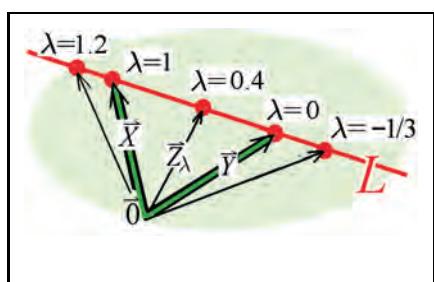
Def (A.2.14).

The line-of-sight function \$G_v\$ sends the three colinear points \$\vec{X}\$, \$\vec{Y}\$, and \$\lambda\vec{X} + (1 - \lambda)\vec{Y}\$ in \$\mathcal{F}^A\$, respectively, to the three colinear points \$G_v(\vec{X})\$, \$G_v(\vec{Y})\$, and \$G_v(\lambda\vec{X} + (1 - \lambda)\vec{Y})\$ in \$\mathcal{F}^B\$.

(4.5.4)

Any line-of-sight function \$G_v\$ preserves proportions since, from (4.5.2), \$L\$ in \$\mathcal{F}^A\$ and \$G_v(L)\$ in \$\mathcal{F}^B\$ are, point-for-point, the same line but with different coordinates (names).

Hence, proportions are preserved in the sense that the mid-point of \$L\$ is sent to the mid-point of \$G_v(L)\$, a point \$\vec{Z}_\lambda\$ that is 0.4 the distance from \$\vec{Y}\$ to \$\vec{X}\$ has an image \$G_v(\vec{Z}_\lambda)\$ that is 0.4 the distance from \$G_v(\vec{Y})\$ to \$G_v(\vec{X})\$, etc. Definition (4.5.5) provides, in mathematical terms, the equation that



characterizes line-preserving and proportion-preserving functions $G_v : \mathcal{F}^A \rightarrow \mathcal{F}^B$.

NOTE: In what follows, we will deal with more general vector spaces \mathcal{F}^A and \mathcal{F}^B that are not necessarily spacetime vector spaces. Hence, for the function $G_v : \mathcal{F}^A \rightarrow \mathcal{F}^B$, we suppress the subscript v and write

G instead of G_v .

(4.5.5)

DEFINITION: A function G from a vector space \mathcal{F}^A to a vector space \mathcal{F}^B PRESERVES PROPORTION and is LINE-PRESERVING if and only if for any fixed vectors $\vec{X} \neq \vec{Y}$ in \mathcal{F}^A where $G(\vec{X}) \neq G(\vec{Y})$,

$$(4.5.5a) \quad G\left(\underbrace{\lambda\vec{X} + (1 - \lambda)\vec{Y}}_{\text{line in } \mathcal{F}^A}\right) = \underbrace{\lambda G(\vec{X}) + (1 - \lambda)G(\vec{Y})}_{\text{line in } \mathcal{F}^B}$$

for all real λ , $-\infty < \lambda < \infty$.

To say that a line-of-sight function preserves straight lines (see (4.5.2)) is to describe a visual *geometric property*. We now show that this geometric property is equivalent to an *algebraic property*—more exactly, the *linear-algebraic property* that is given in the following theorem.

(4.5.6)

THEOREM: *Line-Preserving Functions = Linear Functions.* Given vector spaces \mathcal{F}^A and \mathcal{F}^B and a function $G : \mathcal{F}^A \rightarrow \mathcal{F}^B$ where $G(\vec{0}) = \vec{0}$. Then

(4.5.6a) G is linear in the sense of (A.4.1)

if and only if

(4.5.6b) G preserves straight lines and proportions.

PROOF:

G linear (4.5.6a) (sufficient)	G line-preserving (4.5.6b) (necessary)
--	--

If G is assumed to be lin-

ear then, in Definition (A.4.1a), we can replace \vec{P}_1 with \vec{X} and \vec{P}_2 with \vec{Y} to obtain

$$(4.5.6c) \quad G(\lambda_1 \vec{X} + \lambda_2 \vec{Y}) = \lambda_1 G(\vec{X}) + \lambda_2 G(\vec{Y})$$

for all scalars λ_1 and λ_2 . In (4.5.6c), set $\lambda_1 = \lambda$ and replace λ_2 with $(1 - \lambda)$ to arrive at the form (4.5.5a) for a typical line-preserving function.

G line-preserving (4.5.6b) (sufficient)	\implies	G linear (4.5.6a) (necessary)
--	------------	--

We first note that if G preserves lines, then from (4.5.5), for all scalars λ ,

$$G(\lambda \vec{P}_1 + (1 - \lambda) \vec{P}_2) = \lambda G(\vec{P}_1) + (1 - \lambda) G(\vec{P}_2).$$

Set $\vec{P}_2 = \vec{0}$ and invoke the hypothesis that $G(\vec{0}) = \vec{0}$ to finally obtain

$$(4.5.6d) \quad G(\lambda \vec{P}_1) = \lambda G(\vec{P}_1). \quad \text{scalar } \lambda \text{ "factors out" of } G$$

First, choose some λ not equal to 0 or 1. Then for all vectors \vec{P}_1 and \vec{P}_2 and for all real scalars λ_1 and λ_2 we write

$$\begin{aligned} (4.5.6e) \quad G(\lambda_1 \vec{P}_1 + \lambda_2 \vec{P}_2) &= G\left(\lambda \left[\frac{\lambda_1}{\lambda}\right] \vec{P}_1 + (1-\lambda) \left[\frac{\lambda_2}{(1-\lambda)}\right] \vec{P}_2\right) \\ &= \lambda G\left(\left[\frac{\lambda_1}{\lambda}\right] \vec{P}_1\right) + (1-\lambda) G\left(\left[\frac{\lambda_2}{(1-\lambda)}\right] \vec{P}_2\right) \quad \text{since } G \text{ preserves} \\ &\quad \text{lines (4.5.5a)} \\ &= \lambda \left[\frac{\lambda_1}{\lambda}\right] G(\vec{P}_1) + (1-\lambda) \left[\frac{\lambda_2}{(1-\lambda)}\right] G(\vec{P}_2) \quad \begin{matrix} \text{"factor out"} \\ \left[\frac{\lambda_1}{\lambda}\right], \left[\frac{\lambda_2}{(1-\lambda)}\right] \\ \text{from } G \text{ (4.5.6d)} \end{matrix} \\ &= \lambda_1 G(\vec{P}_1) + \lambda_2 G(\vec{P}_2). \quad \text{numerator/denominator cancellation} \end{aligned}$$

Equating the very first and last terms in this string of equalities proves that G is linear as required by (A.4.1a). The proof is done. ■

(4.5.7)

Note: Necessary vs. Sufficient. The proof of Theorem (4.5.6) consists of showing that the two statements, $P = \text{(4.5.6a)}$ and $Q = \text{(4.5.6b)}$, are logically equivalent in the sense that the truth of one implies the truth of the other. This illustrates the important difference between “*necessary*” and “*sufficient*” since converse implications of the form $P \Rightarrow Q$ and $Q \Rightarrow P$ are, in general, *not* logically equivalent.

(4.6) Exercises

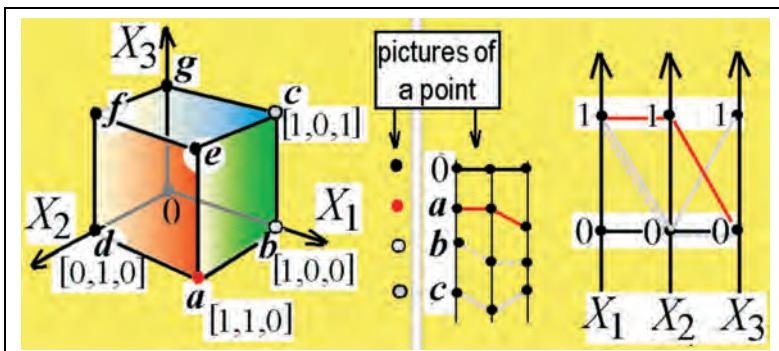
(4.6.1)
Exercise

Cartesian Coordinate Systems. In Figure (4.4.5a), what are the X - T coordinates of the three points indicated as a , b , and b' ?

(4.6.2)
Exercise

Pictures of Higher Dimensions. With Cartesian coordinates, we usually draw mutually perpendicular axes X_1 , X_2 , and X_3 to represent points $[x_1, x_2, x_3]$ in 3-space, as is done in the figure of (18.3.1) and in the left-hand panel of Figure (A).

Figure (A): Two Representations of Points in 3-Space.



However, when the three axes, X_1 , X_2 , X_3 , are drawn vertically and parallel to each other, then, in Figure (A), the points (dots) in the *left-hand* panel are represented as straight lines in the *right-hand* panel.

One advantage of the representation of the right-hand panel of Figure (A) is that we can easily draw spaces of dimension n greater than three. This is done by drawing n parallel vertical axes.

One disadvantage of the right-hand panel of Figure (A) is that familiar shapes in 2-space and 3-space can take on unfamiliar forms. For example, the four corners of the cube, 0 , a , b , and c , which are *points* in the left-hand panel, look like *lines* in the right-hand panel.

.....

As another example, the set of points of a sphere does not have a pleasant “ball-like” shape (*see Exercise (4.6.2c)*).

Note: Even though the triples, $[x_1, x_2, x_3]$, are represented as jagged lines in the right-hand panel of Figure (A), these “lines” are still called “points.”

.....

- (4.6.2a) **Question:** Draw the other four points of the cube, d , e , f , and g , in the manner of the right-hand panel, i.e., using parallel axes.
- (4.6.2b) **Question:** Show that in the right-hand panel, using parallel axes to represent 3-space, the set of all points of a cube fill the band between $[0, 0, 0]$ and $[1, 1, 1]$.
- (4.6.2c) **Question:** Draw any four points $[x_1, x_2, x_3]$ that lie on the surface of the unit sphere in the manner of both the left-hand panel (Cartesian coordinates) and the right-hand panel (with parallel axes).

Hint: A point $[x_1, x_2, x_3]$ on the surface of the unit sphere has the property $x_1^2 + x_2^2 + x_3^2 = 1$.

(4.6.3) Exercise

Film Strip Model of Motion. It can be argued that the graph in the right-hand panel of Figure (4.1.1) is more informative than the “reality” of photographs provided by the film strip (left-hand panel) since photographs record only a *finite* set of distances and times, $[x, t]$. On the other hand, the mathematical model of the right-hand panel fills in the gaps, showing *infinitely* many points $[x, t]$.

For example, the photographs do not tell us the time t_0 when Bernie was at the point 2.6 on the rightward leg of his journey. But (once we know the equation of a straight line in X - T space), the mathematical model shown in the right-hand panel of (4.1.1) gives all $[x, t]$ pairs, including $[2.6, t_0]$.

-
- (4.6.3a) **Question:** In Figure (4.1.1), at what time t_0 was Bernie at point 2.6 while he was traveling rightward?
- (4.6.3b) **Question:** In Figure (4.1.1), at what time t_1 was Bernie at point 2.6 while he was traveling leftward?
- (4.6.3c) **Question:** In Figure (4.1.1), at what distance does Bernie find himself at $t = 18.7$ seconds?
-

(4.6.4)
Exercise

Synchronizing at [0, 0] Why do we not synchronize observers at points other than $[0, 0]$? That is, why do we not have observers in two frames synchronize at time $t = 3.0$?

Hint: Consider the linearity of line-of-sight functions that connect observers in separate frames by a line-of-sight corridor (see (4.5.6)).

(4.6.5)
Exercise

Physics and Geometry. We have seen geometric portraits of motion that are *not* straight lines (4.2.6)) along with graphs that *are* straight lines (Figure (4.1.1)).

- (4.6.5a) **Question:** Show that the physics description of unaccelerated motion (4.2.1) is equivalent to the mathematical notion of a straight line path $t = ax + b$, in the 2-dimensional X - T plane where t is time, x is distance, and a and b are certain constants.

Hint: Physics requires that equal distances are covered in equal times (see Definition (4.2.1)). The mathematical description of a straight line says the slope of its graph, $\Delta t / \Delta x$, on an X - T plane is constant at every point $[x, t]$.



II. Galilean Transformations of Frames



Chapter 5 explores the relation between pairs of frames in a linear algebraic context. A well-defined matrix arises in the line-of-sight connection between Galilean frames.

The over-riding Galilean assumption is that time is universal so that there is one universal clock whose readings are valid for all observers, regardless of their relative motion. Consistent with this property is that the speed of light is infinite.

$$G_v \begin{bmatrix} 1 - v & 0 \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ t_{1A} \end{bmatrix} = \begin{bmatrix} x_2 \\ t_{2B} \end{bmatrix}$$

I see speed v of these F^B coordinates when I stand here in F^A .

(5) Galilean Transformations

(5.1) Key Ideas

We assume frame \mathcal{F}^B moves with constant speed v relative to frame \mathcal{F}^A .

(5.1.1) **From (3.2.1):** The set of all spacetime ordered pairs $\{[x, t]\}$ for inertial frame \mathcal{F} form a mathematical vector space (A.2.1).

From (4.5.6): The line-of-sight function, $G_v : \mathcal{F}^A \rightarrow \mathcal{F}^B$ sending vector space \mathcal{F}^A to vector space \mathcal{F}^B , which takes straight lines to straight lines, where $G_v(\vec{0}) = (\vec{0})$, is necessarily linear (A.4.1b).

From (A.4.6): If the spacetime graph showing overlapping frames \mathcal{F}^A and \mathcal{F}^B is 2-dimensional, then each output vector of the function $G_v : \mathcal{F}^A \rightarrow \mathcal{F}^B$ is the 2×1 column $G_v(X)$ which Theorem (A.4.6), pg. 372, tells us can always be computed by multiplying the input vector \vec{X} on the left by a unique 2×2 matrix $\text{MX}(G_v)$. Specifically,

(5.1.1a)

$$G_v(\vec{X}) = \text{MX}(G_v) * \vec{X} = \underbrace{\left[G_v(\vec{E}_1), G_v(\vec{E}_2) \right]}_{2 \times 2 \text{ MX}(G)} * \underbrace{\vec{X}}_{2 \times 1 \text{ matrix}}$$

where the two columns of $\text{MX}(G_v)$ are the 2×1 columns $G_v(\vec{E}_1)$ and $G_v(\vec{E}_2)$. Moreover, $\vec{E}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\vec{E}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ are the standard basis vectors (A.2.11) of vector space \mathcal{F} .

Our goal is to characterize elements of $2 \times 2 \text{ MX}(G_v)$ of (5.1.1a).

(5.1.2) **NOTE:** Since the equation $G_v(\vec{X}) = \text{MX}(G_v) * \vec{X}$ appears in (5.1.1a) we often think of the matrix $\text{MX}(G_v)$ as being the same as the function G_v . We have already encountered this shortcut in (A.4.8), pg. 373 where it is observed that for any \vec{X} , $\text{MX}(G_v) * \vec{X}$ always equals $G_v(\vec{X})$. Hence, we think of the two notations $\text{MX}G_v$ and G_v as equivalent to each other. Symbolically $\text{MX}(G_v) = G_v$.

(5.1.3) **How Matrices Induce Linear Functions.** As per (A.4.5), display (5.1.1a) states that $\text{MX}(G_v)$ is the 2×2 matrix whose pair of 2×1 columns are $G_v(\vec{E}_1)$ and $G_v(\vec{E}_2)$ where \vec{E}_1 and \vec{E}_2 are the standard basis vectors of \mathbf{R}^2 (A.2.11). It is this mechanism of multiplying a column vector \vec{X} on the left by a matrix $\text{MX}(G_v)$ that tells us how any linear function G_v is implemented.

(5.2) Galilean Spacetime Diagrams

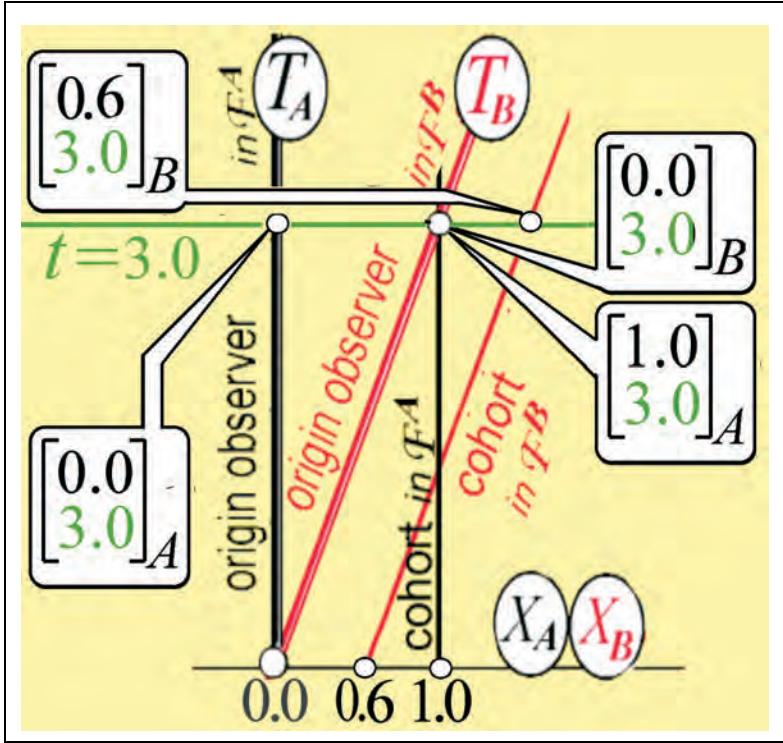
We construct *mathematical* spacetime graphs to model moving *physical* parallel frames (3.3.1), (3.4.4), (3.4.4) that reflect the Galilean assumption that time is the same for all observers in all frames.

Mathematically, the spacetime graph will show that the link between frames — the line-of-sight-corridors — have equal times in any pair of the linked frames, regardless of relative speeds. Specifically, the linking coordinates at the endpoints of line-of-sight corridors of frames \mathcal{F}^A and \mathcal{F}^B , say, have the form $[x, t]_A$ and $[x', t]_B$ with equal time coordinate t .

(5.2.1)

DEFINITION: A GALILEAN GRAPH or DIAGRAM is a spacetime graph of synchronized frames (3.4.5) whose (*horizontal*) X -axes (*and hence, their origins*, $[0, 0]$) overlap. All frames share common (*horizontal*) lines of simultaneous events (see Figure (4.4.4)).

This diagram shows points on worldlines of frames \mathcal{F}^A and \mathcal{F}^B at the common time $t = 3.0$.



(5.3) The Galilean Matrix

(5.3.1)

DEFINITION: A GALILEAN TRANSFORMATION between inertial frames \mathcal{F}^A and \mathcal{F}^B where \mathcal{F}^B has constant speed v relative to \mathcal{F}^A is a function (A.3.2)

$$G_v \left(\begin{bmatrix} x \\ t \end{bmatrix}_A \right) = \begin{bmatrix} x' \\ t \end{bmatrix}_B$$

that links coordinates $[x, t]_A$ and $[x', t]_B$ of frames \mathcal{F}^A and \mathcal{F}^B respectively, where the time coordinates t are always equal regardless of the relative speed v .

(5.3.2)

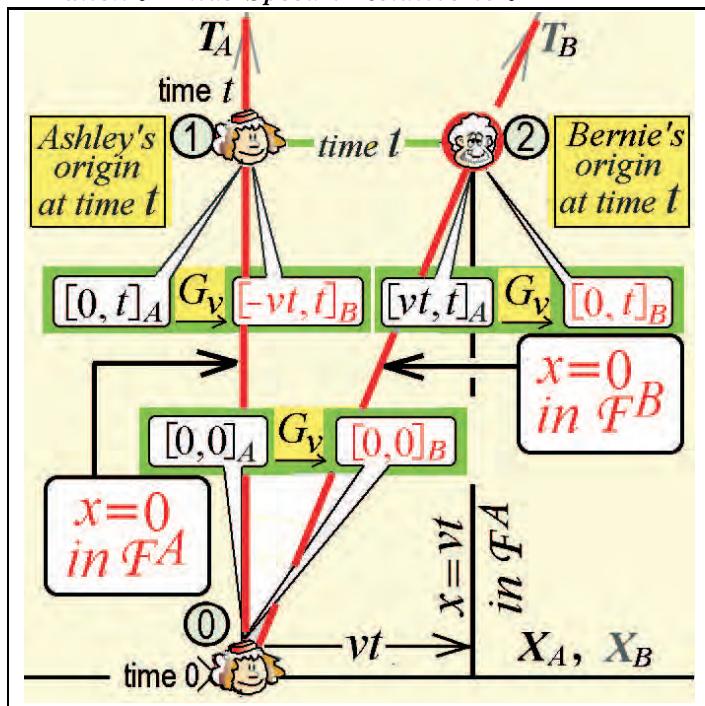
NOTE: Theorem (4.5.6) tells us that G_v is linear in the algebraic sense of (4.5.5) since G_v preserves straight lines (worldlines) and proportions of straight lines.

From Theorem (A.4.2), we can characterize the 2×2 matrix G_v , once we know its values $G_v(\vec{X}_1)$, $G_v(\vec{X}_2)$ on only two basis vectors (A.2.9) $\{\vec{X}_1, \vec{X}_2\}$ of the 2-dimensional X - T spacetime graph.

In fact, if we know the values for $G_v(\vec{X}_1)$ and $G_v(\vec{X}_2)$ and if an arbitrary vector \vec{Y} of \mathbf{R}^2 has the form $\vec{Y} = \alpha_1 \vec{X}_1 + \alpha_2 \vec{X}_2$, then $G_v(\vec{Y}) = \alpha_1 G_v(\vec{X}_1) + \alpha_2 G_v(\vec{X}_2)$ shows how we produce the value of $G_v(\vec{Y})$ from the known values of $G_v(\vec{X}_1)$ and $G_v(\vec{X}_2)$ alone.

(5.3.3)

FIGURE: The Galilean Transformation $G_v: \mathcal{F}^A \rightarrow \mathcal{F}^B$ when \mathcal{F}^B has Speed v Relative to \mathcal{F}^A .



(5.3.4)

THEOREM: (*The Galilean Matrix*) Inertial frame \mathcal{F}^B travels with speed v relative to inertial frame \mathcal{F}^A . Then the Galilean matrix $G_v : \mathcal{F}^A \rightarrow \mathcal{F}^B$ linking coordinates at opposite ends of line-of-sight corridors between \mathcal{F}^A and \mathcal{F}^B is given by

$$(5.3.4a) \quad \begin{bmatrix} 1 & -v \\ 0 & 1 \end{bmatrix} = \text{MX}(G_v) = G_v.$$

NOTE: Linear transformations and their matrices are often used interchangeably in the literature. Hence, we write $\text{MX}(G_v) = G_v$ in (5.3.4a). See (5.1.2) and (A.4.8).

PROOF: The line-of-sight linear function G_v has the following values for points ①, ②, and ③ of Figure (5.3.3):

(5.3.4b)

	\mathcal{F}^A	\mathcal{F}^B
①	$[0, 0]_A$	$[0, 0]_B$
②	$[0, t]_A$	$[-vt, t]_B$
③	$[vt, t]_A$	$[0, t]_B$

Justifying ① of (5.3.4b): The origins of \mathcal{F}^A and \mathcal{F}^B are synchronized at $t = 0$ (3.4.5). At common time $t = 0$, the origins are at opposite ends of a line-of-sight corridor, which is expressed mathematically by the equation $G_v([0, 0]_A) = [0, 0]_B$. Also, at this time, the distance between the origins is zero.

Justifying ② of (5.3.4b): Since \mathcal{F}^A is moving leftward at speed $-v$ relative to \mathcal{F}^B , \mathcal{F}^B sees the origin of \mathcal{F}^A (the T^A -axis) from \mathcal{F}^B coordinates $[-vt, t]_B$. But at time t , the origin of \mathcal{F}^A has \mathcal{F}^A coordinates $[0, t]_A$. Mathematically, the linking of these coordinates at the ends of a line-of-sight corridor is expressed by the equation $G_v([0, t]_A) = [-vt, t]_B$.

Justifying ③ of (5.3.4b): Since \mathcal{F}^B is moving rightward at speed $+v$ relative to \mathcal{F}^A , the observer in \mathcal{F}^A at time t , at coordinates $[vt, t]_A$, looks down the line-of-sight corridor into \mathcal{F}^B and sees the \mathcal{F}^B origin (the T^B -axis) has \mathcal{F}^B coordinates $[0, t]_B$. Mathematically, the linking of these coordinates at the ends of a line-of-sight corridor is expressed by the equation $G_v([vt, t]_A) = [0, t]_B$.

Now (5.3.4b) assures us that we have the values of linear function $G_v : \mathcal{F}^A \rightarrow \mathcal{F}^B$ on two basis vectors of \mathcal{F}^A which, as noted in (5.3.2), suffices to characterize the 2×2 matrix G_v . These two basis vectors are $[0, t]_A$ and $[vt, t]_A$. Thus, from (A.4.2), there is a unique quartet of numbers, $\{a, b, c, d\}$ such that

$$(5.3.4c) \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix} * \begin{bmatrix} 0 \\ t \end{bmatrix}_A = \begin{bmatrix} -vt \\ t \end{bmatrix}_B \quad \text{and} \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix} * \begin{bmatrix} vt \\ t \end{bmatrix}_A = \begin{bmatrix} 0 \\ t \end{bmatrix}_B.$$

Matrix multiplication (A.4.3e) applied to (5.3.4c) generates four linear equations with four unknowns which resolve as

$$\begin{aligned} a &= 1 \\ b &= -v \\ c &= 0 \\ d &= 1 \end{aligned}.$$

This proves (5.3.4a) and hence, Theorem (5.3.4). ■



(5.4) Pattern of the Galilean Matrix

(5.4.1) **An Intuitive Reading of Matrix G_v (5.3.4a).** Figure (5.3.3) and Table (5.3.4b) list some linked pairs of coordinates $[x, t]_A$ and $[x', t']_B$, that lie at opposite ends of line-of-sight corridors connecting two frames of reference. However, once we apply the 2×2 Galilean matrix G_v (5.3.4a), the domain (input) vector, $[x, t]_A$, is considered to be stationary and the range (output) vector, $[x', t']_B$ is perceived by an observer in \mathcal{F}^A to be moving with relative speed v .

This viewpoint is embedded in the matrix itself as summarized in the figure following:

(5.4.1a) **FIGURE:** *Pattern Recognition in Reading the Galilean Matrix*

$$G_v \begin{bmatrix} 1 & -v \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ t_1 \end{bmatrix}_A = \begin{bmatrix} x_2 \\ t_2 \end{bmatrix}'_B$$

I see speed v

of these FB coordinates

when I stand here in \mathcal{F}^A .

Matrix $G_v: \mathcal{F}^A \rightarrow \mathcal{F}^B$ has 1s on the diagonal and a 0 in the lower left hand corner. Whatever appears in the upper right hand corner is always the negative of the speed of \mathcal{F}^B as observed from frame \mathcal{F}^A .

(5.4.2)

NOTE: Time is invariant in all synchronized Galilean frames \mathcal{F}^A , \mathcal{F}^B as indicated in Figure (5.3.3). This is shown linear algebraically (i.e., from the matrix) of Figure (5.4.1a) where we see that for all synchronized frames, t_1 in \mathcal{F}^A equals t_2 in \mathcal{F}^B . Coordinates $[x_1, t_1]_A$ and $[x_2, t_2]_B$ are at opposite ends of a line-of-sight corridor and, according to Definition (5.3.1), $t_1 = t_2$ — for Galilean transformations, time is invariant for all observers.

(5.5) Addition of Speeds via Matrices

In the next theorem, we present an algebraic property of Galilean matrices.

(5.5.1)

THEOREM: (An Algebraic Property of Galilean Matrices)
For a Galilean matrix G and any scalars v and w , we have

(5.5.1a)

$$G_v * G_w = G_{v+w}.$$

PROOF:

$$\begin{aligned}
 G_v * G_w &= \begin{bmatrix} 1 & -v \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & -w \\ 0 & 1 \end{bmatrix} && \text{from (5.3.4a)} \\
 &= \begin{bmatrix} 1 & -(v+w) \\ 0 & 1 \end{bmatrix} && \text{from (A.4.3e) defining} \\
 &&& \text{the matrix product} \\
 &= G_{v+w}. && \text{from Figure (5.4.1a), which} \\
 &&& \text{shows the upper-right corner} \\
 &&& \text{holds the negative of the speed}
 \end{aligned}$$

Property (5.5.1a) is now established, proving the theorem. ■

NOTE that the first line of the proof uses the symbol G_v as a shorthand for the matrix $\text{MX}(G_v)$ as is explained in (5.1.2)

The following theorem confirms our intuition about the way we add speeds of different frames. Specifically, we have:

(5.5.2)

THEOREM: (Galilean Addition of Speeds) Given three frames of reference, \mathcal{F}^A , \mathcal{F}^B , and \mathcal{F}^C , in 1-dimensional space that are synchronized at time $t = 0$ and distance $x = 0$. Given

$$(5.5.2a) \quad \mathcal{F}^A \xrightarrow{G_v} \mathcal{F}^B \quad \text{Observers in } \mathcal{F}^A \text{ see } \mathcal{F}^B \text{ moving at speed } v.$$

$$(5.5.2b) \quad \mathcal{F}^B \xrightarrow{G_w} \mathcal{F}^C \quad \text{Observers in } \mathcal{F}^B \text{ see } \mathcal{F}^C \text{ moving at speed } w.$$

Then

$$(5.5.2c) \quad \mathcal{F}^A \xrightarrow{G_{v+w}} \mathcal{F}^C \quad \text{Observers in } \mathcal{F}^A \text{ see } \mathcal{F}^C \text{ moving at speed } v + w.$$

PROOF: Hypotheses (5.5.2a) and (5.5.2a) say

$$(5.5.2d) \quad G_v([x, t]_A) = [x', t']_B \quad \left\{ \begin{array}{l} [x, t]_A \text{ in } \mathcal{F}^A \\ [x', t']_B \text{ in } \mathcal{F}^B \end{array} \right.$$

and

$$(5.5.2e) \quad G_w([x', t']_B) = [x'', t'']_C \quad \begin{cases} [x', t']_B \text{ in } \mathcal{F}^B \\ [x'', t'']_C \text{ in } \mathcal{F}^C. \end{cases}$$

Substituting $[x', t']_B$ of (5.5.2d) into (5.5.2e), we obtain

$$(5.5.2f) \quad G_w(G_v([x, t]_A)) = [x'', t'']_C \quad \begin{cases} [x, t]_A \text{ in } \mathcal{F}^A \\ [x'', t'']_C \text{ in } \mathcal{F}^C. \end{cases}$$

Now, $G_w(G_v([x, t]_A)) = G_w \circ G_v([x, t]_A)$ from (A.3.6),

$$= G_{v+w}([x, t]_A) \quad \text{and Theorem (5.5.1)}$$

so that (5.5.2f) rewrites itself as $G_{v+w}([x, t]_A) = [x'', t'']_C$ or

$$G_{v+w}: \mathcal{F}^A \rightarrow \mathcal{F}^C$$

which confirms (5.5.2c). The proof is done. ■

The following corollary is a mathematical statement of the symmetry principle (1.6.1).

(5.5.3)

COROLLARY: (*Confirming the Line of Sight*)

$$(5.5.3a) \quad G_v * G_{-v} = I_2,$$

the 2×2 identity matrix.

PROOF: In (5.5.1), set $v = v$ and $w = -v$ to obtain $G_v * G_{-v} = G_0$. But from (5.3.4a) (or (5.4.1a)), $G_0 = I_2$, which proves the corollary. ■

(5.5.4)

NOTE: Proposition (3.4.6e) tells us that, in general, $G_v^{-1} = G_{-v}$, which means that $G_v * G_{-v} = I_2$. Corollary (5.5.3) confirms this result in another way through the use of calculations with the specific matrices G_v and G_{-v} .

(5.6) Addition of Speeds via Areas

(5.6.1)

NOTE: To be dimensionally consistent, as per Chapter (D), pg. 389, we **define** the area in a Minkowski X-T space-time graph as follows: Given horizontal Δx and vertical Δt , then

$$(5.6.1a) \quad \text{Area}(\Delta x, \Delta t) \stackrel{\text{def}}{=} \Delta x(\text{distance}) \times \Delta t(\text{time}) \times \underbrace{1}_{\text{dimension factor}} \frac{1}{\text{time}^2}$$

The reader may reasonably ask why we need to multiply by the factor $1/\text{time}^2$ in (5.6.1a). The reason is that the goal of this section is to provide a graphical way of adding *speeds* by identifying them with appropriate *areas* in a Minkowski diagram. The only purpose of the factor $1/\text{time}^2$, then, is to give the areas appropriate units. Consequently, Definition (5.6.1a) ensures that $\text{Area}(\Delta x, \Delta t)$ has units $\frac{\text{distance}}{\text{time}}$. For simplicity, however, the symbols $(1/\text{time}^2)$, will henceforth be omitted.

The following theorem shows how constant speeds can also be determined from areas of triangles in the spacetime graph.

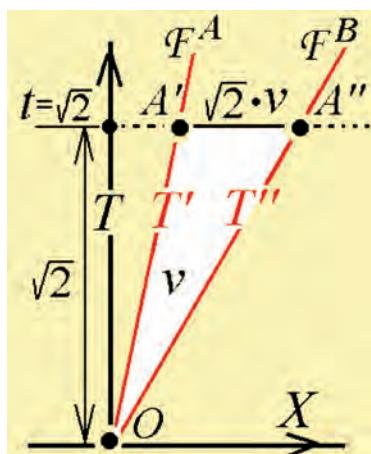
(5.6.2)
Trig

THEOREM: (Relative Speed in Terms of Areas) Given synchronized frames \mathcal{F}^A and \mathcal{F}^B with respective time axes T' and T'' . Let the line($A'A''$) intersect time axes T' and T'' at time $t = \sqrt{2}$. Then frame \mathcal{F}^B has speed v relative to frame \mathcal{F}^A if and only if

$$(5.6.2a) \quad \text{Area}(\triangle OA'A'') = v.$$

PROOF: (Our fundamental strategy is to use the fact that the length of a triangle's base at any time t is the distance between the origins of the two frames at time t .)

Accordingly, with the notation of (A.2.13a), denote the length of line



$\overline{A'A''}$ by $\|\overline{A'A''}\|$, which, according to the figure of (5.6.2), is the length of the base of $\triangle OA'A''$. The length of the base, in turn, corresponds to the distance between the origins of synchronized frames, \mathcal{F}^A and \mathcal{F}^B , after time $t = \sqrt{2}$. But the relative speed between the frames is v if and only if this separation distance, $\|\overline{A'A''}\|$, has the property

$$(5.6.2b) \quad \|\overline{A'A''}\| = \text{speed} \cdot \text{time} = (v \cdot \sqrt{2}).$$

$$\text{Now } \text{Area}(\triangle OA'A'') = \frac{1}{2} \cdot \text{base} \cdot \text{altitude} \quad \text{plane geometry}$$

$$\begin{aligned} &= \frac{1}{2} \cdot \|\overline{A'A''}\| \cdot \sqrt{2} && \text{from figure of (5.6.2),} \\ &= \frac{1}{2} \cdot (v \cdot \sqrt{2}) \cdot \sqrt{2} && \|\overline{A'A''}\| = \text{base}, \\ &= v, && \sqrt{2} = \text{altitude} \\ &&& \text{from (5.6.2b)} \end{aligned}$$

which proves (5.6.2a) and completes the proof. ■

Theorem (5.6.2) provides a geometric way to prove Theorem (5.5.2), as follows:

(5.6.3)

RESTATEMENT OF THEOREM (5.5.2)

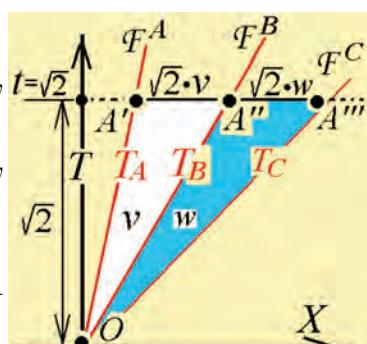
Given:

(5.6.3a) Frame \mathcal{F}^B has speed v relative to \mathcal{F}^A , and

(5.6.3b) frame \mathcal{F}^C has speed w relative to \mathcal{F}^B .

Then.....

(5.6.3c) frame \mathcal{F}^C has speed $v+w$ relative to \mathcal{F}^A .



PROOF: From Theorem (5.6.2),

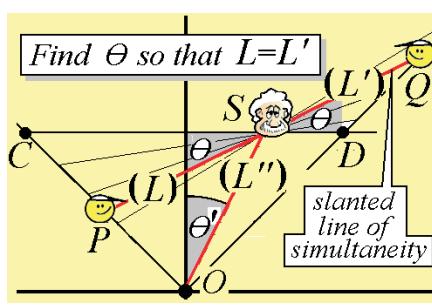
frame \mathcal{F}^B has speed v as seen from frame \mathcal{F}^A if and only if $\text{Area}(\triangle OA'A'') = v$. Similarly, frame \mathcal{F}^C has speed w as seen from frame \mathcal{F}^B if and only if $\text{Area}(\triangle OA''A''') = w$. (See diagram above.) Now the two triangles together form this single triangle, $\triangle OA'A'''$, which has area $v+w$. From Theorem (5.6.2), this area represents

the speed of \mathcal{F}^C relative to \mathcal{F}^A . Having established (5.6.3c), the alternate, geometric proof of Theorem (5.5.2) is done. ■



III. The Speed of Light is Constant

Chapter 6 shows why a pair of events occurring simultaneously at two separate locations in one frame can never occur simultaneously from the point of view of another moving frame.



(6) Constant c in Spacetime

(6.1) Minkowski Spacetime Diagrams

In the definition following, we extend the notion of a spacetime diagram (4.1.2) to a Minkowski diagram by adding the time-paths (worldlines) of leftward traveling photons \leftarrow and rightward traveling photons \rightarrow (e.g., Figure (6.1.2)).

(6.1.1)

DEFINITION: The MINKOWSKI DIAGRAM

- is a spacetime graph of all points (x, t) for one or more synchronized (3.4.5a) frames of reference moving with constant velocity, each of which is represented in spacetime by its own X and T axes.
- At time $t = 0$ two photons are emitted from point $x = 0$ — one traveling leftward and the other traveling rightward. Relative to any inertial frame, \mathcal{F} , the world lines of these synchronized photons form straight lines emanating from the origin.

(6.1.2)

FIGURE: Minkowski Diagram for 2 Observers and 2 Photons

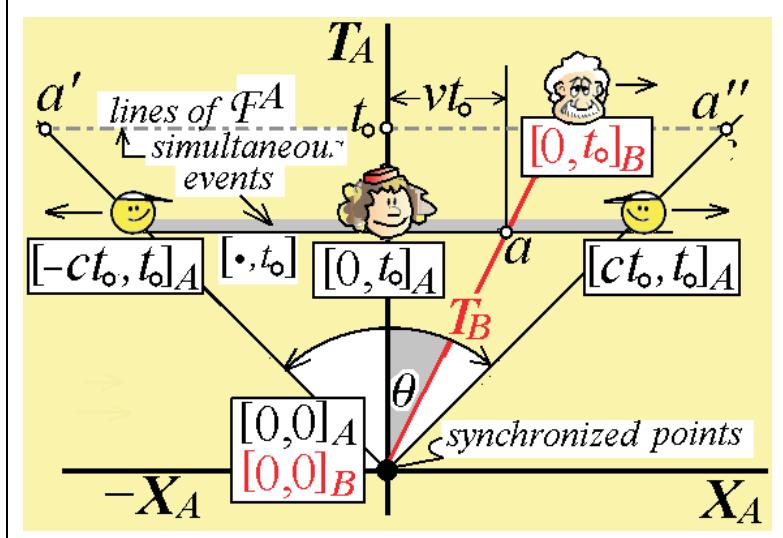


Diagram (6.1.2) shows perpendicular X_A and T_A axes for Ashley's frame. There are four synchronized worldlines (see Def. (3.4.5a)) namely,

- The two worldlines of the synchronized leftward and rightward traveling photons, respectively.
- Ashley's vertical worldline in frame \mathcal{F}^A which coincides with T^A , the \mathcal{F}^A time axis,
- Bernie's slanted worldline in frame \mathcal{F}^B which coincides with T^B , the \mathcal{F}^B time axis that forms angle θ with the vertical T_A . All frames must have their T -axes (the worldline of $x = 0$) overlap at opposite ends of a line-of-sight corridor at common time $t = 0$. This is the mutual synchronization requirement of (3.4.5) and (3.4.3).

(6.1.3)

NOTE: As shown in Figure (4.4.4), lines of simultaneous events within a frame \mathcal{F}^A are drawn parallel to the X -axis of that frame. In Galilean/Newtonian relativity, one master (universal) clock serves all observers in all frames—simultaneous events at time t_0 for all frames \mathcal{F}^A , \mathcal{F}^B , \mathcal{F}^C , ..., can always be represented as straight lines of simultaneous $t = t_0$ events that are pairwise parallel (see Figure (5.3.3)).

However, in Section (6.3), we will see that if the speed of light c is the same in all frames, then simultaneous events at time t_0 in various frames, can no longer be represented as straight lines that are mutually parallel.

(6.2) Constant c and Simultaneity

(6.2.1) Synchronized Observers are Always Equidistant from Synchronized Photon Pairs.

The Setup: Since a Minkowski diagram will require two synchronized photons, there will always be (at least) three actors who are simultaneously synchronized at $x = 0$ and time $t = 0$ (3.4.5a), namely,

- (1) an observer, who is traveling at speed $\pm v$ (which may be zero),
- (2) a leftward moving photon $\leftarrow \text{Photon}$ traveling at speed $-c$,
- (3) a rightward moving photon $\rightarrow \text{Photon}$ traveling at speed $+c$.

Display (6.2.2) is a reformulation of (1.5.2b), which says the speed of light is the same for all observers:

(6.2.2)

Constant Speed of Light and Equidistance. *An observer at constant speed arrives at source point $x = 0$ and time $t = 0$ from which leftward and rightward traveling synchronized (2.2.5) photons are emitted.*

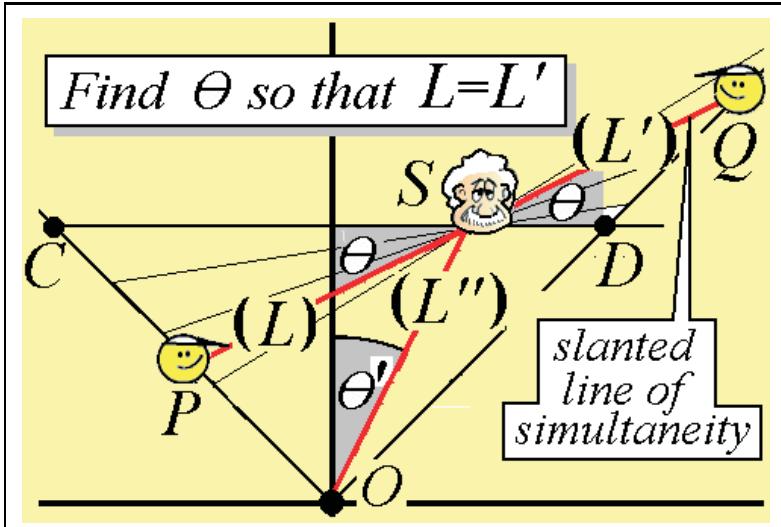
To say the speed of light (of the two photons) is absolute (the same for all observers), is to say that regardless of relative speeds of the observer and the source point $x = 0$, the observer is, at all times, SIMULTANEOUSLY EQUIDISTANT from each of the two photons. That is, at any time, the distances from the observer to each to the two photons will be the same.

Ashley's Simultaneous Events. In Figure (6.1.2), a horizontal line that captures Ashley's position on her T_A -axis also intersects the slanted worldlines at a' and a'' where, for any time t_0 , the two photons are at equal distances from Ashley. Hence, for any time t_0 , Ashley's lines of simultaneity in Figure (6.1.2) are horizontal and parallel to her X_A -axis.

We shall now see why Bernie's line of simultaneity (*the line on which each point shows the same time*) can never be horizontal.

(6.2.3)

Bernie's Simultaneous Events.



This figure shows Bernie at the pivot point S on a swiveling line \overline{PSQ} of simultaneity. Each point on \overline{PSQ} has the same time coordinate in Bernie's frame. The swiveled line forms an angle θ with the horizontal. For all angles θ , the lengths L and L' on line \overline{PSQ} are the respective distances between Bernie and the left and right photons. (*This follows from (4.4.1a) and (4.4.3) since lines of simultaneity like \overline{PSQ} represent (a portion of) the X-axis at a particular time.*)

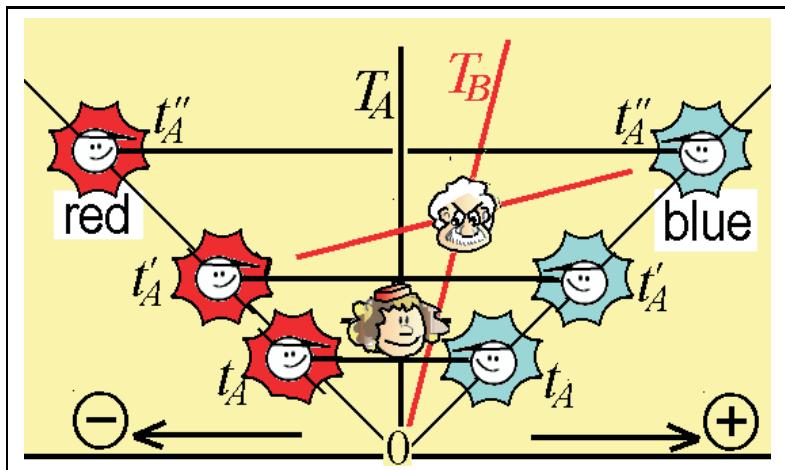
Hence, *The only way* Bernie at speed v is always at the midpoint between the two photons is when line \overline{PSQ} is swiveled to the exact angle $\theta \neq 0$ at which $L = L'$. This appropriately-swiveled line becomes **Bernie's line of simultaneity which cannot be horizontal**. This easily follows from Figure (6.2.3) where a horizontal line \overline{PSQ} ($\theta = 0$) shows $L > L'$.

(6.3) How Constant c Destroys Simultaneity

More accurately, we should label this section, “*How Constant c Destroys Universal Simultaneity.*” With the Minkowski graph in hand, we have a visual model that explains how two observers who are in motion with respect to each other can not see the same events at the same time according to their previously synchronized and identical clocks.

(6.3.1)

Minkowski Diagram for Synchronized Photons with Two Inertial Frames



To fix ideas, consider the following four synchronized actors who, by Definition (3.4.5a), pg. 62, all appear at their respective origins $x = 0$ at common time $t = 0$:

- A red photon traveling leftward in the negative direction,
- A blue photon traveling rightward in the positive direction,
- Origin observer Ashley who is stationary at her origin $x = 0$ in inertial frame \mathcal{F}^A ,
- Origin observer Bernie who is stationary at his origin $x = 0$ in inertial frame \mathcal{F}^B while his frame travels rightward at constant speed relative Ashley's frame \mathcal{F}^A .

Figure (6.3.1) shows the world lines (time paths) of the red and blue photons moving to the northwest and northeast respectively as time increases. Ashley's world line grows vertically and Bernie's world line grows in a northeasterly direction.

Since the speed of light is the same for all observers, both Ashley in \mathcal{F}^A and Bernie in \mathcal{F}^B are always equal distances from the red and blue photons (Ashley and Bernie *started* at equal distance zero from the two photons at time zero so the equal distances will persist thereafter).

Recall that lines of simultaneity — straight lines where time is the same for all points on the line — will be horizontal for Ashley's frame \mathcal{F}^A and on a positive angle for Bernie's frame \mathcal{F}^A (6.2.3). Now both Ashley and Bernie will see the two photons simultaneously according to clocks within their respective frames \mathcal{F}^A and \mathcal{F}^B . Bernie is shown in the figure to see the red photon at time t'_A and the blue photon at time t''_A according to Ashley's clock (*not his clock*) where $t'_A < t''_A$. Although the sighting (appearance) of the two photons is internally simultaneous within each frame, the appearance of the two photons in Bernie's frame \mathcal{F}^B occurs at different times $t'_A < t''_A$ according to Ashley's clock.

Two events that are simultaneous in one frame can never be simultaneous in another frame as long as the frames are in constant motion relative to each other. Figure (6.3.1) has shown why a constant value of c for all observers destroys common simultaneity.

Summary

This chapter presents a prime example of how the fusion of mathematics, cartoons, and reality brings us to an unexpected feature of reality and a new level of understanding. We first encountered the fundamental property of special relativity—that the speed of light is constant for all observers (1.5.2b). With the rephrasing of (6.2.2), we then established the cartoon of Figure (6.2.3) and applied mathematical analysis to lines of simultaneity that correspond to reciprocal slopes. This analysis further led us to the illustration of (6.3.1) that produced a new observation, namely,

- (6.3.2) *If two observers are in motion relative to each other, then two events that are simultaneous for one can never be simultaneous for the other.*

Property (6.3.2) is not likely to be revealed through our five senses alone.

(6.4) Exercise

(6.4.1)

The Swiveled Line Theorem.

(Click [here](#) to see the animation.)



Assume the speed of light $c = 1$ so that the worldlines of the two photons of Figure (6.2.3) are at angles $\pm 45^\circ$, i.e., the lines have slopes ± 1 . Then show that:

Exercise

- If v is the speed of Bernie's frame \mathcal{F}^B relative to Ashley's frame \mathcal{F}^A , then $v = \tan(\theta')$.
- In order to swivel line \overline{PSQ} to the exact angle θ so that distances $L = L'$, then necessarily $\theta = \theta'$ and all three lengths L , L' , and L'' are equal to each other.



IV. Lorentz Transformations of Frames

Chapter 7 explores the relation between pairs of frames in a linear algebraic context (Compare with the description of Chapter 5.) A well-defined matrix arises in the line-of-sight connection between Lorentzian, relativistic frames. The overriding relativistic assumption is that time is not universal—it varies with each observer. Consistent with this property is that c , the speed of light, is finite and does not vary with either the relative speeds of the source or of any observer.

$$\frac{1}{\sqrt{1-(\gamma c)^2}} \begin{bmatrix} 1 & -\gamma v \\ -\gamma v/c^2 & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ t_1 \end{bmatrix}_A = \begin{bmatrix} x_2 \\ t_2 \end{bmatrix}_B$$

I see speed v of these coordinates when I stand here.

Chapter 8 studies the track of origins of moving frames that trace out a well-defined hyperbola in the Minkowski X-T diagram. This hyperbola in relativistic physics corresponds to a straight line in classical Galilean/Newtonian physics.

(7) Lorentz Transformations

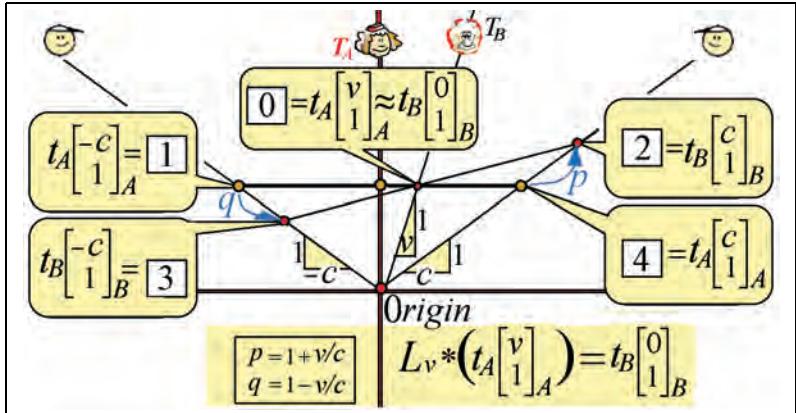
The Setup for Figure (7.1.2)

(7.1) The Lorentz Matrix

(7.1.1) **PREVIEW:** In the Minkowski diagram (7.1.2) following, the worldlines are plotted for four players:

- Ashley at the origin in inertial frame \mathcal{F}^A whose worldline is vertical,
- Bernie at the origin in synchronized frame \mathcal{F}^B (3.4.5a) moving with relative speed v . Bernie's worldline coincides with T_B , the time axis of \mathcal{F}^B .
- Leftward and a rightward traveling photons whose worldlines contain the line segments from $[1]$ to $[3]$ and from $[2]$ to $[4]$, respectively.

(7.1.2) **FIGURE: Lorentz Linking of Synchronized Frames**



The values $p = (1 + v/c)$ and $q = (1 - v/c)$ in Figure (7.1.2) will be established in (7.1.3f).

Line-of-Sight Coordinates of Point $[0]$, Fig. (7.1.2). Ashley and Bernie, who are fixed at their respective origins $x = 0$, are

synchronized at time $t = 0$ in each frame. Therefore, after time t_A in Ashley's frame \mathcal{F}^A , Bernie travels vt_A units to Ashley's right in frame \mathcal{F}^A . At this time t_A , Bernie in \mathcal{F}^B is seen by an observer in \mathcal{F}^A from his/her \mathcal{F}^A coordinates $[vt_A, t_A]_A = t_A[v, 1]_A$.

At the other end of the line-of-sight corridor in \mathcal{F}^B , Bernie is positioned in \mathcal{F}^B at coordinates $[0, t_B]_B$ for some \mathcal{F}^B time t_B . (*Bernie always stands at his own origin $x = 0$.*)

The times t_A and t_B at the opposite ends of the line-of-sight corridor are not assumed to be equal. We can say, however, that the line-of-sight linkage of coordinates at Point $\boxed{0}$ of Figure (7.1.2),

$$t_A \begin{bmatrix} v \\ 1 \end{bmatrix}_A \quad \text{with} \quad t_B \begin{bmatrix} 0 \\ 1 \end{bmatrix}_B$$

is expressed by the Lorentz transformation equation

$$(7.1.2a) \quad t_A L_v * \begin{bmatrix} v \\ 1 \end{bmatrix}_A = t_B \begin{bmatrix} 0 \\ 1 \end{bmatrix}_B.$$

Coordinates Reflecting Constant Speed of Light c . All four traveling actors, namely, (i) Ashley, (ii) Bernie, (iii) the leftward traveling photon, and (iv) the rightward traveling photon, find themselves at point $x = 0$ at the common synchronization time, $t = 0$.

It follows that at time t_A in \mathcal{F}^A , both left and right photons are at a distance $\pm t_A c$ from the \mathcal{F}^A origin where Ashley is always located. Figure (7.1.2) illustrates the \mathcal{F}^A coordinates at time t_A of the two photons as

$$(7.1.2b) \quad \text{Point } \boxed{1} = t_A \begin{bmatrix} -c \\ 1 \end{bmatrix}_A, \quad \text{and} \quad \text{Point } \boxed{4} = t_A \begin{bmatrix} c \\ 1 \end{bmatrix}_A.$$

Similarly, at arbitrary time t_B in \mathcal{F}^B , both left and right photons are $\pm t_B c$ units from the \mathcal{F}^B origin where Bernie is always located. Figure (7.1.2) shows the \mathcal{F}^B coordinates at time t_B of the two photons, namely,

$$(7.1.2c) \quad \text{Point } \boxed{3} = t_B \begin{bmatrix} -c \\ 1 \end{bmatrix}_B, \quad \text{and} \quad \text{Point } \boxed{2} = t_B \begin{bmatrix} c \\ 1 \end{bmatrix}_B.$$

(7.1.2d)

Note: As shown in Figures (6.2.3) and (7.1.2), the line of constant time t_B in frame \mathcal{F}^B cannot be horizontal if we require equal distances from the origin to the photons at equal times.

The coordinates of Figure (7.1.2) are the basis for the following theorem that establishes the Lorentz line-of-sight transformation.

(7.1.3)

Theorem: (The Lorentz Matrix) An inertial frame \mathcal{F}^B has speed v relative to a synchronized frame \mathcal{F}^A . The line-of-sight Lorentz linear transformation $L_v : \mathcal{F}^A \rightarrow \mathcal{F}^B$ has the matrix form

$$(7.1.3a) \quad L_v = \gamma_v \begin{bmatrix} 1 & -v \\ -v/c^2 & 1 \end{bmatrix}$$

where the Lorentz Factor γ_v (and Shrinkage Factor σ_v) are defined by

$$(7.1.3b) \quad \gamma_v \stackrel{\text{def}}{=} \frac{1}{\sqrt{1 - (v/c)^2}} = \frac{1}{\sigma_v} \quad (\text{see (1.8.2)}).$$

Proof: The strategy is to find the values of the linear transformation $L_v : \mathcal{F}^A \rightarrow \mathcal{F}^B$ on just the two vectors $\{[-c, 1]_A, [c, 1]_A\}$ which happen to form a basis of the two-dimensional spacetime graph (7.1.2). With these two vector values in hand, the full Lorentz 2×2 matrix is determined as per (A.4.2).

- **Eigenvector-Eigenvalues of the Lorentz Transformation:** (see Figure (7.1.2).) Multiply Point $\boxed{1}$ of \mathcal{F}^A by a (shrinking) factor $q < 1$ so that it “slides” down its worldline toward the origin to meet and align with Point $\boxed{3}$ of \mathcal{F}^B . Thus, we have the pair of overlapping coordinates

$$(7.1.3c) \quad \underbrace{q t_A [-c, 1]_A}_{q \times \boxed{1}} \leftrightarrow \underbrace{t_B [-c, 1]_B}_{\boxed{3}}$$

that are at opposite ends of a line-of-sight corridor that runs between frames \mathcal{F}^A and \mathcal{F}^B . Similarly, multiply Point $\boxed{4}$ by an (expanding) factor $p > 1$ so that it “slides” up its worldline away from the origin to align with Point $\boxed{2}$. That is, the \mathcal{F}^A and \mathcal{F}^B coordinates

$$(7.1.3d) \quad \underbrace{p t_A [c, 1]_A}_{p \times \boxed{4}} \leftrightarrow \underbrace{t_B [c, 1]_B}_{\boxed{2}}$$

are at opposite ends of a line-of-sight corridor that runs between \mathcal{F}^A and \mathcal{F}^B . But to be paired at opposite ends of a line-of-sight corridor is to be linked by the Lorentz transformation L_v as follows:

(7.1.3e)

$$L_v \left(\begin{bmatrix} -c \\ 1 \end{bmatrix}_A \right) = \frac{t_B/t_A}{q} \begin{bmatrix} -c \\ 1 \end{bmatrix}_B \quad \text{and} \quad L_v \left(\begin{bmatrix} c \\ 1 \end{bmatrix}_A \right) = \frac{t_B/t_A}{p} \begin{bmatrix} c \\ 1 \end{bmatrix}_B$$

where $[-c, 1]_A$ and $[c, 1]_A$ are eigenvectors of L_v with respective eigenvalues $(t_B/t_A)/q$ and $(t_B/t_A)/p$ (see Definition (A.5.1)).

- **Values of numbers (scalars) p and q in (7.1.3e).** The \mathcal{F}^A coordinates of midpoint $\boxed{0}$ are $t_A[v, 1]_A$ which is the average of the endpoints, $\boxed{2}$ and $\boxed{3}$, in \mathcal{F}^A coordinates, which we read from the left hand sides of (7.1.3c) and (7.1.3d). The average $(\boxed{3} + \boxed{2})/2 = \boxed{0}$ in \mathcal{F}^A coordinates implies

$$\underbrace{q t_A \begin{bmatrix} -c \\ 1 \end{bmatrix}_A}_{\boxed{3}} + \underbrace{p t_A \begin{bmatrix} c \\ 1 \end{bmatrix}_A}_{\boxed{2}} = \underbrace{2 t_A \begin{bmatrix} v \\ 1 \end{bmatrix}_A}_{2 \boxed{0}} \implies \begin{aligned} (p - q) &= 2v/c \\ (p + q) &= 2 \end{aligned}$$

which gives us the values

$$(7.1.3f) \quad \boxed{\begin{aligned} p &= (1 + v/c) \\ q &= (1 - v/c) \end{aligned}} \quad \text{which} \quad \boxed{\begin{aligned} \frac{1}{p} - \frac{1}{q} &= \frac{-2(v/c)}{1 - (v/c)^2} \\ \frac{1}{p} + \frac{1}{q} &= \frac{2}{1 - (v/c)^2} \end{aligned}} \quad \text{in turn} \quad \text{implies} \quad \bullet$$

- **Using the Decomposition Theorem (A.5.4):** Since (7.1.3e) provides a pair of eigenvectors and eigenvalues for L_v we use the

decomposition theorem (A.5.4), pg. 375 to write

$$\begin{aligned}
 L_v &= S * D * S^{-1} \\
 &= \begin{bmatrix} -c & c \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} (t_A/t_B)/q & 0 \\ 0 & (t_A/t_B)/p \end{bmatrix} * \frac{-1}{2c} \begin{bmatrix} 1 & -c \\ -1 & -c \end{bmatrix} \\
 &= \frac{(t_A/t_B)}{2} \begin{bmatrix} \left(\frac{1}{p} + \frac{1}{q}\right) & c\left(\frac{1}{p} - \frac{1}{q}\right) \\ \frac{1}{c}\left(\frac{1}{p} - \frac{1}{q}\right) & \left(\frac{1}{p} + \frac{1}{q}\right) \end{bmatrix} \\
 (7.1.3g) \quad &= \frac{t_A/t_B}{(1 - (v/c)^2)} \begin{bmatrix} 1 & -v \\ -v/c^2 & 1 \end{bmatrix} \quad \text{from (7.1.3f).}
 \end{aligned}$$

• **Symmetry and Inverses.** L_v is the line-of-sight function where frame \mathcal{F}^B has speed v relative to frame \mathcal{F}^A . By symmetry (1.6.1), if we view the line-of-sight corridor from the opposite direction, then L_{-v} is the line-of-sight function where frame \mathcal{F}^A has speed $-v$ relative to frame \mathcal{F}^B . That is, L_{-v} is the inverse of L_v (see also Proposition (3.4.6e)), which is to say

$$\underbrace{\frac{(t_B/t_A)}{1 - (v/c)^2} \begin{bmatrix} 1 & v \\ v/c^2 & 1 \end{bmatrix}}_{L_{-v}} = \underbrace{\frac{1}{(t_B/t_A)} \begin{bmatrix} 1 & v \\ v/c^2 & 1 \end{bmatrix}}_{L_v^{-1}}$$

replace v with $-v$ in L_v of (7.1.3g) apply (A.4.4d) to L_v in (7.1.3g)

Equating the scalar multiples of the (equal) matrices above yields the time dilation formula

$$(7.1.3h) \quad (t_B/t_A) = \sqrt{1 - (v/c)^2}. \quad \{ \text{note: } t_B \leq t_A \}$$

Substituting the value of (t_B/t_A) from (7.1.3h) into (7.1.3g) finally yields (7.1.3a) which ends the proof. ■



(7.2) Pattern of the Lorentz Matrix

- (7.2.1) **An Intuitive Reading of Matrix L_v (7.1.3a).** Once we apply the 2×2 Lorentz matrix L_v (7.1.3a), the domain frame, \mathcal{F}^A , containing the (input) vector, $[x, t]_A$, is considered to be stationary and the range frame, \mathcal{F}^B , containing the (output) vector, $[x', t']_B$, is seen to be moving with speed v relative to \mathcal{F}^A .

All this information is embedded in the Lorentz matrix itself as summarized in Figure (7.2.2), a counterpart to the pattern (5.4.1a) of the Galilean matrix.

(7.2.2) **FIGURE: Pattern of the Lorentz Matrix (7.1.3a)**

$$L_v = \frac{1}{\sqrt{1-(v/c)^2}} \begin{bmatrix} 1 & -v \\ -v/c^2 & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ t_1 \end{bmatrix}_A = \begin{bmatrix} x_2 \\ t_2 \end{bmatrix}_B$$

I see speed v

of these coordinates

when I stand here.

Lorentz matrix $L_v : \mathcal{F}^A \rightarrow \mathcal{F}^B$ is the product of the number (scalar) $\gamma_v = 1/\sqrt{1 - (v/c)^2}$ and the 2×2 matrix with 1s on the diagonal and scalars $-v$ and $-v/c^2$ in the northeast and southwest corners respectively. The speed v appears twice in the 2×2 matrix and once in the radical term, $\sqrt{1 - (v/c)^2}$. The matrix corners always contain the negative of the speed of \mathcal{F}^B as observed from frame \mathcal{F}^A .

(7.3) The Lorentz Sum of Speeds

We define a special addition between any pair of numbers, v and w .

(7.3.1)

DEFINITION: For numbers v and w , the symbol $v \oplus w$ denotes their RELATIVISTIC SUM or LORENTZ SUM and is defined to be

$$(7.3.1a) \quad v \oplus w \stackrel{\text{def}}{=} \frac{v + w}{1 + (vw/c^2)}$$

where c is the speed of light, and the numerator $(v + w)$ and denominator $(1 + (vw/c^2))$ are the ordinary arithmetic sums.

(7.3.2)



The **odd-looking sum** (7.3.1a) may seem to be unmotivated. However, as we shall see in Theorem (7.3.3), the sum $v \oplus w$ is exactly what is needed to describe relative speeds between moving frames. Also, we have already noted in (1.13.3), pg. 39, that it is this very sum $v \oplus w$ of (7.3.1a) that resolves the Pea-Shooter Paradox.

(7.3.3)

THEOREM: (An Algebraic Property of Lorentz Matrices)

For Lorentz matrices L_v and L_w where v and w are arbitrary speeds (or scalars), we have

(7.3.3a)

$$L_v * L_w = L_{v \oplus w}$$

where $v \oplus w$ is the Lorentz sum (7.3.1).

PROOF: From (5.3.4a), we have

$$L_v * L_w = \frac{1}{\sqrt{1 - v^2/c^2}} \begin{bmatrix} 1 & -v \\ -v/c^2 & 1 \end{bmatrix} * \frac{1}{\sqrt{1 - w^2/c^2}} \begin{bmatrix} 1 & -w \\ -w/c^2 & 1 \end{bmatrix}$$

which, after matrix multiplication (A.4.3e) that entails some disagreeable algebra, allows us to continue the equality with

$$= \frac{1}{\sqrt{1 - \left[\frac{(v+w)/c}{1+(vw/c^2)} \right]^2}} \begin{bmatrix} 1 & -\left[\frac{v+w}{1+(vw/c^2)} \right] \\ -\frac{1}{c^2} \left[\frac{v+w}{1+(vw/c^2)} \right] & 1 \end{bmatrix}.$$

$$= \frac{1}{\sqrt{1 - [(v \oplus w)/c]^2}} \begin{bmatrix} 1 & -(v \oplus w) \\ -(v \oplus w)/c^2 & 1 \end{bmatrix} \quad \begin{array}{l} \text{from (7.3.1)} \\ \text{defining} \\ (v \oplus w) \end{array}$$

$$= L_{v \oplus w}. \quad \begin{array}{l} \text{from Figure (7.2.2), which shows} \\ \text{the pattern for the three entries} \\ \text{that contain the speed } (v \oplus w) \end{array}$$

Property (7.3.3a) is now established, proving the theorem. ■

(7.4) Addition of Speeds via Matrices

The following theorem, which justifies the form for $v \oplus w$ based on the Lorentz matrix transformation, is the relativistic counterpart to Theorem (5.5.2).

(7.4.1)

THEOREM: (Lorentzian Addition of Speeds) Given three frames of reference, \mathcal{F}^A , \mathcal{F}^B , and \mathcal{F}^C , in 1-dimensional space that are synchronized at time $t = 0$ and distance $x = 0$. Given

$$(7.4.1a) \quad \mathcal{F}^A \xrightarrow{L_v} \mathcal{F}^B. \quad \begin{array}{l} \text{Observers in } \mathcal{F}^A \text{ see } \mathcal{F}^B \\ \text{moving at speed } v. \end{array}$$

$$(7.4.1b) \quad \mathcal{F}^B \xrightarrow{L_w} \mathcal{F}^C. \quad \begin{array}{l} \text{Observers in } \mathcal{F}^B \text{ see } \mathcal{F}^C \\ \text{moving at speed } w. \end{array}$$

Then

$$(7.4.1c) \quad \mathcal{F}^A \xrightarrow{L_{v \oplus w}} \mathcal{F}^C. \quad \begin{array}{l} \text{Observers in } \mathcal{F}^A \text{ see } \mathcal{F}^C \\ \text{moving at speed } v \oplus w \\ (7.3.1a). \end{array}$$

PROOF: Hypotheses (7.4.1a) and (7.4.1b) say

$$(7.4.1d) \quad L_v([x, t]_A) = [x', t']_B \quad \left\{ \begin{array}{l} [x, t]_A \text{ in } \mathcal{F}^A \\ [x', t']_B \text{ in } \mathcal{F}^B \end{array} \right.$$

and

$$(7.4.1e) \quad L_w([x', t']_B) = [x'', t'']_C \quad \left\{ \begin{array}{l} [x', t']_B \text{ in } \mathcal{F}^B \\ [x'', t'']_C \text{ in } \mathcal{F}^C \end{array} \right.$$

Substituting $[x', t']_B$ of (7.4.1d) into (7.4.1e), we obtain

$$(7.4.1f) \quad L_w(L_v([x, t]_A)) = [x'', t'']_C. \quad \left\{ \begin{array}{l} [x, t]_A \text{ in } \mathcal{F}^A \\ [x'', t'']_C \text{ in } \mathcal{F}^C. \end{array} \right.$$

$$\text{Now, } L_w(L_v([x, t]_A)) = L_w \circ L_v([x, t]_A) \quad \text{from (A.3.6)}$$

$$= L_{v \oplus w}([x, t]_A). \quad \begin{array}{l} \text{from (A.3.6)} \\ \text{and Theorem (7.3.3)} \end{array}$$

so that (7.4.1f) rewrites itself as $L_{v \oplus w}([x, t]_A) = [x'', t'']_C$ or $L_{v \oplus w} : \mathcal{F}^A \rightarrow \mathcal{F}^C$ which confirms (7.4.1c). The proof is done. ■

- (7.4.2) **The sum $v \oplus w$ does not have properties** we might expect of ordinary, familiar addition. For example, it is **not** true that for all u, v , and for any scalar α ,

$$(7.4.2a) \quad (\alpha u) \oplus v = u \oplus (\alpha v) = \alpha(u \oplus v). \quad \text{Exercise (7.6.1)}$$

We do have, however, the following structure:

- (7.4.3)

PROPOSITION: *There are only three instances for which the Lorentz sum, $v \oplus w$ (7.3.1a), and the ordinary arithmetic sum, $v + w$, agree, namely,*

$$u \oplus v = u + v \quad \text{if and only if} \quad \begin{cases} u = 0, & \text{or} \\ v = 0, & \text{or} \\ u + v = 0. \end{cases}$$

PROOF: Exercise (7.6.3) ■

- (7.4.4)

NOTE: *If u and v are speeds, then the Lorentz sum $u \oplus v$ of (7.3.1a) shows that speeds of material objects can never exceed c . That is, $-c < u, v < c$. The straightforward proof is left as Exercise (7.6.2).*

(7.5) Addition of Speeds via Areas

In Theorem (7.5.6) following, we present the relativistic counterpart of Theorem (5.6.2) that measures speeds of one frame relative to another by the addition of areas.

Results in this section depend on properties of the particular hyperbola

- $y^2 - x^2 = 1$ that is symmetric about the Y -axis where
- the hyperbola asymptotes are the $\pm 45^\circ$ lines as measured from the X -axis.

The $\pm 45^\circ$ slopes of the asymptotes force the speed of light to be $c = 1$ (see Def. (1.13.5a), pg. 39). In the special case where $c = 1$, we have from (B.3.5a) that relative speeds between frames have the form $v = \tan(\theta) = \tanh(r_{A,B})$. This is the first time the speed c is associated with the trigonometric quantity, $\tan(\theta)$ (see (B.1.1c), pg. 376 for definition of the hyperbolic function $\tanh(\cdot)$.)

(7.5.1)

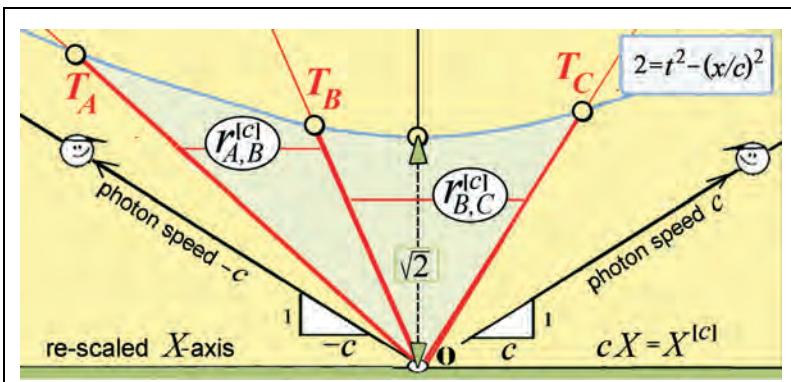


Rescaling the X -axis: To pass from the special case $c = 1$ to more general and arbitrary units for c , we rescale the original X -axis (where $c = 1$) by replacing each x with cx (where the value of c can now be arbitrary). The rescaling now allows for non-unity units for c such as $c = 2.997\,924\,58 \times 10^8$ meters/sec and $c = 186,000$ miles/sec.

We interpret the special-case Figure (B.3.4), pg. 381 as a Minkowski spacetime diagram where the T_A and T_B axes represent the worldlines of the origins of synchronized inertial frames \mathcal{F}^A and \mathcal{F}^B , respectively. The two $\pm 45^\circ$ asymptotes will represent the worldlines of leftward and rightward traveling photons that are simultaneously released at time $t = 0$ from the synchronized origins.

(7.5.2)

FIGURE: worldlines of Two Photons (Leftward and Rightward) and Three Origins of Synchronized Inertial Frames



(7.5.3)

NOTATION: Unscaled and Scaled Quantities. As noted in (7.5.1), the X -axis $X^{[c]}$ of the Minkowski diagram of (7.5.2) is constructed

by replacing each x of the original X -axis (*where $c = 1$*) with the value cx . Then

(7.5.3a)

Figure (7.5.2)	Unscaled	Scaled
$c=1$	$c>0$	
Axes	x	\rightarrow
<i>Speed of one frame relative to another</i>	v	\rightarrow
<i>Triangular area bounded by hyperbola, T_A and T_B-axes</i>	$r_{A,B}$	\rightarrow
	$r_{A,B}^{[c]}$	

(7.5.4)

NOTE: In the notation of (7.5.3) above, $v^{[1]}$ and $r_{A,B}^{[1]}$ coincide with the speed v and area $r_{A,B}$ of the unscaled graph, i.e., $v^{[1]} = v$ and $r_{A,B}^{[1]} = r_{A,B}$. We abandon this somewhat awkward superscript notation $v^{[c]}$ and $r_{A,B}^{[c]}$ when the units for c are clearly established and we no longer need to compare diagrams with differently scaled X -axes.

(7.5.5)
Calculus

PROPOSITION: *In Figure (B.3.4), replace the unscaled axis X with the scaled axis $X^{[c]}$ in which each $x \in X$ is replaced with $cx \in X^{[c]}$. Then*

(7.5.5a) *Hyperbola $2 = t^2 - x^2$ is transformed (rescaled) into the hyperbola $2 = t^2 - (x^{[c]}/c)^2$. Moreover,*

(7.5.5b)
$$v^{[c]} = cv \quad \text{and} \quad r_{A,B}^{[c]} = cr_{A,B}$$

where $v^{[c]}$ and $r_{A,B}^{[c]}$ are defined in (7.5.3a).

PROOF: From the first line of Table (7.5.3a), we replace $x^{[c]}$ with cx in the scaled equation $2 = t^2 - (x^{[c]}/c)^2$ to obtain the unscaled equation $2 = t^2 - x^2$, which proves (7.5.5a).

Speed of \mathcal{F}^B relative to \mathcal{F}^A is $v^{[c]} = (x_1^{[c]} - x_0^{[c]})/(t_1 - t_0)$ where $[x_1^{[c]}, t_1]_A$ and $[x_0^{[c]}, t_0]_A$ are line-of-sight coordinates in \mathcal{F}^A from which a particular point of \mathcal{F}^B is observed at \mathcal{F}^A times t_1 and t_0 .

$$v^{[c]} = \frac{(x_1^{[c]} - x_0^{[c]})}{(t_1 - t_0)} = \frac{(c x_1 - c x_0)}{(t_1 - t_0)} = c \frac{(x_1 - x_0)}{(t_1 - t_0)} = c v,$$

which proves part of (7.5.5b). Now suppose the areas in Figure (7.5.2) are defined by a double integral over region R . In particular, since $x^{[c]} = c x$,

$$r_{A,B}^{[c]} = \iint_R dx^{[c]} dt = \iint_R c dx dt = c \iint_R dx dt = c r_{A,B}$$

Statements (7.5.5a) and (7.5.5b) are fully established, which ends the proof. ■

(7.5.6) **Theorem: Sums of Areas as Speeds.** Given the scaled space-time graph (7.5.2) with general units for c where the areas $r_{A,B}^{[c]}$ and $r_{B,C}^{[c]}$ denote the areas (under the hyperbola $2 = t^2 - (x/c)^2$) between the T_A - T_B pair of axes and the T_B - T_C pair of axes, respectively.

If $v^{[c]}$ denotes the speed of frame \mathcal{F}^B relative to frame \mathcal{F}^A , and $w^{[c]}$ is the speed of frame \mathcal{F}^C relative to frame \mathcal{F}^B , then

$$(7.5.6a) \quad v^{[c]} = c \tanh\left(\frac{1}{c} r_{A,B}^{[c]}\right), \quad w^{[c]} = c \tanh\left(\frac{1}{c} r_{B,C}^{[c]}\right).$$

The speed of frame \mathcal{F}^C relative to frame \mathcal{F}^A is

$$(7.5.6b) \quad v^{[c]} \oplus w^{[c]} = \frac{v^{[c]} + w^{[c]}}{1 + (v^{[c]} w^{[c]})/c^2}.$$

PROOF: Proving (7.5.6a).

$$\begin{aligned} v^{[c]} &= c v && \text{from (7.5.5b)} \\ &= c \tanh(r_{A,B}) && v = \tanh(r_{A,B}) \text{ from (B.3.5a)} \\ &= c \tanh\left(\frac{1}{c} r_{A,B}^{[c]}\right) && r_{A,B}^{[c]}/c = r_{A,B} \text{ from (7.5.5b).} \end{aligned}$$

A similar argument, using $r_{B,C}$ instead of $r_{A,B}$, establishes (7.5.6a).

.....

Proving (7.5.6b). Using (7.5.6a), the speed of frame \mathcal{F}^C relative to frame \mathcal{F}^A is determined by the “double” area $r_{A,C}^{[c]}$ of Figure (7.5.2). To echo the notation of (1.13.3), we denote this speed of \mathcal{F}^C relative to \mathcal{F}^A by $(v^{[c]} \oplus w^{[c]})^{[c]}$. Then

$$\begin{aligned}
 v^{[c]} \oplus w^{[c]} &= c \tanh(r_{A,C}^{[c]}/c) && \text{from (7.5.6a)} \\
 &= c \tanh(r_{A,C}) && r_{A,C}^{[c]}/c = r_{A,C} \text{ from (7.5.5b)} \\
 &= c \tanh(r_{A,B} + r_{B,C}) && \text{addition of areas} \\
 &= c \frac{\tanh(r_{A,B}) + \tanh(r_{B,C})}{1 + \tanh(r_{A,B}) \tanh(r_{B,C})} && \begin{matrix} \text{property (B.1.3e)} \\ \text{of } \tanh(x+y) \end{matrix} \\
 &= c \frac{v+w}{1+vw} && \text{from (B.3.5a), pg. 381} \\
 &= \frac{c \left(\frac{v^{[c]}}{c} + \frac{w^{[c]}}{c} \right)}{1 + (v^{[c]}/c)(w^{[c]}/c)} && \text{from (7.5.5b)} \\
 &= \frac{v^{[c]} + w^{[c]}}{1 + (v^{[c]}w^{[c]})/c^2}. && \text{cancel } c \text{ in numerator}
 \end{aligned}$$

Having established (7.5.6a) and (7.5.6b), the proof is done. ■

(7.6) Exercises

(7.6.1) Exercise **Properties of $u \oplus v$.** Show examples that prove (7.4.2a).

(7.6.2)
Exercise

Property of $u \oplus v$. Show that for all numbers (speeds) u and v , we have $-c < u \oplus v < c$.

(7.6.3)
Exercise

Three cases of $u \oplus v$. Prove Proposition (7.4.3)



(8) The Hyperbola of Time-Stamped Origins

(8.0.4)



One of the most revolutionary results in modern science is Einstein's use of hyperbolic geometry to connect space and time. The explanation will require that we go through a series of theorems (no royal road). Nevertheless, we invite the reader to join us along this challenging yet rewarding path.

(8.1) Invariance of Minkowski Length

As noted on page 83, non-relativistic Galilean physics assumes that time is universal, which implies there is one universal clock whose readings are valid for all observers, regardless of their relative motion.

This universal clock physical property is interpreted mathematically by saying that time is invariant under the Galilean line-of-sight transformation (5.4.2), pg. 90, $G_v : [x, t]_A \rightarrow [x', t]_B$. That is to say, the time entries of the input coordinate pair $[x, t]_A$ and the output coordinate pair $[x', t]_B$ are always equal.

Similarly, we may ask what quantity might be invariant under the Lorentz transformation $L_v : [x, t]_A \rightarrow [x', t']_B$.

A first glance might not reveal an obvious invariant quantity for L_v . However, Theorem (8.1.2) will show a special “length” that remains intact for the linked coordinates $[x, t]_A$ and $[x', t']_B$. We define this “length” now:

(8.1.1)

DEFINITION: For any event $[x, t]$ for any frame \mathcal{F} in X - T space-time, the MINKOWSKI LENGTH is defined to be the quantity

$$(8.1.1a) \quad M([x, t]) \stackrel{\text{def}}{=} c^2 t^2 - x^2 \leftarrow \{\text{units} = \text{distance}^2\}$$

where c is the speed of light in a vacuum.

Invariance of Minkowski Lengths Between Pairs of Frames.

Note that (8.1.1) defines $M([x, t])$ for a point (event) $[x, t]$ in *any* inertial frame \mathcal{F} . The following theorem shows that if two points $[x, t]_A$ and $[x', t']_B$ are in different frames \mathcal{F}^A and \mathcal{F}^B and lie on opposite ends of a line-of-sight corridor, then they share the same Minkowski length.

(8.1.2)

THEOREM: (Lorentz Invariance) *Given the Lorentz transformation $L_v: \mathcal{F}^A \rightarrow \mathcal{F}^B$ (7.1.3a) If $L_v([x, t]_A) = [x', t']_B$, then the Minkowski length M of (8.1.1) is invariant for both line-of-sight points $[x, t]_A$ and $[x', t']_B$ in that*

(8.1.2a)

$$L_v([x, t]_A) = [x', t']_B \quad \text{implies} \quad M([x, t]_A) = M([x', t']_B).$$

PROOF: (As usual, we write vectors of \mathbf{R}^2 either as horizontal row vectors or as vertical column vectors, depending on the context.) Consider the following chain of equalities:

$$L_v \left(\begin{bmatrix} x \\ t \end{bmatrix}_A \right) = \gamma_v \underbrace{\begin{bmatrix} 1 & -v \\ -v/c^2 & 1 \end{bmatrix}}_{L_v \text{ from (7.1.3a)}} * \begin{bmatrix} x \\ t \end{bmatrix}_A = \underbrace{\begin{bmatrix} \gamma_v(x - vt) \\ \gamma_v(-vx/c^2 + t) \end{bmatrix}}_{\text{from (A.4.3e)}}_B = \underbrace{\begin{bmatrix} x' \\ t' \end{bmatrix}}_{\text{from (8.1.2a)}}_B.$$

where $\gamma_v = 1/\sqrt{1 - (v/c)^2}$ (7.1.3b). It is the final (rightmost) equality that tells us

$$(8.1.2b) \quad x' = \gamma_v(x - vt) \quad \text{and} \quad t' = \gamma_v \left(\frac{-vx}{c^2} + t \right).$$

With the specific form (8.1.2b), algebraic manipulation shows that $(ct')^2 - (x')^2 = (ct)^2 - (x)^2$, which is to say, the Minkowski lengths, $M([x', t']_B)$ and $M([x, t]_A)$, as defined in (8.1.1), are equal. This confirms statement (8.1.2a) and ends the proof. ■

(8.2) The Time-Stamped Origins Theorem

One way to view the time axis T of any frame \mathcal{F} is to view each of its coordinates $[0, t]$ as the origin $x = 0$ with a time-stamp t . More formally, we state

(8.2.1)

DEFINITION: For time t in frame \mathcal{F} we say that $[0, t]$ is the ORIGIN OF \mathcal{F} WITH TIME-STAMP t .

NOTE: The set of all time-stamped origins of a single frame \mathcal{F} is T -axis of \mathcal{F} .

(8.2.2)

NOTATION: (*Moving Frames and Coordinates*) Given a fixed frame \mathcal{F}_0 with coordinates denoted $[x, t]_0$. Then any frame moving with constant speed v relative to frame \mathcal{F}_0 is denoted by \mathcal{F}_v and its coordinates are written in the form $[x', t']_v$.

Using the terminology of (8.2.1), we have a useful hyperbolic relation among a set of t_0 -stamped origins for infinitely many frames (all origins are stamped at fixed time t_0).

(8.2.3)

THEOREM: (*The Time-Stamped Origins Theorem*)

The graph of frame \mathcal{F}_0 is drawn with perpendicular X_0 and T_0 axes. For each speed v , \mathcal{F}_v denotes the synchronized frame with speed v relative to \mathcal{F}_0 .

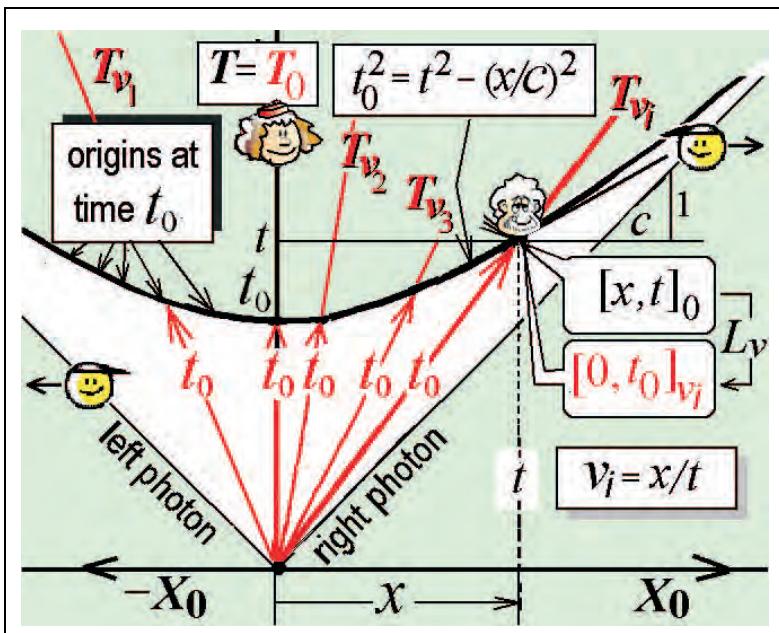
Suppose the time-stamped origin $[0, t_0]_v$ in \mathcal{F}_v corresponds to $[x, t]_0$ in frame \mathcal{F}_0 . That is, $L_v([0, t_0]_0) = [x, t]_v$.

Then

for any fixed time, $t_0 > 0$, and all speeds, $-c < v < c$, the set of all time-stamped origins, $[0, t_0]_v \in \mathcal{F}_v$, forms the hyperbola

$$(8.2.3a) \quad t_0^2 = t^2 - (x/c)^2,$$

where x and t are coordinates in any of the frames \mathcal{F}_v .



PROOF: Let the time-stamped origin, $[0, t_0]_v \in \mathcal{F}_v$ and coordinates $[x, t]_0$ lie at opposite ends of a Lorentz transformation line-of-sight corridor. That is, for $[x, t]_0$ in \mathcal{F}_0 and $[0, t_0]$ in \mathcal{F}_v , we have $L_v([x, t]_0) = [0, t_0]_v$. Then, from (8.1.2a), $[0, t_0]$ and $[0, t_0]_v$ have the same Minkowski length (8.1.1a), or

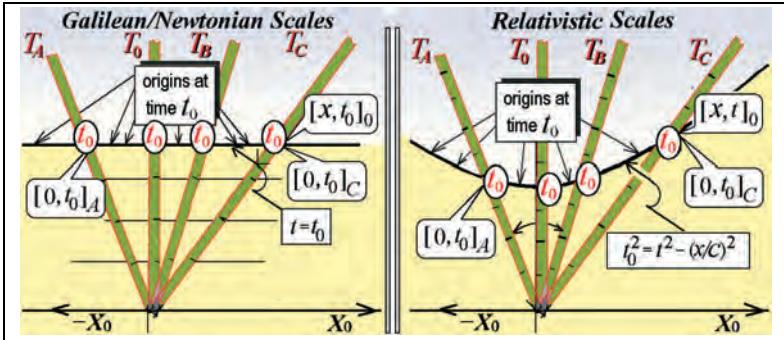
$$M([x, t]_0) = c^2 t^2 - x^2 \text{ and } M([0, t_0]_v) = c^2 t_0^2$$

for variables x and t in \mathcal{F}_v . Having established the equality $(ct_0)^2 = (ct)^2 - x^2$ (equivalently, $t_0^2 = t^2 - (x/c)^2$), (8.2.3a) is proved. ■

(8.3) Interpreting the Time-Stamped Origins Theorem

In Section (7.5) we showed that the hyperbola $t^2 - (x/c)^2 = k^2$ was useful in providing a geometric interpretation of the formula $v \oplus w$, the relative addition of speeds v and w . The following theorem provides a physical interpretation of this hyperbola.

(8.3.1)

**FIGURE: Galilean vs. Relativistic Time Axis Scales**

In the Galilean and relativistic spacetime diagrams above, each frame with relative speed v has its time axis forming a certain angle with the vertical time axis.

In the **Galilean case** (left panel of Fig. (8.3.1)), the various origins $[0, t_0]_0$, $[0, t_0]_A$, $[0, t_0]_B$, and $[0, t_0]_C$, at time t_0 lie on the horizontal line $[x, t_0]$, irrespective of relative speeds between frames \mathcal{F}^A , \mathcal{F}^B , \mathcal{F}^C , etc., where x and t_0 are coordinates in frame \mathcal{F}_0 with perpendicular X_0 and T_0 axes.

In the **relativistic case** (right panel of Fig. (8.3.1)), the various origins $[0, t_0]_0$, $[0, t_0]_A$, $[0, t_0]_B$, $[0, t_0]_C$, at time t_0 lie on the hyperbola $t_0^2 = t^2 - (x/c)^2$ where x and t are coordinates in frame \mathcal{F}_0 with perpendicular X_0 and T_0 axes (8.2.3a).

For both the Galilean (left panel) and relativistic case (right panel) of Figure (8.3.1), each time axis can be viewed as a ruler where the distance from 0 to t_0 varies when it is measured in the \mathcal{F}_0 frame having perpendicular X_0 and T_0 axes.

(8.4) Tangent Lines of Simultaneity

When the spacetime graph for frame \mathcal{F} has perpendicular axes X and T , then the synchronized frame \mathcal{F}' , moving with speed v with respect to \mathcal{F} has its time axis, T' , forming a certain angle with the vertical T axis.

The following theorem describes the slopes and positions of lines of simultaneity (at fixed time t') which we already know represent the X -axis at time t' of the moving frame \mathcal{F}' (see (4.4.3), pg. 73).

(8.4.1)
Calculus

THEOREM: (Tangent Lines = Simultaneous Events)

Given a spacetime graph of frame \mathcal{F} with perpendicular X and T axes, and a synchronized frame \mathcal{F}' , moving with relative speed v , with distance axis X' and time axis T' .

Then

(8.4.1a) $1/v$ is the slope of time axis T' , and

(8.4.1b) v/c^2 is the slope of any and all straight lines of simultaneous events in frame \mathcal{F}' . The X' -axis is the set of $t' = 0$ events in \mathcal{F}' .

In particular, the line of simultaneous events at any time t' in \mathcal{F}' , expressed in \mathcal{F}' coordinates x, t , is the tangent line to the hyperbola

$$(8.4.1c) \quad (t')^2 = t^2 - (x/c)^2.$$

For all times t' in \mathcal{F}' , tangent line (8.4.1c) has slope v/c^2 and intersects the vertical T -axis of \mathcal{F} at the \mathcal{F} coordinates $[0, t']$.

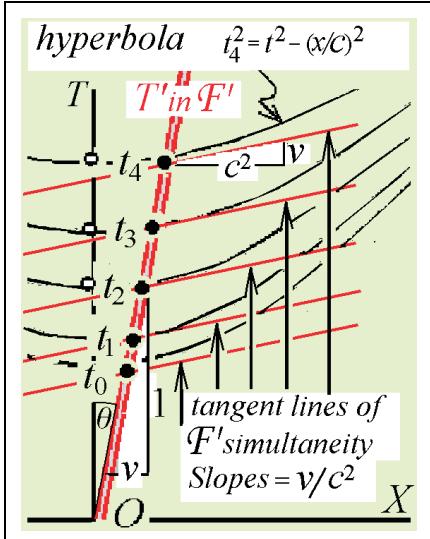
PROOF:

Proof of (8.4.1a). The speed of frame \mathcal{F}' as seen from frame \mathcal{F} is $v = \Delta x / \Delta t$ while slope of lines in the X - T plane is given by $\Delta t / \Delta x$. This proves (8.4.1a).

Proof of (8.4.1b). Points of constant (simultaneous) time t' in \mathcal{F}' have form $[x', t']$ where x' varies and time t' is fixed. The Lorentz line-of-sight transformation $L_v : [x, t] \rightarrow [x', t']$ (7.1.3a) tells us that

$$\underbrace{\gamma_v \begin{bmatrix} 1 & -v \\ -v/c^2 & 1 \end{bmatrix}}_{L_v} \begin{bmatrix} x \\ t \end{bmatrix} = \begin{bmatrix} \gamma_v(x - vt) = x' \\ \gamma_v(t - vx/c^2) = t' \end{bmatrix} \quad \text{in frame } \mathcal{F}'$$

where, from (7.1.3b), $\gamma_v = 1/\sqrt{1 - (v/c)^2}$.



Relative to the frame for the X - T graph, there are infinitely many frames each moving with speed v , where $-c < v < c$. If we fix a time, t_i , then the plot of each of the origins (with age t_i) of the infinitely many frames, yields the hyperbola $t_i^2 = t^2 - (x/c)^2$. This graph shows hyperbolas of origins of ages $t_0 < t_1 < t_2 < t_3 < t_4$. For one frame \mathcal{F}' , we plot its set of origins at all times (its time axis T'). The five tangent lines shown at the intersections of T' , one for each t_i , is the line of simultaneous events that occur at time t_i in frame \mathcal{F}' .

Now the second entry, $\gamma_v(t - vx/c^2) = t'$, is constant. Solving for t , we produce the following straight-line form $t = mx + b$ in \mathcal{F} variables x and t :

$$t = \left(\frac{v}{c^2}\right)x + \frac{t'}{\gamma_v}.$$

This line in the X - T plane has slope (v/c^2) , which proves (8.4.1b).

Proof of (8.4.1c). Solve for dt/dx via implicit differentiation of the hyperbola $(t')^2 = t^2 - (x/c)^2$ to obtain $0 = 2t - 2(x/c^2)(dx/dt)$. This equation implies the slope (derivative) dt/dx at any $[x, t]$ in \mathcal{F} is

$$\frac{dt}{dx} = \frac{(x/t)}{c^2} = \left(\frac{v}{c^2}\right) \quad (\text{when } x = vt)$$

where $[x, t]$ lies on the T' axis in graph of frame \mathcal{F} only when $x = vt$. Substitution of $x = vt$ into the previous equation shows that the slope $dt/dx = (v/c^2)$ which is exactly what is required by (8.4.1b). This justifies (8.4.1c) which ends the proof. ■



(8.5) Exercises

(8.5.1) Exercise

Use of Lorentz transformation: The diagram of (12.2.2) shows four coordinates that are listed in the table to the right — the two \mathcal{F}^A coordinates for P and Q' and the two \mathcal{F}^B coordinates of P' and Q .

Fill in the four missing coordinates in the table (*marked **) that lie at opposite ends of the line-of-sight corridors of the given coordinates.

<i>Point</i>	<i>Frame</i>	\mathcal{F}^A	\mathcal{F}^B
P		$[0, t_0]_A$	*
Q'		$[d_0, t_0]_A$	*
P'		*	$[-d_0, t_0]_B$
Q		*	$[0, t_0]_B$

Hint: Use the Lorentz transformation $L_v : \mathcal{F}^A \rightarrow \mathcal{F}^B$ (equivalently, $L_{-v} : \mathcal{F}^B \rightarrow \mathcal{F}^A$).

(8.5.2) Exercise

(1) From the coordinates of the table (8.5.1) show that $P' = \sigma_v P$ and $Q' = \sigma_v Q$ where $\sigma_v = \sqrt{1 - (v/c)^2}$ (1.8.2).

(2) Show how the results of (1) prove that moving clocks run slower than stationary clocks by a factor of σ_v .

Hint: Recall that measuring moving clocks requires two readings of the same moving clock at different times (F.12.5b).

V. Graphic Resolution of the Paradoxes

Chapter 9 The Accommodating Universe. *An intergalactic traveler apparently travels faster than light.*

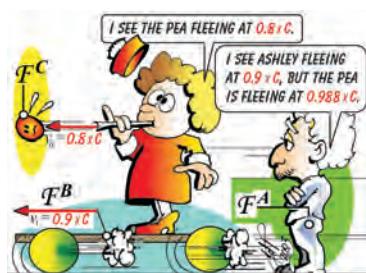
Chapter 10 The Comparison Paradoxes. *Any two observers see each other's moving clock run slower than his own clock.*

Chapter 11 The Twin Paradox. *A (stationary) Earthbound twin sees her (moving) spaceship twin age more slowly than Earth residents. Yet, the (stationary) spaceship twin does not see the (moving) Earthbound twin age more slowly than spaceship residents.*



Chapter 12 The Train-and-Tunnel Paradox. *A pair of lightning bolts strike simultaneously when viewed by observers on the ground but they are viewed as sequential lightning bolts when viewed by passengers on the train.*

Chapter 13 The Pea-Shooter Paradox. *A ball on a moving train has speed $0.8c$ relative to the train while the train, as seen from the embankment, has speed $0.9c$. From the train embankment, the speed of the ball is measured to be $0.988c$ instead of the “expected” speed of $0.8c + 0.9c = 1.7c$.*



Chapter 14 Bug-Rivet Paradox. *The collision of a rivet with a metal plate shows that the rivet cannot remain rigid as it accelerates out of an inertial frame. Moreover, if rigidity is assumed, then the sequence of cause and effect can be reversed—cause can result after its effect.*



(9) The Accommodating Universe Paradox

(9.1) Preview

- (9.1.1) **Summary of the Accommodating Universe, Section (1.9).** If a skateboarder in frame \mathcal{F}^S races on a straight-line track in \mathcal{F}^E towards the finish line, then the faster he runs, the shorter the track becomes from his point of view in frame \mathcal{F}^S —the Universe is *accommodating* in the sense that the faster he goes, the shorter the track becomes from his point of view and, hence, the closer the finish line is to the runner.

The paradox lies in the statement that if the length of the track can be made arbitrarily small as viewed from fame \mathcal{F}^S , then the skateboarder can complete the run in an arbitrarily short amount of time—one year, say. If the at-rest track length in frame \mathcal{F}^E is 100 light years, then a photon requires 100 years to reach the finish line. This implies that the photon may require *more* time than the runner does to run on the same track.

The specious reasoning springs from using clocks from separate frames to time the skateboarder and the photon. The Minkowski diagram (9.2.1) following offers a graphical analysis.

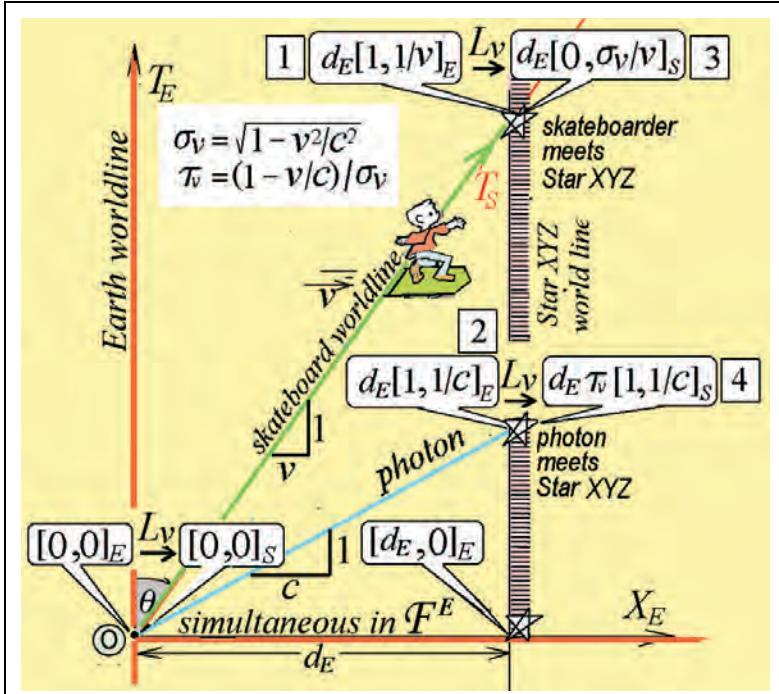
(9.2) Setup for the Minkowski Diagram

Figure (9.2.1) shows four worldlines:

- The line of a photon with slope $1/c$ that indicates the photon is traveling rightward at speed c .
- In frame \mathcal{F}^E , the vertical line of Earth which rests at $x = 0$ and the vertical line of Star XYZ which is at a distance d_E from Earth and rests at $x = d_E$.

- In frame \mathcal{F}^S , the line of the skateboarder has slope $1/v$. This means the skateboarder travels with speed v relative to Earth's frame \mathcal{F}^E .

(9.2.1)

FIGURE: The Accommodating Universe

\mathcal{F}^E Coordinates. Relative to the Earth's frame, \mathcal{F}^E , the skateboarder moves rightward with speed v and a photon travels with speed c over a distance d_E to Star XYZ. The distance for both journeys is d_E so the times in \mathcal{F}^E are d_E/v and d_E/c , respectively. Figure (9.2.1) lists these \mathcal{F}^E coordinates as 1 and 2 at Star XYZ.

\mathcal{F}^S Coordinates. The Lorentz transformation, $L_v([x, t]_E) = [x', t']_S$, converts coordinates $[x, t]_E$ in \mathcal{F}^E to line-of-sight coordinates $[x', t']_S$ in \mathcal{F}^S . These values of Figure (9.2.1) are summarized as follows:

(9.2.1a)

At Star XYZ	\mathcal{F}^E coordinates	\mathcal{F}^S coordinates
skateboarder	$\boxed{1} \quad d_E \begin{bmatrix} 1 \\ 1/v \end{bmatrix}_E$	\rightarrow
photon	$\boxed{2} \quad d_E \begin{bmatrix} 1 \\ 1/c \end{bmatrix}_E$	$\boxed{3} \quad d_E \begin{bmatrix} 0 \\ \sigma_v/v \end{bmatrix}_S$
		$\boxed{4} \quad d_E \tau_v \begin{bmatrix} 1 \\ 1/c \end{bmatrix}_S$

where

$$(9.2.1b) \quad L_v = \frac{1}{\sigma_v} \begin{bmatrix} 1 & -v \\ -v/c^2 & 1 \end{bmatrix}, \quad \sigma_v = \sqrt{1 - (v/c)^2} \text{ from (7.1.3b).}$$

(9.3) Resolving the Accommodating Universe

(9.3.1)



With unsound reasoning, we concluded in (1.9.5), that the skateboarder is traveling “faster than light” because his journey takes only 1% of the time that light requires to traverse the same path from Earth to Star XYZ . We neglected to mention that we were clocking a race with two separate participants (*photon and skateboarder*) using clocks from two separate frames (*Earth frame \mathcal{F}^E and skateboarder frame \mathcal{F}^S*).

The Minkowski diagram (9.2.1) shows the worldline of the photon below that of the skateboarder. This means the photon collision with Star XYZ occurs before (*below*) the skateboarder collision with Star XYZ .

Analytically, we may compare the travel times that appear as the second coordinate entries of the column vectors in Table (9.2.1a). If we choose these times either from the \mathcal{F}^E column (*vectors $\boxed{1}, \boxed{2}$*) or from the \mathcal{F}^S column (*vectors $\boxed{3}, \boxed{4}$*), we will see that in both cases,

$$(9.3.1a) \quad \frac{\text{skateboarder time}}{\text{photon time}} > 1.$$

(9.3.2)

Stephen Hawking implies that the Accommodating Universe Paradox (9.2.1) will help us ensure our long-term survival. In a 13 June 2006 press conference in Hong Kong, he declared that humankind must colonize outer space.

How can a crew of humans complete voyages to planets that may lie millions of light years from Earth? The answer lies in being able to travel fast enough according to the Accommodating Universe Paradox, which says that at sufficiently high speeds, both *distance* and *travel time* from Earth to the desired planet can be as small as desired. In this way, a trip that requires a few million years (*as measured on Earth*) could be achieved within a human lifetime (*as measured on the spaceship*).

There are at least two major problems.

- First, after we determine the appropriate planet of a remote solar system, there is the challenge of how to generate the enormous amount of energy required for this high-speed trip.¹
- Second, even if the trip takes only a few spaceship months, it will take millions of Earth years to complete.

So if the travelers forget something and go back, they will have aged a year or two, but the civilization that sent them will have long vanished. The memory of the colonizing effort will probably be gone. The spaceship travelers will be strangers in a strange land. (*To see why this is so, consult Figure (9.2.1), or the resolution of the Twin Paradox at (1.11.3), (1.11.4), pg. 31, and (12.3), pg. 158.*)

¹The question of how the human race might best survive in the long run was raised by Hawking on the Internet when the people at Yahoo! invited Hawking, as one of ten celebrities, to submit a question online at answer.yahoo.com. This was part of their “Ask the Planet” campaign. Hundreds of replies were submitted although Hawking claimed to have “known the answer” in advance.

(9.4) Exercises

(9.4.1)
Exercise

Computing \mathcal{F}^E Time Ratio Using times from vectors $\boxed{1}$ and $\boxed{2}$ in \mathcal{F}^E coordinates of Table (9.2.1a), show that the ratio of (9.3.1a) is greater than 1.

(9.4.2)
Exercise

Computing \mathcal{F}^S Time Ratio Using times from vectors $\boxed{3}$ and $\boxed{4}$ in \mathcal{F}^S coordinates of Table (9.2.1a), show that the ratio of (9.3.1a) is greater than 2.



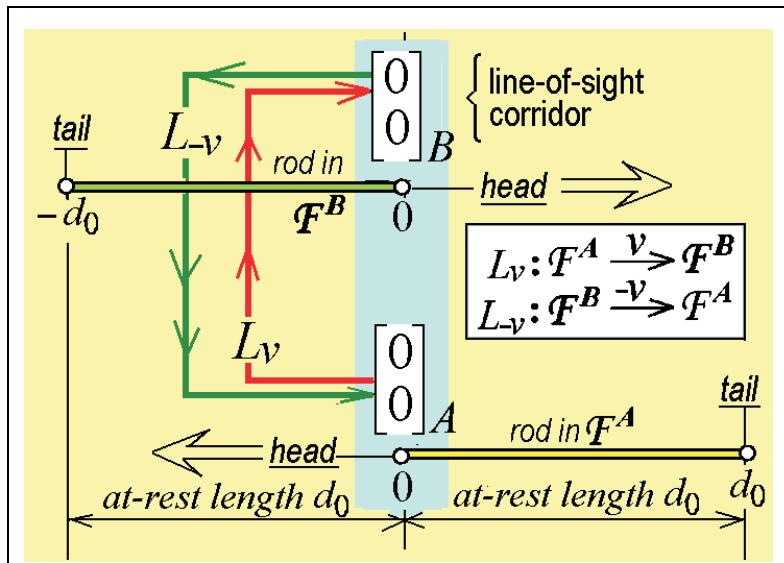
(10) The Length-Time Comparison Paradoxes

(10.1) An Overview of the Paradoxes

(10.1.1) We have two inertial reference frames, \mathcal{F}^A and \mathcal{F}^B , with respective observers A and B that are at rest. Observer A sees Observer B move with speed v and B sees A move with speed $-v$. Both A and B have identical clocks and rulers. Then the first results, (1.7.2a) and (1.7.2b), along with the illustrations of (1.10.4a) and (1.10.4b) lead to the following observations:

- **Both moving clocks run slow.** Each observer (*on the platform, or in the railway car*) sees the other's moving clock run slower than his/her own clock.
- **Both moving rulers shrink.** Each observer sees the other's moving ruler as shorter than his/her own ruler.

(10.1.2) **FIGURE: Synchronized Parallel Rods**



The rod (ruler) in inertial frame \mathcal{F}^B has positive rightward speed v relative to frame \mathcal{F}^A . The rod in inertial frame \mathcal{F}^A has negative leftward speed $-v$.

relative to frame \mathcal{F}^B . The frames \mathcal{F}^A and \mathcal{F}^B are synchronized in the sense of (3.4.5), pg. 61.

(10.1.3)

Question: How can both bullet statements of (10.1.1) be true?

Answer: They can not. However, if we replace “Each observer” with “A pair of observers in each frame,” then, as we shall see in Theorem (10.2.1), there is no inconsistency. For example, from (stationary) frame \mathcal{F}^A , we declare the length of a moving rod that lives in (moving) frame \mathcal{F}^B , to be the \mathcal{F}^A distance between two (stationary) observers in frame \mathcal{F}^A who simultaneously stand at the positions of the head and the tail of the moving rod. (Also see Figures (2.4.2) and (2.4.3), pg. 51.)

Since frames \mathcal{F}^A and \mathcal{F}^B are synchronized (3.4.5), pg. 61, we enlist the aid of two other cohort members, one in each frame so that we can measure

(10.1.4)

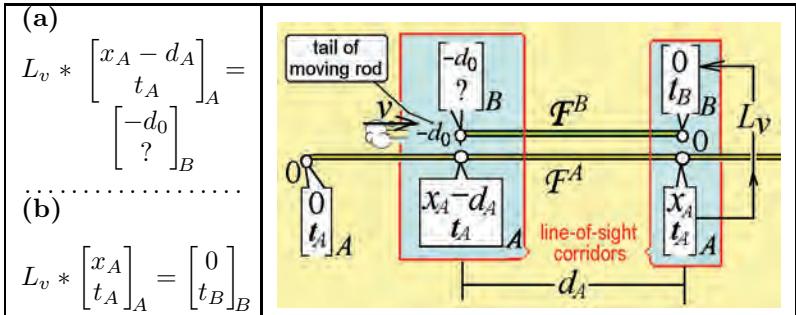
- (a) the ticking rate of the moving clock,
- (b) the speed of the rod in the other moving frame (see (2.4.2), pg. 51), and
- (c) the length of the rod in the other moving frame (see (2.4.3)).

Specifically, choose any two arbitrary times $t_A > 0$ in \mathcal{F}^A and $t_B > 0$ in \mathcal{F}^B . Then, in each frame, choose the (stationary) cohorts with the following coordinates:

$$\begin{bmatrix} x_A - d_A \\ t_A \end{bmatrix}_A, \quad \begin{bmatrix} x_A \\ t_A \end{bmatrix}_A \text{ in } \mathcal{F}^A \quad \text{and} \quad \begin{bmatrix} x_B - d_B \\ t_B \end{bmatrix}_B, \quad \begin{bmatrix} x_B \\ t_B \end{bmatrix}_B \text{ in } \mathcal{F}^B$$

who, through their respective line-of-sight corridors, observe the end-points of the moving rods—the end-points are marked 0 and $\pm d_0$. The two cohorts are depicted in Figure (2.4.3). Moreover, the as yet unknown distances between the two cohorts are d_A in \mathcal{F}^A and d_B in frame \mathcal{F}^B . (See Figures (10.1.5) and (10.1.10) where d_A is the length of \mathcal{F}^B 's rod as measured in \mathcal{F}^A , and similarly, d_B is the length of \mathcal{F}^A 's rod as measured in \mathcal{F}^B .)

(10.1.5) **FIGURE:** Line-of-sight Coordinates in \mathcal{F}^B Where Frame \mathcal{F}^A is Fixed, Relative Speed of Frame $\mathcal{F}^B = v$.



(10.1.6) **Note:** The column vector $\text{col}[-d_0, ?]_B$ appears in Figure (10.1.5) with the symbol “?” since the time at the tail of the rod in \mathcal{F}^B (i.e., at $x = -d_0$) does not need to be calculated.

(10.1.7) **Calculating Quantities of (10.1.4) as Measured in \mathcal{F}^A .** These results follow from the vector equations in the left hand panel of Figure (10.1.5). Specifically, equation (10.1.5)(b) takes the form

$$\underbrace{\frac{1}{\sigma_v} \begin{bmatrix} 1 & -v \\ -v/c^2 & 1 \end{bmatrix}}_{\text{Lorentz matrix } L_v} * \begin{bmatrix} x_A \\ t_A \end{bmatrix}_A = \begin{bmatrix} 0 \\ t_B \end{bmatrix}_B$$

which gives us the two equations

$$(10.1.7a) \quad x_A - vt_A = 0 \quad \text{which implies } \begin{cases} x_A = vt_A \\ v = x_A/t_A \end{cases}$$

(i) $t_A - \frac{v}{c^2}x_A = \sigma_v t_B$.

Substituting $x_A = vt_A$ from (10.1.7a) into equation (i) above, we obtain

$$(10.1.7b) \quad t_A \underbrace{\left(1 - \frac{v^2}{c^2}\right)}_{\sigma_v^2} = \sigma_v t_B \quad \text{which implies } \boxed{\sigma_v t_A = t_B}.$$

Finally, equation (10.1.5)(a) takes the form

$$\underbrace{\frac{1}{\sigma_v} \begin{bmatrix} 1 & -v \\ -v/c^2 & 1 \end{bmatrix}}_{\text{Lorentz matrix } L_v} * \begin{bmatrix} x_A - d_A \\ t_A \end{bmatrix}_A = \begin{bmatrix} -d_0 \\ * \end{bmatrix}_B$$

which, after equating the top entry of the left and right hand sides, gives us the single equation

$$x_A - d_A - vt_A = -\sigma_v d_0.$$

Substitute $x_A = vt_A$ from (10.1.7a) in the equation above to obtain

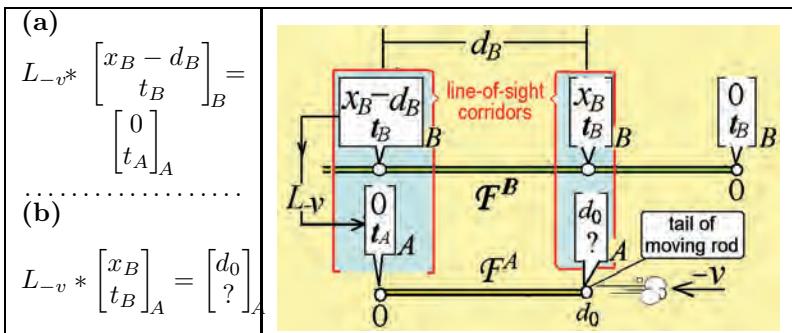
$$(10.1.7c) \quad \boxed{d_A = \sigma_v d_0}.$$

- (10.1.8) **When frame \mathcal{F}^A is fixed, the boxed equations** of (10.1.7a), (10.1.7b), and (10.1.7c) answer the respective questions of (10.1.4)(a), (10.1.4)(b), and (10.1.4)(c). That is, from frame \mathcal{F}^A , the speed of frame \mathcal{F}^B is measured to be v and the clock in \mathcal{F}^B ticks at a rate that is only σ_v times the rate of all \mathcal{F}^A clocks ($\sigma_v t_A = t_B$) and the length d_A of the moving rod as measured in stationary \mathcal{F}^A is σ_v times the at-rest length d_0 ($d_A = \sigma_v d_0$).

This is no optical illusion. From frame \mathcal{F}^A , the length of the rod in (*moving*) frame \mathcal{F}^B is measured to be $\sigma_v d_0$ which is shorter than its at-rest length d_0 .

- (10.1.9) **NOTE:** The values of times t_A and t_B in Figure (10.1.5) and the analysis that follows are not necessarily equal to the values of t_A and t_B in Figure (10.1.10) and the analysis that will follow.

- (10.1.10) **FIGURE: Line-of-sight Coordinates in \mathcal{F}^A Where Frame \mathcal{F}^B is Fixed, Relative Speed of Frame $\mathcal{F}^A = -v$.**



(10.1.11) **Note:** As in Note (10.1.6), the column vector $\text{col}[d_0, ?]$ in Figure (10.1.10) contains the symbol “?” since the time at the tail of the rod in \mathcal{F}^A (i.e., at $x = d_0$) does not need to be calculated.

(10.1.12) **Calculating Quantities of (10.1.4) When \mathcal{F}^B is Fixed.** These results follow from the vector equations in the left hand panel of Figure (10.1.10). Specifically, equation (10.1.5)(a) takes the form

$$\underbrace{\frac{1}{\sigma_v} \begin{bmatrix} 1 & v \\ v/c^2 & 1 \end{bmatrix}}_{\text{Lorentz matrix } L_{-v}} * \begin{bmatrix} x_B - d_B \\ t_B \end{bmatrix}_A = \begin{bmatrix} 0 \\ t_A \end{bmatrix}_A$$

which gives us the two equations

(10.1.12a)

$$\begin{aligned} x_B - d_B + vt_B &= 0 && \text{which implies} \quad \left\{ \begin{array}{l} (x_B - d_B) = -vt_B \\ -v = (x_B - d_B)/t_B \end{array} \right. \\ \text{(ii)} \quad t_B + \frac{v}{c^2}(x_B - d_B) &= \sigma_v t_A. \end{aligned}$$

Substitute $(x_B - d_B) = -vt_B$ from (10.1.12a) into equation (ii) above to obtain

$$(10.1.12b) \quad t_B \underbrace{\left(1 - \frac{v^2}{c^2}\right)}_{\sigma_v^2} = \sigma_v t_A \quad \text{which implies} \quad \boxed{\sigma_v t_B = t_A}.$$

Finally, Equation (10.1.10)(b) takes form

$$\underbrace{\frac{1}{\sigma_v} \begin{bmatrix} 1 & v \\ v/c^2 & 1 \end{bmatrix}}_{\text{Lorentz matrix } L_{-v}} * \begin{bmatrix} x_B \\ t_B \end{bmatrix}_B = \begin{bmatrix} d_0 \\ * \end{bmatrix}_A$$

which, after equating the top entry of the left and right hand sides, gives us the single equation

$$x_B + vt_B = \sigma_v d_0.$$

Substitute $x_B = d_B - vt_B$ from (10.1.12a) in the equation above to obtain

$$(10.1.12c) \quad \boxed{d_B = \sigma_v d_0}$$

(10.1.13) **With frame \mathcal{F}^B fixed, the boxed equations of (10.1.12a),**

(10.1.12b), and (10.1.12c) answer the respective questions of (10.1.4)(a), (10.1.4)(b), and (10.1.4)(c). That is, the speed of frame \mathcal{F}^A is measured to be $-v$ and the clock in moving frame \mathcal{F}^A ticks at a rate that is only σ_v times the rate of all \mathcal{F}^B clocks ($\sigma_v t_B = t_A$) and the length d_B of the moving rod as measured in stationary \mathcal{F}^B is σ_v times the at-rest length d_0 ($d_B = \sigma_v d_0$).

Again, similar to (10.1.8), this is no optical illusion. From frame \mathcal{F}^B , the length of the rod in (*moving*) frame \mathcal{F}^A is measured to be $\sigma_v d_0$ which is shorter than its at-rest length d_0 .

(10.2) Resolving the Mutual Length-Time Paradoxes

The groundwork has now been laid for the following

(10.2.1)

THEOREM: Mutual Length-Shrinking and Time-Slowing:

- Frame \mathcal{F}^B moves with observed speed v relative to frame \mathcal{F}^A (frame \mathcal{F}^A has speed $-v$ relative to frame \mathcal{F}^B).
- Inertial frames \mathcal{F}^A and \mathcal{F}^B have rods with at-rest length d_0 .
- $\sigma_v = \sqrt{1 - (v/c)^2}$.

Then

(10.2.1a) Observers in each frame measure the ticking rate of moving clocks in the other frame to be $\sigma_v \leq 1$ times the ticking rate of clocks in the observer's frame.

(10.2.1b) Observers in each frame measure the moving ruler in the other frame to have length $\sigma_v d_0 \leq d_0$. It is important to note that only lengths along the direction of motion are affected. Lengths that are perpendicular to the direction of motion remain the same.

PROOF of (10.2.1a)

From (10.1.7b): If an observer jumps into frame \mathcal{F}^A so that frame \mathcal{F}^B is seen to have relative speed v , then the moving clocks in \mathcal{F}^B tick at a rate that is σ_v times the rate of all synchronized stationary

clocks in \mathcal{F}^B .

Frpm (10.1.12b): Conversely, if an observer jumps into frame \mathcal{F}^B instead, so that frame \mathcal{F}^A has relative speed $-v$, then the moving clocks in \mathcal{F}^A tick at a rate that is σ_v times the rate of all synchronized stationary clocks in \mathcal{F}^A .

This proves hypothesis (10.2.1a).

PROOF of (10.2.1b)

From (10.1.7c): If an observer jumps into frame \mathcal{F}^A so that frame \mathcal{F}^B is seen to have relative speed v , then a moving rod with at-rest length d_0 in \mathcal{F}^B is measured from stationary frame \mathcal{F}^A to have length $\sigma_v d_0$.

Frpm (10.1.12c): Conversely, if an observer jumps into frame \mathcal{F}^B so that frame \mathcal{F}^A is seen to have relative speed $-v$, then a moving rod with at-rest length d_0 in \mathcal{F}^A is measured from stationary frame \mathcal{F}^B to have length $\sigma_v d_0$.

This proves hypothesis (10.2.1b) and the theorem is done. ■



(10.3) Summary

- (10.3.1) **Time and Distance Comparison Paradoxes Are Resolved.** The misleading questions of Section (10.1) suggest that there is a single observer measuring the length of a moving rod, or the ticking rate of a moving clock. In this way, the illusion of a contradiction is created. To measure lengths of moving rods, two observers in each (stationary) frame are necessary. Figures (2.4.3), pg. 51, shows that there is no inconsistency. To measure speeds of moving frames (*although one observer may suffice*), Figure (2.4.2) shows that two observers at a known distance between them may also be used to measure the speed of a moving object.



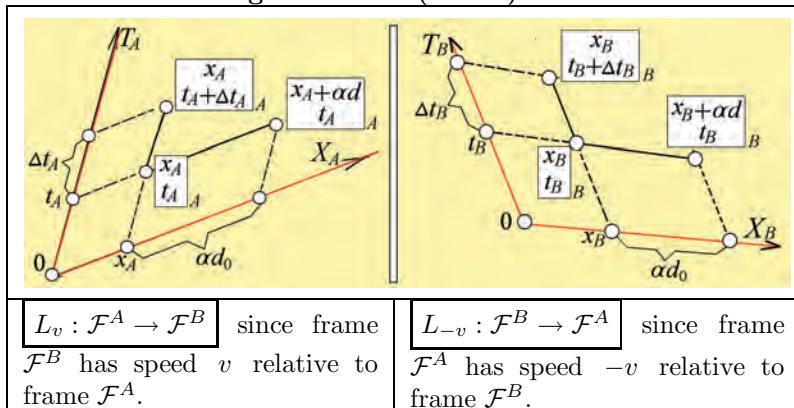
(10.4) Exercises

- (10.4.1) Exercise** **Modified Proof of Theorem (10.2.1).** Theorem (10.2.1) uses three cohort observers to measure the slower ticking rates of moving clocks and the shorter length of moving rods. Prove this theorem using only two cohort observers in each frame.

Hint: In the current proof, there are three cohorts in each frame. The origin observer in the stationary frame (who views the head of the moving rod) is joined by two cohorts who, at a later arbitrary time, view the tail and the head of the moving rod. To reduce the number of cohorts, let the origin observer take on the role of the cohort that views the tail of the moving rod. (The price of reducing the work force is that the later observations of the rod's tail and head can not be made at an arbitrary time.)

- (10.4.2) Exercise** **Mutual Shrinkage of Moving Rods—Slowing of Moving Clocks:** Conditions (10.2.1a) and (10.2.1b) were proved using the physical ruler-clock model of an inertial frame. In this exercise, prove these statements ((10.2.1a) and (10.2.1b)) using the Lorentz transformation and the space time diagrams (10.4.2a) below.

(10.4.2a) Space Time Diagrams for the Mutual Shrinkage Theorem (10.2.1)



Hint: As an example, we show that a rod with at-rest length d_0 in moving frame \mathcal{F}^A is measured in stationary frame \mathcal{F}^B to have length $\sigma_v d_0$. As indicated in the second panel of Figure (10.4.2a), frame \mathcal{F}^B

is stationary and sees frame \mathcal{F}^A moving with speed $-v$. The points of interest are:

$$(10.4.2b) \quad \vec{P} = \begin{bmatrix} x_B \\ t_B \end{bmatrix}_B \quad \text{starting point in } \mathcal{F}^B$$

$$(10.4.2c) \quad \vec{P} + \begin{bmatrix} \alpha d_0 \\ 0 \end{bmatrix}_B = \begin{bmatrix} x_B + \alpha d_0 \\ t_B \end{bmatrix}_B \quad \text{point that is } \alpha d_0 \text{ units from } \vec{P} \text{ at } \mathcal{F}^B \text{ time } t_B$$

$$\vec{P} + \begin{bmatrix} 0 \\ \Delta t_B \end{bmatrix}_B = \begin{bmatrix} x_B \\ t_B + \Delta t_B \end{bmatrix}_B \quad \text{point that is } \Delta t_B \text{ units later than } \vec{P} \text{ in } \mathcal{F}^B.$$

The \mathcal{F}^A line-of-sight coordinates are, respectively, $L_{-v} * \vec{P}$ and $L_{-v} * \left(\vec{P} + \begin{bmatrix} \alpha d_0 \\ 0 \end{bmatrix}_B \right)$ (see (7.1.3)). The vector difference is

(10.4.2d)

$$L_{-v} * \vec{P} - L_{-v} * \left(\vec{P} + \begin{bmatrix} \alpha d_0 \\ 0 \end{bmatrix}_B \right) = L_{-v} * \left(\begin{bmatrix} \alpha d_0 \\ 0 \end{bmatrix}_B \right) = \frac{1}{\sigma_v} \begin{bmatrix} \alpha d_0 \\ \alpha d_0 v / c^2 \end{bmatrix}_A.$$

Interpretation: Now the topmost entry of the final vector of (10.4.2d) is $\alpha d_0 / \sigma_v = d_0$ since d_0 is the at-rest distance in \mathcal{F}^A between the head and tail of the moving rod. Consequently, $\alpha = \sigma_v$. Replacing α with σ_v in Equation (10.4.2c), we see that the distance between the line-of-sight corridors in frame \mathcal{F}^B is $\alpha d_0 = \sigma_v d_0$. That is, the head-to-tail distance of the moving rod in \mathcal{F}^A is measured in stationary \mathcal{F}^B to be only $\sigma_v d_0$.



(11) The Twin Paradox

(11.1) An Overview of the Paradox

A graphical analysis of Figure (11.3.1) completely resolves the twin paradox (*Section (1.11)*) which we restate as follows:

- (11.1.1) • At time 0 years, one twin leaves Earth traveling at an outbound speed of $0.8c$. After 25 years by the Earth clock, the traveling twin instantly boards a returning spaceship which travels in the opposite direction, at an inbound speed of $0.8c$. After 25 more years, or at time 50 years by the Earth clock, the traveling twin returns to Earth to greet his 50 year-old twin while he, himself, is only 30 years old.
- (11.1.2) • **The Apparent Contradiction.** There is a *symmetry*. During the journey, each twin “sees” the (moving) clock of the other twin run slower than his own (stationary) clock (see (1.7.2a)). Thus, the spaceship twin may conclude that the earthbound twin with the slower (moving) earthbound clock will be younger. Likewise, the earthbound twin may conclude that the spaceship twin with the slower (moving) spaceship clock will be younger when they meet.

That each twin is simultaneously younger than the other is clearly not possible: Hence, the paradox.

(11.2) A Simplifying Assumption

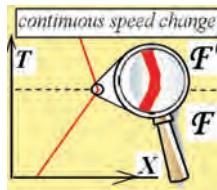
- (11.2.1)  **Special Relativity requires frames to have constant speeds relative to each other.** Hence, when the traveling twin switches from the outgoing spaceship (*at speed v*) to the inbound spaceship (*at speed $-v$*), we assume an instantaneous transfer that requires zero transfer time which we denote $\Delta t_{tfr} = 0$.

However, in (11.5), we apply **Einstein's Theory of General Relativity** to analyze the non-zero time of transfer $\Delta t_{\text{tfr}} > 0$. When $\Delta t_{\text{tfr}} \rightarrow 0$, then we see that *both* special

relativity and general relativity predict that the transferring twin measures a 32-year forward jump into the future. of the Earth clock's readings as he passes from the outbound spaceship to the inbound spaceship. (*Computations for the general relativity confirmation will be found in (11.5).*)

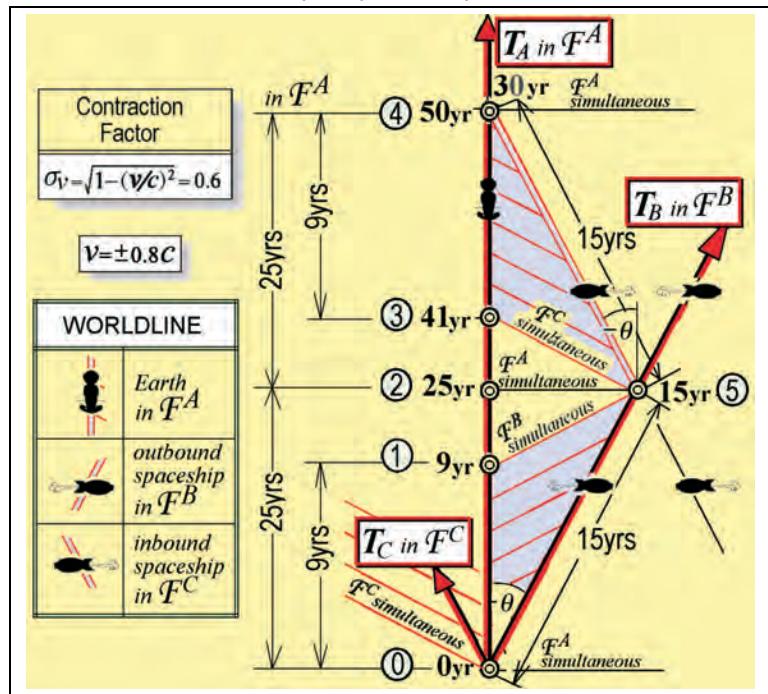
(Note: Instantaneous speed change was also assumed in Figure (1.11.2c) of the twin paradox.)

An instantaneous change of frame and speed approximates the continuous speed change in the diagram.



(11.3) Setup for the Minkowski Diagram

(11.3.1)

**FIGURE:** Three Frames of Reference for the Twin Paradox

(11.4) Resolving the Twin Paradox

(11.4.1)

To fix ideas, we retain the numerical values of Section (1.11) where the relative speed between Earth and the spaceship is $\pm 0.8c$ so that the Lorentz contraction factor (1.8.2) is

$$(11.4.1a) \quad \text{when } c = 1, \quad \sigma_v = \sqrt{1 - (0.8c)^2} = 0.6.$$

With σ_v of (11.4.1a) we have:

(11.4.1b)

Whenever a stationary clock reads time t , and is simultaneously compared (through a line-of-sight corridor) to a synchronized clock moving with speed v that reads time t' , then

$$v = 0.8c \text{ implies } t' = (\sigma_v \cdot t) = 0.6 \cdot t.$$

(11.4.2) **Traveling with the Earthbound Twin in \mathcal{F}^A :*****Worldline of earthbound twin in \mathcal{F}^A :***The vertical T_A axis containing points ①-④.***Simultaneous events in \mathcal{F}^A :***

Along horizontal lines (see Theorem (8.4.1)).

(11.4.2a) *Earth in \mathcal{F}^A Sees Outbound Times in \mathcal{F}^B*

STATIONARY \mathcal{F}^A TIME	OUTBOUND SPACESHIP TIME ($0.6 \times \mathcal{F}^A$ Time)
① $t_A = 0$ yrs.	① $t_B = 0$ yrs. in \mathcal{F}^B (<i>synchronization</i>)
② $t_A = 25$ yrs.	⑤ $t_B = 15$ yrs. in \mathcal{F}^B
② $t_A = 25$ yrs.	⑤ $t_C = 15$ yrs. in \mathcal{F}^C (\mathcal{F}^C <i>clock reset</i>)
④ $t_A = 50$ yrs.	④ $t_C = 30$ yrs. in \mathcal{F}^C

①→①: The Earthbound twin and the spaceship twin at $x = 0$ are synchronized at time $t_A = 0$ in \mathcal{F}^A and $t_B = 0$ in \mathcal{F}^B . At this moment, the spaceship twin accelerates instantly to constant relative speed $v = 0.8c$.

②→⑤: At Earth time $t_A = 25$ yrs. in \mathcal{F}^A , a cohort of the Earthbound twin (*who is also in \mathcal{F}^A*) peers inside the spaceship in \mathcal{F}^B that is traveling at speed $v = 0.8c$ and reads $t_B = 0.6 \times 25 = 15$ yrs. on its onboard slow-running clock (see (11.4.1b)).

②→⑤ (*continued*): An infinitesimal instant after $t_A = 25$ years on Earth in \mathcal{F}^A , the spaceship twin switches spaceships, changing from the outbound frame \mathcal{F}^B to an inbound frame \mathcal{F}^C which has relative speed $v = -0.8c$. The clock aboard the incoming spaceship in \mathcal{F}^C must be re-set to read its new coordinated time, $t_C = t_B = 15$ yrs.

④→④ The clock on the inbound spaceship has been running slow at 0.6 times the rate of the Earth clock. Thus, at Earth time $t_A = 50$ yrs., the spaceship clock reads $t_B = 0.6 \times t_A = 0.6 \times 50 = 30$ yrs. (see (11.4.1b)). *The spaceship twin has returned 20 years younger than the Earthbound twin.*

(11.4.3)

Traveling with the Spaceship Twin in \mathcal{F}^B :***Worldline of outbound twin in \mathcal{F}^B :***

The slanted $\nearrow T_B$ axis which forms angle θ with the vertical axis that also contains the points ① and ⑤.

Simultaneous events in \mathcal{F}^B :

All slanted lines \swarrow parallel to line ①-⑤, that is, all lines with slope $\tan(\theta)$ (see Theorem (8.4.1)).

(11.4.3a) Spaceship in \mathcal{F}^B Sees Earth Clock Times in \mathcal{F}^A

STATIONARY \mathcal{F}^B TIME	MOVING EARTH TIME ($0.6 \times \mathcal{F}^B$ Time)
① $t_B = 0$ yrs. \rightarrow	① $t_A = 0$ yrs. in \mathcal{F}^A (<i>synchronization</i>)
⑤ $t_B = 15$ yrs. \rightarrow	① $t_A = 9$ yrs. in \mathcal{F}^A

$\overbrace{\hspace{15em}}^{15 \text{ yrs. in } \mathcal{F}^B}$ $\overbrace{\hspace{15em}}^{9 \text{ yrs. in } \mathcal{F}^A}$

①→①: The Earthbound twin and the spaceship twin at $x = 0$ are synchronized at time $t_A = 0$ in \mathcal{F}^A and $t_B = 0$ in \mathcal{F}^B . At this moment, the spaceship twin takes off and accelerates instantly so that the Earth's relative (constant) speed is measured to be $v = -0.8 c$.

⑤→①: At spaceship time $t_B = 15$ yrs., the (cohort of the) spaceship twin reads time $t_A = 0.6 \times 15 = 9$ yrs. (11.4.1b) on the slowly running Earth clock in \mathcal{F}^A that has observed speed $v = -0.8 c$.

(11.4.3b) Spaceship Twin Accelerates into a New Frame:

The spaceship twin, traveling at speed $v = 0.8 c$ as measured from Earth in \mathcal{F}^A , decides at time $t_B = 15$ yrs. in frame \mathcal{F}^B , to instantaneously change to an incoming spaceship in frame \mathcal{F}^C that has speed $v = -0.8 c$ relative to Earth.

It is now clear why the frames for each the twins are not equivalent or symmetric. While the Earth twin spends 50 Earth years in only one inertial frame, \mathcal{F}^A , the traveling spaceship twin lives in several frames:

- an inertial outbound frame \mathcal{F}^B ,
- an inertial inbound frame \mathcal{F}^C , and, for an extremely brief time,
- *non-inertial* frames during the “instant” transition from \mathcal{F}^B to \mathcal{F}^C when strong forces eventually change the worldline of the spaceship twin from line ①-⑤ in inertial frame \mathcal{F}^B to the line ⑤-④ in inertial frame \mathcal{F}^C (see Figure (11.3.1)).

As the time of the transition becomes shorter and shorter, these forces felt by the traveling twin become arbitrarily large and the intermediate reference frames deviate more and more from being inertial.

As (11.3.1) indicates, the transition from inertial frame \mathcal{F}^B to inertial frame \mathcal{F}^C severely changes the slope of the lines of simultaneity from positive $v = \tan(\theta)$ to negative $-v = -\tan(\theta)$. All lines of simultaneity in \mathcal{F}^C are parallel to line ⑤-③.

(11.4.3c) A Jump into the Future:

Special relativity: As Figure (11.3.1) confirms, when spaceship time $t_B = 15$ yrs. in frame \mathcal{F}^B , a cohort (*also in frame \mathcal{F}^B*) simultaneously reads the moving Earth in \mathcal{F}^A time as $t_A = 0.6 \times t_B = 0.6 \times 15 = 9$ yrs. (see (11.4.1b)). An infinitesimal moment later, when the traveling twin is in frame \mathcal{F}^C , another cohort (*in frame \mathcal{F}^C*) simultaneously reads the moving Earth time in \mathcal{F}^A as $t_A = 41$ yrs., an advance of 32 years!

General relativity: Section (11.5) has computational details.

(11.4.4) Traveling with the Spaceship Twin in \mathcal{F}^C :

Worldline of inbound twin in \mathcal{F}^C :

The slanted ↘ line that contains the points ⑤ and ④.

Simultaneous events in \mathcal{F}^C :

All slanted lines parallel to line ⑤-③, that is, all lines with slope equal to $-\tan(\theta)$ (see Theorem (8.4.1)).

(11.4.4a) Spaceship in \mathcal{F}^C Sees Earth Clock Times in \mathcal{F}^A

STATIONARY \mathcal{F}^C TIME	MOVING EARTH TIME
⑤ $t_C = 15$ yrs.	→ ③ $t_A = 41$ yrs. in \mathcal{F}^A
④ $t_C = 30$ yrs.	→ ④ $t_A = 50$ yrs. in \mathcal{F}^A
$\overbrace{\quad}$ 15 yrs. pass in \mathcal{F}^C	$\overbrace{\quad}$ 9 yrs. pass in \mathcal{F}^A

⑤→③: The stationary clock in the newly-entered spaceship is reset to read $t_C = 15$ yrs. Simultaneously, the (cohort of the) Earth-bound twin in \mathcal{F}^A sees the moving Earth clock time to be $t_A = 41$ yrs.



(11.4.4b) Why Earth time is $t_A = 41$ yrs. when $t_C = 15$ yrs. Symmetry in Figure (11.3.1) and Table (11.4.2a) tell us that until the spaceship returns to Earth, the spaceship clock will run for 15 more years in \mathcal{F}^C and the Earth clock will run for $0.6 \times 15 = 9$ more years in \mathcal{F}^A . Since the Earth clock will read 50 yrs. at the end of the journey, we calculate backwards 9 years to see that Earth time reads $t_A = 50 - 9 = 41$ years at the start of the returning journey.

④→④: The inbound spaceship journey ends, during which time, as discussed above, 15 Earth years and 9 spaceship years pass. The Earth clocks reads 50 yrs. and the inbound spaceship clock reads 30 yrs.

This concludes the analysis of the Twin Paradox using the tools of special relativity.



(11.5) General Relativity Confirmation

General relativity predicts and quantifies the 32-year jump mentioned in (11.4.3c). We analyze the non-linear portion of the diagram of (11.2.1) where acceleration of the traveling twin disallows the special relativity rules of inertial frames. Here is the overview.

(11.5.1) Calculus One of the central tenets of general relativity (GR) is that within a sufficiently small region, the effects of acceleration are indistinguishable from those produced by an appropriate gravitational force. Forces that are time-independent and relatively weak (*as is the case with the Twin Paradox*), can be described as follows:

For $\vec{r} = [x, y, z]$ in three-dimensional space \mathbf{R}^3 , there is a real-valued function $\Phi(x, y, z)$ called the GRAVITATIONAL POTENTIAL. When this potential function Φ is differentiable (16.3.4), we construct the vector-valued GRADIENT, **grad** of Φ , given by

$$(11.5.1a) \quad \text{grad}(\Phi(\underbrace{x, y, z}_{\vec{r} \in \mathbf{R}^3})) = \left[\frac{\partial \Phi(x, y, z)}{\partial x}, \frac{\partial \Phi(x, y, z)}{\partial y}, \frac{\partial \Phi(x, y, z)}{\partial z} \right].$$

EXAMPLE: If $\Phi(x, y, z) = x^2y + z + \sin(z)$,
then $\text{grad}(\Phi(x, y, z)) = [2xy, x^2, 1 + \cos(z)]$.

EXAMPLE: If point $\vec{r} = x$ lies on the one-dimensional X -axis, then the gravitational potential function $\Phi(x)$ is a function of the single variable x and, according to (11.5.1a),

$$(11.5.1b) \quad \text{grad}(\Phi(x)) = \Phi'(x), \text{ the derivative of } \Phi \text{ at point } x.$$

.....

The gravitational force, on a mass m at point \vec{r} is the three-dimensional vector

$$(11.5.1c) \quad \vec{F}(\vec{r}) = -\text{grad}(\Phi(\vec{r})) \cdot m.$$

Finally, Newton's classic relation between force and acceleration on mass m at a point \vec{r} is given by

$$(11.5.1d) \quad \vec{F}(\vec{r}) = m \cdot \vec{a}(\vec{r}).$$

(11.5.2) **Calculus** **Another result from GR** is that clocks slow down in a region where gravitational effects are strongest. (*Equivalently, clocks speed up as they pass to weaker regions of gravitational effects. It is this interpretation that is used in (24.5.3), pg. 347, in the discussion of clocks aboard GPS satellites*).

Quantitatively, if an observer at position \vec{r}_1 where the gravitational potential is $\Phi(\vec{r}_1)$ finds that an event occurs over a time interval $\Delta t(\vec{r}_1)$, and another observer at a different position \vec{r}_2 where the gravitational potential is $\Phi(\vec{r}_2)$ finds the same event occurs over a time interval $\Delta t(\vec{r}_2)$, then

$$(11.5.2a) \quad \Delta t(\vec{r}_1) = \left(1 + \frac{\Phi(\vec{r}_2) - \Phi(\vec{r}_1)}{c^2}\right) \Delta t(\vec{r}_2).$$

At point \vec{r} , a final relation between force $\vec{F}(\vec{r})$ on a mass m due to acceleration $\vec{a}(\vec{r})$ is given by Newton's classic formula (11.5.1d).

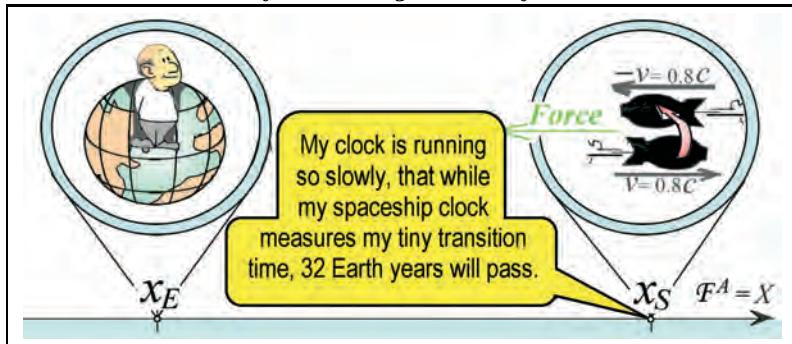
|| **Note:** Although these equations can be derived in a straightforward manner, doing so would take us too far afield.

The time has come to see how these equations justify the 32-year

jump (11.4.3c) in the Twin Paradox.

(11.5.3)
Calculus

FIGURE: Transition of Traveling Twin After 25 Earth Years



The one-dimensional X -axis contains Earth, considered as a point at $x = x_E$, the straight-line paths of the outbound and inbound spaceships and the jump point $x = x_S$ where the traveling twin changes from outbound Spaceship B to inbound Spaceship C. As the twin jumps from one spaceship to the other at $x = x_S$, the change in time on the spaceship clocks is $\Delta t(x_S)$ while the time change on Earth clocks is $\Delta t(x_E)$. (These time changes are related in (11.5.2a).) See Figure (1.3.2), pg. 13.

(11.5.4)
Calculus

NOTATION: As shown in Figure (11.5.3), all vectors belong to the 1-dimensional X -axis. Hence, we eliminate vector notation for force, velocity, acceleration, and position. Hence, we write v instead of \vec{v} for velocity vectors, a instead of \vec{a} for acceleration, and we write x instead of \vec{r} for position vectors.

Our goal is to find $\Delta t(x_E)$ the change in Earth time during the small, almost instantaneous, time interval $\Delta t(x_S) \rightarrow 0$ of the jump of the traveling twin from outbound Spaceship B to inbound Spaceship C.

Suppose the jump from the outbound to the inbound spaceship occurs at an arbitrary point x over spaceship time $\Delta t(x)$. During this (short) transfer time, the speed of the traveling twin changes from constant $+v$ (*outbound*) to constant $-v$ (*inbound*), which is a total change of $\Delta v = -2v$. Since the *acceleration* a is the change in velocity per unit time, we have

$$\begin{aligned}
 (11.5.4a) \quad -2v/\Delta t(x) &= a & v \text{ constant} \Rightarrow a \text{ constant} \\
 &= F(x)/m & \text{solve for } a \text{ in (11.5.1d)} \\
 &= -\mathbf{grad}(\Phi(x)) & \text{solve for } F/m \text{ in (11.5.1c)} \\
 &= -\Phi'(x). & \text{from (11.5.1b)}
 \end{aligned}$$

Since $-2v/\Delta t(x) = a$ is constant for all points x , we set the jump point in the first term of (11.5.4a) at $x = x_S$. We then compute the definite integral of the (equal) first and last terms of Equation (11.5.4a), taking the limits from $x = x_E$ to $x = x_S$, to obtain

$$2 \int_{x_E}^{x_S} (v/\Delta t(x_S)) dx = \int_{x_E}^{x_S} \Phi'(x) dx,$$

which computes to be

$$(11.5.4b) \quad 2(v/\Delta t(x_S))(x_S - x_E) = \Phi(x_S) - \Phi(x_E).$$

Finally, in Equation (11.5.2a), set $\vec{r}_1 = x_E$ and $\vec{r}_2 = x_S$. Now replace the term $\Phi(x_S) - \Phi(x_E)$ of (11.5.2a) with the left hand side of (11.5.4b) above so that (11.5.2a) takes the form

$$\Delta t(x_E) = \left(1 + \frac{2v(x_S - x_E)}{\Delta t(x_S) c^2} \right) \Delta t(x_S).$$

Since the 25-year journey of the spaceship at speed $v = 0.8c$ covers a distance $(x_S - x_E) = v \cdot 25$ years, the equation above takes the form

(11.5.5)
Calculus

$$\begin{aligned}
 \Delta t(x_E) &= \Delta t(x_S) + 50 \left(\frac{v}{c} \right)^2 \text{ years} \\
 &= \Delta t(x_S) + 32 \text{ years} & \text{since } v = 0.8c \\
 &\longrightarrow 32 \text{ years, as } \Delta t(x_S) \longrightarrow 0.
 \end{aligned}$$



(11.6) Exercises

(11.6.1) Exercise

Actual clock time of inbound spaceship. We know that the outbound spaceship in \mathcal{F}^B traveling at constant speed $v = 0.8c$ relative to Earth will have an onboard clock time of 15 years when it arrives at point ⑤ of Figure (11.3.1). At this moment, the Earth clock (simultaneously) reads 25 years.

Question: What is the clock time on board the inbound spaceship in \mathcal{F}^C when it arrives at the same point ⑤? (*This is the clock that will be reset to read “15 yrs.” by the twin who transfers from the outbound spaceship.*)

Hint:

Background. The clocks of origin observers at $x = 0, t = 0$ for the three frames, \mathcal{F}^A (Earth), \mathcal{F}^B (outbound spaceship), and \mathcal{F}^C (inbound spaceship), are synchronized as per Definition (3.4.5a). The earthbound twin and the departing outbound spaceship are Origin Observers, so their clocks are synchronized. But the inbound spaceship is a cohort, not the Origin Observer in \mathcal{F}^A .

Analysis. At point ⑤, the \mathcal{F}^A $[x, t]$ coordinates are $[v \times 25\text{yrs.}, 25\text{yrs.}]$. Use the Lorentz transformation, L_{-v} , to link these coordinates to $[x'', t'']$, the distance-time coordinates in \mathcal{F}^C . That is,

$$[v \times 25, 25] \stackrel{L_{-v}}{\equiv} [x'', t'']$$

in which case, $t'' = (1 + v^2)^{-1} 25 / \sigma_v$.

(11.6.2) Exercise

Generalizing Exercise (11.6.1). In Exercise (11.6.1), replace speed $0.8c$ with an arbitrary speed $v_0 < c$. Let the two spaceships meet when the earthbound twin is t_0 years old. Then, in Figure (11.3.1), the simultaneous line-of-sight times at point ⑤ for the three frames are:

In \mathcal{F}^A on Earth, $t = t_0$.

In \mathcal{F}^B on the outbound spaceship, $t = \sigma_v t_0$.

In \mathcal{F}^C on the inbound spaceship, $t = (1 + v_0^2)^{-1} t_0 / \sigma_v$.

- (11.6.3) Exercise** **Congruent Symmetric Triangles.** Show that in Figure (11.3.1), triangles ⑤-③-④ and ⑤-①-② are congruent.

Hint:

- (1) As Figure (11.3.1) indicates, $\angle \textcircled{1}-\textcircled{2}-\textcircled{5} = \angle \textcircled{3}-\textcircled{4}-\textcircled{5} = \theta$.
 - (2) From Theorem (8.4.1), lines of simultaneity, line ⑤-③, and line ⑤-① form angle θ with the horizontal. Hence, $\angle \textcircled{1}-\textcircled{5}-\textcircled{2} = \angle \textcircled{3}-\textcircled{5}-\textcircled{4} = (90^\circ - 2\theta)$.
 - (3) From the Angle-Side-Angle theorem, $\triangle \textcircled{1}-\textcircled{5}-\textcircled{2}$ and $\triangle \textcircled{3}-\textcircled{5}-\textcircled{4}$ are congruent to each other since they share a common side length ($\text{length}[\textcircled{5}-\textcircled{4}] = \text{length}[\textcircled{5}-\textcircled{2}] = 15$), and two common angles as described in (2) above.
 - (4) Congruency tells us that $\text{length}(\textcircled{2}-\textcircled{1}) = \text{length}(\textcircled{4}-\textcircled{3}) = 9$. If ③ is 9 units from the 50-year mark ④ in \mathcal{F}^A time, then ③ must be at the 41-year mark in \mathcal{F}^A .
-

- (11.6.4) Exercise** **Slow-Fast-Slow Analysis.** Assume at age 0, one twin instantly travels into space at constant speed v and after t_0 years on the outbound leg, as measured by Earth's clock, travels the inbound leg for t_0 more years before returning to Earth. Generalize Figure (11.3.1) to show that for any speed $v = \tan \theta$, the traveling twin, after a round trip requiring $2 \cdot t_0$ years according to an Earth clock, will always return at age $2 \cdot \sigma_v \times t_0$ years. (see (1.8.2) for the definition of σ_v .)

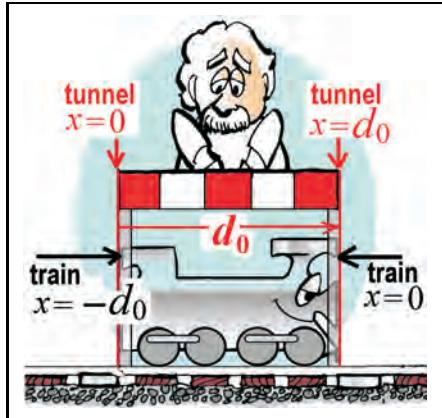
Hint: Analytically, use the Lorentz transformation to link line-of-sight times at points ① through ⑤ between appropriate pairs of frames. Alternately, use trigonometry on the triangles in the diagram to obtain lengths that are interpreted as times on the clocks of frames \mathcal{F}^A , \mathcal{F}^B , and \mathcal{F}^C .

(12) The Train-Tunnel Paradox

(12.1) An Overview of the Paradox

(12.1.1) **FIGURE:** Static Train and Tunnel are of Equal Length.

- The common at-rest length of the tunnel in frame \mathcal{F}^A and the train in frame \mathcal{F}^B is d_0 .
- In \mathcal{F}^A , the origin $x = 0$ is at the left entrance of the tunnel — $x = d_0$ marks the right entrance.
- In \mathcal{F}^B , $x = 0$ is at the right end (the front) of the train — $x = -d_0$ marks the left end (the rear) of the train.

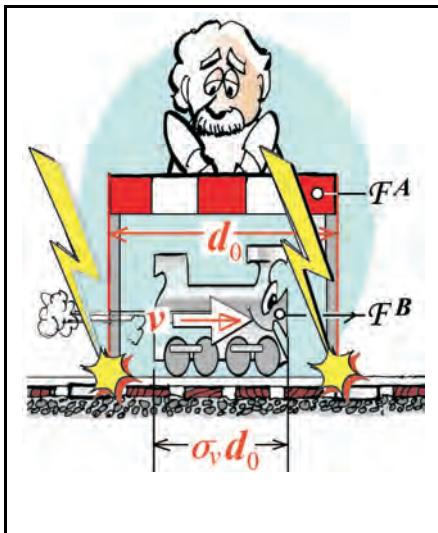


We will use the fact that objects moving with relative speed $\pm v$ shrink in the direction of motion by a factor of σ_v (1.8.2), pg. 20.

(12.1.2) **FIGURE:** Train in Motion Shrinks, Tunnel is Stationary.

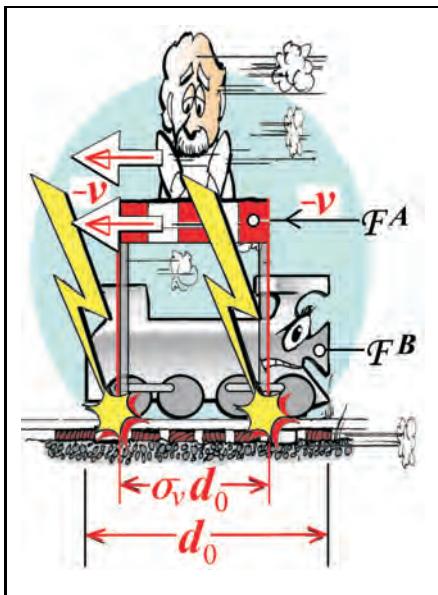
- The frame \mathcal{F}^B of the train has speed v relative to the stationary frame \mathcal{F}^A of the tunnel.

As measured in the tunnel's frame \mathcal{F}^A , the train has reduced length $\sigma_v d_0$. When the (short) train is midway through the tunnel, two bolts of lightning simultaneously strike the ends of the tunnel. Since the train is well inside the tunnel, as observed from \mathcal{F}^A , the train *is spared* any damage from the simultaneous lightning strikes.

(12.1.3) **FIGURE:** Tunnel in Motion Shrinks, Train is Stationary.

- The frame \mathcal{F}^A of the tunnel has speed $-v$ relative to the stationary frame \mathcal{F}^B of train.

As measured in the train's frame \mathcal{F}^B , the tunnel has reduced length $\sigma_v d_0$. When the midpoints of the moving tunnel and stationary train are aligned (overlap), the short tunnel leaves the ends of the train exposed outside the tunnel entrances. If the two bolts of lightning simultaneously strike the ends of the tunnel, then the (too-long) train *is not spared* electrocution by the simultaneous lightning strikes.



(12.1.4)

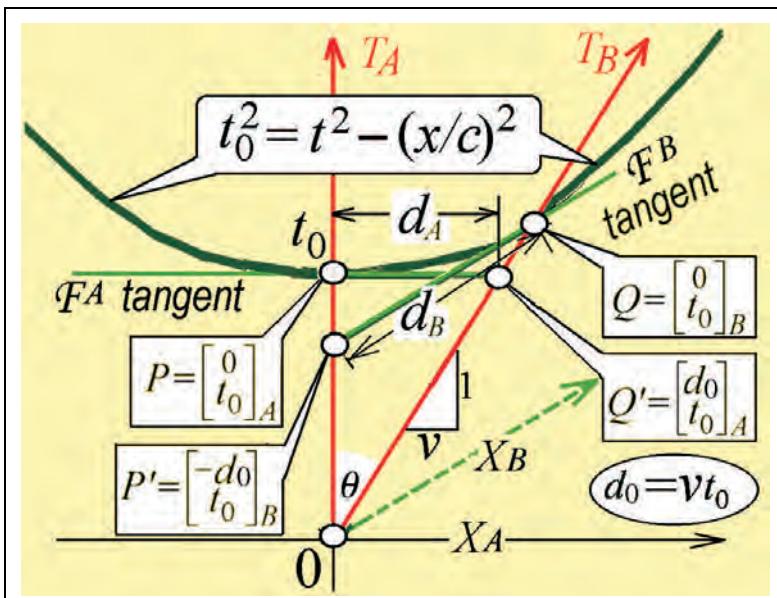
The Paradox: The concluding statements of (12.1.2) and (12.1.3) contradict each other. Is the train electrocuted or is it not?

The Resolution: The lightning strikes that are simultaneously viewed from the tunnel's frame \mathcal{F}^A (12.1.2) are not simultaneous when viewed from the train's frame \mathcal{F}^B . In the final sentence of (12.1.3), there is an invalid use of the adjective "simultaneous."

(12.2) A Distance Lemma

- (12.2.1)  The Minkowski diagram (12.2.2) following shows two lengths d_A and d_B which are, apparently, unequal — the horizontal line \overline{PQ} (*measured in \mathcal{F}^A with length d_A*) seems to be shorter the slanted line $\overline{P'Q}$ (*measured in \mathcal{F}^B with length d_B*). Lemma (12.2.2) shows us that in spite of their apparent inequality on the graph, the actual at-rest distances (*or lengths*) measured within their respective frames are, in fact, equal to each other, that is, $d_A = d_B$.

- (12.2.2) **THEOREM:** *Equal Distances d_0 for Frames \mathcal{F}^A and \mathcal{F}^B*



Given: The Minkowski diagram for the inertial frame \mathcal{F}^A has orthogonal spacetime axes X_A and T_A . The synchronized frame \mathcal{F}^B has speed v relative to \mathcal{F}^A so its time axis T_B has slope $1/v$ and forms an angle θ with the vertical T_A -axis where $\tan(\theta) = v$.

Choose arbitrary time $t_0 > 0$ which then defines distance $d_0 = vt_0$.

Construct

(12.2.2a) The hyperbola $t_0^2 = t^2 - (x/c)^2$.

(12.2.2b) Points P and Q on the hyperbola (12.2.2a) at the intersections of the T_A -axis and the T_B -axis of \mathcal{F}^A and \mathcal{F}^B , respectively.

(12.2.2c) Tangent lines Tangent(P) and Tangent(Q) to the hyperbola (12.2.2a) at the respective points P and Q .

(12.2.2d) Point P' at the intersection of line Tangent(P) and the T_A -axis,

and

Point Q' at the intersection of line Tangent(Q) and the T_B -axis.

THEN

(12.2.2e) $\|\overline{PQ'}\|_A = d_0$ is the length (A.2.18) of the line segment $\overline{PQ'}$ as measured in \mathcal{F}^A coordinates,

(12.2.2f) $\|\overline{P'Q}\|_B = d_0$ is the length (A.2.18) of the line segment $\overline{P'Q}$ as measured in \mathcal{F}^B coordinates,

(12.2.2g) The X_A -axis and the X_B -axis in the Minkowski diagram (12.2.2) are parallel to the line segments $\overline{PQ'}$ and $\overline{P'Q}$, respectively.

PROOF:

The importance of hyperbola (12.2.2a). The Time-Stamped Origin Theorem (8.2.3) tells us that all the points on the hyperbola $[x, t]_A$ in \mathcal{F}^A (where $t_0^2 = t^2 - (x/c)^2$) lie at the \mathcal{F}^A end of a line-of-sight corridor. At the opposite end of each of these line-of-sight corridors lies the time-stamped origin, $[0, t_0]_v$ of frame \mathcal{F}^v , which has relative speed v with respect to \mathcal{F}^A where $-\infty < v < \infty$.

Coordinates of points P and Q (12.2.2b). In our case, the two points P and Q (12.2.2b) whose T -axes intersect the hyperbola

of origins (12.2.2a) have home frame coordinates

$$(12.2.2h) \quad P = \begin{bmatrix} 0 \\ t_0 \end{bmatrix}_A \text{ in } \mathcal{F}^A \quad \text{and} \quad Q = \begin{bmatrix} 0 \\ t_0 \end{bmatrix}_B \text{ in } \mathcal{F}^B.$$

The importance of time t_0 . Synchronization (3.4.5) of frames \mathcal{F}^A and \mathcal{F}^B tells us that at time $t = 0$ in each frame, their respective origins overlap—or are zero distance apart—in the sense that $[0, 0]_A$ and $[0, 0]_B$ are at opposite ends of a line-of-sight corridor (see figure of Theorem (12.2.2)).

The symmetry principle (1.6.1) says that as time increases, the origins speed away from each other at the same rate in each frame—at speed $\pm v$.

(12.2.2i) Hence, at common time t_0 in frames \mathcal{F}^A and \mathcal{F}^B , the origins (i.e., the T_A and T_B axes on the Minkowski diagram) are separated by the usual Euclidean distance $d_0 = vt_0$.

Therefore, coordinates of points P' and Q' (12.2.2b) have the form

$$(12.2.2j) \quad P' = \begin{bmatrix} -d_0 \\ t_0 \end{bmatrix}_A \text{ in } \mathcal{F}^A \quad \text{and} \quad Q' = \begin{bmatrix} d_0 \\ t_0 \end{bmatrix}_B \text{ in } \mathcal{F}^B.$$

Using the definition (A.2.18a), pg. 366, for the usual vector lengths in frames \mathcal{F}^A and \mathcal{F}^B respectively, we calculate the Euclidean distance between the \mathcal{F}^A origin and the \mathcal{F}^B origin at time t_0 in \mathcal{F}^A as follows:

$$\begin{aligned} \|\overline{PQ'}\| &= \|P - Q'\| = \left\| \begin{bmatrix} 0 \\ t_0 \end{bmatrix}_A - \begin{bmatrix} d_0 \\ t_0 \end{bmatrix}_A \right\| \\ &= \sqrt{(0 - d_0)^2 + (t_0 - t_0)^2} = d_0. \end{aligned}$$

Similarly, the calculation of the distance between the \mathcal{F}^A origin and the \mathcal{F}^B origin as observed at time t_0 in \mathcal{F}^B is

$$\begin{aligned} \|\overline{P'Q}\| &= \|P' - Q\| = \left\| \begin{bmatrix} -d_0 \\ t_0 \end{bmatrix}_B - \begin{bmatrix} 0 \\ t_0 \end{bmatrix}_B \right\| \\ &= \sqrt{(-d_0)^2 + (t_0 - t_0)^2} = d_0. \end{aligned}$$

The previous two strings of equalities prove (12.2.2e) and (12.2.2f).

Finally, (12.2.2g) holds true since, according to (4.4.1a), pg. 73, (4.4.3), pg. 73, all lines of events with fixed time t_0 —i.e., lines of

simultaneous events—are, in fact, (a subset of) their respective X axes at time t_0 . The proof is done. ■

(12.3) The Train-Tunnel Minkowski Diagram

(12.3.1) The single Minkowski diagram that illustrates the relative motion of Figures (12.1.2) and (12.1.3), will be based on the Minkowski diagram of Figure (12.2.2), pg. 155 (*repeated on the right*).

This diagram shows the worldlines of a pair of origins at relative speed $\pm v$ after they separate from their line-of-sight overlap (synchronization) at a common coordinate $[0, 0]$. At time t_0 in either frame, the separation of the origins is measured to be $d_0 = v t_0$. This allows us to identify the horizontal line segment $\overline{PQ'}$ (of length d_0) with the tunnel and the slanted line segment $\overline{P'Q}$ (of length d_0) with the train.

(12.3.2) **Note: Why Are There Two d_0 Lengths in Figure (12.2.2)?**

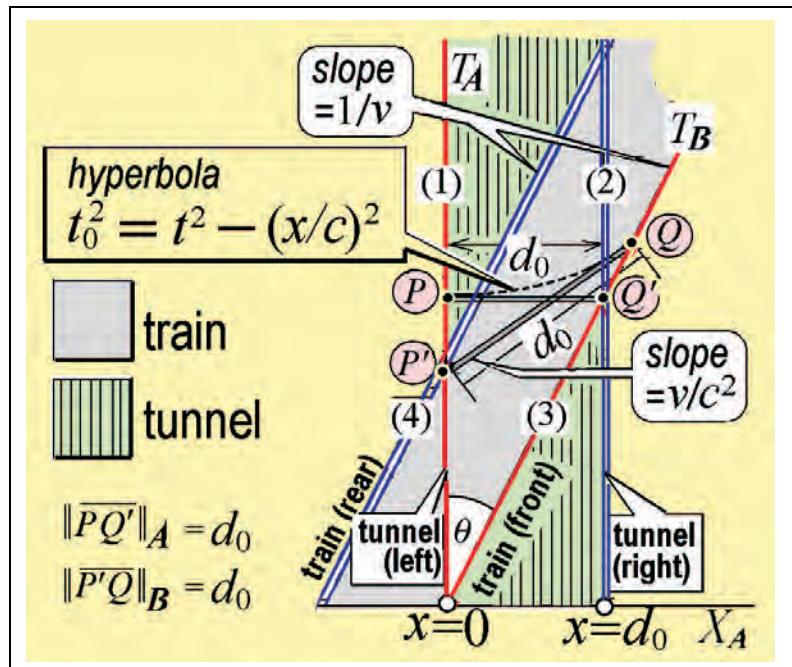
Similar to the figure of (8.2.3), pg. 120, where the scale of the *time* axes change from frame to frame, in Figure (12.3.1), the scale of *length* d_0 changes from frame \mathcal{F}^A to \mathcal{F}^B . This is why the slanted d_0 in \mathcal{F}^B appears longer than the horizontal d_0 in \mathcal{F}^A .

The Minkowski diagram (12.3.3) following is based on Figure (12.3.1). For example, the worldlines of the two origins from Figure (12.3.1)—the left entrance to the tunnel in \mathcal{F}^A and the right front of the train in \mathcal{F}^B —are transferred to Figure (12.3.3) as the vertical line (1) and the slanted line (3). The construction of Figure (12.3.3) continues as follows:

For each of the infinitely many points $0 < x \leq d_0$ of the tunnel floor in \mathcal{F}^A , add its vertical worldline, creating a vertical (green) band of all possible spacetime coordinates for the tunnel. The left and right boundaries of this vertical band are indicated by (1) and (2), respectively.

For the infinitely many points $-d_0 \leq x < 0$ of the train in \mathcal{F}^B , add its vertical worldline, creating a slanted (gray) band of all possible spacetime coordinates for the train. The left and right boundaries of this slanted band are indicated by (4) and (3), respectively.

(12.3.3) **FIGURE:** The Minkowski Diagram for the Train-in-Tunnel Paradox:



NOTE: Consistent with Theorem (8.4.1), the horizontal lines of simultaneity in frame \mathcal{F}^A also have slope of the form (v'/c^2) . In our case, the relative speed v' of \mathcal{F}^A with respect to itself is zero.

(12.4) Explaining Mutual Contraction

The Train-in-Tunnel Paradox addresses an apparent *time anomaly* for the pair of lightning strikes. There is also a question of an apparent *distance anomaly*, namely, when the train is shorter than the tunnel (Figure (12.1.2)), and, at the same time, the tunnel is shorter than the train (12.1.3).

(12.4.1) **Question:** How can both comparisons be true at the same time?

(12.4.2) Figure (12.3.3) shows:



Simultaneity in the tunnel's frame \mathcal{F}^A is always along a single line segment like \overline{PQ} that is horizontal in Figure (12.3.3). Hence, from frame \mathcal{F}^A , the train in \mathcal{F}^B must be viewed along a horizontal line that intersects the train's slanted band of spacetime coordinates. Notice that the part of the horizontal line \overline{PQ} that intersects the slanted band (*between worldlines (4) and (3)*) is shorter than the width of the tunnel (*horizontal distance between worldlines (1) and (2)*). That is,

- from the frame of the tunnel \mathcal{F}^A , the moving train in \mathcal{F}^B is shorter than the width of the tunnel.

Simultaneity in the train's frame \mathcal{F}^B is always along a single line segment like \overline{PQ} that has slope (v/c^2) in Figure (12.3.3). Hence, the tunnel (*which is in \mathcal{F}^A*) must be viewed along the slanted line $\overline{P'Q}$ which is the length of the train in its home frame \mathcal{F}^B . From this point of view (*from \mathcal{F}^B*) the train extends beyond the green band—the train is longer than the tunnel.

- from the frame of the train \mathcal{F}^B , the moving tunnel in \mathcal{F}^A is shorter than the length of the train.

(12.5) Resolving the Train-Tunnel Paradox

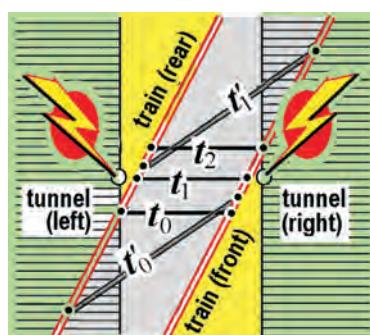
(12.5.1)

FIGURE: Two Simultaneous Lighting Strikes at Time t_1 in \mathcal{F}^A



The vertical shaded band between the labels, **tunnel (left)** and **tunnel (right)**, represents all points of the tunnel in frame \mathcal{F}^A at all times.

The slanted shaded band between the labels, **train (rear)** and **train (front)**, represents all points of the train in frame \mathcal{F}^B at all times.



Simultaneity in \mathcal{F}^A of the stationary tunnel is represented by events on any single horizontal line. Three such lines in \mathcal{F}^A are shown for times $t_0 < t_1 < t_2$. The **simultaneous** lightning strikes occur at

\mathcal{F}^A time t_1 .

This agrees with Figure (12.1.2).

Simultaneity in \mathcal{F}^B of the stationary train is represented by events on any single slanted line with slope (v/c^2) . Two such lines in \mathcal{F}^B are shown for times $t'_0 < t'_1$. The two **sequential** lightning strikes (*that occur simultaneously at time t_1 in \mathcal{F}^A*) occur in \mathcal{F}^B first at time t'_0 then later at time t'_1 .

This does not agree with Figure (12.1.3).

which falsely shows the two lightning strikes occurring at the same time in the train's frame \mathcal{F}^B .

(12.5.2) In Figure (12.1.3), the lightning strikes are **sequential**. As seen in Figure (12.5.1):

- At time t_0' in \mathcal{F}^B , all points of the train lie on the slanted t_0' line. The front end of the train has entered the tunnel while its rear end is still outside the tunnel. Straight ahead on the slanted t_0' line, the first lightning strike occurs at the tunnel's right entrance.
- Later, at time t_1' in \mathcal{F}^B , all points of the train lie on the slanted t_1' line. The front of the train has exited the right side of the tunnel while its rear end has passed by the left entrance of the tunnel. Behind the train on the slanted t_1' line, the second lightning strike occurs at the tunnel's left entrance.

(12.6) Exercises

(12.6.1)

Lin Alg

The Moving Train Shrinks by a Factor of σ_v : In Figure (12.3.3), show that in \mathcal{F}^A coordinates, the length of the intersection of horizontal line segment $\overline{PQ'}$ with the slanted band is $\sigma_v d_0$.

Hint: The \mathcal{F}^A coordinates for $\overline{PQ'}$ have the form $[x, t_0]_A$, $0 \leq x \leq d_0$. The \mathcal{F}^B coordinates for worldline (4) (the left entrance of the tunnel) have the form $[-d_0, t]_B$ where $-\infty < t < \infty$.

Use the Lorentz transform $L_v : \mathcal{F}^A \rightarrow \mathcal{F}^B$ (7.1.3) to find the intersection point $[\hat{x}, t_0]_A$ of $\overline{PQ'}$ and worldline (4) and show that $\hat{x} = (1 - \sigma_v)d_0$, which shows that the intersection of $\overline{PQ'}$ with the slanted band of points on the train has length $\sigma_v d_0$.

The calculations proceed as follows:

The equation $L_v([\hat{x}, t_0]_A) = [-d_0, t]_B$ produces two equations. Eliminate the variable t and solve for $\hat{x} = (1 - \sigma_v)d_0$.

(12.6.2)
Lin Alg

The Moving Tunnel Shrinks by a Factor of σ_v : In Figure (12.3.3), show that in \mathcal{F}^B coordinates, the length of the intersection of slanted $\overline{P'Q}$ with the vertical band is $\sigma_v d_0$.

Hint: See the hint in Exercise (12.6.1).

(12.6.3)
Lin Alg

A Moving Origin Clock on Worldline (3) Slows by a Factor of σ_v : Show that in Figure (12.3.3), the origin clock (3.4.3) whose worldline coincides with worldline (3), moving at speed v relative to frame \mathcal{F}^A slows down by a factor of σ_v .

Hint: Let us use the notation **Clock_3** to denote the origin clock of \mathcal{F}^B whose coordinates coincide with worldline (3). Then the origin clocks of \mathcal{F}^A and \mathcal{F}^B (**Clock_3**) are synchronized at common time $t = 0$. At point $Q' = [d_0, t_0]_A$, clocks in \mathcal{F}^A read time $t = t_0$. Use the Lorentz transformation to read the origin clock **Clock_3** at the other end — the \mathcal{F}^B end — of the line-of-sight corridor. That reading will be $\sigma_v t_0$.

(12.6.4)
Lin Alg

A Moving Origin Clock on Worldline (1) Slows by a Factor of σ_v : Show that in Figure (12.3.3), the origin clock (3.4.3) whose worldline coincides with worldline (1), moving at speed v relative to frame \mathcal{F}^B slows down by a factor of σ_v .

Hint: See the hint in Exercise (12.6.3).

(13) The Pea-Shooter Paradox

(13.1) An Overview of the Paradox

Restatement of the pea-shooter paradox (1.13): You observe an airplane traveling away from you at speed w while the plane launches a spaceship (*in the same direction of the plane*) whose instantaneous speed, as measured by the pilot, is v . Your intuition and experience should tell you that you would measure the spaceship speed to be $v + w$. In fact, the actual spaceship speed is the theoretically accurate Lorentz sum, $v \oplus w$ (7.3.1), which, due to the low speeds involved, is practically indistinguishable from the usual arithmetic sum $v + w$ for speeds v and w .

Our powers of perception are inadequate to distinguish between Lorentzian addition $v \oplus w$ (*which is valid for all speeds*) and the imperceptibly larger quantity $v + w$ we usually experience at *small speeds*. When the measured sum of v plus w for *high* values of v and w disagrees with the familiar formula $v + w$ used for *low* values of v and w , we tend to throw up our hands and announce a *paradox*.

(13.1.1) A pea in frame \mathcal{F}^C has speed $v = 0.8c$ as measured in frame \mathcal{F}^B by the pea-shooter, Ashley. Ashley (and her platform) have speed $w = 0.9c$ as measured in frame \mathcal{F}^A by Bernie. Yet, Bernie does not see the pea moving with the speed we would expect from our experience.

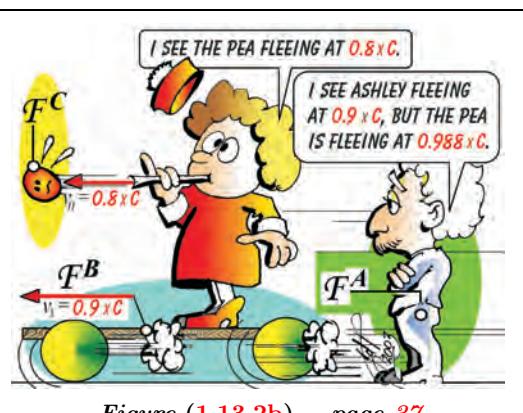


Figure (1.13.2b), page 37

Galilean/Newtonian addition of speeds tells us that

$$(13.1.1a) \quad v + w = 0.8c + 0.9c = 1.7c. \quad \text{classical Galilean addition}$$

Formula (13.1.1a) was justified in (5.5.2) using matrices, and in (5.6.3), using triangles whose areas correspond to speeds.

- (13.1.2) **Pea-Shooter Paradox Already Resolved.** Due to the relativistic constraint that no material object can have a speed greater than the speed of light, $c = 1$, a rule different from (13.1.1a) must apply — this general rule turns out to be the Lorentzian addition

$$(13.1.2a) \quad v \oplus w = \frac{v + w}{1 + vw}, \quad \{-1 < v, w < 1 \text{ implies } v \oplus w < 1\}$$

that has *already* been defined in (7.3.1). Formula (13.1.2a) for relativistic addition was justified in (7.4.1) using matrices.

Similar to the Galilean case (5.6.3), we established the rule for Lorentzian addition in (13.1.2a) using triangles whose areas are closely related to relative speeds.

(13.2) The Fizeau Experiment: Adding Speeds

The experiment of Fizeau offered early verification of (13.1.2a), the relativistic formula for the addition of speeds. Here are the details:

The Democratization of Light

- (13.2.1) **Wave Speed Is Not Affected by Speed of the Wave Source.** Suppose a transmitting medium (e.g., air) is stationary relative to an observer — there is no wind. Then any wave (e.g., the sound of a train whistle) that is transmitted in that medium will have the *same observed speed* at all points along the line of motion even if its source (e.g., the train) is in motion relative to the observer. This is seen from the “skipping stone” model of Figure (16.5.4). Although the source speed does not affect the wave speed, it does affect the measured frequency, the *Doppler effect* that produces a higher pitch for an approaching train and a lower pitch when the train recedes (see Section (16.6)).

- (13.2.2) **Wave Speed Is Affected by Speed of the Medium.** As seen in (16.5.5), the observed speed of the transmitting medium (e.g., wind speed) *is* communicated to the speed of the transmitted wave (e.g., sound of a train whistle).

(13.2.2a) **Since light needs no transmitting medium,** the speed of light in a vacuum is a constant c for all observers, whether they are in motion or not (1.5.2b).

(13.2.2b) **However, when light is transmitted through a medium,** it is subject to the physics of (13.2.2) (*this is the democratization of light*), in which case, the speed of light is reduced to

$$(13.2.2c) \quad v = \frac{c}{n} \quad \left\{ \begin{array}{l} n \text{ (unitless)} \\ \hline \text{Air} & 1.0002926 \\ \text{Diamond} & 2.417 \\ \text{Water} & 1.33 \end{array} \right.$$

where $n > 1$ is the unitless REFRACTIVE INDEX of the medium.

- (13.2.3) **The Refractive Indices Are Approximate** since their values depend on the frequency of the transmitted light and the temperature of the transmitting medium. Consequently, the wave speed $v = c/n$ (13.2.2c) relative to the *stationary* medium is approximately equal to $c/1.33 \approx 140,000$ miles/second.

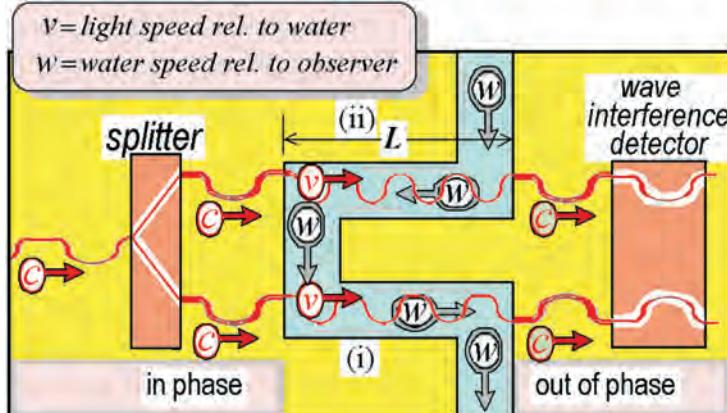
Due to (16.5.5), pg. 211, the observed speed of light in *moving* water is greater than (less than) $c/1.33$ if the water flow is in the same (opposite) direction of the light.

The Setup for Fizeau's Experiment

- (13.2.4) **To describe how Fizeau designed his experiment** for measuring the pea-shooter effect, or the addition of speeds (1.13.2), we must define the reference frames for the observer, the water, and the impeded light beam traveling in water.
- (13.2.5)
- \mathcal{F}^A is the frame of the *observer* (experimenter), which is analogous to Bernie's frame \mathcal{F}^A in the figure of (13.1.1).
 - \mathcal{F}^B is the frame of the transporting *water* which has speed w

relative to frame \mathcal{F}^A . This is analogous to frame \mathcal{F}^B in (13.1.1), the frame of the transporting wagon. From (3.4.2), the water has zero speed relative to its own frame \mathcal{F}^B .

(13.2.6) *How Fizeau Measures Addition of Speeds*



A single light beam enters a splitter and exits as two identical beams that are directed to parallel tubes (i) and (ii). At the entry point of the tubes, the two beams are *in phase*. From the experimenter's viewpoint, water in tube (ii) has speed $-w$ (moving leftward) and diminishes the speed of the opposing light beam. Similarly, water in tube (i) has speed $+w$ (moving rightward) and adds to the speed of the light beam. This difference of speeds implies that at the exit point of the tubes, the light beams are *out of phase* as is seen by the interference patterns of their waves.

(13.2.6a) **Observed Speeds of Light in the Tubes.** The speed of light relative to stationary water is $v = c/1.33$.

- As light travels in tube (i) in the same direction as the water, its observed speed is $v_{\text{fast}} > v = c/1.33$.
- As light travels in tube (ii) against the direction of the water, its observed speed is $v_{\text{slow}} < v = c/1.33$.

(13.2.6b) **Therefore,** there are *different numbers of wavelengths* in each tube. In Fizeau's experiment, it is observed that frequency of the light in each tube is unchanged. Such constant frequency tells us that a slower speed of the transmitted light (*upper tube (ii)*) implies a shorter wavelength and a faster speed for the transmitted light (*lower tube (i)*) implies a longer wavelength (see Figure (13.2.6) and

Definition (16.1.2)).



(13.2.6c) To see intuitively that increased wave speed implies increased wavelength, imagine a moving train (one wavelength) whose front and back ends must pass a fixed point on the platform exactly one second apart. The faster the speed of the train, the longer it must take to meet the one-second requirement.

To see this analytically, note that the formula $v = \lambda \cdot \text{frequency}$ (16.1.2) shows that wave speed v and wavelength λ are proportional. At any instant of time, the different wave speeds in the parallel tubes produce a different number of standing wavelengths (or cycles) as shown in Figure (13.2.6).

Because of the different speeds, the crests and troughs of the two light beams are not aligned when they exit the water-filled tubes. When they recombine, there are regions where they enhance each other and other regions where they cancel each other. The result is a pattern of light and dark fringes that changes as the speed w changes.

- (13.2.7)** **Speeds of Light in Water.** Since the speed of light relative to standing water is $c/1.33$, water traveling in the *same direction* at observed speed w adds to the speed of transmitted light (16.5.5). Non-relativistic Galilean physics holds that the augmented speed is

$$(13.2.7a) \quad v_{\text{fast}} = c/1.33 + w. \quad \begin{matrix} \text{Galilean addition} \\ \text{of speeds} \\ c/1.33 \text{ and } w \end{matrix}$$

Formula (13.2.7a) is not valid! What is observed instead is

$$(13.2.7b) \quad v_{\text{fast}} = c/1.33 + fw \quad \begin{matrix} \text{observed (true) addition} \\ \text{of speeds} \\ c/1.33 \text{ and } w \end{matrix}$$

where f is the unitless FRESNEL COEFFICIENT OF DRAG, so-called for historical reasons.

Data obtained using different liquids gives the following relation between n , the Fresnel refractive index of (13.2.2c), and f , the Fresnel drag of (13.2.7b):

$$(13.2.7c) \quad f = 1 - \left(\frac{1}{n}\right)^2. \quad (If n \approx 1.33 \text{ for water, then } f \approx 0.43.)$$

This reasons for this empirical relation were not understood until Einstein provided the following theoretical basis:

- (13.2.8) **Einstein's Theoretical Basis.** According to special relativity, if the speed of light with respect to the frame of reference where the medium (water) is at rest is c/n , and the speed of the medium with respect to the observer is w , then the speed of light with respect to the observer will be given by

$$(13.2.8a) \quad c/n + fw = c/n \oplus w$$

Evaluating the Lorentzian sum above with $w \ll c$ gives $(c/n) + (1 - 1/n^2)w$ (see Exercise (13.3.2)). This prediction of the special theory then coincides precisely with the observations (13.2.7c)!

(13.3) Exercises

- (13.3.1) **Exercise** **Wave Interference.** The symbols (m) and (m/s) indicate the units of *meters* and *meters/second*. Given the Fizeau apparatus (13.2.6) where the known quantities are:

$$\begin{aligned} L(m) &= \text{tube length}, \\ \lambda(m) &= 570 \times 10^{-9} \text{ meters, the wave length of yellow light} \\ &\quad \text{in a vacuum,} \\ w(m/s) &= \text{observed water speed,} \\ v(m/s) &= c/1.33 = 1/1.33 = \text{the speed of light relative to} \\ &\quad \text{stationary water.} \end{aligned}$$

Suppose the tube length L and water speed w can be adjusted to respective values L_0 and w_0 such that wave interference is maximized as shown in Figure (13.2.6). This occurs when maximum cancellation of the interfering waves occurs. With these known values of L_0 and w_0 , there is sufficient information to determine the observed speeds for light in the two tubes, v_{fast} in tube (i), and v_{slow} in tube (ii).

Hints:

- From (16.1.2), we may set

$$(13.3.1a) \quad \phi_{yellow} = \frac{c(m/s)}{\lambda(m)} = \frac{1}{\lambda}, \quad \text{units } \left(\frac{1}{\text{seconds}} \right)$$

as the frequency of yellow light where the speed of light $c = 1$. The value of frequency ϕ_{yellow} is constant — the color (frequency) does not change as light is observed in running water.

- Assuming the form (13.2.7b), write the speeds of light in the tube's running water as

$$v_{fast} = c/1.33 + \mathbf{f}w$$

for the speed of light in tube (i) and

$$v_{slow} = c/1.33 - \mathbf{f}w$$

for the speed of light in tube (ii). The unknown constant \mathbf{f} will be determined by our later calculations.

- From (16.1.2), the wave length $\lambda_{fast} = v_{fast} \cdot \phi_{yellow}$ for the accelerated light beam in tube (i) and the wave length $\lambda_{slow} = v_{slow} \cdot \phi_{yellow}$ for the impeded light beam in tube (ii). Accordingly, the numbers of wavelengths of light in tubes (i) and (ii) respectively, are

$$N_{fast} = \frac{L_0}{\lambda_{fast}} = \frac{L_0 \phi_{yellow}}{v_{fast}} = \frac{L_0 \phi_{yellow}}{(v + \mathbf{f}w_0)},$$

$$N_{slow} = \frac{L_0}{\lambda_{slow}} = \frac{L_0 \phi_{yellow}}{v_{slow}} = \frac{L_0 \phi_{yellow}}{(v - \mathbf{f}w_0)}.$$

- The unitless number $(N_{slow} - N_{fast})$ is the difference between the number of waves in the two tubes. We find that

$$(N_{slow} - N_{fast}) = 2L_0 \phi_{yellow} \frac{\mathbf{f}w_0}{v^2}$$

$$= 2 \frac{L_0}{\lambda} \frac{\mathbf{f}w_0}{v^2}. \quad \begin{aligned} &\text{set } \phi_{yellow} = 1/\lambda \text{ from (13.3.1a)} \\ &\text{and ignore } (\mathbf{f}w_0)^2 \approx 0 \end{aligned}$$

- In the equation above, set $(N_{slow} - N_{fast}) = 1/2$ so that wave interference is maximized. After L_0 and w_0 have been set to maximize wave interference, solve for the single unknown, \mathbf{f} , which then determines the desired speeds v_{fast} and v_{slow} .

(13.3.2)
Exercise

Deriving the Fresnel Drag Coefficient f . Show that (13.2.8) suffices to derive the empirical value of f (13.2.7c) for *all* refraction indices n .

Hint: Set $c = 1$ so that $c/n \oplus w = 1/n \oplus w$. Using the form (7.3.1) for the Lorentz sum, equation (13.2.8) reduces to

$$f = (1 - 1/n^2)/(1 + w/n).$$

Since water speed w is much less than light speed, 1 ($w \ll 1$), the $w/n \approx 0$ term in the denominator can be ignored.



(14) The Bug-Rivet Paradox

NOTE: To simplify calculations, we set the speed of light $c = 1$.

(14.1) The Minkowski Diagram

(14.1.1)



In frame \mathcal{F}^A , two **vertical plates**, each with horizontally aligned holes, have left surfaces d_0 units apart. A rivet in \mathcal{F}^B with at-rest shank length d_0 travels rightward towards the two aligned holes which are large enough so that the shank but not the head will pass through. A bug is in front of the hole in the right plate.

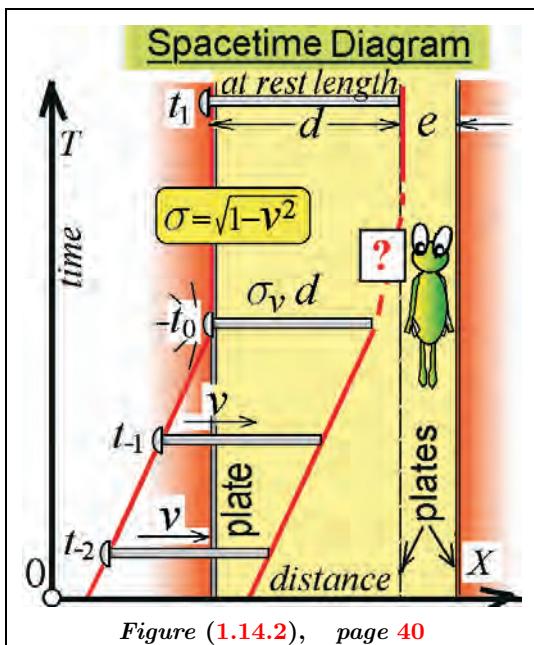


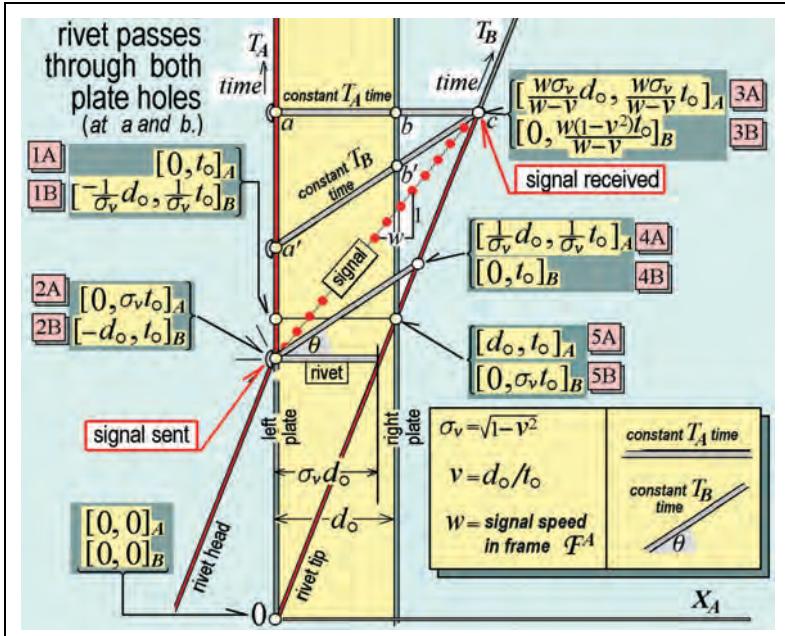
Figure (1.14.2), page 40

Viewed from the (stationary) plates, the moving rivet has length $\sigma_v d_0 < d_0$ — hence, the bug is not injured when the rivet head collides with the left plate. **But viewed from the (stationary) rivet**, the moving plates are separated by only $\sigma_v d_0 < d_0$ units — the bug would be squashed after the rivet collides.

(14.1.2) **Question:** Is the bug squashed or not?

Answer: Analyzing the Minkowski diagram (14.1.3), we shall see why the bug cannot be spared.

(14.1.3) **FIGURE: Minkowski Diagram for Rivet Expanding Beyond Its At-rest Length**



Just before the rightward-traveling rivet hits the left plate, all its atoms move at speed v . The moment the rivet head hits the left plate, it stops. But the rest of the rivet *cannot* stop simultaneously since the message “I’ve hit a plate” cannot travel instantaneously from the head of the rivet to the rightward tip. In fact, a signal is transmitted from the head to the rest of the rivet through the atomic bonds between the atoms of the rivet. Although the signal may travel quickly by human standards (*at speed w relative to the plates in \mathcal{F}^A*), it cannot travel faster than $c = 1$. The rivet tip will keep on moving at speed v until the signal finally reaches it.

Here is the analysis:

(14.1.4) **Setup of the Minkowski Diagram (14.1.3):**

(14.1.4a) **Axes for \mathcal{F}^A (parallel plates frame)** are drawn in the Minkowski graph (14.1.3) so that its time axis, T_A , is vertical and its distance axis, X_A , is horizontal. The point $x = 0$ in \mathcal{F}^A lies on the left surface of the left plate.

The units of the axes are normalized so that c , the speed of light in a vacuum in \mathcal{F}^A , equals 1.

(14.1.4b) **Axes for \mathcal{F}^B (*rivet frame*)** are drawn in the Minkowski diagram (14.1.3) so that its time axis, T_B , forms an angle θ with the vertical. By symmetry (1.6.1), $-v$ is the speed of the plates in \mathcal{F}^A as seen from the rivet's frame, \mathcal{F}^B .

The distance axis, X_B of \mathcal{F}^B , (not shown in Figure (14.1.3)), forms an angle θ with the horizontal. The angle θ is required since $v = \tan(\theta)$ is the speed of the rivet in \mathcal{F}^B as seen from the plates' frame, \mathcal{F}^A .

The point $x = 0$ in \mathcal{F}^B lies on the right-hand tip of the rivet.

The units of the axes are normalized so that c , the speed of light in a vacuum in \mathcal{F}^B , equals 1.

(14.1.4c) **Lines of simultaneity (*constant time*)** in frame \mathcal{F}^A lie on horizontal lines that are necessarily parallel to the X_A -axis.

All lines of simultaneity (*constant time*) of frame \mathcal{F}^B lie on lines that form an angle θ with the horizontal that are necessarily parallel to the X_B axis.

(14.1.4d) **Vertical plates (*in frame \mathcal{F}^A*)** are parallel metal plates with horizontally aligned holes large enough to allow the shank of a rivet to pass through but not large enough for the rivet head. The at-rest distance between the left surfaces of the two plates is d_0 .

(14.1.4e) **The rivet (*in frame \mathcal{F}^B*)** has a shank with at-rest length d_0 . The speed of the inertial frame \mathcal{F}^B of the rivet is $v = \tan(\theta)$ relative to the inertial frame \mathcal{F}^A of the plates. Hence, from the symmetry principle, the speed of the frame \mathcal{F}^A of the plates is $-v$ relative to the frame \mathcal{F}^B of the rivet.

(14.1.4f) **Synchronization:** The left surface of the left-hand plate in \mathcal{F}^A and the right tip of the rivet in frame \mathcal{F}^B are synchronized (3.4.5a) in Figure (14.1.3) at $[0, 0]_A$ and $[0, 0]_B$.

(14.1.4g) **Time** $t_0 = d_0/v$ where v and d_0 are given in (14.1.4e).

We can see that $v = d_0/t_0$ from the Minkowski diagram (14.1.3) since the slope of the worldline of the rivet is $1/v$. Specifically, the point $[d_0, t_0]_A$ 5A lies on the worldline as do the origins, $[0, 0]_A$ and $[0, 0]_B$, due to synchronization (14.1.4f).

.....

(14.1.4h) **The Lorentz Transformation**, L_v , connects the line-of-sight coordinates between frames \mathcal{F}^A and \mathcal{F}^B where frame \mathcal{F}^B has speed v relative to frame \mathcal{F}^A (7.1.3a), pg. 106, (7.2.2), pg. 109.

Specifically, for $\sigma_v = \sqrt{1 - v^2}$, an observer in \mathcal{F}^A at time t with coordinates $[x, t]_A$ is linked (*via a line-of-sight corridor*) to an observer in \mathcal{F}^B with coordinates $[x', t']_B$, if and only if

$$(*) \quad \underbrace{\frac{1}{\sigma_v} \begin{bmatrix} 1 & -v \\ -v & 1 \end{bmatrix}}_{L_v} * \begin{bmatrix} x \\ t \end{bmatrix}_A = \begin{bmatrix} x' \\ t' \end{bmatrix}_B \quad \text{and} \quad \underbrace{\frac{1}{\sigma_v} \begin{bmatrix} 1 & v \\ v & 1 \end{bmatrix}}_{L_{-v}} * \begin{bmatrix} x' \\ t' \end{bmatrix}_B = \begin{bmatrix} x \\ t \end{bmatrix}_A.$$

(14.2) Coordinates in the Minkowski Diagram

The coordinates 1A, 1B, 2A, 2B, ..., 5A, 5B of Figure (14.1.3) are justified through the following chain of ideas. We start with

1A = $[0, t_0]_A$: On the T_A -axis of \mathcal{F}^A , choose t_0 in accordance with (14.1.4g). The Minkowski coordinate in frame \mathcal{F}^A is $[0, t_0]_A$.

1B = $[-\frac{1}{\sigma_v}d_0, \frac{1}{\sigma_v}t_0]_B = L_v([0, t_0]_A)$ in accordance with (14.1.4h).

.....

4B = $[0, t_0]_B$ are the coordinates in \mathcal{F}^B of the intersection of the hyperbola $t_0^2 = t^2 - x^2$ and the T_B axis (the worldline of the rivet tip).

$$\boxed{4A} = \left[\frac{1}{\sigma_v} d_0, \frac{1}{\sigma_v} t_0 \right]_A = L_{-v}([0, t_0]_B) \text{ in accordance with (14.1.4h).}$$

.....

$\boxed{5A} = [d_0, t_0]_A$ is the coordinate of the point on the left surface of the right plate which (at \mathcal{F}^A time t_0) is d_0 units distance in \mathcal{F}^A from $\boxed{1A}$, the point on the left surface of the left plate.

$$\boxed{5B} = [0, \sigma_v t_0]_B = L_v[d_0, t_0]_A \text{ in accordance with (14.1.4h).}$$

.....

$\boxed{2B} = [-d_0, t_0]_B$ is the coordinate of the point on the left surface of the left plate which (at \mathcal{F}^B time t_0) is $-d_0$ units distance in \mathcal{F}^B from $\boxed{4B}$, the point on the left surface of the right plate.

$$\boxed{2A} = [0, \sigma_v t_0]_A = L_{-v}[-d_0, t_0]_B \text{ in accordance with (14.1.4h).}$$

.....

$\boxed{3A} = \left[\frac{w\sigma_v}{w-v} d_0, \frac{w\sigma_v}{w-v} t_0 \right]_A$ are the spacetime coordinates of the rivet tip when it first receives the shock wave signal that has been traveling rightward along the shank at constant speed $w \leq c = 1$. The worldline of this signal (represented by red dots in (14.1.3)) starts at the point $\boxed{2A}$, $\boxed{2B}$ and has slope $1/w$.

$$\boxed{3B} [0, \frac{w(1-v^2)}{w-v} t_0]_B = L_v \left[\frac{w\sigma_v}{w-v} d_0, \frac{w\sigma_v}{w-v} t_0 \right]_A \text{ in accordance with (14.1.4h).}$$



(14.2.1)

THEOREM: (*Rivet Overshoot Beyond the Second Plate*)

The rivet in frame \mathcal{F}^B travels at speed v as measured from frame \mathcal{F}^A of the vertical plates. Once the rivet head collides with the left plate at point $[2A]$, $[2B]$ of Figure (14.1.3), a shock wave is propagated through the rivet shank at speed w relative to the vertical plates, where $v < w \leq c$. Then

(14.2.1a) The shock wave signal reaches the rivet tip at the point in Figure (14.1.3) denoted by $c = [3A]$, $[3B]$.

Suppose the extended shank length is measured from the left surface of the right-hand plate to the rivet tip. Then

$$(14.2.1b) \quad \frac{\text{extended shank length}}{\text{total shank length}} = 1 - \frac{1 - (v/w)}{\sigma_v},$$

which is invariant for both frames, \mathcal{F}^A and \mathcal{F}^B .

PROOF: The coordinates for point $c = [3A]$, $[3B]$ in Figure (14.1.3) have already been justified on page 175 as the point at which the shock wave arrives at the rivet tip. This establishes (14.2.1a).

Note that Figure (14.1.3) gives us both \mathcal{F}^A and \mathcal{F}^B coordinates of the event when the collision shock wave first arrives at the rivet tip. Using just the \mathcal{F}^A coordinates for points a , b , and $c = [3A]$, of Figure (14.1.3), we obtain

(14.2.1c) Time $t = (w\sigma_v)/(w-v)t_0$ is the common (simultaneous) time in frame \mathcal{F}^A at points a , b , and $c = [3A]$. From Figure (14.1.3), we see that their x coordinates a_x , b_x , and c_x are given by

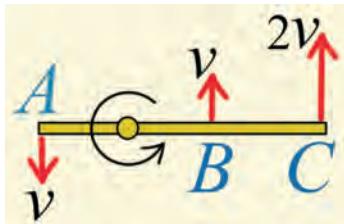
point	x coordinate
c	$c_x = \frac{(w\sigma_v)}{(w-v)}d_0$
b	$b_x = d_0$
a	$a_x = 0$

$$\text{which implies } \frac{c_x - b_x}{c_x - a_x} = 1 - \frac{1 - (v/w)}{\sigma_v}.$$

This expression is exactly the ratio of the extended shank length to the total shank length. That is, (14.2.1b) is established and the theorem is proved. ■

(14.2.2) **NOTE:** The ratio (14.2.1b) is Frame-Invariant. Table (14.2.1c) considers the ratio in \mathcal{F}^A coordinates. Exercise (14.4.1) arrives at the same ratio using \mathcal{F}^B coordinates.

(14.2.3) **NOTE:** Special relativity is incompatible with rigidity. There are examples showing that rigidity forces certain particles to travel faster than c , the speed of light. One such example is a stiff rod pivoting between points A and B each of which rotates at speed $v = (2/3)c$. If the rod is rigid then the point C must rotate at twice the speed, viz., $2 \times (2/3)c = (4/3)c$. For the speed at C to be less than c , the rod has to bend so that point C “lags” behind.



(14.2.4) **NOTE:** The Future Affects the Present If Perfect Rigidity is Allowed. If we assume that bodies remain rigid under acceleration, then the rivet can never expand beyond its at-rest length, d_0 , and, consequently, the future must affect the present. Details are revealed in Exercise (14.4.4).



If w is the speed of the shock wave as measured in frame \mathcal{F}^A of the plates, then what is the shock wave speed as measured from \mathcal{F}^B , a passenger on the rivet?

We now present an alternate geometric proof that follows immediately from reading the coordinates of Figure (14.1.3).

(14.2.5)

THEOREM: (Relative Speeds of the Shockwave): The rivet in frame \mathcal{F}^B travels at speed v as measured from frame \mathcal{F}^A of the vertical plates. Once the rivet head collides with the left plate at point $[2A]$, $[2B]$ of Figure (14.1.3), a shock wave is propagated through the rivet shank at speed w as measured in frame \mathcal{F}^A of the vertical plates, where $v < w \leq c$.

Then

w_{rivet} , the speed of the shockwave, as measured in frame \mathcal{F}^B of the rivet (or by a passenger on the rivet), is

$$(14.2.5a) \quad w_{\text{rivet}} = \frac{w + (-v)}{1 + w(-v)}.$$

PROOF: We compute the shock wave speed, $w_{\text{rivet}} = \Delta x_A / \Delta t_A$ in \mathcal{F}^A coordinates and, redundantly, $w_{\text{rivet}} = \Delta x_B / \Delta t_B$ in \mathcal{F}^B coordinates.

Using the orthogonal X_A and T_A axes in Figure (14.1.3),

$$\begin{aligned} \Delta x_A &= (1/\sigma_v)d_0 && \text{from } [4A], [2A], \\ \Delta t_A &= \left(\frac{w\sigma_v}{w-v} - \frac{1}{\sigma_v} \right) t_0 && \text{from } [3A], [4A]. \end{aligned} \Rightarrow \frac{\Delta x_A}{\Delta t_A} = \frac{w + (-v)}{1 + w(-v)}.$$

In Figure (14.1.3), the X_B -axis is parallel to the line formed between points $[2A, B]$ and $[4A, B]$. The T_B -axis is shown on the figure. Then

$$\begin{aligned} \Delta x_B &= d_0 && \text{from } [4B], [2B], \\ \Delta t_B &= \left(\frac{w(1-v^2)}{w-v} - 1 \right) t_0 && \text{from } [3B], [4B]. \end{aligned} \Rightarrow \frac{\Delta x_B}{\Delta t_B} = \frac{w + (-v)}{1 + w(-v)}.$$

In both frames \mathcal{F}^A and \mathcal{F}^B (14.2.5a) is confirmed, which ends the proof. ■

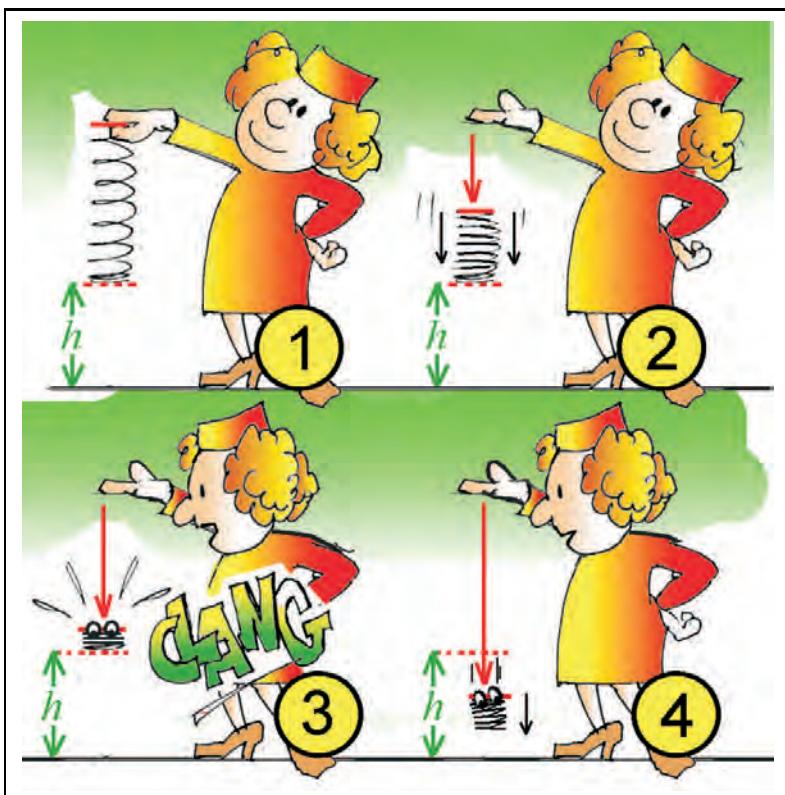
(14.3) The Slinky Connection

The previous sections show that when a collision occurs between a plate and the head of a rivet, then information about the collision must travel from the rivet head along the shank and finally to the rivet tip before the tip “knows” that it must slow down.

In this section, we cite other situations in which a non-zero time is required for a message to propagate through a medium before a response to that message results.

- In the first instance (*Figure (14.3.1)*), a Slinky toy is suspended from its top. Once the top is released, the bottom of the Slinky will continue to feel an upward tension (*and remain fixed*) as long as the ring above it is stretched. Once the Slinky fully collapses, the bottom ring feels no more upward tension and the whole unit falls to the ground. (*Thanks to DAVID ELLERMAN for this example.*)
- The second example (*14.3.2*) illustrates that the loss of gravitational pull of the sun would not be felt on Earth for almost nine minutes.

(14.3.1) **FIGURE:** *Suspended Slinky is Released from the Top While Its Bottom Remains Fixed in Space*



(14.3.1a) **Question:** What happens to the lower end of a suspended Slinky when the top end is released?¹

(14.3.1b) **Answer:** As indicated in Figure (14.3.1), [Panels ①, ②] the bottom of the Slinky remains suspended at the initial height h from the ground — until the collapsing wave from the falling Slinky top reaches the bottom of the Slinky®. At this moment [Panel ③], the following message reaches the Slinky bottom: “From now on, there is no longer an upward tension equal to the downward force of gravity. You must fall!” When this message is received [Panel ④], the bottom of the Slinky joins the top in falling to the ground.

¹For a video demonstration, search on YouTube.com for “Slinky Drop Answer”

(14.3.1c) The Analogy of the Rivet and the Slinky Toy: In Figure (14.1.3), the rivet tip has constant speed $v > 0$ and is in an inertial frame until the collision wave (message) reaches it at point [3A], [3B]. In Figure (14.3.1), the Slinky bottom also has constant speed $v = 0$ in an inertial frame *until* the message that there is no upward tension reaches from the top to the bottom at point ③.

(14.3.2)



The Disappearing Sun: *Time is required even for the “message” of gravity waves.* The earth is held in its orbit by centripetal force from the sun much as a pail of water, being swung by a rope, is held in its circular orbit by the centripetal pull of the rope. However, gravity waves from the sun require some eight minutes (at the speed of light) to reach the earth.

Now if the sun and all its mass were to instantly disappear, this would “cut the gravity rope,” causing the earth to fly off its circular orbit on a straight-line tangent nine minutes *after* the sun’s disappearance — this is the time required for the earth to “know” that there is no longer any gravitational pull from the sun. Afterwards, the earth would continue on its straight-line path, destined to travel forever into a void of cold darkness.

(14.4) Exercises

**(14.4.1)
Exercise**

(Extending Theorem (14.2.1)). Show that the ratio (14.2.1b) which was shown for \mathcal{F}^A coordinates is also valid for points a', b', c in \mathcal{F}^B . Specifically, from Figure (14.1.3), show that in \mathcal{F}^B coordinates,

$$(14.4.1a) \quad \frac{\text{extended shank length}}{\text{total shank length}} = \frac{c_x - b'_x}{c_x - a'_x} = 1 - \frac{1 - (v/w)}{\sigma_v}$$

where a'_x, b'_x, c_x are the X_B coordinates of a, b, c . (*Equation (14.4.1a) can also be established using T_B coordinates a'_t, b'_t, c_t in \mathcal{F}^B .*)

(14.4.2)
Exercise

Under the hypotheses of Theorem (14.2.1), show that for all shock wave speeds w , where $v \leq w \leq c = 1$, there is (at least) one moment after the rivet-plate collision in which the rivet shank length exceeds its at-rest length, d_0 .

Hint: Formula (14.2.1b) is derived for the moment the shock wave reaches the rivet tip which finally gets the message to leave its inertial frame \mathcal{F}^B and to decelerate. The question, then, is to produce a shock wave speed w for which (14.2.1b) is zero. But this implies that $w = (1/v) + \sqrt{(1/v)^2 - 1}$, which is greater than 1, the speed of light.

(14.4.3)
Exercise

(*When Shock wave Speed = c=1*) Show that if either $w = 1$ or $w_{\text{rivet}} = 1$ for the shock wave speed of Theorems (14.2.1) and (14.2.5), then both values are equal to 1. (*This reflects the fact that any wave traveling at the speed of light c in one frame has speed c in all frames.*

Hint: Apply Equation (14.2.5a).

(14.4.4)
Exercise

(*When the Future Affects the Present*) Let the rivet shank have at-rest length d_0 as measured in the frame of the vertical plates, \mathcal{F}^A . Using the Minkowski diagram coordinates (14.1.3), pg. 172, show that if the rivet shank length is always less than or equal to its at-rest length, d_0 , then the future must affect the past.

Hint: Figure (14.1.3) shows that the rivet collides with the left plate at point $2A$ (\mathcal{F}^A coordinates) = $2B$ (\mathcal{F}^A coordinates). From the Slinky principle, the collision signal information (the red dots in Figure (14.1.3)) require time to reach the rivet tip. In this case, the rivet must expand beyond its at-rest length until the collision signal arrives at the tip.

Hence, the only way to maintain rigidity is to have the rivet and plates meet at point $2A = 2B$ when their relative speed to each other is zero. This says the plates receive a rightward force or the rivet receives a leftward force in order to decelerate to relative speed zero before the actual collision. In other words, the future collision causes deceleration in the present.

- (14.4.5) **(*Applying the Pea-Shooter Paradox*)** Use the Lorentzian sum \oplus (7.3.1) directly to prove (14.2.5a).

Hint: Use the Pea-Shooter paradox as a template (see (1.13.2b)). Let w equal the speed of the shock wave (the pea) as observed from \mathcal{F}^A , the frame of the vertical plates (Bernie on solid ground). Then w is also equal to: The speed v of the rivet (Ashley on the train) as observed from \mathcal{F}^A “plus” the speed w_{rivet} of the shock wave (the pea) as observed from \mathcal{F}^B , the frame of the rivet (Ashley on the train). Formally, then,

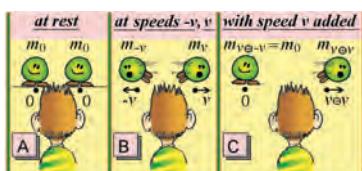
$$(14.4.5a) \quad w = v \oplus w_{\text{rivet}},$$

so that using (1.13.3) to solve for w_{rivet} in (14.4.5a) yields (14.2.5a).



VI. Energy and Mass

Chapter 15 $E=mc^2$ This celebrated equation is almost an afterthought once the connection is made between the pea-shooter paradox and conservation of momentum.



$$(15) \quad E = mc^2$$

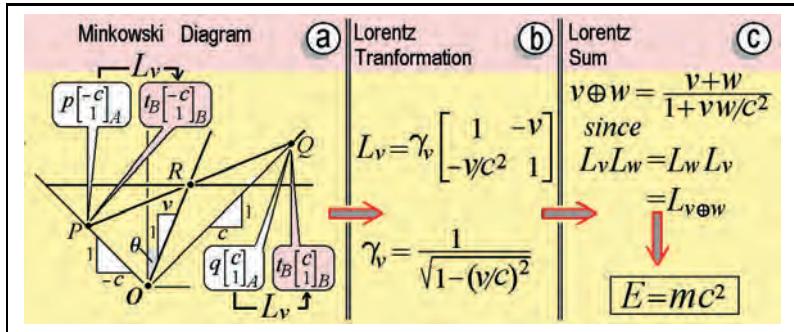
(15.1) How We Came to This Place

It might seem that our analyses of the paradoxes through Minkowski diagrams have little to do with $E = mc^2$, the relation of a mass m at rest and its energy equivalent E .

As it turns out, the structure underlying the pea-shooter paradox is also essential to showing that $E = mc^2$. The structure in question is the Lorentzian addition of speeds (7.3.1).

(15.1.1)

FIGURE: (Recapitulation: Three Steps to the Lorentz Sum)



- **PANEL ①:** *Coordinates for the endpoints P and Q are found in two frames of reference: $P = p[-c, 1]$ and $Q = q[c, 1]$ relative to the frame \mathcal{F}^A with perpendicular X and T axes, while $P = p[-c, 1]'$ and $Q = q[c, 1]'$ relative to the frame \mathcal{F}^B with the T -axis lying on \overline{OR} .*
- **PANEL ②:** *The Lorentz transformation, a 2×2 matrix, is defined on the basis vectors (points) P , Q , and hence, is fully defined everywhere (the importance of the basis property for linear transformations follows from Theorem (A.4.2), pg. 369). The matrix has for its domain the frame of reference from which another frame is seen to be moving with relative speed v .*
- **PANEL ③:** *If frame \mathcal{F}^A sees frame \mathcal{F}^B move with speed v , then we have $L_v : \mathcal{F}^A \rightarrow \mathcal{F}^B$. If frame \mathcal{F}^B sees frame \mathcal{F}^C move with speed w , then $L_w : \mathcal{F}^B \rightarrow \mathcal{F}^C$. Since the matrix product $L_w * L_v : \mathcal{F}^A \rightarrow \mathcal{F}^C$*

$(L_w * L_v$ is defined in (A.4.3e)) is equivalent to $L_{v \oplus w} : \mathcal{F}^A \rightarrow \mathcal{F}^C$, where the Lorentz sum, $v \oplus w$, is given in Panel C, then frame \mathcal{F}^A sees frame \mathcal{F}^C move with speed $v \oplus w$.

We will see in this chapter that if speeds must be added with the Lorentz sum, then necessarily $E = mc^2$.

(15.2) Speed-Dependent Mass: an Intuitive View

(15.2.1)



Speed-dependence in the notation. The mass of a particle viewed at speed zero (at rest) is denoted m_0 . The mass of the same particle viewed at speed v is denoted m_v . Our Galilean/Newtonian intuition tells us that there is no need to change notation for a mass just because it happens to be in motion, since mass does not change with observed speed—or does it? (In contrast, we accept the concept from introductory calculus that mass can change over time as is shown by the time-dependent “petri dish equation” $m_t = m_0 e^{kt}$. Einstein’s brilliant insight revealed that mass also changes according to its observed speed v).¹ In any case, there is no harm in using the (ornamented) notation

(15.2.1a) $m_v \stackrel{\text{def}}{=} \text{mass of particle at speed } v$.

Hence, its rest mass, the mass at speed 0, is denoted m_0 .

Intuition misleads us in another way. Let us empathize for a moment with a point mass that might ask itself, *If an observer measures my mass as equal to m_0 when I am observed at speed $v = 0$, then why is it that I have a different mass just because the observer now sees my speed to be $v > 0$?*

The Two-Particle Setup

(15.2.2) **Preliminaries:** To acknowledge that a mass m is being observed

¹At time $t = 0$, the initial population of bacteria in a petri dish has mass m_0 . If at any time t , population growth is proportional to the present population, m_t , then in terms of differential equations, this statement translates to $dm_t/dt = km_t$ for all time $t > 0$ for some constant k . The solution is $m_t = m_0 e^{kt}$.

while it has speed v , we denote it by m_v . Therefore, m_0 denotes the **at-rest mass** at observed speed zero.

(15.2.3)

DEFINITION: The MOMENTUM OF A PARTICLE with rest mass m_0^i at velocity \vec{v}_i is defined to be $m_{\vec{v}_i}^i \cdot \vec{v}_i$ (For clarity in equations such as (15.2.3a), we sometimes use the square bracket notation $[m_{\vec{v}_i}^i] \cdot \vec{v}_i$). The MOMENTUM OF THE SYSTEM of two particles with rest masses m_0^1 and m_0^2 when observed at respective velocities \vec{v}_1 and \vec{v}_2 is defined by

$$(15.2.3a) \quad [m_{\vec{v}_1}^1] \cdot \vec{v}_1 + [m_{\vec{v}_2}^2] \cdot \vec{v}_2 \stackrel{\text{def}}{=} [m_{\vec{v}_1}^1 + m_{\vec{v}_2}^2] \cdot \vec{v}_{sys},$$

where \vec{v}_{sys} is the SYSTEM VELOCITY which, intuitively, treats the two particles as a single mass entity.

If all motion occurs in the same direction or its opposite, then **conservation of momentum** gives us the

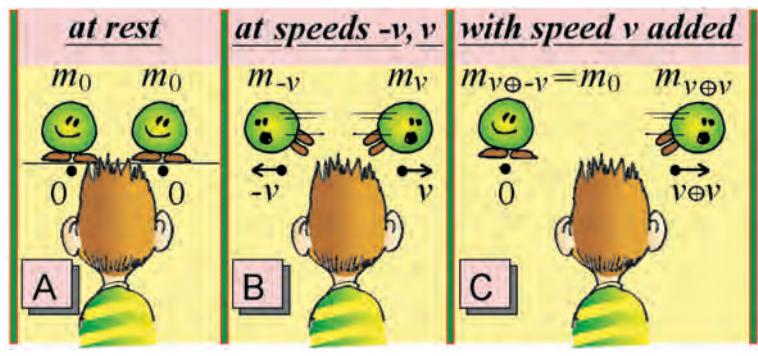
(15.2.3b) SYMMETRY PRINCIPLE, $m_{-v} = m_v$

which holds that both speeds, $+v$ and $-v$, when applied to any mass m_0 , produce the same observed speed-dependent mass m_v .

To see why this is so, imagine two ice-skaters, each of mass m_0 facing each other on frictionless ice. Since they each have speed $v = 0$, the system of two ice-skaters has total momentum zero. Next, they push each other away so that one ice-skater has speed $-v$ and the other has speed v . Now, the system momentum (*which is conserved to be zero*) has form $(m_{-v})(-v) + (m_v)(v) = (-m_v)(v) + (m_v)(v) = (-m_{-v} + m_v)(v) = 0$ which implies (15.2.3b).

We will apply the momentum equation (15.2.3a) to the special case of two equal rest masses that initially have equal and opposite speeds $\pm v$. Accordingly, we shall use the the following figure that summarizes the states of these particles at various speeds. Note that we are obliged to add speeds relativistically according to the Lorentz formula (7.3.1) $v \oplus w = (v + w)/(1 + vw/c^2)$.

(15.2.4) FIGURE: A Simplified Two-Particle System



- PANEL **A** OF FIGURE (15.2.4). Two particles are observed, each with the same at-rest mass m_0 . Each particle has zero momentum $mass \cdot speed = [m_0] \cdot 0 = 0$ so that from (15.2.3a), we compute the sum and obtain the system momentum and the system speed v_{sys} as follows:

$$(15.2.4a) \quad [m_0] \cdot 0 + [m_0] \cdot 0 = [m_0 + m_0] \cdot \underbrace{0}_{v_{sys}}.$$

- PANEL **B** OF FIGURE (15.2.4). The left and right masses, m_{-v} and m_v , say, have equal and opposite speeds, $-v$ and v . The respective values of their momentums are also equal and opposite. That is,

$$(15.2.4b) \quad \underbrace{([m_{-v}] \cdot -v)}_{\substack{\text{left} \\ \text{momentum}}} = - \underbrace{([m_v] \cdot v)}_{\substack{\text{right} \\ \text{momentum}}}.$$

We now obtain the following sum that produces the system momentum and the system speed v_{sys} (15.2.3a):

$$([m_{-v}] \cdot -v) + ([m_v] \cdot v) = [m_{-v} + m_v] v_{sys}.$$

From (15.2.4b), we may substitute $-([m_v] \cdot v)$ for $([m_{-v}] \cdot -v)$ above to obtain

$$(15.2.4c) \quad \underbrace{-([m_v] \cdot v) + ([m_v] \cdot v)}_0 = [m_{-v} + m_v] v_{sys}.$$

Since $m_{-v} > 0$ and $m_v > 0$, (15.2.4c) implies $v_{sys} = 0$.

- PANEL **C** OF FIGURE (15.2.4). Speed v is added to the entire 2-particle system of Panel **B** which means the system speed $v_{sys} = 0$ of (15.2.4c) is augmented by speed v to become

$$(15.2.4d) \quad \text{new } v_{sys} = (\underbrace{0}_{\text{old } v_{sys}} \oplus v) = v. \quad \text{from (7.4.3)}$$

Since the speed of each particle is augmented by v units, the left particle, formerly at speed $-v$, acquires the new speed $-v \oplus v = 0$ (see Proposition (7.4.3)) so its new momentum is $[m_0] \cdot (-v \oplus v) = [m_0] \cdot 0 = 0$. Similarly, the right particle, formerly with speed v , acquires added speed v to produce the new speed $v \oplus v$. This speed accounts for a new momentum of $[m_{v \oplus v}] \cdot (v \oplus v)$. In (15.2.3a), replace v_1 with 0 and v_2 with speed $v \oplus v$. This allows us to set $m_{v_1}^1 = m_0$ and $m_{v_2}^2 = m_{v \oplus v}$, which yields

$$(15.2.4e) \quad \underbrace{[m_0] \cdot 0 + [m_{v \oplus v}] \cdot v \oplus v}_{0} = [m_0 + m_{v \oplus v}] \cdot v_{sys} \\ = [m_0 + m_{v \oplus v}] \cdot v. \quad \text{from (15.2.4d)}$$

The stage is set for the following

(15.2.5)



THEOREM: Mass Increases with Increased Speed. Given a particle with rest mass m_0 . If m_w represents the measurement of this mass when observed at speed w , then

$$(15.2.5a) \quad m_w > m_0.$$

PROOF: Consider a system of two particles, each with identical rest mass m_0 so that all the equations of Figure (15.2.4) apply. In particular, among the three equal quantities in (15.2.4e), set only the first and last quantities equal to each other and solve for $m_{v \oplus v}$ to obtain

$$(15.2.5b) \quad m_{v \oplus v} = m_0 \frac{v}{(v \oplus v) - v} = m_0 \frac{1 + v^2}{1 - v^2} \quad \text{from (7.3.1)}$$

$$> m_0. \quad \text{since } \frac{1 + v^2}{1 - v^2} > 1$$

Note that (15.2.5b) justifies (15.2.5a) once we set $w = v \oplus v$. This ends the proof. ■

.....
COROLLARY: At speed $w = v \oplus v$, a particle with rest mass m_0 has increased mass m_w growing to mass m_w where

$$(15.2.5c) \quad m_w = m_0 / \sqrt{1 - w^2}.$$

PROOF: Exercise (15.5.3). ■

(15.3) Equivalence of Mass and Energy

(15.3.1) The power series expansion for infinitely differentiable $f(x)$ is
 Calculus
 needed

$$(15.3.1a) \quad f(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \cdots + \frac{f^{(k)}(0)}{k!}x^k + \cdots,$$

where $f'(0)$, $f''(0)$, ..., $f^{(k)}(0)$, ... represent the first, second, ..., and k th derivative of the function f evaluated at 0. In particular, the k th derivative evaluated at 0 is

$$(15.3.1b) \quad f^{(k)}(0) = \frac{1 \cdot 3 \cdot 5 \cdots (2k-1)}{2^k} \quad \text{when } f(x) = \frac{1}{\sqrt{1-x}}.$$

In (15.3.1a), set $x = (v/c)^2$ and replace the derivative expressions with (15.3.1b) to obtain the series

$$(15.3.1c) \quad \frac{1}{\sqrt{1-(v/c)^2}} = 1 + \frac{1}{2} \left(\frac{v}{c}\right)^2 + \underbrace{\frac{3}{8} \left(\frac{v}{c}\right)^4 + \frac{5}{16} \left(\frac{v}{c}\right)^6 + \cdots}_{O(v^4)}$$

where $O(v^4)$ indicates a sum of v^k terms where $k \geq 4$. (At this point, units of speed v can be arbitrary (cm/sec, miles/hour) so it is not necessary that $c = 1$.)

In (15.3.1a), replace $v < 1$ with $(v/c) < 1$ —this releases us from the special case where we write $c = 1$. Now recapture this new form of (15.3.1a) by multiplying (15.3.1c) through by $m_0 c^2$ to obtain the equalities

(15.3.1d)

$$m_v c^2 = \frac{m_0 c^2}{\sqrt{1 - (v/c)^2}} = m_0 c^2 + \underbrace{m_0 \frac{v^2}{2}}_{\substack{\text{Newtonian} \\ \text{kinetic energy}}} + m_0 O(v^4)$$

$$= m_0 c^2 + \underbrace{E(v)}_{\substack{\text{relativistic} \\ \text{kinetic energy}}} .$$

(15.3.2) **Interpreting (15.3.1d).**

Is the power series (15.3.1d) a mathematical invention designed to torture calculus students like some irrelevant chess game of the mind where symbols on paper only obey certain abstract rules? Or is there a valid connection with reality as we (think) we know it? Let us see how a certain marker term brings the series (15.3.1d) from the calculus textbooks into the world of physical reality.

Series (15.3.1d) represents energy. It is the classical kinetic energy term, $m_0 v^2/2$, in (15.3.1d) that spreads its type (energy) to all other terms in the series. Informally, the energy attribute of $m_0 v^2/2$ forces all other terms to have the same attribute.

Why $E = m_0 c^2$. We write $E(v)$ to denote the sum of all the terms of (15.3.1d) on the right hand side except for the term, $m_0 c^2$. This sum, $E(v)$, which depends on v , can only be interpreted as a general energy term that arises from the *energy required to bring the particle from a state of rest to a speed of v* . To see this, set $v = 0$ and $E(v)$ goes away. That is, $E(v)$ appears only when motion is induced, which means a force and distance, or energy, is invested into the particle in order to get it into motion.

Therefore, the left hand side of (15.3.1d) represents the sum of the at-rest energy, $m_0 c^2$, which is clearly not zero, and $E(v)$, the energy necessary to bring the particle from speed $v = 0$ to speed $v \neq 0$. Since the term $m_0 c^2$ does not change with v , we conclude

$E = m_0 c^2$ is the rest energy
of a particle that has rest mass m_0 .

(15.4) A Numerical Example

(15.4.1) *Masses Under the Effects of a Force Cannot Stay Constant.*

The following example shows how (15.2.3a), the definition of momentum using Lorentzian addition of speeds, will not tolerate constant mass with increased speed (due to a force).

(15.4.1a) Set mass $m^1 = 1\text{kg}$ with speed $v_1 = -0.8$ and mass $m^2 = 2\text{kg}$ with speed $v_2 = 0.4$.

(We need not write the units for speeds which are, after all, unitless ratios of form v/c). Then from (15.2.3),

(15.4.1b)

$$\underbrace{1\text{kg} \cdot (-0.8) + 2\text{kg} \cdot (0.4)}_{0\text{kg}} = \underbrace{1\text{kg} + 2\text{kg}}_{3\text{kg}} \cdot w_{old}. \quad \begin{matrix} \text{which implies} \\ w_{old} = 0 \end{matrix}$$

Suppose the masses remain constant after the system speed has increased its value from $w_{old} = 0$ to $w = 0.8$. This means we must add $w = 0.8$ to the speeds $v_1 = -0.8$, $v_2 = 0.4$, and $w_{old} = 0$ in (15.4.1b) (using relativistic addition \oplus (7.3.1)) to obtain

(15.4.1c)

$$\underbrace{1\text{kg} \cdot \underbrace{(-0.8 \oplus 0.8)}_0 + 2\text{kg} \cdot \underbrace{(0.4 \oplus 0.8)}_{0.909}}_{m^1} = \underbrace{1\text{kg} + 2\text{kg}}_{3\text{kg}} \cdot \underbrace{(0 \oplus 0.8)}_{w_{new} = 0.8}$$

which gives us the absurd equation

$$(15.4.1d) \quad \underbrace{1\text{kg} \cdot 0}_{0\text{kg}} + \underbrace{2\text{kg} \cdot 0.909}_{1.818\text{kg}} = \underbrace{3\text{kg} \cdot 0.8}_{2.4\text{kg}}.$$

Standard proof-by-contradiction (G.4.5) tells us we must reject our hypothesis that masses remain constant with augmented speed.

To say the same thing more formally, note that in reasoning from (15.4.1a) to (15.4.1d), we find that the constant-mass assumption leads to a false conclusion, namely,

(15.4.1e) $\{ \text{Constant mass} \} \Rightarrow \{ 1.818 = 2.4 \}$ (*False*).

We can never use valid reasoning where a true hypothesis leads to a false conclusion. Therefore, in (15.4.1e), our hypothesis *Constant mass*, which leads to a false conclusion, must be (*False*).

- (15.4.2) **What is the non-absurd resolution to (15.4.1d)?** The following table gives the values of

$$\tilde{m}^1 = m_{v_1 \oplus w}^1 \quad \text{and} \quad \tilde{m}^2 = m_{v_2 \oplus w}^2,$$

for system speed $w = 0.8$ and the resulting momentum.

(15.4.2a) **TABLE: Momentum P at system speeds $v_{sys}=0$ and $v_{sys}=0.8$**

	$v_{sys} = w = 0$	P	$v_{sys} = w = 4/5$	P
<i>Speed</i> <i>Mass</i>	$v_1 = -4/5$	$-4/5$	$v_1 \oplus w = 0$	0
	$m_{v_1}^1 = 1 \text{ kg}$		$m_{v_1 \oplus w}^1 = 3/5 \text{ kg}$	
<i>Speed</i> <i>Mass</i>	$v_2 = 2/5$	$4/5$	$v_2 \oplus w = 10/11$	4.0
	$m_{v_2}^2 = 2 \text{ kg}$		$m_{v_2 \oplus w}^2 = 22/5 \text{ kg}$	
<i>Total P</i>		0	<i>Total P</i>	

- (15.4.3) **Two Computations for Momentum.** When $v_{sys} = w = 4/5$, Table (15.4.2a) gives us the momentum as the sum of each particle's momentum, namely,

$$(15.4.3a) \quad M = \underbrace{m_{v_1 \oplus w}^1}_{(3/5)\text{kg}} \cdot \underbrace{v_1 \oplus w}_0 + \underbrace{m_{v_2 \oplus w}^2}_{(22/5)\text{kg}} \cdot \underbrace{v_2 \oplus w}_{10/11} = 4.0.$$

But from the same table, we also have the momentum of the “single point” particle, $m_{v_1 \oplus w}^1 + m_{v_2 \oplus w}^2$, namely,

$$(15.4.3b) \quad M = \left[\underbrace{m_{v_1 \oplus w}^1}_{(3/5)\text{kg}} + \underbrace{m_{v_2 \oplus w}^2}_{(22/5)\text{kg}} \right] \cdot \underbrace{w}_{4/5} = 4.0.$$

The rest masses are

$$m_0^1 = m_{v_1}^1 \cdot \sqrt{1 - v_1^2} = 1\text{kg} \cdot \sqrt{1 - (-4/5)^2} = (3/5)\text{kg}$$

and

$$m_0^2 = m_{v_2}^2 \cdot \sqrt{1 - v_2^2} = 2kg \cdot \sqrt{1 - (2/5)^2} = (2/5)\sqrt{21}kg.$$

The increased masses at the speeds $v_i \oplus w$ are

$$m_{v_1 \oplus w}^1 = m_{-4/5 \oplus 4/5}^1 = m_0^1 = (3/5)kg$$

where Proposition (7.4.3) justifies $-4/5 \oplus 4/5 = 0$, and

$$m_{v_2 \oplus w}^2 = m_{2/5 \oplus 4/5}^2 = m_{10/11}^2 = (22/5)kg.$$

(15.4.4) The sum of the momentums of two particles equals the momentum of the sum (of particle masses), as is confirmed in (15.4.3a) and (15.4.3b). In Exercise (15.5.4), the reader constructs a program that generalizes Table (15.4.2a) by allowing two arbitrary mass inputs, an arbitrary speed $v_1 < 1$, and arbitrary system speed $w < 1$. Since the user has choice of masses and speeds, this program should give the reader a sense of how masses increase with speed.

(15.5) Exercises

(15.5.1) An Alternate Proof for $E = mc^2$. The force F is defined as the rate of change of the momentum p :

Calculus

$$F = \frac{dp}{dt}.$$

The work $W_{i \rightarrow f}$ required from an initial position x_i to a final position x_f is defined as the line integral of the force ²

$$W_{i \rightarrow f} = \int_{x_i}^{x_f} F dx$$

Then

²In more than one dimension this is the line integral along a path; in general the work will depend on the path. Forces for which the work is path-independent are called conservative forces.

$$\begin{aligned}
W_{i \rightarrow f} &= \int_{x_i}^{x_f} F dx \\
&= \int_{x_i}^{x_f} \frac{dp}{dt} dx \\
&= \int_{p_i}^{p_f} v dp \quad (\text{set } v = \frac{dx}{dt} \text{ in the previous integral}) \\
&= p v|_i^f - \int_{v_i}^{v_f} p dv \quad p = \frac{m_0 v}{\sqrt{1 - v^2/c^2}} \\
&= \left[\frac{m_0 v^2}{\sqrt{1 - v^2/c^2}} + m_0 c^2 \sqrt{1 - v^2/c^2} \right]_i^f \\
&= \left[\frac{m_0 c^2}{\sqrt{1 - v^2/c^2}} \right]_i^f \\
&= [m_v c^2]_i^f.
\end{aligned}$$

Finally, the work is also the change in the kinetic energy T :

$$W_{i \rightarrow f} = T_f - T_i$$

that implies $T = m_v c^2 + \text{constant}$. The kinetic energy vanishes at zero speed so

$$T = m_v c^2 - m_0 c^2 \Rightarrow m_v c^2 = T + m_0 c^2$$

that defines the rest energy $E_0 = m_0 c^2$ and the total energy $E = m_v c^2$:

$$E = E_0 + T.$$

Note that in terms of the momentum

$$E = \sqrt{p^2 c^2 + m_0^2 c^4},$$

which is valid even for particles with zero rest mass (the expression in terms of v is ambiguous because we let $m_0 \rightarrow 0$ while at the same time $v \rightarrow c$).

These definitions are useful because E is conserved. In any process where particles are free in the initial and final states the sum of the energies E (in terms of the momenta and rest masses) is conserved.

- (15.5.2) Exercise** **Derivatives.** Show that for $f(x) = (1-x)^{-1/2}$, the k th derivative $f^{(k)}(0)$ has the form

$$f^{(k)}(0) = \frac{(1 \cdot 3 \cdot 5 \cdots (2k-1))}{2^k}.$$

- (15.5.3)** **Mass Increases with Speed.** Show that



$$m_{v \oplus v} = m_0 \frac{1+v^2}{1-v^2} = m_0 \frac{1}{\sqrt{1-(v \oplus v)^2}}.$$

Hint: Definition (7.3.1) allows replacement of $v \oplus v$ with its equivalent, $v \oplus v = 2v/(1+v^2)$.

- (15.5.4) Exercise** **A Useful Spreadsheet.** Devise a spreadsheet, or a program in a language of your choice (e.g., C, C++, Matlab, Maple, Mathematica) to generate tables of the form (15.4.2a).

Hint: Construct your program with the following input/output properties:

USER INPUT:

(I₁) The first mass $m_{v_1}^1$, and its unitless speed $v_1 < 1$.

(I₂) The second mass $m_{v_2}^2$. The program computes $v_2 < 1$ so that momentum $m^1 v_1 + m^2 v_2 = 0$ is guaranteed (see first column of Table (15.4.2a)).

(I₃) The arbitrary “add-on” speed $w < 1$.

PROGRAM OUTPUT:

(O₁) Speed v_2 which guarantees that for the two-particle system, the initial momentum is zero (mentioned in (I₂)).

(O₂) The two at-rest masses, $m_0^i = m_{v_i}^i \sqrt{1-v_i^2}$, where $i = 1, 2$.

(O₃) The Lorentz sums, $v_1 \oplus w < 1$ and $v_2 \oplus w < 1$.

(O₄) The two masses with initial speeds, v_1 and v_2 , increased by w ,

namely, $m_{v_i \oplus w}^i = m_0^i / \sqrt{1 - (v_i \oplus w)^2}$, where $i = 1, 2$.

- (O₅) The momentum sum of the two individual particles,
 $[m_{v_1 \oplus w}^1] \cdot v_1 \oplus w + [m_{v_2 \oplus w}^2] \cdot v_2 \oplus w$.
- (O₆) The system momentum (*due to system speed $v_{sys} = w$*),
 $[m_{v_1 \oplus w}^1 + m_{v_2 \oplus w}^2] \cdot w$.

CHECK:

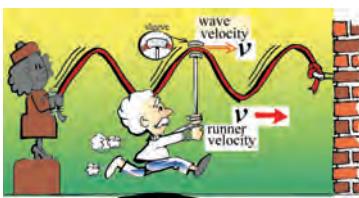
Outputs (O₅) and (O₆) should always be equal.



VII. The Mathematics of Waves and Light

Chapter 16 develops the mathematical model of propagated waves since light exhibits the properties of a propagated wave under certain circumstances.

Chapter 17 offers details of some of early and ingenious methods for measuring c , the finite speed of light. As recently as the time of Galileo and Newton, it had not been proven that the speed of light was finite. But today, using a microwave oven at home, you can measure the speed of light in a 20-second experiment (Exercise (17.6.6d), pg. 237).



(16) The Nature of Waves

Preview

A fundamental starting point of Einstein's paper, [12], is that the speed of light is the same for all observers, whether or not the light source is moving. Equivalently (and non-intuitively), the speed of light is a constant 186,000 miles per second (in a vacuum) whether or not the observer is moving (see (1.5.4)).

Why should we believe such a non-intuitive idea? The following chapters provide the answer.

The constant speed of light, independent of frame of reference (point of view) was predicted decades before Einstein's paper, by Maxwell's equations which connected electricity, magnetism, and optics. These equations also predicted that the speed of all electromagnetic (*EM*) waves propagated in a vacuum was 3×10^8 meters/second—exactly the speed of light! These results are presented in Chapters (18) through (23).

(16.1) Propagated Waves (Calculus Required)

(16.1.1) **Anatomy of a Wave:** We consider a shape defined by a real-valued function, g , to be a **wave** traveling in one dimension if it has the following properties:

(1) **Wave length λ :** At any fixed time, t_0 , its shape, $y = g(x, t)$ in the X - Y plane replicates itself over all successive intervals of length λ . For any x , such intervals take the form

$$(16.1.1a) \quad \underbrace{[x + k\lambda, x + (k + 1)\lambda]}_{\text{length } \lambda}, \quad k = 0, \pm 1, \pm 2, \dots$$

Formally stated, for fixed time, t_0 , for all x , and all integers k , we have $g(x, t_0) = g(x + k\lambda, t_0)$

(2) **Wave speed v :** The wave travels to the right, in the positive direction, with speed v . Consider a wave $g(\cdot, t_0)$ at fixed time t_0 . As explained in Section (16.3), at wave speed v , after time Δt ,

the full wave will shift to the right a distance $v\Delta t$. That is,

$$(16.1.1b) \quad \text{For all } x, g(x, t_0) = g(x + v\Delta t, t_0 + \Delta t).$$

Equation (16.1.1b) says that for all x , the y -value of wave g at x_0 and time t_0 is the same as the y -value of g at $x + v\Delta t$ (*the rightward shifted distance*) at the later time $t_0 + \Delta t$.

(16.1.2)

DEFINITION: In a traveling wave with WAVELENGTH λ and SPEED v , we define the PERIOD P to be the time for one wavelength (one cycle) to pass an arbitrary point. The FREQUENCY f is the number of full cycles of function g that pass any fixed point in one second. Period P , frequency f , wavelength λ , and speed v , are related by the formulas

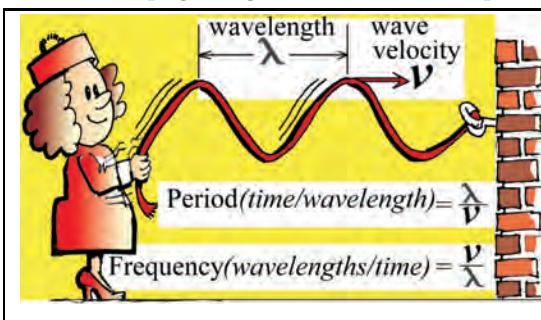
$$(16.1.2a) \quad P = \frac{\lambda}{v}, \quad f = \frac{v}{\lambda}, \quad \text{and} \quad P = \frac{1}{f}.$$

NOTE how the units of (16.1.2a) are consistent with Appendix (D). Since the unit for λ is distance, d , and speed, v , is in units of d/t , the period, $P = \lambda/v$ is in units of t alone. This is the time for one full cycle of g to pass an arbitrary point. Also, frequency, f , is in units of $1/t$, or “per-second,” which is the number of cycles of g that pass an arbitrary point per second.

We analyze a traveling wave in a rope that is tied at one end.

(16.1.3)

FIGURE: Propagating a Wave in a Rope



Ashley oscillates the loose end of a rope to generate a traveling wave that has velocity v . If Ashley keeps the tension constant, but increases the “hand speed,” or energy to the rope

by oscillating it faster, the wave will not travel any faster or slower. The wave length, λ , and the period, P , will decrease, but the wave speed, v , will be constant. Equivalently, the frequency, f , will increase, while the wave speed remains the same.

(16.1.4)

THEOREM: (*Horizontal wave speed is independent of vertical “hand speed.”*) One end of a rope is tied to a wall while the other (loose) end oscillates vertically so as to generate a wave that travels toward the wall (as illustrated in Figure (16.1.3)). The rope

- has linear density δ (ounces per inch, say) and
- is pulled with constant horizontal tension T (pounds of force, say).

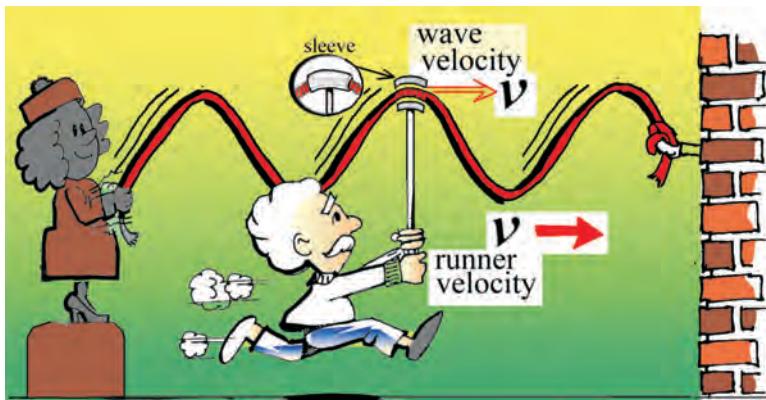
Then, if gravitational pull on the rope is ignored, the speed v of the propagated wave is

$$(16.1.4a) \quad v = \sqrt{T/\delta}.$$

PROOF: Assume the wave travels to the right as indicated in Figure (16.1.3). To measure the rightward wave speed v , it will be useful for Bernie, the runner, to keep pace with the wave, running to the right at identical speed, v . Hence, from Bernie’s point of view, the rope coils are going leftward with speed $-v$ while the captured wave peak appears stationary with speed 0 as seen in Figures (16.1.5) and (16.1.6)).

(16.1.5)

FIGURE: The Wave Peak Captured in a Curved Sleeve.

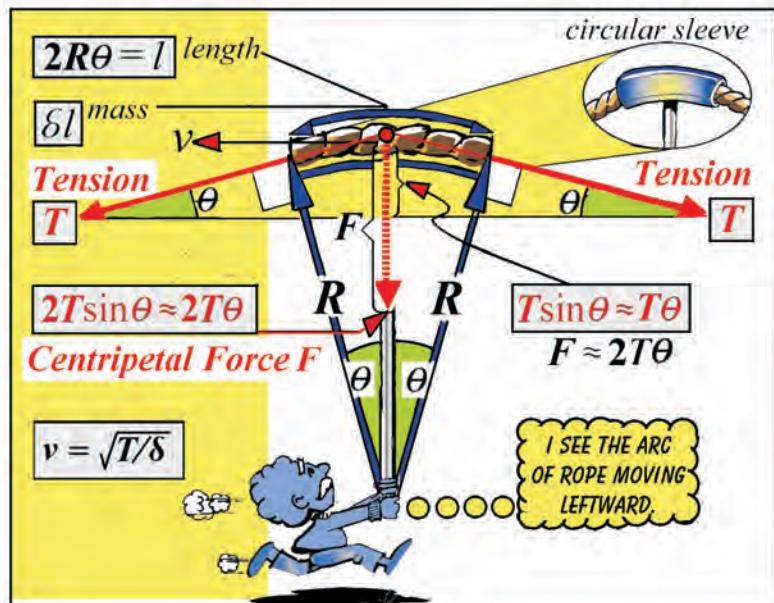


Bernie runs alongside the wave at the same speed of the propagated wave, capturing the wave peak in a curved tubular sleeve. The rope does not touch the sleeve’s inside surface.

There are two ways to calculate the centripetal (vertical downward) force illustrated in Figure (16.1.6): (i) with Newton's law $F = ma$ as seen in (16.1.6a)-(16.1.6c), or (ii) by decomposing the almost-horizontal tension force T as in (16.1.6d). The two analyses are merged to produce the conclusion (16.1.6e).

(16.1.6)

Trig needed

FIGURE: Two Ways to Measure Centripetal Force

First Analysis of Centripetal Force, F . When the angle θ is small enough, both the rope segment and the sleeve approximate a small circular arc with radius R . At this small scale, the rope segment with length $l = 2R\theta$ and linear density δ gives us the following value of the rope segment mass:

$$(16.1.6a) \quad m = (l)\delta = (2R\theta)\delta \quad \{ \text{mass of rope} \} .$$

The rope in Figure (16.1.6) runs through the sleeve leftward on a circular path. From Newton's law, any mass on a circular arc of radius R with constant speed v is a result of a (centripetal) acceleration

$$(16.1.6b) \quad a = \frac{v^2}{R} \quad \left\{ \begin{array}{l} \text{see Exercise (16.7.1)} \\ \text{for a proof of this fact} \end{array} \right\} .$$

If we take the mass m from equation (16.1.6a) and the (circular) acceleration a from equation (16.1.6b), then we can write the centripetal force, $F = ma$, as

$$(16.1.6c) \quad F = \underbrace{\frac{2R\theta\delta}{1}}_m \times \underbrace{\frac{v^2}{R}}_a = 2\theta\delta v^2.$$

Second Analysis of Centripetal Force, F . Now the centripetal force F of (16.1.6c) is also obtained from Figure (16.1.6) by decomposing the opposing tension vectors with length T into their horizontal and vertical components. The horizontal components are equal and opposite so they cancel.

The vertical components of the tension vectors, on the other hand, both point downward and reinforce each other. They each have length $T\sin(\theta)$, so their downward sum is $2\sin(\theta)$. Since for small θ we have $\sin(\theta) \approx \theta$, we may write

$$(16.1.6d) \quad F = 2T\theta$$

for the centripetal force on the rope segment where T is the initial tension on the ends of the rope.

Finally, equate F of (16.1.6c) with F of (16.1.6d) to obtain

$$(16.1.6e) \quad v = \sqrt{T/\delta}$$

which is the equation (16.1.4a) we set out to prove. ■

Note: Exercise (16.7.2) asks you to confirm that the dimensions of velocity—distance over time—result from an analysis of the right-hand side of (16.1.4a) which is expressed only in terms of the tension force, T , and the linear mass-density, δ , of the rope.

(16.2) Speed of Rope Wave is Constant

Your intuition might tell you that the speed of the wave propagated will increase if you oscillate the rope faster (increase the frequency). Not so. As long as the tension T is constant, the speed of the traveling wave remains the same.

Light waves, or more generally, electromagnetic waves, share a similar property. The speed of light in a vacuum is constant, no matter what its frequency (color).

A rope is composed of matter and light is not. Nevertheless, they share the common property that their speed of propagation does not depend on the frequency carried by the wave.

A guitar string is an example of how the tension, T , affects the frequency. According to (16.1.4a), tightening the string (increasing tension) will increase the string's wave speed v which corresponds to a higher frequency f , or a higher pitch.

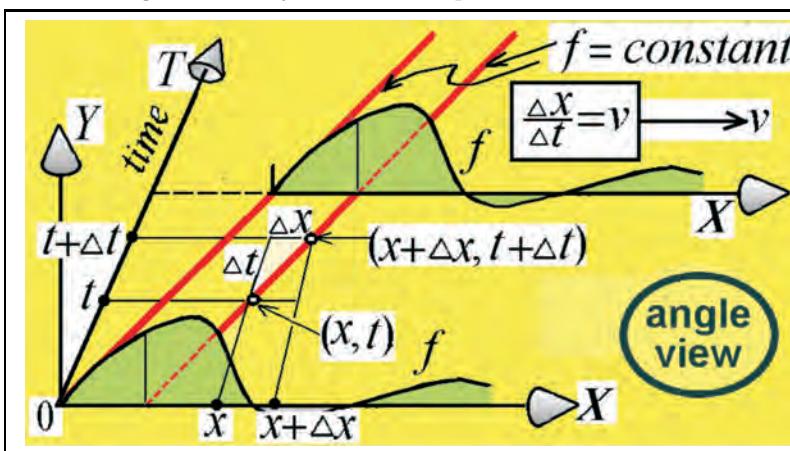
(16.3) Shapes Traveling in One Dimension

Now that we see that a wave in a rope tends to maintain a constant speed v , we characterize that differential equation, called the wave equation (16.4), that describes shapes traveling at speed v in one direction.

- (16.3.1) **Notation for Curves Traveling to the Right:** What is the proper notation to describe how a real-valued function, f , of one variable, travels to the right along the X -axis over a distance Δx units in time Δt ? (The speed of rightward travel is then necessarily $\Delta x/\Delta t = v$.)

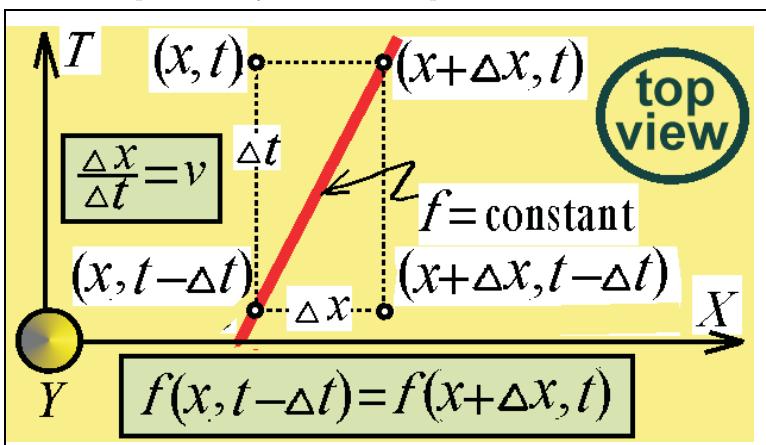
We begin with the following graphs:

(16.3.2) **FIGURE: Rightward Traveling Curve:**
Angle View of X-Y-T Graph.



- The Y values, $f(x, t)$, of the moving curve form a surface over the X - T SPACETIME plane. Since each point of the curve travels rightward on the X -axis with speed v , every vertical Y value of the surface is constant on a velocity line traced in the X - T plane. Two such velocity (red) lines of constant value are shown.

(16.3.3) **FIGURE: Rightward Traveling Curve:**
Top View of X-Y-T Graph.



- This is the “floor” shown in Figure (16.3.2). The Y -axis points up and out of the page. As we see from the graph, the lines in the X - T plane

over which the surface values for f are constant, require that

$$(16.3.3a) \quad f(x, t - \Delta t) = f(x + \Delta x, t), \text{ whenever } \frac{\Delta x}{\Delta t} = v.$$

NOTE: To achieve the form of (16.3.3a) which appears in Figure (16.3.3), replace t with $t - \Delta t$ in Figure (16.3.2).

(16.3.4)

DEFINITION: Given two-variable function $f(x, t)$. Then the limits

$$(16.3.4a) \quad \frac{\partial}{\partial x} f(x, t) \stackrel{\text{def}}{=} \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x, t) - f(x, t)}{\Delta x} \text{ and}$$

$$(16.3.4b) \quad \frac{\partial}{\partial t} f(x, t) \stackrel{\text{def}}{=} \lim_{\Delta t \rightarrow 0} \frac{f(x, t + \Delta t) - f(x, t)}{\Delta t}$$

are the PARTIAL DERIVATIVES OF f at point (x, t) with respect to the variables x and t , respectively.

(16.3.5)

MOTIVATION: Definition (16.3.4) introduces symbols $\partial f / \partial t$ and $\partial f / \partial x$, the partial derivatives of $f(x, t)$, which are the usual derivatives, df/dx and df/dt , with all variables of f fixed except x and t , respectively. Use of the “curly” ∂ instead of the ordinary d is a reminder that function f has more than one variable.

(16.3.6)

NOTE: Equivalent variations of symbol $\frac{\partial}{\partial x} f(x, t)$ (16.3.4a) are

$$\frac{\partial f}{\partial x}, \quad \partial f / \partial x, \quad \partial f(x, t) / \partial x, \quad \partial / \partial x f(x, t).$$

(16.3.7)

LEMMA: For the function (moving wave) $f(x, t)$ traveling rightward in time with speed v , the partial derivatives $\partial f / \partial x$ and $\partial f / \partial t$ are related as follows for all (x, t) :

$$(16.3.7a) \quad \frac{\partial}{\partial x} f(x, t) = -\frac{1}{v} \frac{\partial}{\partial t} f(x, t),$$

or, in symbolic short-hand,

$$\frac{\partial}{\partial x} = -\frac{1}{v} \frac{\partial}{\partial t}$$

where existence of the rightward traveling function f , as t increasing, is assumed on both sides of the equal sign.

PROOF: For our two-variable function $f(x, t)$,

$$\begin{aligned}
 (16.3.7b) \quad \frac{\partial}{\partial x} f(x, t) &\approx \frac{f(x + \Delta x, t) - f(x, t)}{\Delta x} \quad \text{Def. (16.3.4a) of } \frac{\partial}{\partial x} \\
 &= \frac{f(x, t - \Delta t) - f(x, t)}{\Delta x = v \Delta t} \quad \text{Equation (16.3.3a)} \\
 &= -\frac{1}{v} \cdot \frac{f(x, t - \Delta t) - f(x, t)}{-\Delta t} \quad \text{Factor out } -\frac{1}{v} \\
 &\approx -\frac{1}{v} \cdot \frac{\partial}{\partial t} f(x, t), \quad \text{Def. (16.3.4b) of } \frac{\partial}{\partial t}
 \end{aligned}$$

which ends the proof. ■

Differences such as $f(x + \Delta x, t) - f(x, t)$, which we see in (16.3.7b), occur often enough to warrant a more compact notation, which we now present.

(16.3.8) **NOTE:** For a real-valued function, $f(x, t)$ at a point, (x, t) , in the spacetime plane, we use the symbols

$$(16.3.8a) \quad \Delta f_x \stackrel{\text{def}}{=} f(x + \Delta x, t) - f(x, t) \quad \text{when } t \text{ is held fixed,}$$

$$\Delta f_t \stackrel{\text{def}}{=} f(x, t + \Delta t) - f(x, t) \quad \text{when } x \text{ is held fixed.}$$

This notation is used in the following proposition that says: If f increases (or decreases) as x increases, then f necessarily decreases (or increases) as time t increases. More formally, we have

(16.3.9)

PROPOSITION: Given a real-valued function traveling to the right over the X -axis with speed v . for any point (x, t) in spacetime,

$$(16.3.9a) \quad \text{sign}(\Delta f_t) = -\text{sign}(\Delta f_x).$$

That is,

$$\Delta f_x > 0 \text{ implies } \Delta f_t < 0 \text{ and } \Delta f_x < 0 \text{ implies } \Delta f_t > 0.$$

PROOF: Write

$$\frac{\Delta f_x}{\Delta x} = \frac{f(x + \Delta x, t) - f(x, t)}{\Delta x} \quad \text{Notation (16.3.8)}$$

$$\approx \frac{\partial}{\partial x} f(x, t) \quad \text{Def. (16.3.4a) of } \frac{\partial}{\partial x}$$

$$= -\frac{1}{v} \frac{\partial}{\partial t} f(x, t) \quad \text{Equation (16.3.7a)}$$

$$\approx -\frac{1}{v} \frac{f(x, t + \Delta t) - f(x, t)}{\Delta t} \quad \text{Def. (16.3.4b) of } \frac{\partial}{\partial t}$$

$$= -\frac{1}{v} \frac{\Delta f_t}{\Delta t}. \quad \text{Notation (16.3.8)}$$

Since $\Delta x > 0$ and $\Delta t > 0$, the first and last entries in this string of equalities tell us that $\text{sign} \Delta f_x = -\text{sign} \Delta f_t$. The proof is done. ■

(16.3.10) **NOTE:** The “arcane” mathematical fact, (16.3.9a), that the sign of the time change is opposite to that of the distance change, becomes an important physical fact when Maxwell uses it to determine c , the speed of light (Section (23.4)).

(16.4) The Wave Equation in One Dimension

(16.4.1)

Calculus
required

THEOREM: (*One-dimensional wave equation.*) Given f , a sufficiently differentiable real-valued function traveling rightward with constant speed v . Then

$$(16.4.1a) \quad v^2 \frac{\partial^2}{\partial x^2} f(x, t) = \frac{\partial^2}{\partial t^2} f(x, t)$$

where $\frac{\partial^2}{\partial w^2} f = \frac{\partial}{\partial w} \left(\frac{\partial}{\partial w} f \right)$ for $w = x$ or t . This is the

ONE-DIMENSIONAL WAVE EQUATION.

PROOF: The equality

$$(16.4.1b) \quad \frac{\partial}{\partial x} \left(\frac{\partial}{\partial t} f(x, t) \right) = -\frac{1}{v} \frac{\partial}{\partial t} \left(\frac{\partial}{\partial t} f(x, t) \right)$$

is a rewriting of (16.3.7a) in which the function $f(x, t)$ has been replaced with another function, $\frac{\partial}{\partial t} f(x, t)$. If, in (16.4.1b), we replace $\frac{\partial}{\partial t}$ on the left-hand side with $-v \frac{\partial}{\partial x}$ (again, from (16.3.7a)), the result is

$$(16.4.1c) \quad -v \frac{\partial}{\partial x} \left(\frac{\partial}{\partial x} f(x, t) \right) = -\frac{1}{v} \frac{\partial}{\partial t} \left(\frac{\partial}{\partial t} f(x, t) \right),$$

which, when multiplied through by $-v$, yields the desired wave equation (16.4.1a). ■

(16.4.2)

Solutions to the Wave Equation. The preceding development assumes only that the real-valued function, f , travels to the right over the 1-dimensional X -axis. If we impose periodicity (the shape of f repeats over every interval of length L), or certain initial conditions (specifying the shape of f over the $[0, l]$ interval at time $t = 0$), or specific boundary conditions (for all time t , $f(x) = 0$ for $x = 0, L, 2L, \dots$), then specialized techniques exist for finding the appropriate periodic solutions.

(16.5) Wave Propagation: The Skipping Stone Model

Suppose an individual, the “source” of a ball, is stationary (source speed = $v_s = 0$). The source individual throws the ball at a speed of 10 mph relative to herself (ball speed = $v_w = 10$ mph). A stationary observer reckons the speed of the ball to be $w_w = 10$ mph.

Now suppose the source moves at speed 3 mph (source speed = $v_s = 3$ mph). In this case, the stationary observer sees the ball moving with speed (source speed + ball speed = $v_s + v_w = 3$ mph + 10 mph) = 13 mph. In other words, the speed of a particle is affected by the speed of its emitting source.

In contrast, the speed of a propagated wave is not affected by the speed of its emitting source.

.....

Before developing this invariant property of propagated waves, we introduce the following notation.

(16.5.1)

NOTATION: We will use the following notation:

- v_w is the speed of a wave relative to its carrying medium.
- v_s is the speed of a wave's source relative to the medium.
- v_m is the speed of the carrying medium of the wave relative to an observer.

Now instead of a ball being thrown from a moving platform, consider a sound wave (a train whistle) being “thrown” from a moving train whistle. Assume there is no wind so the speed of air, the carrying medium, is $v_m = 0$.

The source of the sound wave, the train whistle, has speed $v_s = 100$ mph relative to an observer. The speed of sound waves in air, relative to the observer, is $v_w = 750$ mph. Will the sound of a train’s whistle reach us at $v_s + v_w = 850$ mph? The answer is no.

Unlike particles, wave speeds are not affected by the speed of the wave source. As long as the observer and the air (*the wave's medium*) are not moving relative to each other, the speed, v_s , of the sound's *source* does not affect the speed, v_w , of the wave relative to the medium. In other words,

(16.5.2)

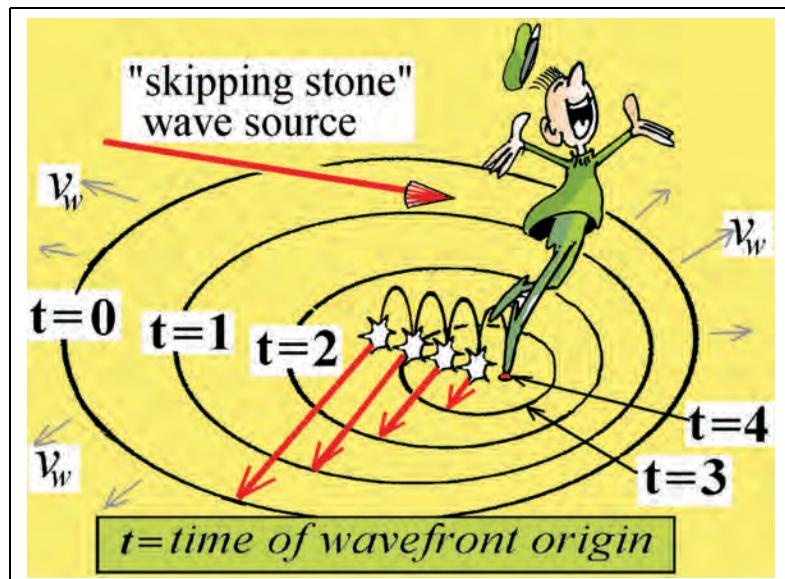
PROPOSITION: The speed of a wave through a stationary medium is constant no matter what the speed of its source is.

To see why this is so, consider the “skipping stone model” for propagated waves.

(16.5.3) **The speed of a wave source v_s relative to the medium is not transmitted to the speed of the propagated wave v_m relative to the medium.** For motivational purposes, Figure (16.5.4) presents a simplified model of circular ripples or wave fronts (*instead of spherical ripples*) propagated on a 2-dimensional space or “table-top” by a dancer whose every step produces a new expanding ripple from the point of contact.

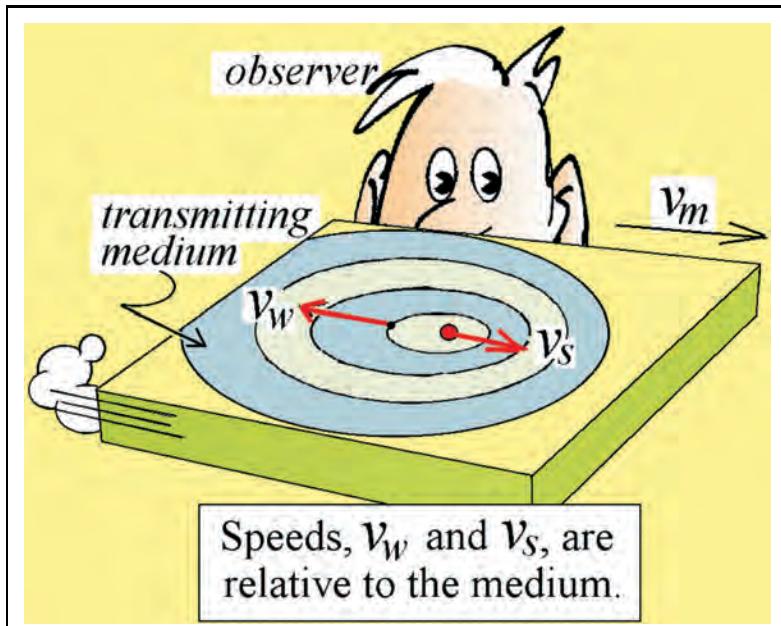
(16.5.4)

FIGURE: The Skipping Propagator of Waves.



Imagine the dancer’s floor with infinitely many uniformly distributed buttons, each of which, when pressed, generates one expanding ripple. All ripples in the medium (table-top) expand at the same speed, v_w from the center. Whenever the dancer taps any one of the fixed buttons, that button doesn’t care how fast the dancer is prancing overhead—the expansion of the radius of each disk relative to its generating button is always v_w .

- (16.5.5) The speed of the medium—not the source—is transmitted to the speed of the propagated wave. We have noted in (16.5.3) that to an outside observer, the speed of a wave, v_w , is always the same regardless of the speed of the wave source, v_s —provided that v_m , the speed of the transmitting medium (air) is zero relative to the observer.



Now if the medium is moving at speed $v_m \neq 0$ relative to the observer, then the entire medium-wave-source “unit” is, likewise, carried along at speed v_m relative to the observer. Thus, $v_m + v_s$ is the source speed, and $v_m + v_w$ is the wave speed, relative to the observer. The relative speed of the medium *is* transmitted to the speed of the wave (and to the speed of the source).

(16.5.5a)

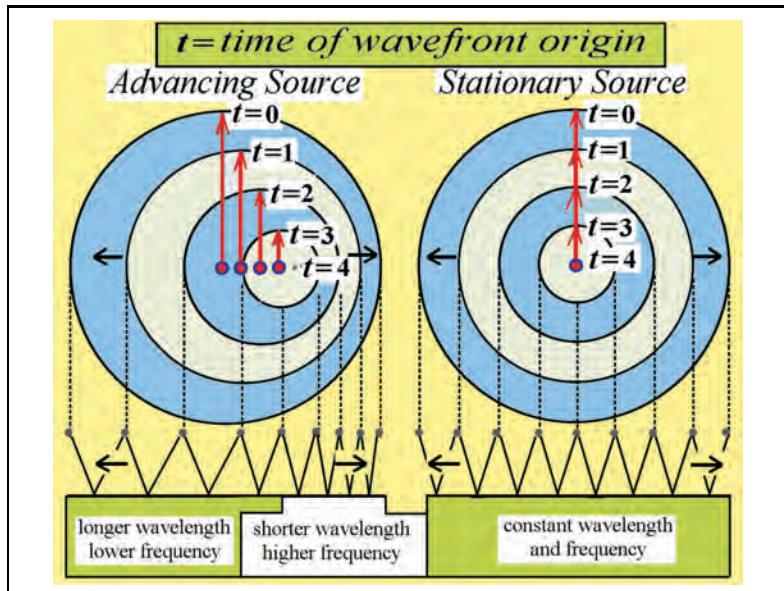
Note: Since the medium speed v_m and the source speed v_s are tiny compared to the speed of light, c , there is a minuscule difference between the Galilean/Newtonian arithmetic sum $v_m + v_s$ and the true Lorentzian or relativistic sum, $v_m \oplus v_w$ (1.13.3), (7.3.1), whose value, as we shall see, was experimentally confirmed by Fizeau (13.2).

(16.6) The Doppler Effect in Spacetime

- (16.6.1) **A Qualitative View of Frequency Shift.** If the speed of the waves (ripples) does not change with the speed of the source, then does the source speed have any observable effect at all?

To fix ideas, and for convenience, we may think of the waves as sound waves traveling through air. As Figures (16.5.4) and (16.6.2) illustrate, an observer who sees (more exactly, hears) the source approaching will perceive an increase of the frequency (the ripples bunch up) while an observer who sees the wave source departing, will hear a decreased frequency (the ripples spread out).

- (16.6.2) **FIGURE: Shifting Ripple-Centers Affect Frequency**



From experience, we know that as a train's whistle approaches, we hear one pitch, but as soon as the train passes by, we hear another, lower pitch.

(16.6.3)

DEFINITION: If f is the frequency of a wave with a stationary source, then the increase or decrease of wave frequency f induced by relative speed of the source is known as the DOPPLER EFFECT.

(16.6.4)



As we shall see, the listener will hear only two distinct pitches, (16.6.6a), from the whistle if the train has constant speed on a 1-dimensional track and the observer is standing in the train's path. (To ensure that the observer actually gets to hear the second tone, we suppose he or she stands very close to, but not on, the track. However, this deviation from the 1-dimensional ideal will produce two almost-constant tones separated by a brief, continuous slide from the higher to the lower pitch.)

(16.6.5)

A Quantitative View of Frequency Shift. The following theorem quantifies the value of the two tones that are heard by a stationary observer from a wave-emitting source traveling at constant speed, v_s .

(16.6.6)

THEOREM: (Doppler Shift: Fixed Observer, Moving Source) A wave has speed v_m relative to a transmitting medium that is stationary with respect to an observer. If the observer measures the wave source to have speed $v_s = 0$, then the observed frequency is f_0 . If the observer measures the wave source to have speed $v_s > 0$, then the observed frequency is

$$(16.6.6a) \quad f' = \begin{cases} f_0 \cdot \frac{v_m}{v_m - v_s} & \text{while the source} \\ & \text{approaches the observer} \\ f_0 \cdot \frac{v_m}{v_m + v_s} & \text{while the source} \\ & \text{recedes from the observer.} \end{cases}$$

PROOF: In Figure (16.6.6c), $P = 1/f_0$ is the period of the wave (time for one wavelength to pass a given point).

At time $t = 0$, the wave crest (wave maximum) emanates from source Point (0).

At time $t = P$, as shown in the top view in Figure (16.6.6c), the wave crest, traveling with speed v_m , has spread from Point (0) to a

circle with radius $distance = speed \times time = v_m \times P$. Points (a) and (a') lie on this circle.

Also at time $t = P$, the source Point(0) has traveled to Point (b) with speed v_s . The source will now transmit another wave crest (*maximum*). The distance traveled by the source is $distance = speed \times time = v_s \times P$.

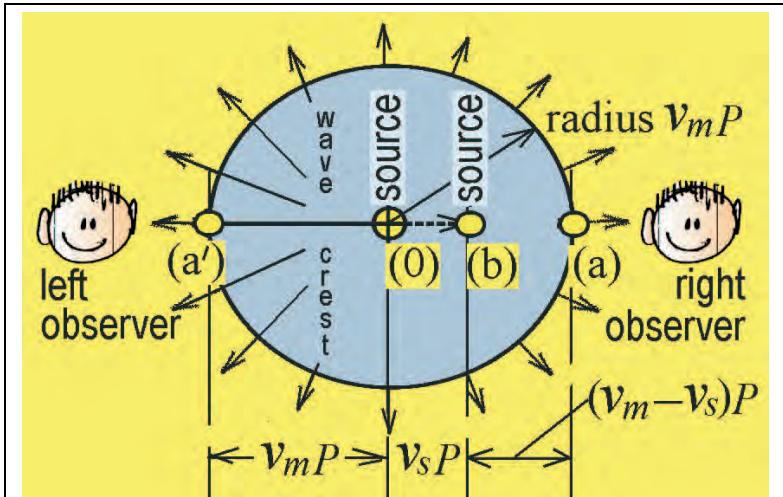
Figure (16.6.6c) shows a full wave length between Points (b) and (a), and another full wave length between Points (b) and (a') since these points represent the positions of the wave crests at time $t = P$. As shown in Figure (16.6.6c), the distance between Points (b) and (a), i.e., a wavelength, is $\lambda_{right} = (v_m - v_s) \times P$. From Definition (16.1.2), the frequency as seen by the **right observer**, is

(16.6.6b)

$$f'_{right} = \frac{\text{wave speed}}{\text{wavelength}} = \frac{v_m}{\lambda_{right}} = \frac{v_m}{(v_m - v_s)P} \frac{1}{P} = \frac{v_m}{(v_m - v_s)} f_0.$$

This proves the first equation of (16.6.6a).

(16.6.6c) **FIGURE: Top View of Wave Crest Propagated After P units of Time**

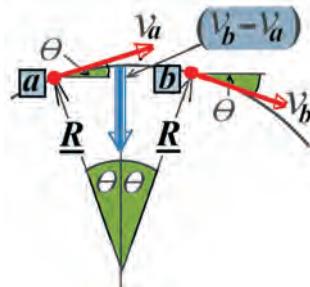


To prove the second equation of (16.6.6a), we establish that the wavelength $\lambda_{left} = (v_m + v_s) \times P$ and then argue as we did in (16.6.6b). This is left as Exercise (16.7.4). The proof is done. ■

(16.7) Exercises

(16.7.1) Exercise

$v^2/R = \text{Centripetal Acceleration}$. A particle travels clockwise with speed $v = v_a = v_b$ on an arc with radius R as it subtends an angle of 2θ . It takes t seconds to travel from point a to point b . If the acceleration is defined as $[\text{vector}(v_b) - \text{vector}(v_a)]/t$, then show that the acceleration is directed toward the center and has magnitude v^2/R .



(16.7.2) Exercise

Dimensional Consistency. Show that (16.1.4a) is dimensionally consistent. That is, both quantities,

$$v \quad \text{and} \quad \sqrt{T/\delta},$$

have common dimension, namely, *distance/time*.

Hint: The dimension of the force of tension, T , is (mass \times acceleration) and the dimension of linear density, δ , is (mass/length).

(16.7.3) Exercise

Speed and rpm. Given that a wheel has radius R and is turning at ρ rpm (revolutions per minute). What is the speed of a point on the circumference?



(16.7.3a) Why, in Exercise (16.7.3), do we use symbols R and ρ rather than specific numbers like “7.3 feet” or “2,300 rpm?” The answer is that whether symbols or specific numbers are used, the time and energy required to solve this problem are identical since the manipulations in each case are identical. The advantage of using symbols is that one solution provides a valid template for infinitely many cases, each case corresponding to one of the infinitely many values for R and ρ . But a solution for particular numbers such as “7.3 feet” and “2,300 rpm”, gives you that solution and no other.

(16.7.4) Exercise

Following the pattern of (16.6.6b), prove the second equation of (16.6.6a).

(17) Measuring the Speed of Light

(17.1) Early Thoughts on the Speed of Light

(17.1.1) Why Is c Used to Represent the Speed of Light in a Vacuum?



One credible reason is that c is the first letter of CELERITAS, a Latin word that means “speed.” However, from Maxwell’s writings in 1865 through Einstein’s papers in 1905, the letter V —not c —was used for the speed of light.

During this time, the symbol, c , had been waiting in the wings on the stage of history. As early as 1856, the letter c , as related to the speed of light, was used by Weber and Kohlrausch in their work on an electrodynamics force law. Eventually, as the speed of light came to be recognized as a unique, universal constant, different from any other speed, it was awarded its own symbol, c , as opposed to the more general, V . (*More of this story can be found in an essay by PHILIP GIBBS on the website of JOHN BAEZ whose URL is <http://math.ucr.edu/home/baez/physics/Relativity/SpeedOfLight/c.html>.*)

(17.1.2) The Uniqueness of c .

A fundamental starting point of Einstein’s paper, [12], is that the speed of light is the same for all observers, whether or not they or the light source are moving with respect to each other (see (1.5.2b), pg. 16).

Constant wave speed for a moving source is no surprise. The skipping stone model of wave propagation (16.5.4) shows how waves in a stationary medium always have constant speed regardless of the speed of the source. If the source of sound (like a moving train whistle) does speed up, then the effect is to change the perceived *frequency*, not the *speed* of the sound wave. An observer measures the speed of the train’s sound to be constant—no matter what the speed of the train—as long as the medium (air) is not moving relative to the observer.

Constant wave speed for a moving observer is a surprise. The speed of sound from a moving train whistle will change as the speed of the carrying medium (air) changes relative to the observer.

Windy days produce different measures for the speed of sound.

- (17.1.3) **Waves and their transmitting media.** Sound waves are expressed as traveling variations of pressure in a medium (air, for example). Water waves correspond to traveling variations in the weight of their medium (water). Intuition, therefore, suggests that light, as a wave, also needs a medium to propagate, just as sound requires air; the appropriate medium through which light was supposed to propagate was called the LUMINESCENT AETHER.

But Einstein went against tradition when he proposed that light was a different kind of wave that needed *no transmitting medium at all*. In fact, existence of the aether was generally accepted until Einstein's work in 1905 [12]. Moreover, said Einstein, the speed of light was absolute — independent of the state of motion of either observer or source (see (1.5.4)).

- (17.1.4) **Why should we believe the non-intuitive idea** that the speed of light does not depend on the speed of either the source or the observer?

Even though the value of c , the speed of light in a vacuum, was shown by Maxwell's equations in 1864 (21.7.3), pg. 300, to be independent of frames of reference, and even though this property was theoretically assumed by Einstein in his seminal paper of 1905, experimental confirmation is always required before any idea of physics can be fully accepted. Indeed, from the Michelson-Morley experiment of 1887 (17.5.3) to the present day, an untold number of experiments have confirmed the independence of the value of c with increasing accuracy.

(17.2) Rømer: The Speed of Light is Finite

Similar to the discussion of (17.1.4), we may ask why we should believe that the speed of light is only finite when our day-to-day experience and intuition tell us otherwise. On one hand, intuition tells us that sound waves have finite speed. After all, we perceive the cracking or pinging sound of a baseball hitting a bat *after* we see the ball being hit. On the other hand, there is no event in our daily

experience for which we sense a noticeable time delay with respect to light waves.

What experimental evidence could convince us to override our intuition?

- (17.2.1) **Unexpected Evidence.** In 1675, Danish astronomer, OLE (Olaus) RØMER (1644-1710) observed that about every 42 hours, Io, the innermost of Jupiter's satellites (moons), would regularly disappear and reappear from behind the planet. Rømer set out to graph Io's times between disappearances, based on the implication:

- (17.2.2) **Rømer's Hypothesis.**

Hypothesis: Light speed is infinite	\implies	Conclusion: We have instant observation of Io.
--	------------	---

- (17.2.3) **Logic of the Scientific Method.** In the interpretation of Rømer's diagram (17.2.4) we shall later see that the conclusion (*right hand side*) of the implication (17.2.2) is FALSE, which means the hypothesis (*left hand side*) must also be FALSE. This follows since a TRUE hypothesis can never imply a FALSE conclusion. In other words, the implication, (TRUE \implies FALSE), is never allowed.

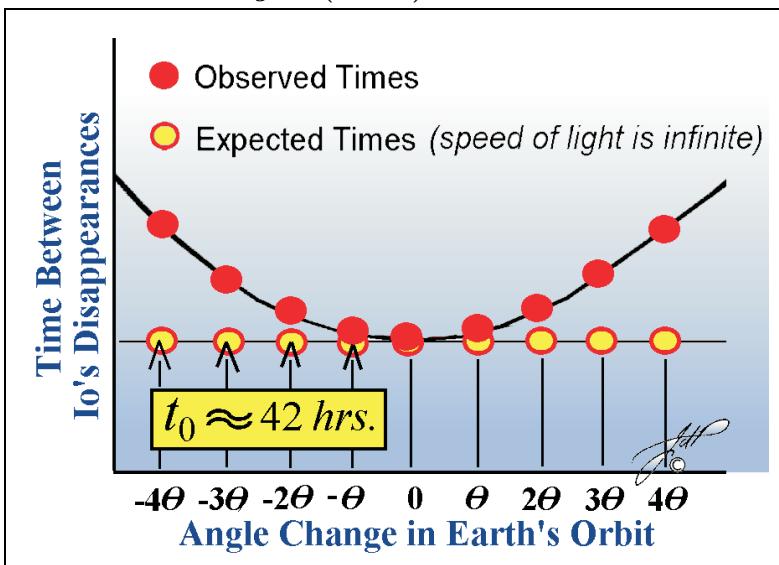
But to say the hypothesis of (17.2.2) is FALSE is to say that the speed of light is bounded or finite.

Here is how this analysis proceeds:

If the conclusion of (17.2.2) is TRUE, then the image of Io is observed instantly which says that the observed time between two disappearances must be a constant t_0 hours.

The following graph shows both the experimentally observed and the expected (*constant*) times between disappearances of Io. Equal intervals of Earth time correspond to the equal angular increments θ of the Earth's orbit as shown in Figure (17.2.6), pg. 220.

- (17.2.4) **FIGURE:** *Times Between Events (Observed and Expected).*
Also See Figure (17.2.6).



If times between disappearances of Io were a constant $t_0 \approx 42$ hours, then the observations would be plotted on the graph with the open circles ○. In fact, the intervals between the disappearances increase and decrease at different times of the year, as indicated by the solid circles, ●.

- (17.2.5)

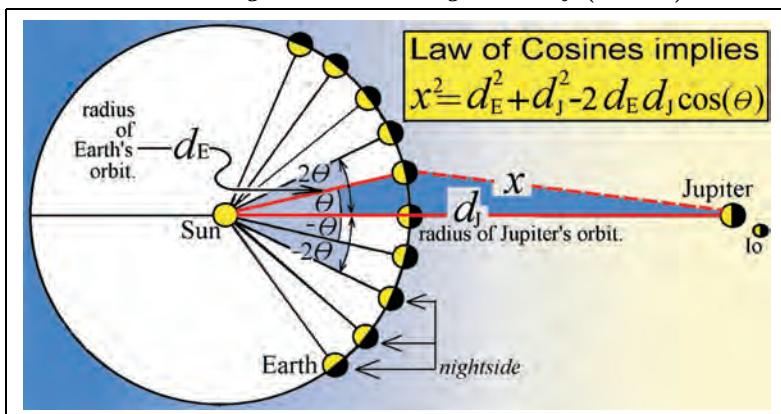


It is fortunate that Io has such a short 42-hour “year” or period about Jupiter. The briefer the orbital period of a satellite, the more “disappearance events” occur in an Earth year and hence, more data can be recorded.

Rømer was in the right laboratory at the right time. We can only speculate — if Io had not existed, would Rømer have invested time in Jupiter’s slower satellites?

Without Io, what orbital periods would Rømer have to work with? Here is the list: Europa = 3.6 days, Ganymede = 7.2 days, and Callisto = 16.7 days.

(17.2.6)
Trig
required

FIGURE: Rømer Diagram Predicting Data of (17.2.4).

Since light requires a finite (non-zero) time to travel between any two points, it is the varying distances of Io from Earth — due to Earth's changing positions in orbit—that account for the varying times between Io's disappearances behind Jupiter.

Discussion. Simplifying assumptions are made with respect to the analysis of Figure (17.2.4). It is assumed that

- Jupiter is stationary since the Earth orbits the sun much more quickly than does Jupiter.
- All planetary and satellite orbits are circular.
- The distance between Io and Jupiter is so much smaller than the distance from Jupiter to Earth that one can say that, as viewed from Earth, Jupiter and Io are seen almost at the same spot.
- After each 42-hour interval, there is a measurable *change* in distance between Io and Earth in its circular orbit.

Rømer went further. Based on the diagram (17.2.6), he computed the finite value $c \approx 220,000$ km/sec,¹ the speed of light necessary to account for the observed curved path of Figure (17.2.4) as opposed to the straight-line path that would be caused by a value $c = \infty$.

(17.3) Fizeau Measures the Speed of Light

The inaccuracy of Rømer's estimate of 220,000 km/sec for the speed of light was due to uncertainties in the distance to Jupiter and the

¹The more accurate value of c is 299,792 km/sec (*Rømer did not use these modern units*).

timing of Io's eclipses. Hence, a more controlled methodology of measurement was required.

- (17.3.1) **Measuring the Speed of Light.** In 1849, a French physicist, Hippolite Louis Fizeau (1819–1896), devised an experiment that, for the first time, provided an accurate estimate for c , the speed of light.

Fizeau's mirrors and spinning toothed wheel, illustrated in Figure (17.3.2), worked as follows:

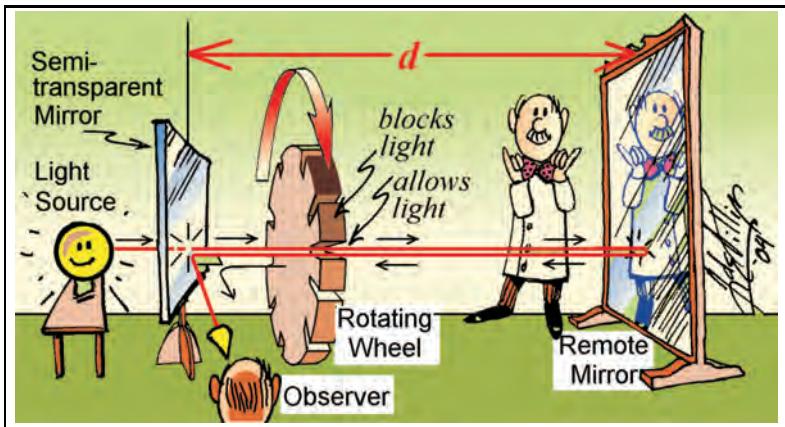
- (1) Light is emitted from the light source.
- (2) Light passes through the semi-transparent mirror, and through a gap between two cogs on the spinning wheel.
- (3) Light reflects off the remote mirror and returns to the spinning wheel. The speed of the wheel must be adjusted so that it spins fast enough to guarantee that the returning light does not hit a cog, but passes successfully through the next space between two cogs.
- (4) When the returning light successfully passes through a space, it reflects off the semi-transparent mirror at a 90° angle to finally reach the eye of an observer.

We describe steps (1)-(4) mathematically using variables:

- ρ , the revolutions per minute of the wheel,
 N , the number of teeth (and slots) on the wheel,
 d , the distance from the wheel to the remote reflecting mirror.

|| **Note:** The distances from the semi-transparent mirror to the wheel, and from the semi-transparent mirror to the observer, are tiny compared to the distance d .

(17.3.2)

FIGURE: How Fizeau Measured the Speed of Light.

Film Strip Diagram for Fizeau's Experiment: The film strip following interprets Figure (17.3.2), giving a frame-by-frame account of the paths of photons 0, 1, 2, 3, 4, after they are emitted from the light source.

(17.3.3)

FIGURE: Time Sequence for the Fizeau Experiment

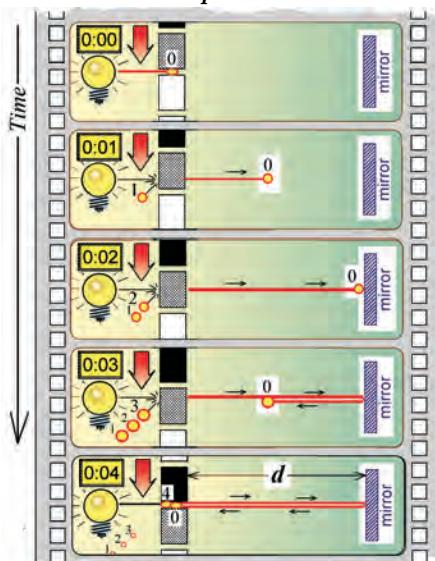
Time **0:00**: Photon₀ passes through the slot between gray and white cogs.

Time **0:01**: Photon₁ is blocked by gray cog while Photon₀ continues towards the remote mirror.

Time **0:02**: Photon₂ is blocked, Photon₀ hits the remote mirror.

Time **0:03**: Photon₃ is blocked, Photon₀ makes its return from the remote mirror.

Time **0:04**: If the wheel is adjusted to spin at just the



right rpm, then Photon₀'s leftward return occurs exactly when the next open slot (the one between the gray and black teeth) arrives. (Simultaneously, Photon₄ exits rightward through the slot, starting a journey similar to that

of Photon₀.) Finally, Photon₀ will complete its return by reflecting 90° off the semi-transparent mirror of Figure (17.3.2) and into the eye of the observer.

- (17.3.4) **Computing the Speed of Light with Fizeau's Wheel.** It is time to be specific. Here are the numerical aspects alluded to in the discussion of the Scientific Method.

t_{RT} , the Time of a Photon's Round Trip. For constant speeds, $\text{time} = \text{distance}/\text{speed}$. Apply this formula to Figure (17.3.2), and we see that the time for the photon's round trip is

$$(17.3.5) \quad \left. \begin{array}{c} \text{minutes per} \\ \text{round trip} \end{array} \right\} = \frac{2d}{c} = t_{\text{RT}},$$

where c is the speed of light, and d is the distance from the toothed wheel to the reflecting mirror.

t_{slot} , Time for One Slot to Replace Another: If R is the radius of the toothed wheel and there are N teeth, then the distance from a point on one slot to the next slot is $2\pi R/N$.

- (17.3.6) **NOTE:** In the following discussion, we convert from one set of units to another by multiplying expressions by appropriate forms of 1 (*For example*, 1 = (4 quarts/1 gallon).) and then canceling certain units as though they were numbers. This popular technique among engineers is developed in Chapter (D), pg. 389.

As an application of this unit-converting technique, let ρ (rev/min) denote the angular speed of the rotating wheel in revolutions per minute, denoted rpm with units (rev/min). Let minutes (denoted by min) be the units of time. Then t_{slot} , the time required by any slot on the wheel to rotate to the position of its neighboring slot, is

$$(17.3.7) \quad \begin{aligned} \frac{1}{\rho} \left(\frac{\text{min}}{\text{rev}} \right) &= \frac{1}{\rho} \left(\frac{\text{min}}{\text{rev}} \right) \underbrace{\frac{1}{2\pi R} \left(\frac{\text{rev}}{\text{ft}} \right)}_1 \underbrace{\frac{2\pi R/N}{1} \left(\frac{\text{ft}}{\text{slot}} \right)}_1 \\ &= \frac{1}{\rho N} \left(\frac{\text{min}}{\text{slot}} \right) \\ &= t_{\text{slot}} \left(\frac{\text{min}}{\text{slot}} \right). \end{aligned}$$

Fizeau's experiment requires that the angular speed ρ of the toothed wheel should be chosen so that a photon returns to the wheel when the *next* open slot takes the position of the photon's exit slot.

We require t_{RT} of (17.3.5) to equal t_{slot} of (17.3.7), using exactly the units above. Once we set $t_{\text{RT}} = t_{\text{slot}}$, we will obtain the value of c . That is, if ρ is the proper number of revolutions per minute of a wheel with N teeth, which is also d units distant from a reflecting mirror, then

$$(17.3.8) \quad (t_{\text{RT}} = t_{\text{slot}}) \implies (c = 2d\rho N).$$

Note: Although it might challenge our intuition, the mathematical analysis leading to Equation (17.3.8) shows that the radius R of the wheel is irrelevant.

Note: Fizeau calculated that $c = 323,000 \text{ km/sec}$ which is only 4.5% higher than the currently accepted value of $c = 299,792 \text{ km/sec}$.

(17.4) de Sitter: c Independent of Source Speed

In 1913, WILLEM DE SITTER (1872–1934) showed that the speed of light, c , is independent of the speed of the source. This is consistent with (16.5.4) which illustrates why the speed of a propagated wave does not depend on the speed of its source.

To do this, he assumed the opposite (proof by contradiction (G.4.5)), namely, that c is increased by a slingshot effect of the source if the source is advancing toward the observer; the speed of light is decreased if the source is running away (receding) from the observer.

Note: Whenever we say that a photon is “*thrown*” from a light source, we will assume (falsely, as it happens) that the photon's speed increases if it is thrown toward us, and its speed decreases if the photon is thrown away from us.

Where does one obtain light sources that significantly advance or recede? The answer is in a binary star system where two stars travel opposite each other in a common, near-circular orbit. In such

an orbit, each star spends half the time advancing toward Earth and the rest of the time receding from Earth.

With the appropriate geometry, light “thrown” from a single binary star at *three different positions in its orbit* will arrive at Earth simultaneously!

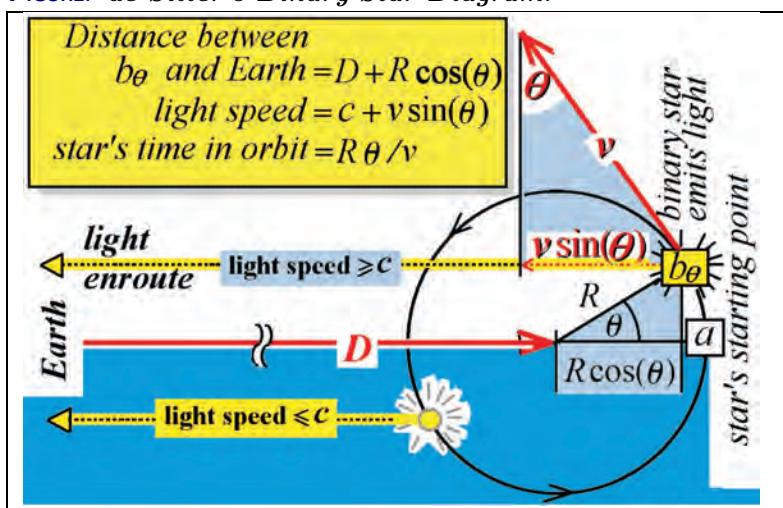
Here is why we receive three photons simultaneously on Earth if the slingshot effect on photons is valid:

Assume that light is susceptible to the slingshot effect whereby its speed is c plus (or minus) the speed of its source. Then is it possible that a photon of light from a binary star that is transmitted *later* will reach Earth *before* a photon had been thrown or transmitted earlier. See Exercise (17.6.4).

Here is the diagram with which we resolve this question:

(17.4.1)
Trig
required

FIGURE: de Sitter's Binary Star Diagram.



Starting at point $[a]$, a star in orbit carries a bundle of photons. For the sake of the illustration, the star emits the single photon, p_{θ} , precisely when angle θ is reached. Photon p_{θ} then proceeds on its journey to Earth.

The travel time, T_θ , for each photon, p_{θ} , is equal to

- The star's time in orbit, starting at point $[a]$ (at angle 0 degrees

and time $t = 0$) and ending at point b_θ (at angle θ), during which time it “carries” photon ph_θ ,

plus

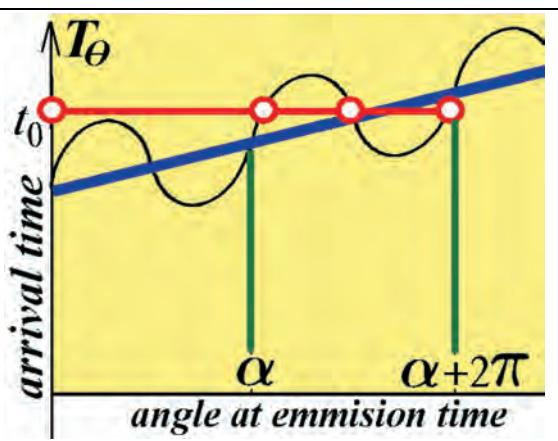
- The time that the photon ph_θ requires to travel from release point, b_θ , to Earth.

In mathematical terms, the total time it takes the photon ph_θ to travel from point a to Earth is

$$(17.4.2) \quad T_\theta = \underbrace{\frac{R}{v} \theta}_{\begin{array}{l} \text{ph}_\theta \text{ time} \\ \text{in orbit} \end{array}} + \underbrace{\frac{D + R \cdot \cos \theta}{c + v \cdot \sin \theta}}_{\text{ph}_\theta \text{ time in space}}.$$

(17.4.3) **FIGURE:** Graphing (17.4.2): How One Star Can Produce Three Simultaneous Images.

The oscillating graph of the function T_θ (17.4.2) shows a single arrival time, t_0 , for three separate photons, ph_{θ_1} , ph_{θ_2} , ph_{θ_3} , that are thrown from a star at positions, θ_1 , θ_2 , and θ_3 . A photograph taken on Earth at receiving time, t_0 ,



should simultaneously record these three images, which are known as DE SITTER’S GHOSTS.

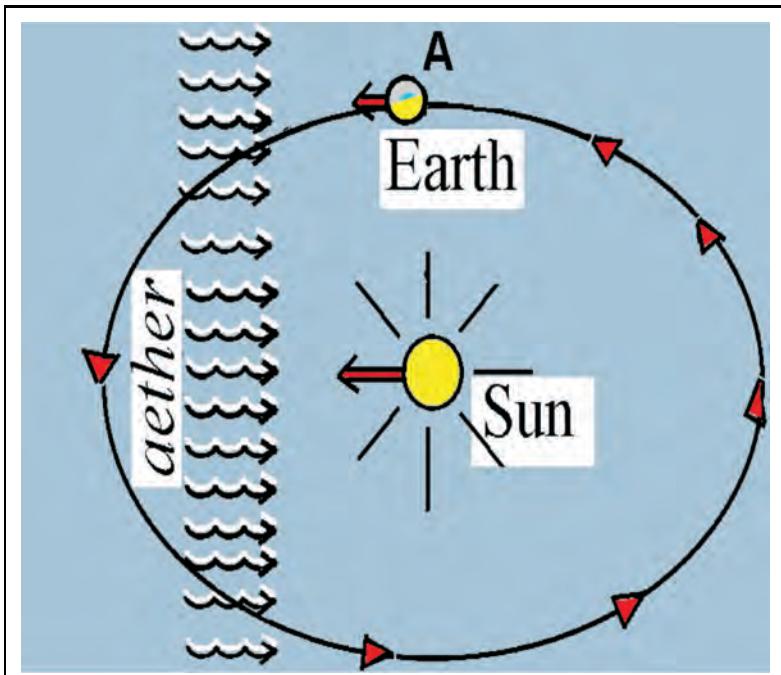
An analysis of formula T_θ (17.4.2) shows that the distance D from Earth to the star system center must be sufficiently large to produce the correspondingly large amplitude of the periodic part of (17.4.2) that is shown in Figure (17.4.3). If the speed of light is increased or decreased by the speed of its “slingshot source,” then faster light thrown from the top half of a star’s orbit (Figure (17.4.1)) but far away, should reach Earth at the same time as slower light that is cast

from the bottom half of the orbit but at a closer distance to Earth.

- (17.4.4) **No Ghosts Ever Appear—Prediction (17.4.3) Fails.** Though no “ghosts” appear in de Sitter’s images, this does not guarantee absolutely that the speed of the source has no effect on the speed of light — such a guarantee would require infinite experimental accuracy. The independence of light speed from the speed of its source is, rather, the best analysis or description that could be made within the bounds of scientific accuracy of the time.

(17.5) Michelson-Morley's Happy Failure

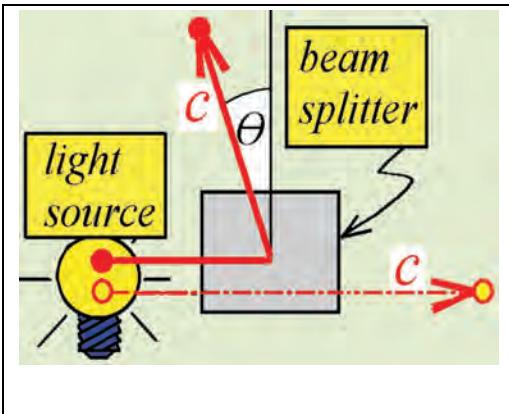
- (17.5.1) **FIGURE: Sun and Earth in an Aether Stream**



When Maxwell’s theory of electromagnetism was completed in the 19th century, it was assumed that light, like other waves, needed a medium in which to propagate. This LUMINIFEROUS AETHER, or AETHER for short, fills the entire visible universe; otherwise we would not be able to see all the universe.

Everything in the Universe, including Earth, was supposed to move in this cosmic sea of aether. Just as when cycling on a calm day, when we feel the wind on our faces, the Earth should, if the aether existed, experience an “*aether wind*.” Even if the Sun is moving with respect to the aether, Earth’s ever-changing direction of motion in its orbit would ensure that effects of the aether wind would be measurable, at least for part of any year.

(17.5.2)

FIGURE: The Light-Beam Splitter.

The theoretical beam splitter, with an appropriate configuration of internal mirrors, accepts a horizontal beam from a light source and reflects half the beam onto an almost-vertical path, while it allows the other half to pass horizontally. The beam splitter can

aim the vertical beam at any angle θ . (If $\theta = 0$, then the vertical beam is perpendicular to the horizontal beam.)

(17.5.3)

Michelson's River Metaphor. It is said that Michelson was talking to his daughter and on a napkin drew a diagram for the famous experiment. Michelson reasoned that from the point of view of the Earth, the Earth was fixed and it was the aether that was moving. (Figure (17.5.1) shows the Earth moving leftward relative to the aether. From the Earth's point of view, however, the Earth is stationary and the aether wind is moving left to right.)

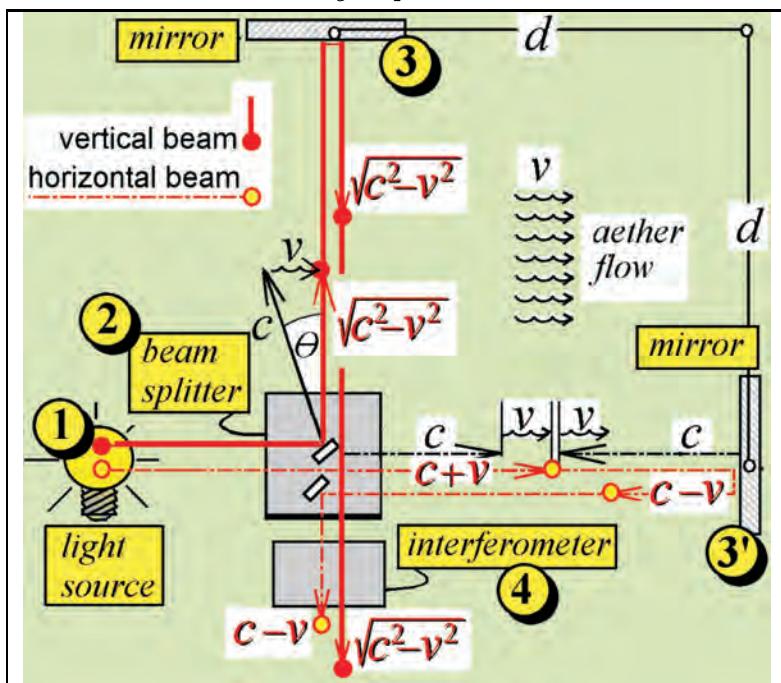
In fact, one could regard the Earth as a boat in an aether river. With this model, a light beam in flight could be directed to travel upstream, downstream, or cross stream. Since light traveling in the aether was assumed to behave like a boat in a stream, light would be slowed down going against the stream, and it would speed up going with the stream.

Michelson had an ingenious way to split a single beam in two, sending one on an upstream-downstream journey, while the other half trav-

eled cross stream. Both beams were reflected back to their starting point so that their respective speeds could be compared.

But Michelson did not clock the actual speeds of the twin light beams. Rather, he devised an INTERFEROMETER that was sensitive to the shift in the wave patterns of the two beams. If there is no shift, then the peaks and troughs (*valleys*) of the twin waves would add together to produce a brighter, reinforced wave. This is called *constructive interference*. If there is a shift, as when the peak of one wave overlaps the trough of the other, then we get interference or cancellation of the waves which produces a decrease in the brightness. This is called *destructive interference*.

(17.5.4)

FIGURE: *Michelson-Morley Experiment.*

We diagram a slight variation of the Michelson Morley experiment which compares the aether-drag of a crosswind beam against the drag of an upstream-downstream light beam. Earth is at point A of Figure (17.5.1), pg. 228. Relative to a stationary aether, Earth is moving leftward at speed v . But relative to a stationary Earth, the aether “wind” moves rightward at speed v .

This superposition of two similar waves produces a light-dark pattern known as the *interference pattern*.

The surprising result of the Michelson-Morley experiment was that the two superimposed beams of Figure (17.5.4) never showed the destructive interference expected from a shift due to the aether drag on the upstream-downstream beam. No matter what the direction of the interferometer, the split beams always ended their respective journeys in perfect overlapping alignment with each other as we shall see in (17.5.7).

Details of the Michelson-Morley Calculations

Point ①-② With starting speed c , a beam of light is emitted horizontally, or downstream, parallel to the aether wind. It enters the beam splitter of Figure (17.5.2).

A vertical and a horizontal beam exit the beam-splitter.

Path ② -③ *The vertical beam* exits the beam splitter and, like a boat going cross stream, leans into the aether stream at an angle θ . This is to ensure a resulting cross stream path that is truly perpendicular to the horizontal flow of the aether. The Pythagorean Theorem tells us that the resulting vertical beam has diminished speed $\sqrt{c^2 - v^2} < c$. Time = distance/speed implies the travel time of the vertical beam is

$$(17.5.4a) \quad t_{\uparrow} = \frac{d}{\sqrt{c^2 - v^2}}.$$

.....

Path ② -③ *The horizontal beam* exits the beam splitter with speed increased by v , the speed of the aether wind.² As the horizontal beam is riding with the flow of the aether wind, its augmented speed is now $c+v$. The relation, time = distance/speed, says the rightward horizontal travel time is

$$(17.5.4b) \quad t_{\rightarrow} = \frac{d}{c + v}.$$

.....

²For simplicity, we ignore the (small) distance between the light source and the beam splitter so that all speed computations can begin at the beam splitter.

Path ③ -④ *The vertical beam* is reflected from the mirror and it returns on its cross stream path toward the interferometer with its original speed $v_{\downarrow} = \sqrt{c^2 - v^2}$. Using time = distance/speed, the time for the vertical return trip is

$$(17.5.4c) \quad t_{\downarrow} = \frac{d}{\sqrt{c^2 - v^2}}.$$

.....

Path ③ -④ *The horizontal beam*, on its return is now opposing the aether wind whose speed is v . The diminished horizontal speed of the light beam is now $c-v$. The relation, time=dist./speed, confirms that the time for the leftward horizontal return trip is

$$(17.5.4d) \quad t_{\leftarrow} = \frac{d}{c-v}.$$

.....

Point ④ *The vertical and horizontal beams* meet in the interferometer where their wave patterns are compared. If one beam is slower than the other, a wave interference pattern will result.

(17.5.5) **Comparing Times of Travel.** Now that the four legs of the journeys are established in (17.5.4a)-(17.5.4d), we can compare the time, $(t_{\leftarrow} + t_{\rightarrow})$, of the horizontal round trip to $(t_{\uparrow} + t_{\downarrow})$, the time of the vertical round trip. In fact, the exact ratio is

$$(17.5.6) \quad \begin{aligned} \frac{(\text{horizontal time})}{(\text{vertical time})} &= (t_{\rightarrow} + t_{\leftarrow}) / (t_{\uparrow} + t_{\downarrow}) \\ &= (\frac{d}{c+v} + \frac{d}{c-v}) / (\frac{d}{\sqrt{c^2 - v^2}} + \frac{d}{\sqrt{c^2 - v^2}}) \\ &= (\frac{2cd}{c^2 - v^2}) / (\frac{2d}{\sqrt{c^2 - v^2}}) \\ &= \frac{1}{\sqrt{1 - (v/c)^2}} = \frac{1}{\sigma_v} \geq 1. \end{aligned}$$

(17.5.7) **The Null Observed Result.** There was no wave interference or shift between the vertical and horizontal light beams. Michelson

and Morley rotated the apparatus, trying to find the best alignment with the aether, but observation always showed no apparent time difference between the two beams.

-
- (17.5.8) **Fitzgerald's Ad Hoc Fix.** GEORGE FITZGERALD (1851-1901) observed that under the aether model, the horizontal round trip always takes longer than the time for the vertical round trip³. Hence, to make the times equal, he simply assumed that distances along the direction of motion are contracted (presumably by the pressure of the aether). This effect is different from the one we have discussed: Fitzgerald assumed the contraction was independent of the reference frame.

For example, an astronaut in an accelerating spaceship would be increasingly uncomfortable and eventually die as her organs become more and more compressed, no matter how mild the acceleration. This independence of acceleration is completely different from the predictions of the Special Theory.

In sum, Fitzgerald had the right conclusion, saying that distance shrinks in the direction of motion. But he had no theoretical basis. The theoretical basis was finally provided when Einstein showed that length contraction was a consequence of the speed of light being the same for all observers.

-
- (17.5.9)  **An Intuitive Observation.** Assuming a light-carrying aether exists, then (17.5.6) shows analytically that it always takes less time to complete a cross stream round trip (vertical time) than an upstream-downstream round trip (horizontal time). This agrees with intuition since a light beam directed upstream might have to spend a good deal of time confronting a high-speed aether current. For example, if the opposing aether current has speed $v = c$, then the light beam, relative to its source, would have a speed of zero and the light would *never* complete the upstream leg of the journey.

See Exercise (17.6.5) for speeds due to a general angle α of the splitter arms relative to the aether wind. This case is more general than the horizontal and vertical angles of Figure (17.5.4).

-
- (17.5.10) **Is the Aether Theory Dead?** Interpretation of data is always a challenge. Michelson could not dismiss the existence of the aether

³For simplicity, we assume equal vertical and horizontal distances $d = d_v = d_h$. We could generalize the result by allowing for different distances $d_v \neq d_h$.

since, he reasoned, *some* sort of medium had to carry light—or else we must accept the non-intuitive statement that *nothing is doing something*.

(17.5.11)



Counterfeit Models: The aether model was consistent with observed phenomena up to but not including the Michelson-Morley experiment. It is interesting to note that no less a thinker than René Descartes produced a true result (*Snell's Law*) from a false assumption, an act that put him at odds with Pierre Fermat (see (24.6), pg. 348).

(17.5.12)

Michelson-Morley Table of Hypotheses.

<i>H</i> : Hypothesis	⇒	<i>C</i> : Deduced Conclusion
<ul style="list-style-type: none"> • The speed of the source is added to the speed of transmitted light (light travels through the aether like a boat being carried by a river). <p style="text-align: center;">and</p> <ul style="list-style-type: none"> • The aether has speed v relative to a light source. 	⇒	<p>If light beam A travels distance D on an upwind-downwind round trip and beam B travels the same distance on a crosswind trip, then the two beams will always exhibit an observable wave phase difference due to the predicted ratio of travel times (beam A over beam B) equal to</p> $\frac{1}{\sqrt{1 - (v/c)^2}}.$

(17.6) Exercises

**(17.6.1)
Exercise**

A Jovian Over-Simplification. Assume Jupiter orbits the sun so slowly that it may be considered stationary. Then, using optical (as opposed to radio) instruments, approximately how many days of the year can an observer on Earth view Jupiter and its satellites? (See Figure (17.2.6).)

(17.6.2)
Exercise

Viewing Jupiter. From Figure (17.2.6), what relative positions of Earth and Jupiter ensure that Jupiter is not visible in the night time sky?

(17.6.3)
Exercise

The Twelve-Year Jupiter Cycle. Given that Jupiter's orbital period is about twelve years (actually, 11.86 years), then how many days of the year can an observer on Earth view Jupiter and its satellites? (See Exercise (17.6.1).)

(17.6.4)
Exercise

Light Emitted Late Arrives Earlier. Assume that the speed of light is $c + v$ if the source of the light has speed v . Under this assumption, show that a "slingshot effect" will allow a photon from a binary star to arrive at Earth before a photon that is thrown or transmitted earlier.

Specifically, a binary star in its orbit passes point P_1 after it passes point P_0 . Show that if the distance D from the center of the binary star system to Earth is large enough, then light emitted from P_1 (the later point) reaches Earth *before* the light emitted from P_0 (the earlier point). (**Hint:** To simplify the values of $\cos \theta$ and $\sin \theta$ in (17.4.2), pg. 227, set the earlier point at angle $\theta = -90^\circ$ and the later point at angle $\theta = +90^\circ$.)

(17.6.5)
Exercise

Generalized Splitter. Generalizing Figure (17.5.4) Show that if the aether speed is v and the angle α is the angle of the object relative to the aether, then

$$v_{\uparrow} = \sqrt{c^2 - (v \cos \alpha)^2} + v \sin \alpha$$

$$v_{\downarrow} = \sqrt{c^2 - (v \cos \alpha)^2} - v \sin \alpha$$

$$v_{\rightarrow} = \sqrt{c^2 - (v \sin \alpha)^2} + v \sin \alpha$$

$$v_{\leftarrow} = \sqrt{c^2 - (v \sin \alpha)^2} - v \sin \alpha.$$

(17.6.6)
Background
Measuring
the Speed
of Light at
Home

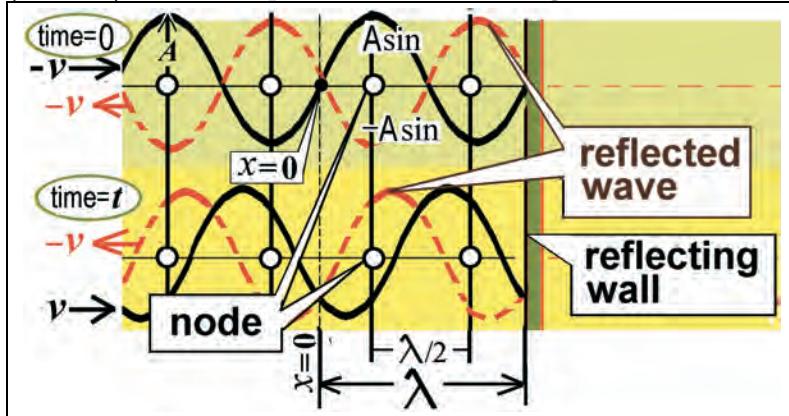
In your Kitchen: Measuring the Speed of Light with Your Microwave Oven (Introduction). Your home microwave oven produces electromagnet microwaves that travel at the speed of light in air, which is very close to c , the speed of light in a vacuum. The waves are reflected back and cancel or strengthen parts of the original wave. In the special case where the sum of the waves has fixed points of zero value we have a so-called standing wave that we now define.

(17.6.6a)

DEFINITION: When a wave f in one direction is added to another wave g traveling in the opposite direction so that the sum $f + g$ produces a wave with stationary zero values at regular distances for all times, then the points at which $f + g = 0$ are called **NODES** (or “cold spots”) and the wave $f + g$ is called a **STANDING WAVE**.

Construction of a Standing Wave: In the figure following, the rightward-traveling (solid line) wave produces the reflected leftward-traveling (dotted line) wave. The small circles indicate the nodes.

(17.6.6b) FIGURE: Components of a Standing Wave:



- A sinusoidal microwave (solid line) moves rightward with speed v while the reflected wave (dashed line) travels leftward with speed $-v$.
- Nodes, or “cold spots” lie $\lambda/2$ units apart where λ is the full wavelength. At all other points, the standing wave varies in time.
- Interlaced halfway between adjacent nodes are “hot spots”—points where the fluctuating standing wave periodically achieves its maximum intensity. The distance between adjacent hot spots is also $\lambda/2$.

Exercise:
Trig
Required

(17.6.6c) Show that the equation for the time-dependent standing wave of Figure (17.6.6b) is

$$(\dagger) \quad w(x, t) = -2A \cdot \cos\left(\frac{2\pi}{\lambda}x\right) \sin\left(\frac{2\pi}{\lambda}vt\right), \quad \text{where } v \text{ is the wave speed.}$$

Hint: If $\pm v$ are the speeds of the moving waves, then for any time t , the equations for the time-dependent sinusoidal waves of Figure

(17.6.6b)—the original wave f and the reflected wave g —are

$$(\dagger\dagger) \quad f(x, t) = A \cdot \sin\left(\frac{2\pi}{\lambda}(x - vt)\right) \quad \text{right-shifted solid-line wave at time } t$$

$$g(x, t) = -A \cdot \sin\left(\frac{2\pi}{\lambda}(x + vt)\right) \quad \text{left-shifted dashed-line wave at time } t.$$

The sum $f + g = w$ for f and g defined in $(\dagger\dagger)$ gives us the equation for the standing wave $w(x, t)$ as described in (\dagger) . ■

Therefore from (\dagger) , the fixed nodes (cold spots) are defined independent of t whenever $w(x, t) = 0$. That is, whenever $\cos(\frac{2\pi}{\lambda}x) = 0$, or when $x = \pm\frac{1}{4}\lambda, \pm\frac{3}{4}\lambda, \pm\frac{5}{4}\lambda, \dots$ as is indicated by the small circles \circ in Figure (17.6.6b). Hence, the interlaced hot spots occur when $\cos(\frac{2\pi}{\lambda}x)$ is at its maximum or minimum value ± 1 , or when $x = 0, \pm\frac{1}{2}\lambda, \pm\frac{2}{2}\lambda, \pm\frac{3}{2}\lambda, \dots$

Exercise:

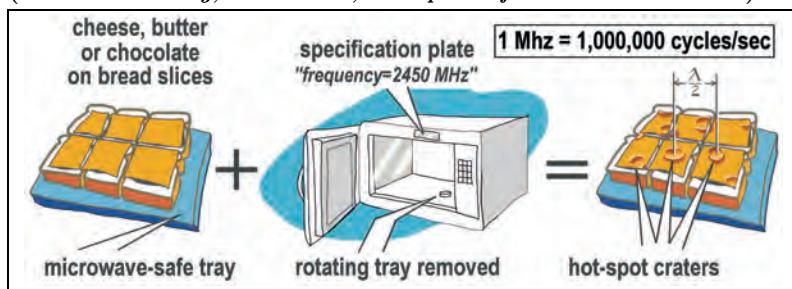
Microwave

Oven

Required

(17.6.6d) Measure c , the Speed of Light at Home

(More accurately, measure v , the speed of microwaves in air.)



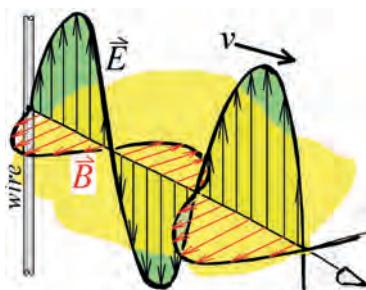
Hint: Cover a microwave-safe tray with slices of bread with butter, cheese, or chocolate on top. Place the tray in a microwave oven for about twenty seconds with the rotating tray removed! Measure the distance d between the hot-spot craters caused by the standing microwaves. The distance $d = \lambda/2$ is one half the microwave wave length $\boxed{\lambda}$. From the information plate of your microwave oven, obtain the wave frequency \boxed{f} . Finally, to calculate $v \approx c$, use the equation $v = \boxed{f \cdot \lambda}$ from (16.1.2a), pg. 200.



VIII. Maxwell's Equations

Chapters (18)–(23) present the mathematical modeling techniques that Maxwell used in developing his four equations that unified the theories of electricity, optics, and magnetism.

Section (23.4) of chapter (23) shows how Maxwell, using just two laboratory measurements, predicted the value of c , the speed of electromagnetic radiation in a vacuum, and also identified light as a particular type of electromagnetic radiation. These predictions have now been verified experimentally to enormous accuracy, though it took decades to develop the technology to provide such accurately measured and convincing evidence.



(18) Maxwell's Mathematical Toolkit

(18.1) Preface

- (18.1.1)** **Calculus will be required** **Things to Come:** We present the mathematical machinery used by Maxwell to develop his four equations (21.7.3a)–(21.7.3d). This requires a significant amount of calculus, much of which will be reviewed in this chapter.

Use of calculus may seem daunting but the rewards are great. Maxwell's equations have stood the test of time, as they describe not only the motion of charged particles, properties of magnets and electromagnetic waves at *all* frequencies, but they apply to phenomena in the quantum regime as well.

The language of mathematics is essential—and there does not seem to be a simpler path.

- (18.1.2)**



Note: *Maxwell's equations (21.7.3), pg. 300 will be and should be applauded as a great achievement in mathematical modeling. However, we must be aware that sometimes, apparently reliable models can be based on false premises. See (17.5.11), pg. 234 for a spectacular example.*

(18.2) Language and Proportionality

(18.2.1)

DEFINITION: Given a set of numbers X , each of which is associated with a number Y . We say number X is PROPORTIONAL to its related number Y if there is a number k that is constant with respect to X and Y , where for all X and its related Y ,

$$(18.2.1a) \quad X = kY.$$

(k may depend on variables other than X and Y .)

Motivation for (18.2.1). To say “ X is proportional to Y ,” is to say if X doubles, then so does its related number Y ; if X triples, then so does Y , and so on. In other words, the quotient X/Y is a constant k for all X and its related Y . But

(18.2.1b) $X/Y = k$ is algebraically equivalent to $X = kY$.

From equations (18.2.1a) and (18.2.1b), we can say more:

(18.2.2)

THEOREM: If a quantity X is proportional to both Y and Z , then X is proportional to the product YZ . (For example, the area A of a rectangle is proportional to both length L and width W . Hence, A is proportional to the product LW .)

PROOF: Exercise (18.9.1). ■

NOTE: Theorem (18.2.2) will prove to lie at the foundation of Coulomb's Law (19.2.4b).

(18.2.3)

DEFINITION: We say X is INVERSELY PROPORTIONAL to Y whenever X is proportional (18.2.1) to $1/Y$, the inverse of Y . That is,

$$(18.2.3a) \quad X = k/Y \text{ for some constant } k$$

that is called the CONSTANT OF PROPORTIONALITY.

(18.2.4) Examples:

(18.2.4a)

We observe the set of related numbers, (F, m) , where force F is used to accelerate mass m by a fixed amount, $a \text{ ft/sec}^2$. To say that the force F is always proportional to mass m is to state Newton's Second Law, $F = ma$, where a is a constant of acceleration.

(18.2.4b)

To say the force F is *inversely* proportional to the square of distance r , is to say F is proportional to $(1/r^2)$. That is, $F = k/r^2$. Such a proportionality is called an INVERSE SQUARE LAW.

(18.2.4c)

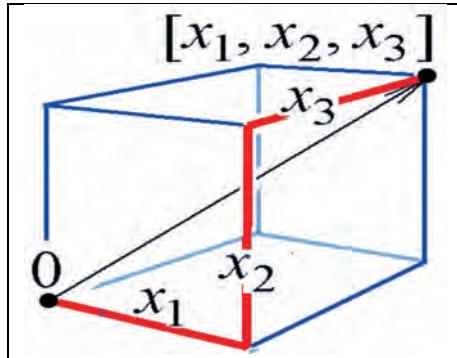
Consider the list of related numbers, (A, L) and (A, W) , where A is the area of a rectangle with length L and width W . Since $A = LW$, then, consistent with Definition (18.2.1), we declare that the area A of a rectangle is proportional to both the length L and the width W . However, area A is also proportional to a third number, the product, LW . In other words, $A = kLW$. (See Exercise (18.9.1).)

(18.3) 1D Lengths & 2D Areas as 3D Vectors

(18.3.1)



A new interpretation of vectors. We first encountered ordered triples, $[x_1, x_2, x_3]$ in (A.2.3) as a way to numerically locate points in 3-space within a 3D (*3-dimensional*) frame of reference. Arrows can be used to represent such ordered triples, or vectors.

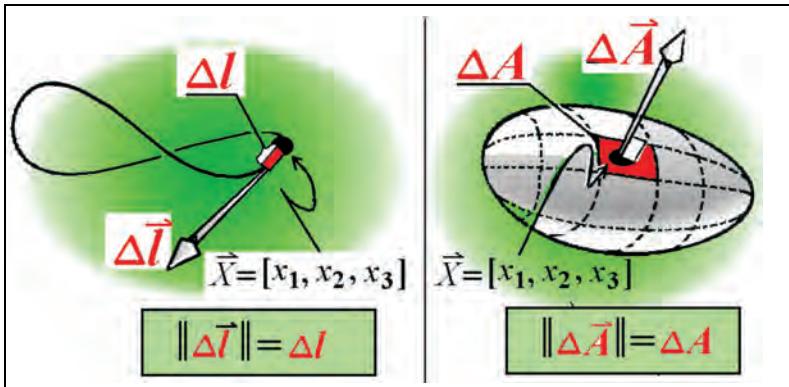


In the following, we use arrows to graphically represent physical quantities of *lengths* and *areas* in 3-dimensional space.

(18.3.2) Tiny, (i.e., differential) pieces of 1D arc lengths and 2D surfaces can be described by 3D vectors only if there is sufficient smoothness—1D arcs are not severely crinkled and 2D surfaces are not very crumpled. Here are the details:

- LEFT PANEL OF FIGURE (18.3.3): LENGTHS AS VECTORS. Arc segment of length Δl is assigned a 3-dimensional arrow, $\Delta \vec{l}$, whose tail is placed at a point in the segment, whose length is Δl , and whose direction is tangent to the arc. Either this orientation or its negative (a rotation of 180°) is declared to define the *positive* direction of $\Delta \vec{l}$.
- RIGHT PANEL OF FIGURE (18.3.3): AREAS AS VECTORS. A region of area ΔA is taken small enough so that it is essentially flat—much the way we view the curved surface of a lake, a small area of Earth's surface, as flat. The area ΔA is assigned a 3-dimensional arrow, $\Delta \vec{A}$, whose tail is placed at a point in the region, whose length is ΔA , and whose direction is NORMAL, or perpendicular to the area. Either this orientation or its negative (a rotation of 180°) is declared to define the *positive* direction of $\Delta \vec{A}$. (*Vectors describing surface orientation will be defined in (18.4.7).*)

(18.3.3) **FIGURE: 1D-Lengths and 2D-Areas as 3D-Vectors, $\Delta \vec{l}$ and $\Delta \vec{A}$.**



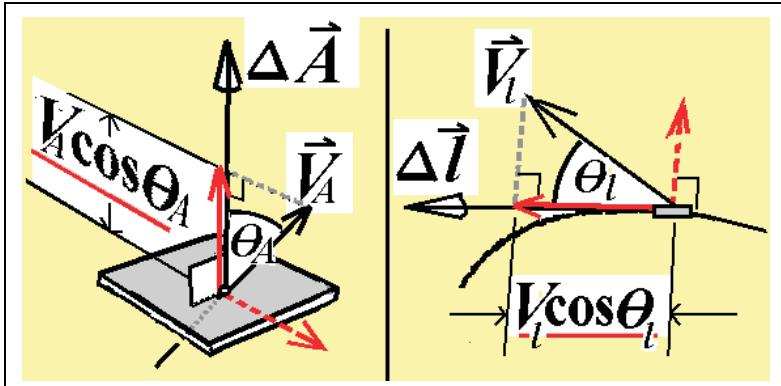
The notations for vectors and their lengths are given by Table (18.3.4).

(18.3.4) **Vector Representations of Length and Area**

	Vector	Vector Magnitude	Vector Direction
Length	$\Delta \vec{l}$	$\Delta l = \ \Delta \vec{l}\ $	tangent to curve
Area	$\Delta \vec{A}$	$\Delta A = \ \Delta \vec{A}\ $	normal to surface

Physical experiments show that for vectors \vec{V} in \mathbf{R}^3 emanating from points on a surface or an arc length, it is useful to consider their components or projections onto the vectors $\Delta \vec{A}$ or $\Delta \vec{l}$ of (18.3.4). This motivates the following discussion:

The setup: The vector \vec{V}_A in \mathbf{R}^3 emanates from a point on the surface element ΔA and forms an angle θ_A with the normal vector, $\Delta \vec{A}$. Similarly, the vector \vec{V}_l is situated at a point on the differential line segment Δl , forming an angle θ_l with the tangent vector $\Delta \vec{l}$.

(18.3.5)
Trig

NOTE: The components of arbitrary vectors, \vec{V}_A and \vec{V}_l , on the area vector, $\Delta \vec{A}$, and length vector, $\Delta \vec{l}$, (18.3.3) are called the PROJECTION of vectors \vec{V}_A and \vec{V}_l , respectively.

The *lengths* of the projection vectors, as illustrated in Definition (18.3.5), are given by the table:

$\vec{V}_A \qquad \vec{V}_l$ $V_A \cos \theta_A \quad V_l \cos \theta_l$	{Vector} {Length of Projection}
---	------------------------------------

(18.4) Orientations of Lines and Surfaces

(18.4.1)



Motivation: There will follow a considerable number of results about lines and surfaces which culminate in the four famous equations of Maxwell (21.7.3a)-(21.7.3d). The importance of these equations is indicated by the fact that Einstein acknowledged and applauded Maxwell's work in the very first paragraph of his paper of 1905 [12] in which the ideas of special relativity were first introduced.

At first glance, these mathematical results might seem to have little to do with special relativity. However, the time spent in understanding the mathematics in this chapter will bear fruit when used in subsequent chapters. In addition, we hope the reader will recognize that there is an intrinsic elegance that, in itself, justifies the effort to understand the tools of this chapter.

(18.4.2)

Inside and Outside Regions. Where Section (18.3) described the lengths of a tiny line segment at a *point*, and the area of an infinitesimal surface element at a *point*, we now describe the orientation in space of the entire line and surface.

First, we are interested in lines and surfaces that are *closed*, a term we now define in an intuitive, non-rigorous way:

(18.4.3)
Calculus
required

DEFINITION: A non-intersecting 1-dimensional line in space that closes back on itself, like a string whose two ends meet, is called a CLOSED LINE. Similarly, a non-intersecting 2-dimensional surface that closes in on itself, like a potato skin that covers a finite-volume potato, is called a CLOSED SURFACE.

|| See Figure (18.3.3) for an illustration of a closed line and a closed surface.

(18.4.4)

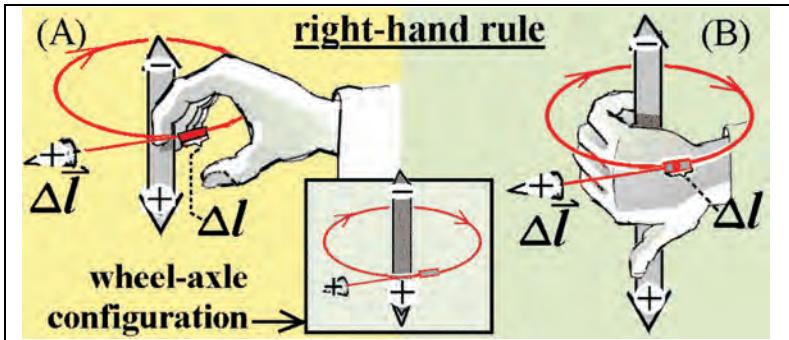


Closed non-intersecting 1-dimensional lines (loops) are called JORDAN CURVES, after the French mathematician MARIE ENNEMOND CAMILLE JORDAN (1838–1922). Like a closed 1-dimensional circle, 1-dimensional Jordan curves define an infinite *exterior* (outside region) and a finite *interior* (inside region). This may seem obvious, but a rigorous proof is a rather tough mathematical nut. In fact, the proof by Jordan himself was in error.

Like the 2-dimensional surface of a sphere, any closed non-intersecting “potato-skin” area in space defines an infinite outside region and a finite inside region.

(18.4.5)

FIGURE: Wheel-Axle Orientation of Closed Loops.



A closed loop and a line that enters its enclosed area is a **wheel-axle configuration**. It is an experimental fact that a magnet moving through a loop generates a current through the loop. The direction of the current depends on whether the magnet's north pole or the south pole enters first. The right-hand rule applied to the wheel-axle configuration provides a mathematical means of describing the leading pole of the magnet (north = thumb) and the direction of the current (clockwise = direction of the right-hand fingers).

- (18.4.6) The Right-Hand Rule:** If the positive direction of one component (wheel or axle) is known, then the positive direction of the other is uniquely determined according to the following:

- **PANEL (A) OF (18.4.5) Four Fingers are Inside the Loop:** *The four fingers point in the positive direction of the straight wire (axle) while the thumb points in the positive direction of the loop (wheel).*
- **PANEL (B) OF (18.4.5) Thumb is Aligned with the Straight Wire:** *The thumb points in the positive direction of the wire (axle) while the four fingers point in the positive direction of the loop (wheel).*

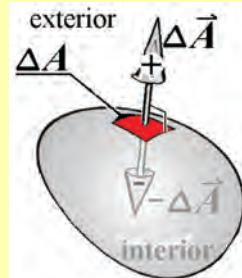
Motivation: The Right-Hand Rule is devised to uniquely relate the direction in which we travel along a *closed curve* (*defined in (18.4.3)*) to the “direction” of the enclosed area. If we imagine that we travel in a certain direction on the closed curve—the boundary of a small area ΔA —then we can uniquely define a positive (right-hand) direction of the vector $\Delta \vec{A}$ whose length $\|\Delta \vec{A}\|$ numerically coincides with the area of ΔA as is shown in the next section.

.....

- (18.4.7)**

DEFINITION: At each point on a closed surface, the POSITIVE DIRECTION of the 3-dimensional area vector, $\Delta \vec{A}$

- is perpendicular to the tangent plane at that point of the surface and
- points away from the enclosed interior, that is, toward the exterior of the closed surface from which it emanates.



NOTE: Table (18.3.4) marks the distinction between Δl and $\vec{\Delta l}$ in Figure (18.4.5), and between ΔA and $\Delta \vec{A}$ in Figure (18.4.7).

(18.5) Vectors Modeling Reality

(18.5.1) **Perception and Mathematics.** Vectors and their components are observed in real life.



For example, observers at different positions will each see different vector components of an automobile's velocity vector, \vec{X} , as Figure (18.5.1a) shows.

- A viewer above the auto (i.e., the reader) sees a 50 mph speed for the skidding car,
- Bernie perceives the car to be moving 40 mph to his left, and
- Ashley observes a 30 mph drift to her right.

In Figure (18.5.1a), Bernie and Ashley each see perpendicular *components* (A.2.1), \vec{X}_1 and \vec{X}_2 , respectively, of the vehicle's velocity vector, \vec{X} , which, by the Pythagorean theorem, has value $\vec{X} = 50$ mph. It is the overhead viewer, the reader, who sees the full velocity vector \vec{X} and its components.

(18.5.1a) **FIGURE:** Perceived Vector Components of Velocity

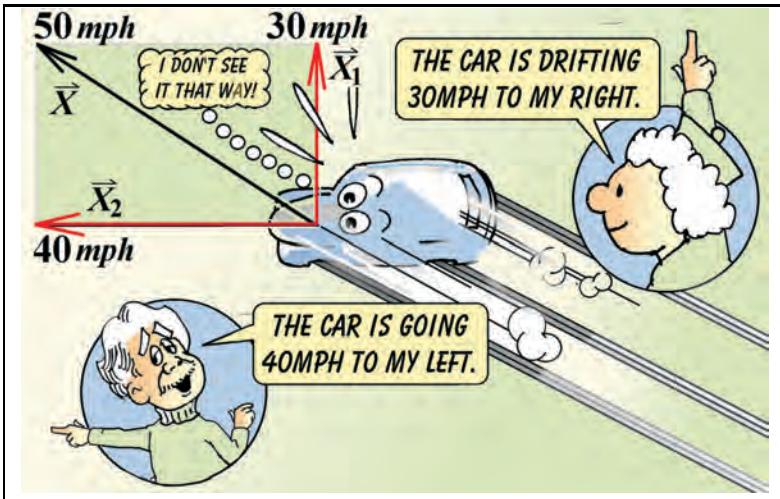


Figure (18.5.1a), representing a 2-dimensional situation, illustrates a mathematical structure—**vector decomposition**—that accounts for three differently perceived speeds of 30 mph, 40 mph, and 50 mph, for the same automobile. We state (without proof) that any

2-dimensional vector \vec{X} has two component vectors, \vec{X}_1 and \vec{X}_2 (see (A.2.4b)) such that $\vec{X} = \vec{X}_1 + \vec{X}_2$, where \vec{X}_1 and \vec{X}_2 form a 90° angle with each other.

In Figure (18.5.1a), the lengths of the vector \vec{X} and its mutually perpendicular (A.2.6b) pair of components, \vec{X}_1 , and \vec{X}_2 , are

$$X = 50 \text{ mph}, X_1 = 30 \text{ mph}, \text{ and } X_2 = 40 \text{ mph}.$$

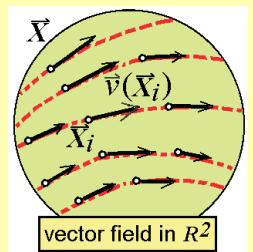
- (18.5.2) **Vector Fields:** Given a region X in space \mathbf{R}^2 or \mathbf{R}^3 . At each point (or vector) \vec{X}_i in the region X , the state of the point \vec{X}_i may be described by an associated vector $\vec{v}(\vec{X}_i)$ that also lies in \mathbf{R}^2 or \mathbf{R}^3 .

For example, the state of a particle at point \vec{X}_i might be described by an associated vector $\vec{v}(\vec{X}_i)$ which could be the *velocity* vector of that particle, or a *force* vector acting on that particle due to gravity or magnetism.

A region X with this structure $\vec{X}_i \rightarrow \vec{v}(\vec{X}_i)$ is called a *field*. More formally,

- (18.5.3)

DEFINITION: Given a region X in either \mathbf{R}^2 or \mathbf{R}^3 where for each point (vector) \vec{X}_i of space X , there is an associated vector $\vec{v}(\vec{X}_i)$ that also lies in \mathbf{R}^2 or \mathbf{R}^3 . Then the region X is called a VECTOR FIELD. Vectors $\vec{v}(\vec{X}_i)$ are called the FIELD VECTORS of X that are tangent to (red dotted) lines that are called FIELD LINES.



- (18.5.4)

|| **NOTE:** The norms $\|\vec{v}(\vec{X}_i)\|$ of the field vectors $\vec{v}(\vec{X}_i)$ are not necessarily constant along red-dotted field lines.

(18.6) Inner and Cross Products

We first encountered the inner product $\langle \vec{X}, \vec{Y} \rangle$ as the coordinate-dependent summation given by Definition (A.2.6). We now present an equivalent definition that is more geometrically based in that $\langle \vec{X}, \vec{Y} \rangle$ depends on lengths of vectors $\vec{X}, \vec{Y} \in \mathbf{R}^3$ and the angle between them. Consistent with the notation of Table (18.3.4), $X = \|\vec{X}\|$ and $Y = \|\vec{Y}\|$, we have

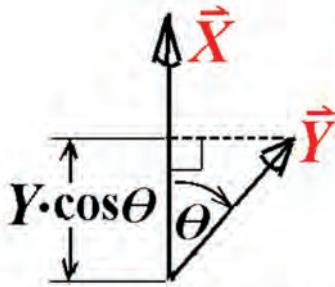
(18.6.1)

DEFINITION: Given the 3-dimensional vectors \vec{X} and \vec{Y} with respective lengths X and Y . Then the INNER PRODUCT, $\langle \vec{X}, \vec{Y} \rangle$, is the number

(18.6.1a)

$$\langle \vec{X}, \vec{Y} \rangle = X \cdot \cos \theta \cdot Y$$

where θ is the angle between vector \vec{X} and vector \vec{Y} .



$$\langle \vec{X}, \vec{Y} \rangle = X \cos \theta Y$$

$$\|\vec{X}\| = X$$

$$\|\vec{Y}\| = Y$$

Beyond 3-Dimensional Reality. Classical trigonometry defines the *cosine* of an angle (*formed by a side and the hypotenuse of a right triangle*) to be the ratio of the adjacent side length to the hypotenuse length.

Now, thanks to the general definition of inner product (*the left-hand side of (18.6.1a)*), which is equivalent to the coordinate-based definition (A.2.6), we can go beyond familiar 3-dimensional space to define the cosine of angles formed by vectors in higher dimensions.

For example, angles that are formed “between” vectors in spaces \mathbf{R}^k of *any* dimension, $k = 1, 2, 3, \dots$ also have a cosine—this extended definition of cosine does not *contradict*, but *agrees with* the familiar 2-dimensional version (A.2.6).

Using (18.6.1a) we unambiguously define

$$(18.6.1b) \quad \cos \theta \stackrel{\text{def}}{=} \frac{\langle \vec{X}, \vec{Y} \rangle}{\|X\| \|Y\|}$$

for any pair of vectors $\vec{X}, \vec{Y} \in \mathbf{R}^k$. Using this number $-1 \leq \cos \theta \leq 1$, we find the angle θ by using the usual 2-dimensional cosine. For example, if (18.6.1b) tells us that for vectors \vec{X} and \vec{Y} in 400-dimensional space, the cosine of the included angle is $\cos \theta = 0.5$, then we would declare that the angle between these vectors is $\theta = 30^\circ$.

A special case. As shown in Definition (A.2.6b) and Figure (A.2.6d), vectors \vec{X} and \vec{Y} are *perpendicular* (*form an angle of 90°*) if and only if $\langle \vec{X}, \vec{Y} \rangle = 0$.

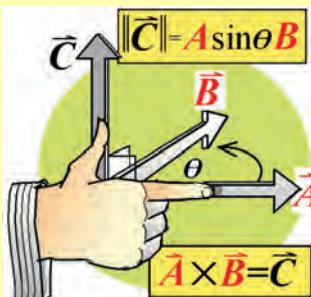
(18.6.2)

NOTE: Two Right-Hand Rules The Right-Hand Rule of (18.4.6) in 3-dimensional space \mathbf{R}^3 defines the “direction” of an enclosed area to be the *direction of travel* along its closed boundary.

In (18.6.3) following, we will define a different right-hand rule, one that determines the positive direction for a new product $\vec{A} \times \vec{B}$ in \mathbf{R}^3 for pairs of vectors \vec{A} and \vec{B} in \mathbf{R}^3 .

(18.6.3)

DEFINITION: Right-Hand Rule and Cross Product $\vec{A} \times \vec{B}$.



Vectors \vec{A} and \vec{B} form an angle θ .

1. Point the index finger of your right hand in the direction of \vec{A} , a vector with length A .
2. Face your palm toward vector \vec{B} whose length is B .
3. The cross-product, $\vec{C} = \vec{A} \times \vec{B}$, is the vector that is parallel to your thumb and perpendicular to the plane formed by \vec{A} and \vec{B} .

(18.6.3a) The length $\|\vec{C}\|$ is $C = A \sin(\theta) B$.

(18.6.4) **Notes on the Angle θ :**

- Trig $\theta = 90^\circ$ When the angle $\theta = 90^\circ$, we have $\sin(\theta) = 1$ which, from (18.6.3a), says that the length $(\vec{A} \times \vec{B}) = A \cdot B$, its maximum value.
- Trig $\theta = 0$ Geometrically, when $\theta = 0$ then \vec{A} and \vec{B} are colinear. Also, since $\sin(\theta) = \sin(0) = 0$, the cross-product $\vec{A} \times \vec{B} = \vec{0}$.

- (18.6.5) **The previous statement tells us:** The cross product of non-zero factors (vectors) *can* produce a zero result (when $\theta = 0^\circ$). Contrast this with the fact that the product of ordinary non-zero numbers *can never* be zero.

- (18.6.6)  **The cross product and the inner product** are both products of a pair of vectors. However, only for vectors \vec{X} and \vec{Y} in 3-dimensional space \mathbf{R}^3 , is the *cross product* $\vec{X} \times \vec{Y}$ defined. Moreover, this cross product is another vector in \mathbf{R}^3 . But for a vector pair $\{\vec{X}, \vec{Y}\}$ in any finite-dimensional space (and even for infinite dimensions), the *inner product*, $\langle \vec{X}, \vec{Y} \rangle$, is always a number, i.e., a vector in \mathbf{R}^1 .

(18.7) Riemann Sums and Integrals

One of the most interesting results derived from Maxwell's equations is the deep connection between magnetic and electric phenomena. For example, as mentioned in (18.4.5), moving a magnet through a wire loop will generate an electric current in the wire. Quantitatively, the current in the wire is determined by the combination of the magnetic field changes on each section of the wire loop.

The mathematical tool for carrying out this "combination" along the wire is called an *integral*. In (18.7.4), we will provide a brief introduction to the definition and properties of integrals.

(18.7.1)

DEFINITION: Let \mathcal{U} be a region in space— \mathcal{U} is either a surface area, A , or an arc, L . Now let \mathcal{U} be decomposed into n separate, mutually exclusive pieces. Then

$$(18.7.1a) \quad \Pi_{\mathcal{U}, \Delta u_i} = \{\Delta u_1, \Delta u_2, \dots, \Delta u_i, \dots, \Delta u_n\}$$

is the notation for this PARTITION of the region \mathcal{U} .

NOTE: The construction of (18.7.1a) is an approximation of a curve with a series of short line segments which, if very short, will closely approximate the original curve. Similarly, with surfaces, we use smaller and smaller polygons to approximate a surface (see Figure (18.7.1c)).

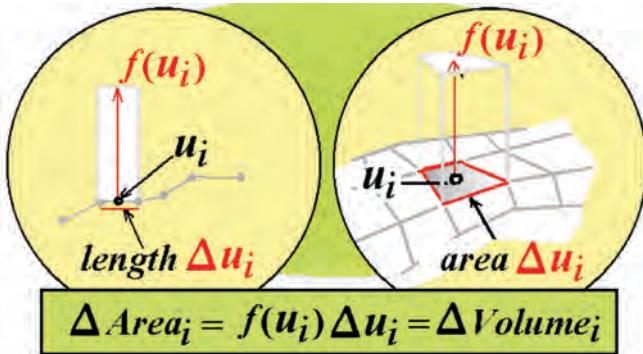
At each point in the region \mathcal{U} , there is defined a real-valued function, $f : \mathcal{U} \rightarrow \mathbf{R}$ (A.3.2) where each Δu_i stands for either an area segment, ΔA_i , or a line segment, Δl_i , accordingly.

Assumption: We will suppose that the Δu_i 's of (18.7.1a), are so small that the function f is virtually constant over each piece, Δu_i —the values of f do not significantly vary over any single Δu_i , or

(18.7.1b) For any two points, u_i and u'_i , in Δu_i , $f(u_i) \approx f(u'_i)$.

Accordingly, at each of the n differential (small) segments, Δu_i , of partition $\Pi_{\mathcal{U}, \Delta u_i}$ (18.7.1a), we may construct the single, unambiguous product,

(18.7.1c)



which we may interpret geometrically as (see Figure (18.7.1c))

- The *ith differential area* of a rectangle over the line segment, Δu_i , whenever region \mathcal{U} is an arc in space, or as
- The *ith differential volume* of a 3-dimensional “box” over surface segment, Δu_i , whenever region \mathcal{U} is a surface in space.

We now have the structure to set the following definition in place:

(18.7.2)

DEFINITION: Given a region \mathcal{U} , which is either a 1-dimensional arc, L , or a 2-dimensional surface A . Let a real-valued function, $f : \mathcal{U} \rightarrow \mathbf{R}$ be defined on \mathcal{U} , which is partitioned into n parts as described in (18.7.1a). Then the Riemann sum relative to this partition, $\Pi_{\mathcal{U}, \Delta u_i}$, is denoted by

$$(18.7.2a) \quad \sum_{\Delta u_i} f(u_i) \Delta u_i \quad \text{or, equivalently,} \quad \sum_{i=1}^n f(u_i) \Delta u_i.$$

(18.7.3)



We need proper notation for the result of taking the limit of Riemann sums (18.7.2a) as the size of the biggest Δu_i goes to zero. (The limit will not depend on the particular partition—all partitions (18.7.1a) will yield the same result (18.7.2a).)

Now if the sum of terms (18.7.2) does not depend on the partition, but only on the region, \mathcal{U} , and the function, f , then what notation is suitable for this two-variable limit? For example, we might choose to represent the Riemann sum

$$\lim_{\max \Delta u_i \rightarrow 0} \sum_{\Delta u_i} f(u_i) \Delta u_i$$

with a symbol such as $\Lambda(\mathcal{U}, f)$. However, we reject this form since it conceals the summation structure of the Riemann sum (18.7.2a). The standard notation for the Riemann sum is

$$(18.7.3a) \quad \int_{\mathcal{U}} f(u) du,$$

which echoes the structure of a Riemann sum in the sense that the symbol \int suggests a “continuous” form of the sum \sum (18.7.2a) where du represents the presence of Δu in the sum.

(18.7.4)
Calculus
needed

DEFINITION: The limit of the sequence of Riemann sums (18.7.2a), as the size of the largest partition piece goes to zero, that is, as $\max\{\Delta u_i\} \rightarrow 0$, is the DEFINITE RIEMANN INTEGRAL,

(18.7.4a)

$$\lim_{\substack{\max \Delta u_i \\ \rightarrow 0}} \sum_{\Delta u_i} f(u_i) \Delta u_i \text{ which is denoted by } \int_{\mathcal{U}} f(u) du.$$

If the region \mathcal{U} is a closed arc (loop) or a closed surface (potato skin), then the integral (18.7.4a) is written with an embellished circle notation,

$$(18.7.4b) \quad \oint_{\mathcal{U}} f(u) du$$

and is called the CLOSED INTEGRAL over \mathcal{U} .

(For the definition of a closed line or surface, see (18.4.3), pg. 244.)

(18.7.5)

For certain special cases, shortcuts have been invented to avoid calculation of the actual limit (18.7.4a).



For example, consider the power functions, $f(x) = x^k$, for a fixed scalar $k \neq -1$. We can compute $\int f(x) dx$ without wrestling with the Riemann sum of (18.7.4a). Thanks to the Fundamental Theorem of Calculus, we remove ourselves from this definition altogether by following precise rules of mechanical manipulation of the fixed exponent, k . In particular, if the region \mathcal{U} is the interval $[a, b]$, this technique produces the tidy answer,

$$\int_{\mathcal{U}} f(x) dx = \int_a^b x^k dx = \frac{b^{k+1} - a^{k+1}}{k + 1}.$$

For more general functions, we are obliged to use the unwieldy limit-based definition (18.7.4a) and stoically deal directly with the Riemann sums (18.7.2a). The desire to devise calculations that defy round-off and accumulation errors (that are due to limitations of computer architecture) is the engine that drives the field of NUMERICAL ANALYSIS.

(18.8) Integrals of the Inner Product

Integrals in the Laboratory. In the study of electromagnetism, there arise integrals that take the form

$$(18.8.1) \quad \int_L \langle \vec{V}_l, d\vec{l} \rangle \quad \text{or} \quad \int_A \langle \vec{V}_l, d\vec{A} \rangle,$$

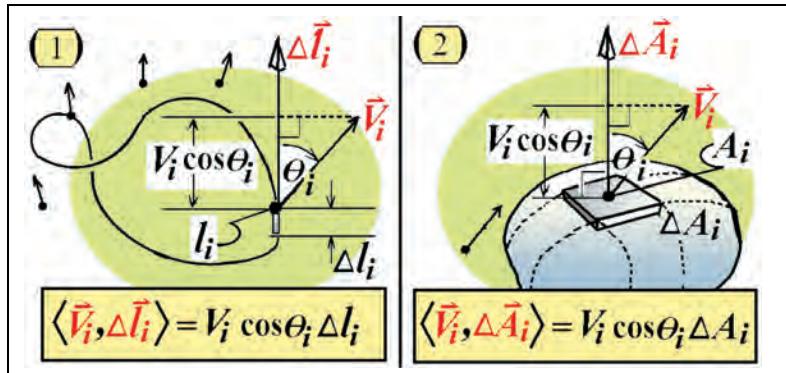
where the 3-dimensional vector \vec{V}_l , a function of \vec{l} , represents either a magnetic field vector \vec{B} or an electric field vector \vec{E} . (*Formal definitions follow in Chapter 19.*)

But the integrals of (18.8.1) do not seem to resemble the Riemann integral form, $\int f(u)du$, given by (18.7.4). That the integrals are, in fact, disguised forms of the Riemann integral (18.7.4a), will now be shown.

The Physical Background for (18.8.1):

Figure (18.8.2), which illustrates this behavior of \vec{V} relative to arcs and surfaces, is the framework for the following discussion that resolves the apparently different forms, (18.7.4a) and (18.8.1), for the Riemann integral.

(18.8.2) **FIGURE: Projection Component of \vec{V} and its Inner Products**



(18.8.3) **The Setup for $\int \langle \vec{V}_l, d\vec{l} \rangle$ of (18.8.1).** See (18.3.3)–(18.3.5).

Choose any point u_i inside the i th differential arc segment Δu_i . The length of the arc is also denoted by the symbol Δu_i , which, in this

case, is also written Δl_i . Through this point u_i , there passes the i th (electric/magnetic) field vector, $\vec{V}(u_i)$, (abbreviated as V_i) which forms an angle $\theta(u_i)$ (abbreviated as θ_i) with respect to the tangent vector, $\Delta \vec{u}_i = \Delta \vec{l}_i$. For physical reasons, we are interested in the product of the following two quantities, (18.8.3a) and (18.8.3b):

$$(18.8.3a) \quad f(u_i) = V(u_i) \cos(\theta(u_i)) = V_i \cos \theta_i,$$

the projection component (18.3.5) of the field vector, \vec{V}_i , and

$$(18.8.3b) \quad \Delta u_i = \Delta l_i,$$

the differential arc length. But the product of (18.8.3a) with (18.8.3b) produces a familiar scalar—the coordinate-free inner product (18.6.1)

$$(18.8.3c) \quad f(u_i) \Delta u_i = V_i \cdot \cos \theta_i \cdot \Delta l_i = \langle \vec{V}_i, \Delta \vec{l}_i \rangle.$$

Summing on the first and third quantities of (18.8.3c) over all $\Delta u_i = \Delta l_i$ of the arc L gives us two forms of the Riemann sum (18.7.2)

$$(18.8.3d) \quad \sum_{\Delta u_i \text{ in } L} f(u_i) \Delta u_i = \sum_{\Delta l_i \text{ in } L} \langle \vec{V}_i, \Delta \vec{l}_i \rangle$$

whose respective limits, as $\max \Delta u_i \rightarrow 0$ (as $\max \Delta l_i \rightarrow 0$) are written

$$(18.8.3e) \quad \int_{arc L} f(u) du = \int_{arc L} \langle \vec{V}_l, d\vec{l} \rangle.$$

We conclude, therefore, that the right-hand side of (18.8.3e), the integral of inner products, is a *bona-fide* Riemann integral consistent with Definition (18.7.4).

(18.8.4) The Setup for $\int \langle \vec{V}_l, d\vec{A} \rangle$ of (18.8.1).

We follow (18.8.3a)-(18.8.3e) almost word for word, except that

- The arc length is replaced by a surface area. Instead of Δu_i representing the i th differential arc length (*Panel (1) of Figure (18.8.2)*), the quantity $\Delta u_i = \Delta A_i$ now represents the i th differential surface element (*Panel (2) of Figure (18.8.2)*).
- Instead of being *tangent* to the i th differential arc length $\Delta \vec{l}_i$ (*Panel (1) of Figure (18.8.2)*), the projection component of vector \vec{V}_i is now *perpendicular* to the i th surface element ΔA_i (*Panel (2) of Figure (18.8.2)*). Note that relative to both cases, $\Delta u_i = \Delta l_i$ and $\Delta u_i = \Delta A_i$, the magnitude of the projection component of \vec{V}_i is $V_i \cdot \cos \theta_i$.

(18.8.5)

DEFINITION: The INTEGRAL OF THE INNER PRODUCT $\langle \vec{V}_l, d\vec{l} \rangle$ over an arc L is defined by,

$$(18.8.5a) \quad \int_{arc L} \langle \vec{V}_l, d\vec{l} \rangle \stackrel{\text{def}}{=} \lim_{\max \Delta l_i \rightarrow 0} \left[\sum_{\Delta l_i} \langle \vec{V}_i, \Delta \vec{l}_i \rangle \right].$$

The scalar $\Delta l_i > 0$ is the length of the i th segment in the partition $\Pi_{L, \Delta l_i}$ (18.7.1) of the arc L . The vector $\vec{V}_i = \vec{V}(l_i)$ is the vector \vec{V} evaluated at some point, l_i , in the segment Δl_i . Similarly, the integral of the inner product $\langle \vec{V}_l, d\vec{A} \rangle$ over the area A is defined by

$$(18.8.5b) \quad \int_A \langle \vec{V}_l, d\vec{A} \rangle \stackrel{\text{def}}{=} \lim_{\max \Delta A_i \rightarrow 0} \left[\sum_{\Delta A_i} \langle \vec{V}_i, \Delta \vec{A}_i \rangle \right].$$

The scalar $\Delta A_i > 0$ is the area of the i th segment in the partition $\Pi_{A, \Delta A_i}$ (18.7.1) of the surface A and the vector $\vec{V}_i = \vec{V}(l_i)$ is the vector \vec{V} evaluated at some point l_i in the area segment, ΔA_i .

(18.8.6)

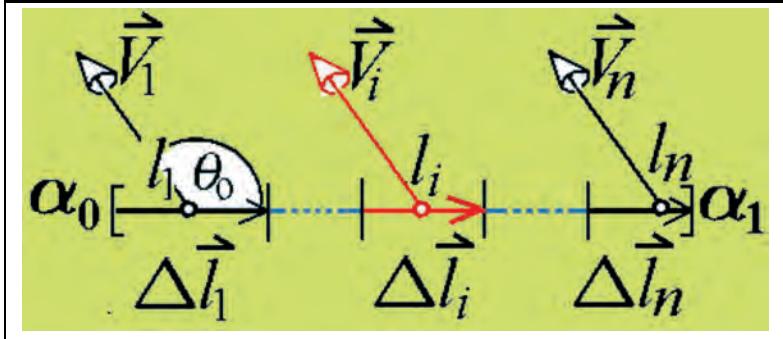
EXAMPLE: *Integrating the inner product over an interval.* We show how Definition (18.8.5a) applies to determine that

$$(18.8.6a) \quad \int_{x=a}^{x=b} \langle \vec{V}_0, d\vec{l} \rangle = V_0 \cos \theta_0 (b - a)$$

where the arc L in (18.8.5a) is the interval $[a, b]$ on the X -axis in \mathbf{R}^3 , and the vector $\vec{V}_0 = V_l$ in \mathbf{R}^3 is constant at every point l in the interval $[a, b]$.

NOTE: *Integrating over the interval $[a, b]$* (18.8.6a), we may use any one of the three (equivalent) forms,

$$(18.8.6b) \quad \int_{x=a}^{x=b} \langle \vec{V}, d\vec{l} \rangle, \int_{[a,b]} \langle \vec{V}, d\vec{l} \rangle, \text{ or } \int_a^b \langle \vec{V}, d\vec{l} \rangle.$$



The partition: Following Definition (18.8.5), the interval $[\alpha_0, \alpha_1]$ is partitioned or cut (18.7.1) into n parts that are represented by the rightward-oriented vectors,

$$(18.8.6c) \quad \Delta \vec{l}_1, \dots, \Delta \vec{l}_i, \dots, \Delta \vec{l}_n,$$

so that

$$(18.8.6d) \quad \sum_{\Delta l_i} \Delta l_i = |\alpha_1 - \alpha_0| > 0. \quad \left\{ \begin{array}{l} \text{the length of the} \\ \text{interval } [\alpha_0, \alpha_1] \end{array} \right.$$

|| **Note:** We use the absolute value notation, $|\alpha_1 - \alpha_0| > 0$, since it may be the case that $\alpha_1 - \alpha_0 < 0$, that is, $\alpha_0 > \alpha_1$.

|| **Note:** We use notation Δl_i to represent both the line segment and the positive length of the line segment.

Now

$$\int_{x=\alpha_0}^{x=\alpha_1} \langle \vec{V}_0, d\vec{l} \rangle = \lim_{\substack{\max \Delta l_i \\ \rightarrow 0}} \sum_{\Delta l_i} \langle \vec{V}_0, \Delta \vec{l}_i \rangle \quad \left\{ \begin{array}{l} \text{from} \\ \text{Definition (18.8.5a)} \end{array} \right.$$

$$= \lim_{\substack{\max \Delta l_i \\ \rightarrow 0}} \sum_{\Delta l_i} V_0 \cos \theta_0 \Delta l_i \quad \left\{ \begin{array}{l} \text{from} \\ \text{Definition (18.6.1a)} \end{array} \right.$$

$$= V_0 \cos \theta_0 \left[\lim_{\substack{\max \Delta l_i \\ \rightarrow 0}} \sum_{\Delta l_i} \Delta l_i \right] \quad \left\{ \begin{array}{l} \text{factor out the} \\ \text{constant } V_0 \cos \theta \end{array} \right.$$

$$= V_0 \cos \theta_0 \lim_{\substack{\max \Delta l_i \\ \rightarrow 0}} |\alpha_1 - \alpha_0| \quad \left\{ \text{from (18.8.6d)} \right.$$

$$= V_0 \cos \theta_0 |\alpha_1 - \alpha_0| \quad \left\{ \begin{array}{l} \text{for constant } |\alpha_1 - \alpha_0|, \\ \lim |\alpha_1 - \alpha_0| = |\alpha_1 - \alpha_0|. \end{array} \right.$$

Equating the first and last terms in this chain of equalities establishes (18.8.6a). It is important to recall that well-defined integrals require that the result is independent of the particular partition, that is, all partitions must yield the same result. ■

(18.9) Exercises

- (18.9.1) Exercise** If X is proportional to numbers Y and Z , then show that X is proportional to their product, YZ .

Hints:

- **The Defining Equations:** To say that the variable X is proportional to the variable Y is to say that

$$X = kY \text{ for some constant } k.$$

For example, if X doubles then Y doubles. If X triples, then so does Y , and so on.

But if there is a third variable, Z , say, then the “constant” k_1 relating X and Y may depend on Z and the constant k_2 relating X and Z may depend on Y . Hence, we denote these “constants” by k_1^Z and k_2^Y , respectively. That is, to say X is proportional to both Y and to Z is to say

$$(18.9.1a) \quad X = k_1^Z Y \quad \text{and} \quad X = k_2^Y Z.$$

where k_1^Z is constant with respect to X and Y , and k_2^Y is constant with respect to X and Z .

- **A Proof:** Divide (18.9.1a) through by YZ to obtain

$$(18.9.1b) \quad \frac{X}{YZ} = \frac{\frac{k_1^Z}{Z}}{\underbrace{f(Z)}_{\text{f}(Z)}} = \frac{\frac{k_2^Y}{Y}}{\underbrace{g(Y)}_{\text{g}(Y)}}.$$

Now the functions $f(Z)$ and $g(Y)$ cannot be equal for all choices of independent variables Z and Y unless they are both equal to the same constant d_0 , say. To see this, note that

$$\frac{d}{dZ} f(Z) = \frac{d}{dZ} g(Y) = 0 \implies f(Z) = g(Y) = d_0 = \text{a constant},$$

which, from (18.9.1b), implies $X = d_0 Y Z$. ■

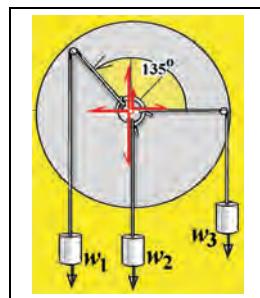
**(18.9.2)
Exercise**

Inner Product Equivalence. Show that definitions (A.2.6) and (18.6.1) are equivalent for vectors \vec{X} and \vec{Y} in 2-dimensional space \mathbf{R}^2 .

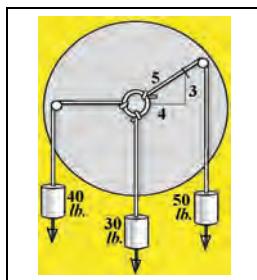
**(18.9.3)
Exercise**

Wheel of Balanced Vectors: (*validating vector decomposition*). The apparatus illustrated will verify that force vectors decompose into their horizontal and vertical components. Hang the weights with a low-friction line that attaches to the center ring, draping the line over the peg to achieve the desired angles. If the forces actually decompose as the mathematical structure indicates, then a stable, balanced center ring occurs when the sum of all forces on it equals zero—that is, when the sum of all *horizontal* components and the sum of all *vertical* components each equals zero.

- (18.9.3a) Given Fixed Angles.** Find the ratio of the weights w_1 , w_2 , and w_3 that produce equilibrium, i.e., the sum of the forces on the center ring is zero.



- (18.9.3b) The 3-4-5 Triangle.** If the three forces, 30 lbs., 40 lbs., and 50 lbs., decompose according to the geometry of the 3-4-5 triangle, then equilibrium on the central ring is achieved.



- (18.9.3c) Forbidden Angles.** At what angles, if any, in Figure (18.9.3a), is it impossible to achieve equilibrium on the central ring, regardless of the value of the weight w_1 ? (**Note:** All weights w_1 , w_2 , w_3 are positive.)

(18.9.4) Integration and Electromagnetism. Show that

Exercise

$$(18.9.4a) \quad \int_A \langle \vec{V}, d\vec{A} \rangle$$

is a Riemann integral (18.7.4). **Hint:** Model your response on the similar argument that follows Figure (18.8.2).

(19) Electric and Magnetic Fields

(19.1) Background

In 1873, the Scottish physicist JAMES CLERK MAXWELL (1831–1879) published *A Treatise on Electricity and Magnetism* in which he extended and unified discoveries of CHARLES-AUGUSTIN DE COULOMB, CARL FRIEDRICH GAUSS (1777–1855), ANDRÉ MARIE AMPÈRE (1775–1836), and MICHAEL FARADAY (1791–1867).

With only four fundamental equations (listed in Section (21.7)), Maxwell was able to integrate the theories of electricity, magnetism, and optics. Moreover, these equations enabled him to make the armchair calculation that the speed of electromagnetic waves in a vacuum was $c = 3 \times 10^8$ meters/second, a number that exactly coincided with the experimentally measured speed of light.

Merely a coincidence? Maxwell thought not.

Maxwell suspected, but could not prove, that light was as much an electromagnetic phenomenon as the magnetic and electric fields he studied in his laboratory—later experiments were to fully confirm his suspicions.

.....

The catalog of electromagnetic (EM) waves was extended in 1888 when HEINRICH RUDOLPH HERTZ (1857–1894) (*whose name “hertz,” abbreviated as (Hz), became used for radio and electrical frequencies*) showed that oscillating electrical charges produced electromagnetic waves (radio waves). Although Hertz failed to measure the speed of electromagnetic waves, later investigators were to confirm that the waves move at the speed of light c .

But what about light whose speed is also c ? Is light a special electromagnetic wave? If oscillating electrical charges produced electromagnetic waves, then which oscillating charges were responsible for the production of light? In 1895, Dutch physicist HENDRIK ANTOON LORENTZ (1853–1928) suggested that the emitted light was generated by oscillating charges within the atoms. In 1896, Lorentz’s student, PIETER ZEEMAN (1865–1943), confirmed this hypothesis. At this

point, light became a valid entry in the catalog of electromagnetic waves. (See also [8], pp. 267–269.)

This chapter will describe ideas of electricity and magnetism that eventually led to Maxwell's marriage of observation and mathematics that united the theory of electricity, magnetism, and optics.

(19.2) Electric Forces: Coulomb's Law

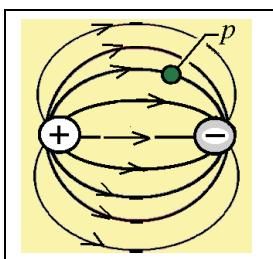
Electrically charged particles have either a plus or a minus charge. An electron, a negatively charged particle, that is placed between a plus and a minus charged particle will experience a force toward the positive particle. (*The positive/negative terminology for electrical charges is due to BENJAMIN FRANKLIN (1706–1790).*)

(19.2.1) || **NOTE:** By convention, we use the single symbol, q , to represent both the physical particle and the charge on that particle.

(19.2.2)

DEFINITION: The collection of all forces in a region experienced by an electron in that region is the FORCE FIELD created by a set of charged particles of the region. The paths within the region that the electrons take as they yield to the attractive force of the positive particle(s) are called the LINES OF FORCE.

(19.2.3) **FIGURE: Rules for Drawing Electric Field Lines**



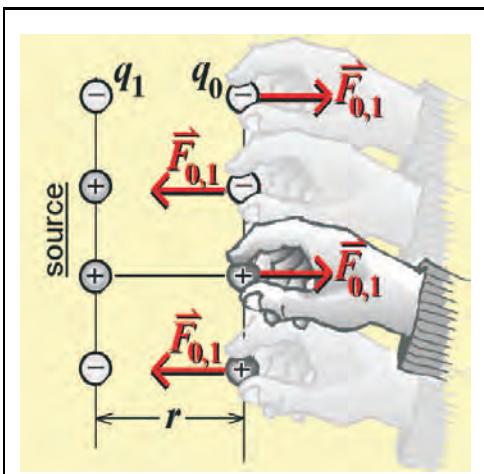
- (1) Lines run from + charges to – charges whenever there are \pm pairs of electrically charged particles.
- (2) Lines never cross.
- (3) The number of lines emanating from any particle is proportional to the charge of that particle.
- (4) A charged particle positioned on a field line experiences a force that is tangent to that field line. Two charges with identical signs repel each other, and two particles of opposite sign attract each other.

A graphical representation. Figure (19.2.3) shows two charged particles with opposite charges. By convention, lines of force (within the force field) always emanate from positively charged particles. An analytical description of the lines of force will follow.

(19.2.4)

FIGURE: Forces Between Charges: Coulomb's law

Two electrically charged particles— q_0 at point p_0 in space and q_1 —are r units apart. Experimental measurements confirm that a test particle q_0 (being held) exerts a resisting force, $\vec{F}_{0,1}$, whose magnitude $F_{0,1}$ is proportional to the charge, q_0 , the charge q_1 , and to $1/r^2$, the inverse square of the distance r . These proportionality relationships, as per Theorem (18.2.2), are expressed mathematically as



$$(19.2.4a) \quad F_{0,1} = k_1 q_0, \quad F_{0,1} = k_2 q_1, \quad \text{and} \quad F_{0,1} = k_3 (1/r^2) \\ \text{for certain proportionality constants } k_1, k_2, k_3.$$

Since (19.2.4a) tells us that $F_{0,1}$ is proportional to q_0 , q_1 and $(1/r^2)$, Theorem (18.2.2) will guarantee that $F_{0,1}$ is proportional to their product, i.e., $F_{0,1} = k_0(q_0)(q_1)(1/r^2)$ for a certain proportionality constant k_0 . This is **COULOMB'S LAW** which has form

$$(19.2.4b) \quad \vec{F}_{0,1} = \underbrace{\left(\frac{1}{4\pi\varepsilon_0} \right)}_{k_0} \frac{q_0 q_1}{r^2} \vec{e}_{0,1}$$

where $k_0 = \left(\frac{1}{4\pi\varepsilon_0} \right)$ is the constant of proportionality and

$$(19.2.4c) \quad \vec{e}_{0,1} = \frac{\vec{r}_1 - \vec{r}_0}{\|\vec{r}_1 - \vec{r}_0\|} \text{ is a unit vector } (\|\vec{e}_{0,1}\| = 1) \text{ pointing}$$

from q_0 to q_1 (see (19.2.4)).

From (A.2.13a), pg. 364, the length of the vector $\vec{F}_{0,1}$ of (19.2.4b) is

$$(19.2.4d) \quad F_{0,1} = \|\vec{F}_{0,1}\| = \left| \frac{1}{4\pi\varepsilon_0} \frac{q_0 q_1}{r^2} \underbrace{\|\vec{e}_{0,1}\|}_1 \right| = \frac{1}{4\pi\varepsilon_0} \frac{q_0 q_1}{r^2}.$$

- (19.2.5) **Permittivity.** If we place known and measurable quantities, q_0 , q_1 , π , r , and $F_{0,1}$, on one side of the equation (19.2.4d), we are left with the single “unknown” quantity ε_0 , the electrical **PERMITTIVITY**, which computes to be

$$(19.2.5a) \quad \varepsilon_0 \stackrel{\text{def}}{=} \frac{1}{4\pi} \frac{q_0 q_1}{r^2 F_{0,1}}.$$

- (19.2.6) **The sign of the product**, $q_0 q_1$, in (19.2.4b) will algebraically point the unit vector $\vec{e}_{0,1}$ in the correct direction: That is, $q_0 q_1 < 0$ implies that the vector $-\vec{e}_{0,1}$ points toward the source particle (*particles with different-signed charges will attract*) while $q_0 q_1 > 0$ implies $+\vec{e}_{0,1}$ points away from the source particle (*particles with same-signed charges will repel*).
- (19.2.7) Permittivity, ε_0 (with the 0 subscript), depends on the *vacuum* in which the force $\vec{F}_{0,1}$ of (19.2.4b) is measured. Measurements of $\vec{F}_{0,1}$ taken in other environments—nitrogen, air, water, etc. produce their own permittivity constants, ε .

The units of permittivity, ε_0 , are

$$(19.2.7a) \quad \frac{q^2}{l^2 \times F} \approx \frac{\text{charge}^2}{\text{length}^2 \times \text{Force}}.$$

What Are the Specific Units for (19.2.7a)? At this point it is not necessary to state a particular system of units since the relation is valid for all unit systems. However, when we later calculate the speed of light in (23.4.3), we will use MKS units, namely,

- (1) charge $q = C$ (Coulombs),
- (2) length $l = \text{mtr}$ (meters),
- (3) force $F = N$ (Newtons).

(19.3) Electric Fields

(19.3.1)



Multiple Source Charges. Coulomb's law (19.2.4b) gives us the force between two particles q_1 and q_0 . But what force is produced by n source charges q_1, q_2, \dots, q_n on a single "target" or test charge q_0 that is fixed at position p_0 (see Figure (19.3.4)) ?

The *experimental* result is that for n charges, the force on charge q_0 at position p_0 is the *vector sum* (A.2.4b) of the electric forces generated by the source charges q_i , each acting independently of the other charges—as if the other charges were not present. This experimentally confirmed vector-sum behavior does not follow automatically from Coulomb's law (19.2.4b).

(19.3.2)

Question: How can we measure the total force exerted by several source charges q_1, q_2, \dots, q_n on a single test charge q_0 ?

Answer: Since each force $\vec{F}_{0,i}$ (19.2.4b) between q_i , $i = 1, 2, \dots, n$, and the test particle q_0 is proportional to both charges q_i and q_0 , we normalize the situation by setting the test particle charge $q_0 = 1$. This normalization ensures that the force computed at point p_0 does not depend on q_0 , which is always equal to 1. Equivalently, we can always divide equation (19.2.4b) through by the value of the target test charge q_0 which then yields the force per unit charge due to a single source charge q_1 , namely,

(19.3.3)

DEFINITION: A TEST CHARGE q_0 is located at a point p_0 . The force-per-unit-charge, or the ELECTRIC FIELD VECTOR at a point p_0 due to a single charge q_1 that lies r units from p_0 is the vector

$$(19.3.3a) \quad \vec{E}_{0,1} \stackrel{\text{def}}{=} \frac{\vec{F}_{0,1}}{q_0} = \frac{q_1}{(4\epsilon_0\pi r^2)} \vec{e}_{0,1}$$

where the direction of the unit vector $\vec{e}_{0,1}$ is determined by q_0 and q_1 in accordance with (19.2.6).

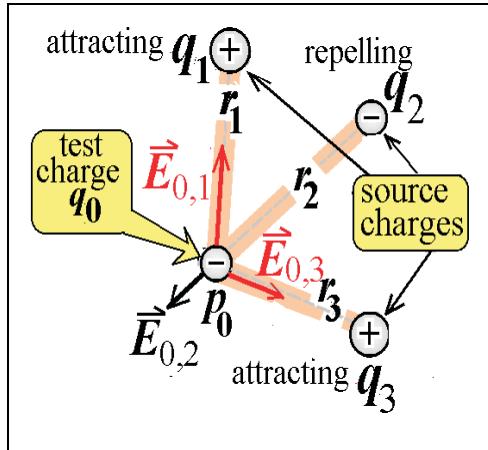
We now extend the quotient of (19.3.3a), which is defined for a single source charge q_1 , to several test charges q_1, q_2, \dots, q_n .

(19.3.4) **FIGURE:** Electric Field Vector due to Many Charges

The Setup: Given n charged (source) particles,

$q_1, q_2, \dots, q_i, \dots, q_n$,
with respective distance
vectors,

$\vec{r}_1, \vec{r}_2, \dots, \vec{r}_i, \dots, \vec{r}_n$
emanating from a
charged test particle
 q_0 situated in space at
point p_0 . Note that
 $r_i = \|\vec{r}_i\|$ as declared in
(A.2.13a).



Definition (19.3.3) describes the electric field vector $\vec{E}_{0,1}$ at point p_0 due to a single source charge q_1 that lies r_1 units distant. Definition (19.3.5) will describe the “composite” electric field vector \vec{E}_0 at point p_0 due to several source charges, $q_i, i = 1, 2, \dots, n$, where the source charge q_i lies r_i units from point p_0 .

(19.3.5)

DEFINITION: The ELECTRIC FIELD VECTOR, \vec{E}_0 at a point p_0 due to n source particles $\{q_i : i = 1, 2, \dots, n\}$ is the vector sum (A.2.4b) of all (force-per-unit-charge) field vectors $\vec{E}_{0,i}$ (19.3.3a) generated by all the source charges in the region, namely,

$$(19.3.5a) \quad \vec{E}_0 \stackrel{\text{def}}{=} \sum_{i=1}^n \vec{E}_{0,i} = \sum_{i=1}^n \frac{\vec{F}_{0,i}}{q_0} = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^n \frac{q_i}{r_i^2} \vec{e}_{0,i}.$$

An ELECTRIC FIELD \vec{E} (no subscript) is the set of all electric field vectors of a region.

Note: The construction of electric field vectors in (19.3.5a) is but one example in nature where mathematical vector addition (A.2.4b) applies.

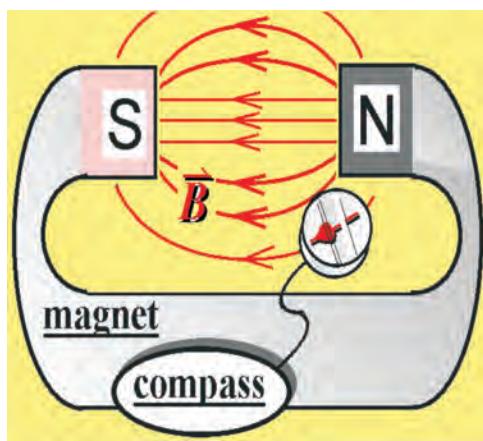
(19.4) Magnetic Fields

Just as we have an electric field \vec{E} (19.3.5) of vectors in 3-space that exert forces on electrically charged particles, there are magnetic fields \vec{B} , represented by vectors in 3-space, that exert forces on *moving* charged particles.

Magnetic fields are confirmed by their effects on moving charges and on other magnets. By using a compass needle as the “other magnet,” we can see the effects of a magnetic field.

- (19.4.1) In sum, an electric field \vec{E} is a vector field that results from nearby charged particles—a MAGNETIC FIELD \vec{B} is a vector field created by moving charged particles or a permanent magnet.

(19.4.2) **FIGURE: Magnetic Field Lines**

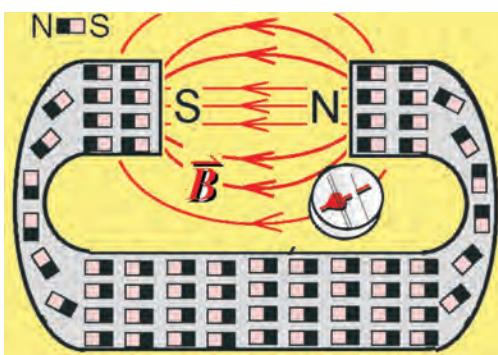


A magnet, with ends marked “S” (South) and “N” (North) is shown with invisible lines, \vec{B} , of magnetic force running from “N” to “S”. The magnetic field lines are revealed when the needle of a (south-seeking) compass is observed to follow the lines as indicated in the diagram. The “N” and “S” notations are arbitrary but universally accepted.

- (19.4.3) The following rules for drawing magnetic field lines are similar to the rules (19.2.3) for drawing electric field lines:

- (1) Lines run from North to South poles.
- (2) Lines never cross.
- (3) The number of lines emanating from any North pole, or ending at any South pole, is proportional to the strength of that pole.
- (4) A magnetic needle experiences a force tangent to the field line. Since like poles repel and opposite poles attract, the North pole of the needle is attracted to an external South pole.

- (19.4.4) **FIGURE: (No Magnetic Monopoles Exist.)**



An object is a magnetic MONOPOLE if it exhibits the properties of either a South pole or a North pole, but not both. However, no magnetic monopoles have yet been observed. No matter how small a piece is taken from a magnet, there is always both a

North end and a South end. Unlike positive and negative electric charges that can exist in isolation from each other, magnetic materials can *not* be decomposed so as to isolate a North pole from a South pole. As Figure (19.4.4) indicates, the magnet is composed of atomic “micro” magnets which, due to their alignment, produce a “macro” magnet.

NOTE: To reveal the direction of the magnetic field, we use a south-seeking compass, as shown in Figure (19.4.2). A standard north-seeking compass becomes south-seeking after the arrowhead of the needle is transferred to the opposite end.

(19.5) Magnetic Forces: Lorentz Forces

- (19.5.1)** How Nature requires the cross product (18.6.3). We know from Coulomb's law (19.2.4) that forces are exerted on charged particles in an *electric* field, whether they are in motion or not. In contrast to this, forces are exerted on charged particles in a *magnetic* field only when they are in motion.



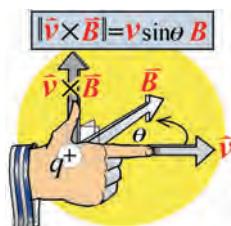
Moreover, the particle motion and the resulting force must both be against the “grain” or lines of the magnetic field. That is, experiments show that the force on a moving charge is perpendicular to both the velocity \vec{v} and the \vec{B} field which provides an example of the cross product (18.6.3) in Nature.

Here are the details.

IN THE WORLD OF MECHANICS: If you jump into a stream whose current runs right to left, then you would feel a force from the right pushing you leftward (downstream).

IN AN ELECTRIC FIELD: If the “stream” is an electric field running right (+) to left (-), and you are a positively charged particle, then, similar to the mechanical condition above, you would feel a “downstream” force pushing you along the field lines in the (+)-to-(−) direction.

HOWEVER, IF YOU ARE A CHARGED PARTICLE IN A MAGNETIC FIELD or stream, and you run (motion is necessary) across the magnetic field lines which run right (N) to left (S), then you would not be pushed leftward, or downstream. Instead, consistent with the geometry of the cross-product, you would be pushed upward, toward the heavens. (*You must be positively charged to experience an upward thrust. A negatively charged particle is destined to be pushed downward, toward—well—the other place.*)

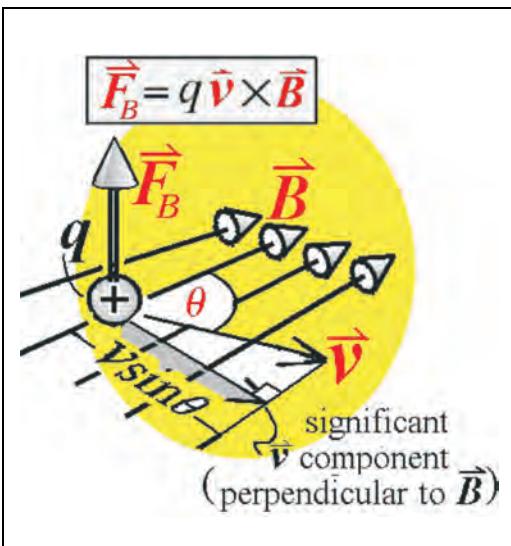


These ideas are codified in (19.5.2a) of the following discussion:

(19.5.2)

FIGURE: Magnetic Force on a Moving Charge

Trig



A particle with charge q travels with velocity \vec{v} at an angle θ with respect to the magnetic field lines. Only the component of \vec{v} that is perpendicular to the magnetic field lines contributes to the induced force, \vec{F}_B , whose direction is determined by the cross product, $\vec{v} \times \vec{B}$. The length of the perpendicular part of \vec{v} is $v \cdot \sin \theta$ (18.6.3) (Figure at left).

In sum, the magnetic force \vec{F}_B induced by a particle with charge q and velocity \vec{v} , passing through a magnetic field line \vec{B} is

$$(19.5.2a) \quad \vec{F}_B = q \vec{v} \times \vec{B} \quad \text{so that, as per (18.6.3),}$$

$$(19.5.2b) \quad \|\vec{F}_B\| = \|q \vec{v} \times \vec{B}\| = q \cdot v \cdot \sin \theta \cdot B$$

where $v = \|\vec{v}\|$, $B = \|\vec{B}\|$ (18.3.4) and particle with charge q travels at an angle θ with respect to the magnetic field line \vec{B} .

As a consequence of (19.5.2a), an electrically charged particle moving through a magnetic field \vec{B} experiences no force if its path is parallel to the field lines of Figure (19.4.2). In this case $\theta = 0$ in Figure (19.5.2) so that using $\sin \theta = 0$ in (19.5.2b), we see that force $\vec{F}_B = 0$. Similarly, the force \vec{F}_B is maximal when the path is perpendicular to the field lines, i.e., $\theta = 90^\circ$ implies $\sin \theta = 1$ in (19.5.2b).

NOTE: In contrast to the magnetic force on a charged particle (19.5.2a), which depends on the speed of the particle, the electrical force on charged particles, as given by Coulomb's law, (19.2.4b), is not affected by particle speeds at all. For two particles, the electrical force between them at any instant of time depends only on their charges, q_0 and q_1 , and the instantaneous distance r between them.

The force between charged particles at a distance r changes in time when the distance r changes. This raises the question of how fast (or instantaneous) the signal must be between particles to effect the force change.

- (19.5.3) **The Lorentz Force.** Forces on a charged particle are due to (1) electric fields generated by neighboring charged particles (Coulomb's law (19.2.4b)) and (2) magnetic fields due to the speed of the particle. Since both electric and magnetic fields can exist simultaneously, we have the Lorentz Law that combines (1) and (2) as the vector sum (A.2.4b):

$$(19.5.3a) \quad \vec{F} = q\vec{v} \times \vec{B} + q\vec{E},$$

where q is the charge moving through the electric field \vec{E} and the magnetic field \vec{B} with (vector) velocity \vec{v} .

NOTE: The coexistence of both electric and magnetic fields allowed for Thomson's discovery of the electron. See Section (19.6).

In the following section, we shall see how J. J. Thomson, inferring from effects alone, manages to "see", and discover, the *electron*.

(19.6) How Thomson Discovers the Electron

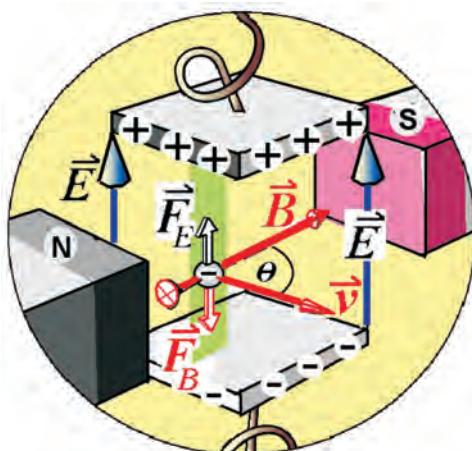
The interplay of the forces induced by the electric field (19.2.4b) and the magnetic field (19.5.2a) led to the discovery of the electron. Here is how it happened.

In 1897, J. J. THOMSON (1856–1940) conducted an experiment at Cambridge University that involved an evacuated cathode ray tube (CRT), much like the picture tube in a television receiver. At one end of the tube, a hot filament emitted charged particles, later determined by Thomson to be electrons. The stream of particles was

channeled through a slit to form a narrow beam that collided with and illuminated a spot on the phosphorescent screen at the other end.

Here follows a description, with schematics, of the crossed-fields apparatus used in Thomson's experiment:

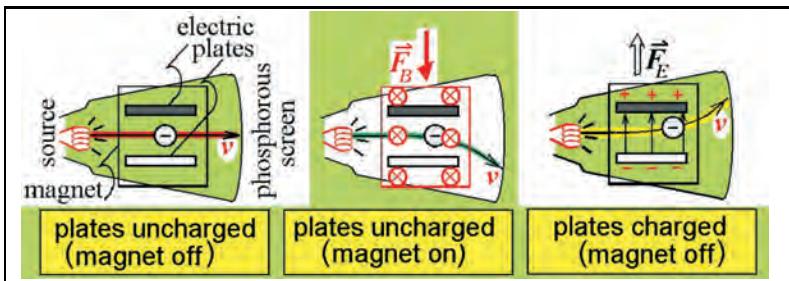
(19.6.1)

FIGURE: Electric and Magnetic Forces on an Electron

- A magnetic field, \vec{B} , runs horizontally (\rightarrow) from the N pole to the S pole.
- An electric field, \vec{E} , runs vertically (\uparrow) from the negative ($-$) plate to the positive ($+$) plate.
- A negatively charged electron, whose path forms an angle θ with the magnetic field, travels horizontally, at speed v . The angle θ is adjusted by rotating the magnet. In this way, the magnetic force on particles with charge q varies with θ according to the formula, $\vec{F}_B = q\vec{v} \times \vec{B} = q \cdot v \sin \theta \cdot B$ (19.5.2a). Due to the Right Hand Rule (19.5.2) for the magnetic field, the force \vec{F}_B will direct *negatively charged* electrons downward. The electric field \vec{E} on the other hand, drives each electron upwards with force \vec{F}_E .

magnetic force on particles with charge q varies with θ according to the formula, $\vec{F}_B = q\vec{v} \times \vec{B} = q \cdot v \sin \theta \cdot B$ (19.5.2a). Due to the Right Hand Rule (19.5.2) for the magnetic field, the force \vec{F}_B will direct *negatively charged* electrons downward. The electric field \vec{E} on the other hand, drives each electron upwards with force \vec{F}_E .

Figure (19.6.2) shows how the trajectory of the electron beam, whose particles have speed v , is affected by the magnetic and electric fields separately.

(19.6.2) **FIGURE: Beam Deflection in Electric and Magnetic Fields**

- LEFT PANEL: Both the electric and magnetic fields are off. The beam travels horizontally to its reference point on the screen.
- MIDDLE PANEL: The electric field is off while the magnetic field, at an angle θ to the path of the beam, is turned on. As in Figure (19.6.1), the “x”s indicate the tail of the magnetic \vec{B} arrow—the magnetic field is directed into the page. From (19.5.2a), the magnetic force \vec{F}_B on the *negatively* charged particle is downward with magnitude $q \cdot v \cdot \sin \theta \cdot B$.
- RIGHT PANEL: The magnetic field is off and the electric field is turned on. The beam is forced upward with force \vec{F}_E due to the attractive force of the positive plate at the top and the repelling force of the negative plate at the bottom.

(19.6.3) **Thomson used the apparatus as follows:** With the two fields turned off (left panel), the position of the undeflected beam on the screen is marked. Once the electric field \vec{E} is turned on (right panel), the position of the deflected beam is measured. Then, the magnetic field is turned on and adjusted until the magnetic force equals that of the electric field and returns the beam to its original, undeflected, position. From the geometry of the cathode ray tube, the deflection distance and the combined forces of the fields \vec{E} and \vec{B} , as governed by the Lorentz equation (19.5.3), Thomson calculated

$$(19.6.3a) \quad v = E/B, \quad \text{the speed } v \text{ of the negative charges, and} \\ m/q, \quad \text{the ratio of mass to electric charge.}$$

The low value of m/q meant that each negatively charged particle, now called an *electron*, had mass much less than 1/10 of one percent of the mass of the smallest known atom, hydrogen. The existence of

such a tiny mass meant that the atom (which means “indivisible”) was not indivisible at all. In fact, Thomson asserted, these ultra-light particles, must be common to, and part of, every atom of every element.

These inferences of Thomson, made in 1897, mark the discovery of the (apparently) invisible electron. Thomson, following the principles of the scientific method ([F.1.1](#)), embraced its reality through its effects.

Thomson received the Nobel prize for physics in 1937.

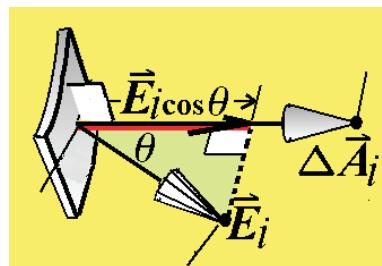
(20) Electricity and Magnetism: Gauss' Laws

(20.1) Flux of Vector Fields

We consider a vector field \vec{E} (18.5.3), pg. 248, which is composed of field vectors including \vec{E}_i , $i = 1, 2, \dots, k$. For example, the individual field vectors \vec{E}_i might be water speeds, electric field vectors, magnetic field vectors, or gravitational field vectors.

As per (18.3.3), we can represent a small element with area ΔA as a vector $\Delta \vec{A}$ whose length is ΔA . That is, $\|\Delta \vec{A}\| = \Delta A$.

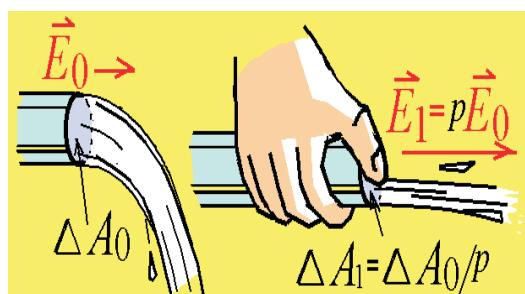
- (20.1.1) **FIGURE: Passing Through a Surface Element.** Our objective is to measure the flow of “ E -stuff” or so-called *flux* passing through a surface. To do this, we only measure the perpendicular component of \vec{E}_i which has length $\|\vec{E}_i \cos(\theta)\|$ where θ is the angle between vectors \vec{E}_i and ΔA_i .



Flow along the surface cannot be said to *pass through* the surface.

- (20.1.2) **EXAMPLE: Varying Area and Constant Flux of a Water Stream**

A 3-dimensional velocity vector field \vec{E} describing the flux (*amount of water volume per unit time that flows in the direction of flow*) of incompressible water flows through a hose with cross-sectional



area $\Delta A_0 = 12 \text{ cm}^2$. The average velocity vector \vec{E}_0 runs perpendicularly to the cross-sectional area with magnitude $\|\vec{E}_0\| = 8 \text{ cm/sec}$. The flux defined in (20.1.3) gives us the rate of volume flow through ΔA_0 , namely,

$$(20.1.2a) \quad \Phi_E(\Delta A_0) = \langle \vec{E}_0, \Delta \vec{A}_0 \rangle$$

$$= \|\vec{E}_0\| \cos(0) \|\Delta \vec{A}_0\| = 8 \frac{\text{cm}}{\text{sec}} \times 1 \times 12 \text{ cm}^2 = 96 \frac{\text{cm}^3}{\text{sec}}$$

Therefore, as shown in Figure (20.1.2), when the exit area ΔA_0 is *decreased* by a factor of p , as when a thumb is held over the hose end, then the speed of flow, \vec{E}_0 , *increases* to $\vec{E}_1 = p\vec{E}_0$. Conversely, if the exit area *increases* by a factor p , then the water speed *decreases* (is divided by) the factor p .

(20.1.3)

DEFINITION OF FLUX: Given vector field \vec{E} (18.5.3) containing a surface area A that is partitioned into k small disjoint surface elements $\Delta A_i, i = 1, 2, \dots, k$. Sub-area ΔA_i is tiny enough so that an individual field vector \vec{E}_i passing through it can be considered to be constant. Then the APPROXIMATE FLUX through the i-th area element ΔA_i is the inner product (A.2.6a)

$$(20.1.3a) \quad \Phi_E(\Delta A_i) \approx \langle \vec{E}_i, \Delta \vec{A}_i \rangle = \|\vec{E}_i\| \cos \theta \|\Delta \vec{A}_i\|$$

where $\|\vec{E}_i\|$ and $\|\Delta \vec{A}_i\|$ are the lengths of vectors \vec{E}_i and $\Delta \vec{A}_i$. The APPROXIMATE FLUX through the entire area A is the sum of the inner products

$$(20.1.3b) \quad \Phi_E(A) \approx \sum_{i=1}^k \langle \vec{E}_i, \Delta \vec{A}_i \rangle = \sum_{i=1}^k \|\vec{E}_i\| \cos \theta \|\Delta \vec{A}_i\|.$$

Through the limit process, we define the TOTAL FLUX through the area A to be

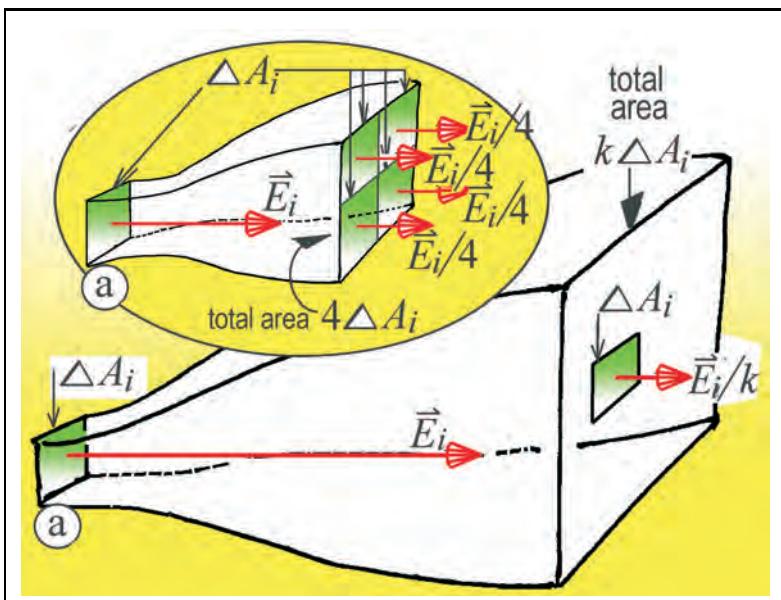
$$(20.1.3c) \quad \Phi_E(A) \stackrel{\text{def}}{=} \int_A \langle \vec{E}, d\vec{A} \rangle$$

$$= \lim_{\max \|\Delta A_i\| \rightarrow 0} \left[\sum_{\Delta A_i} \|\vec{E}_i\| \cos \theta \|\Delta \vec{A}_i\| \right].$$



- (20.1.4) The following example (20.1.5) illustrates a field \vec{E} constrained in a “tubular” region in which flux is constant through each cross-section. This constancy of flux prevails with fields of incompressible fluids, magnetic fields, and electric fields. We may think of our flowing substance acting like incompressible water flowing through a volume.

- (20.1.5) **FIGURE:** *Constant Flux Through Parallel Cross-Sections*



Over a sample area ΔA at ①, field E has field (flow) vector with average value \vec{E}_i . Hence, from (20.1.3a), the flux at the entry area cross-section ① is

$$(20.1.5a) \quad \Phi_E(\Delta A_i) = \|\vec{E}_i\| \cos 0^\circ \|\Delta \vec{A}\| = \|\vec{E}_i\| \cdot 1 \cdot \Delta A.$$

Small Inset Figure (20.1.5). Since the exit area cross-section (*to the right of the entry area ①*) is the sum of four areas

$$\Delta A_i + \Delta A_i + \Delta A_i + \Delta A_i = 4\Delta A_i,$$

the flow vector \vec{E}_i must distribute itself over the four pieces of this larger cross-sectional area. (*Recall from (20.1.4) we use water flow as our model and hence assume there is no source or sink inside the volume and that the flowing substance is incompressible.*)

This means that each of the four areas of size ΔA_i accepts only $(1/4)$ th of the flow, or $\frac{1}{4}\vec{E}_i$. The flux at each quarter of the exit cross-sectional area is $\frac{1}{4}\|\vec{E}_i\|\Delta A$ which implies that the total flux over all four regions is the 4-term sum

(20.1.5b)

$$\underbrace{\frac{1}{4}\|\vec{E}_i\|\Delta A_i + \frac{1}{4}\|\vec{E}_i\|\Delta A_i + \frac{1}{4}\|\vec{E}_i\|\Delta A_i + \frac{1}{4}\|\vec{E}_i\|\Delta A_i}_{\text{Total Flux}} = \|\vec{E}_i\|\Delta A_i.$$

$$\text{Total Flux} = \sum_{i=1}^4 \frac{1}{4}\|\vec{E}_i\|\Delta A_i$$

Larger Diagram of Figure (20.1.5). Since the exit area cross-section (*to the right of the entry area ④*) is $k\Delta A_i$, the flow vector \vec{E}_i must distribute itself over the k pieces of this larger area. This means that each of the k areas of size ΔA_i accepts only $(1/k)$ th of the flow, or $(1/k)\vec{E}_i$. The flux at each smaller piece of the exit area is $(1/k)(\vec{E}_i) \times \Delta A$ which implies that the total flux over all k regions is the k -term sum

(20.1.5c)

$$\underbrace{\frac{1}{k}\|\vec{E}_i\|\Delta A_i + \frac{1}{k}\|\vec{E}_i\|\Delta A_i + \dots + \frac{1}{k}\|\vec{E}_i\|\Delta A_i}_{\text{Total Flux}} = \|\vec{E}_i\|\Delta A_i$$

$$\text{Total Flux} = \sum_{i=1}^k \frac{1}{k}\|\vec{E}_i\|\Delta A_i$$

(20.2) Electric and Magnetic Flux

As per Definition (20.1.3), electrical “throughput,” or *flux* is defined for an arbitrary surface area, A as follows:

(20.2.1)

Calc

DEFINITION: Given a surface area A in space that lies within an electric field \vec{E} . Then the integral

$$(20.2.1a) \quad \Phi_E(A) = \int_A \langle \vec{E}, d\vec{A} \rangle$$

is called the ELECTRIC FLUX through the area A .

- NOTE:** If the integral (20.2.1a) is taken over a closed surface area A (like the “potato surface” shown in the Figure of (18.4.7), pg. 246), then the ornamented integral symbol, \oint_A , is used in place of \int_A .
- NOTE:** the integral (20.2.1a) has been defined in (18.8.5b) of Definition (18.8.5).
-

Similar to Definition (20.2.1), magnetic flux will be defined by (20.2.2) following.

(20.2.2)
Calculus
needed

DEFINITION: Given a surface area A in space that lies within a magnetic field, \vec{B} . Then the integral

$$(20.2.2a) \quad \Phi_B(A) = \int_A \langle \vec{B}, d\vec{A} \rangle$$

is called the MAGNETIC FLUX through the area A .

(20.3) Gauss' Law for Electricity

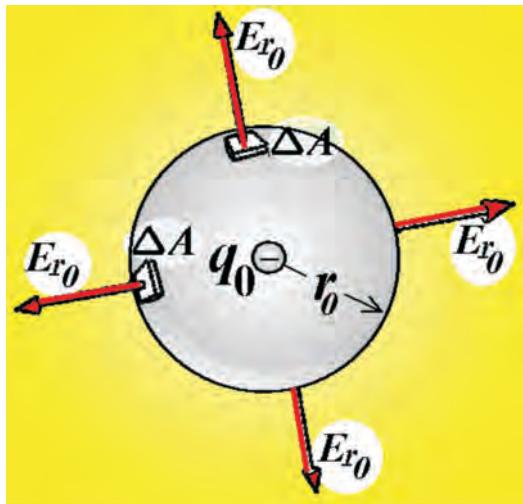
At all points of the surface of a sphere with radius r , $\|\vec{E}_r\| = E_r$ (19.3.5) is constant. This gives us an easy proof of Gauss' law of electricity of the special case where a single source charge q_0 is enclosed by an r -sphere.

(20.3.1)
Calculus
needed

THEOREM: (Gauss' Law for electricity.) Given a closed surface, A , which encloses charged particles whose total charge is q . Then the total enclosed electric charge q_{enc} can be determined from the following form of the electrical flux $\Phi_E(A)$:

$$(20.3.1a) \quad \Phi_E(A) = \oint_A \langle \vec{E}, d\vec{A} \rangle = \frac{q_{enc}}{\varepsilon_0}.$$

PROOF: We only prove the result for the simplest case where A is the closed surface of a sphere of radius r_0 and there is only one charged particle, $q_{enc} = q_0$, at its center. As per (19.3.3a), electric field vectors $\vec{E}_{r,0}$ point radially outward, parallel to the area vector $\Delta \vec{A}$. The electric flux $\Phi_E(A)$ (20.2.1) through this r_0 -sphere has the form



$$\begin{aligned} \oint_A \langle \vec{E}_{r,0}, d\vec{A} \rangle &= \lim_{\max \Delta A_i \rightarrow 0} \left[\sum_{\Delta A_i} \langle \vec{E}_{r,0}, \Delta \vec{A}_i \rangle \right] && \text{from (18.8.5b)} \\ &= \lim_{\max \Delta A_i \rightarrow 0} \left[\sum_{\Delta A_i} E_{r,0} (\cos 0^\circ) \Delta A_i \right] && \text{from (18.6.1)} \\ &= E_{r,0} \cdot \left[\lim_{\max \Delta A_i \rightarrow 0} \sum_{\Delta A_i} \Delta A_i \right] && \left\{ \begin{array}{l} \text{constant } E_{r,0} \text{ factors} \\ \text{out and } \cos(0^\circ) = 1 \end{array} \right. \\ &= \frac{q_0}{4\pi\epsilon_0 r_0^2} \cdot \left[4\pi r_0^2 \right] = \frac{q_0}{\epsilon_0}. && \left\{ \begin{array}{l} \text{from (19.3.5a) with } n = 1 \\ \text{and } \lim \sum \Delta A_i = 4\pi r_0^2 \end{array} \right. \end{aligned}$$

This chain of equalities establishes (20.3.1a) and the theorem is proved for the special case where the closed area A is the surface of the r_0 -sphere. ■

(20.3.2)

NOTE: Coulomb \Rightarrow Gauss Coulomb's Law (19.2.4b) implies Gauss' Law of electricity, Theorem (20.3.1), since the last equality of the proof above depends on (19.3.5a) which depends on Coulomb's Law. Exercise (20.5.2) shows the converse, Gauss \Rightarrow Coulomb.

(20.3.3)

NOTE: Theorem (20.3.1) states that Equation (20.3.1a)—Gauss' law for electricity—holds true even when the closed surface, A , does not define a sphere, and there may be several charged particles at arbitrary points within the closed surface. Exercise (20.5.5) takes you through the computation when the closed surface defines a right circular cylinder and there is a single charged particle at an arbitrary point on the central axis.

(20.4) Gauss' Law for Magnetism

(20.4.1)

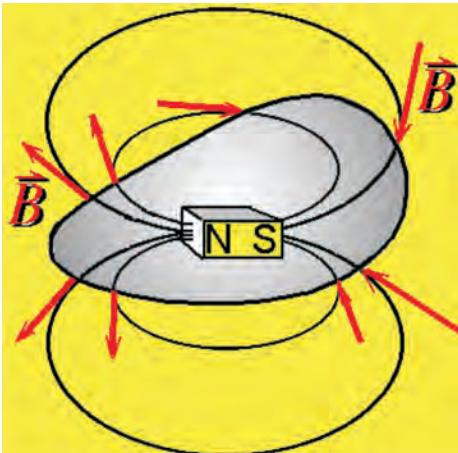
Calc

POSTULATE: (*Gauss' Law for magnetism*). Given a closed surface A that encloses a magnetic source. Then the magnetic flux $\Phi_B(A)$ is zero. That is,

$$(20.4.1a) \quad \Phi_B(A) = \oint_A \langle \vec{B}, d\vec{A} \rangle = 0.$$

Empirical justification:

This postulate, which has never been experimentally contradicted, is equivalent to saying there are no monopoles (see Figure (19.4.4)). That is, as the figure on the right shows, any magnetic source comes equipped with both a North pole and a South pole. Hence, every magnetic field line \vec{B} gives a *positive* reading as it exits the surface and produces a *negative* reading on its return when it re-enters the surface. The positive and negative orientations are a result of Definition (18.4.7) for surface orientation. ■



(20.5) Exercises

(20.5.1)

Exercise

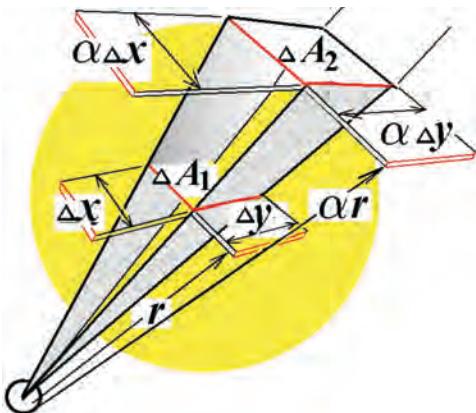
Surface Areas of Maxwell Cones

Given the cone of length r with rectangular top with dimensions Δx and Δy .

- Show that the extension of the cone to a length αr produces rectangular top dimensions $\alpha \Delta x$ and $\alpha \Delta y$.
- Conclude from this that the area,

$$\Delta A = \Delta x \Delta y,$$

of the rectangular top is proportional to the square of length r . (See (18.2.1) for definition of proportionality.)



(20.5.2)

Exercise

Gauss \Rightarrow Coulomb. Show that Gauss' law (20.3.1) implies Coulomb's law (19.2.4b). We have already shown the converse, **Coulomb \Rightarrow Gauss** (see (20.3.2)).

Hint: Gauss' Law $\oint_A < \vec{E}, d\vec{A} > = \frac{q_{enc}}{\epsilon_0}$.

Since $4\pi r^2$ is the surface area A of a sphere of radius r , and the charge q_{enc} is at its center, physical symmetry requires that $\|\vec{E}\|$, the length of \vec{E} , is constant at each point of the sphere's surface. Since \vec{E} and differential area vector $d\vec{A}$ are both perpendicular to the sphere's surface (see Figure (18.3.3)),

$$| < \vec{E}, d\vec{A} > | = \| \vec{E} \| \cdot | d\vec{A} |.$$

The constant $\|\vec{E}\|$ then factors out of the integral implying that

$$\| \vec{E} \| \times (\text{sphere area}) = q_{enc}/\epsilon_0,$$

which is a form of Coulomb's Law.

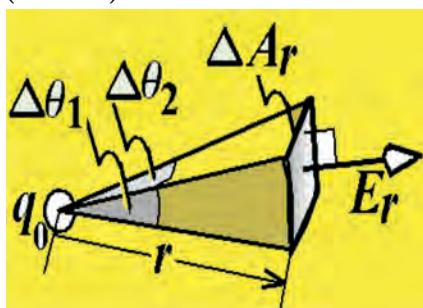
(20.5.3) Motivation for Decomposing E_{r_0} into exactly k^2 Shorter Pieces:
Exercise

At each distance, kr_0 , from the source charge, q_0 , the field line of strength E_{r_0} could have been decomposed into one, three, seven, or *any* number of pieces. The choice to cut E_{r_0} into k^2 equal pieces is motivated by the fact that at radius r_0 , the area is multiplied by a factor of k^2 .

In fact, at distance r_0 , the field line pierces an initial area of size ΔA_{r_0} (see Figure (20.1.5)). Then, at distance $2r_0$, the area grows to size $4\Delta A_{r_0} = 2^2\Delta A_{r_0}$; at distance $3r_0$, the area grows to size $3^2\Delta A_{r_0}$. In general, at distance kr_0 , the area grows to size $k^2\Delta A_{r_0}$.

To provide each area segment with a field line, then, consistent with rules (19.2.3), we cut the field line of magnitude E_{r_0} into k^2 (shorter) pieces and give one of these pieces to each area segment of size ΔA .

(20.5.3a) **FIGURE: A Cone Intersecting a Sphere.**



A charge q is at the center of a sphere with radius r . Four rays from the center define a four-faced cone whose intersection with the sphere's surface forms a four-sided rectangle with area ΔA_r . The angles θ_1 and θ_2 are sufficiently small so that the surface element approximates a flat rectangle. Moreover, the field vector \vec{E}_r is perpendicular to the surface element.

With the construction of Figure (20.5.3a), the product $(E_r \cdot \Delta A_r)$ is independent of r —the product is the same for any cone that is a multiple (or stretching) of the original cone centered at q . This is the content of the next theorem.

(20.5.4) **THEOREM: (*Product $E_r \cdot \Delta A_r$ independent of r .*)** Given the intersection of the sphere centered at charge q_0 and the cone of Figure (20.5.3a). Then the product

$$(20.5.4a) \quad E_r \cdot \Delta A_r = \frac{q_0 \cdot \Delta\theta_1 \cdot \Delta\theta_2}{4\pi\epsilon_0},$$

which depends on the vertex angles, $\Delta\theta_1$, $\Delta\theta_2$, the charge q_0 , and the constant ϵ_0 , is, therefore, independent of r .

PROOF: For very small angles, $\Delta\theta_1$, $\Delta\theta_2$, the intersection of the cone and the sphere in Figure (20.5.3a) closely approximates a rectangle with sides $r \cdot \Delta\theta_1$ and $r \cdot \Delta\theta_2$. Hence, the area of this surface element is $\Delta A_r = r^2 \cdot \Delta\theta_1 \cdot \Delta\theta_2$. Now use the form of Coulomb's law given by (19.3.5) to see that $E_p = E_r = q_0/4\pi\epsilon_0 r^2$. The product of E_r and ΔA_r gives us (20.5.4a), which proves the theorem. ■

Particular radius r_0 . Choose as a reference the cylinder intersecting the sphere with radius r_0 . At this point, the electric field has magnitude E_{r_0} and the surface has value ΔA_{r_0} . The product of these two quantities, therefore, computes to be

$$(20.5.4b) \quad E_{r_0} \cdot \Delta A_{r_0}.$$

General radius r . The electric field is proportional to $1/r^2$ (18.2.1). Thus, at every point (area) that is r units from the charge has value

$$(20.5.4c) \quad (k_1)/r^2 = \underbrace{(r_0^2 E_{r_0})}_{k_1} / r^2.$$

The surface area r units from the charge is proportional to r^2 . Hence, the area has value

$$(20.5.4d) \quad (k_2)r^2 = \underbrace{(\Delta A_{r_0}/r_0^2)}_{k_2} r^2.$$

Therefore, the product of the field magnitude and the surface area at an arbitrary radius r computes to be $E_{r_0} \cdot \Delta A_{r_0}$, exactly the same as (20.5.4b), the product at our starting radius, r_0 .

From (19.3.5), we see that the electric field E_r decreases algebraically in proportion (18.2.1) to $1/r^2$. At the same time, the surface area A_r

of a sphere of radius r increases with r^2 . More generally, the area, ΔA_r , say, of any subsection of the sphere's surface also increases in proportion to r^2 .

The (area \times field) product is constant at all distances r from the center. This follows since the area of the cone is proportional to r^2 and the field is proportional to $1/r^2$. Hence, their product ($\delta A_r \cdot E_r$) is constant for all $r > 0$.

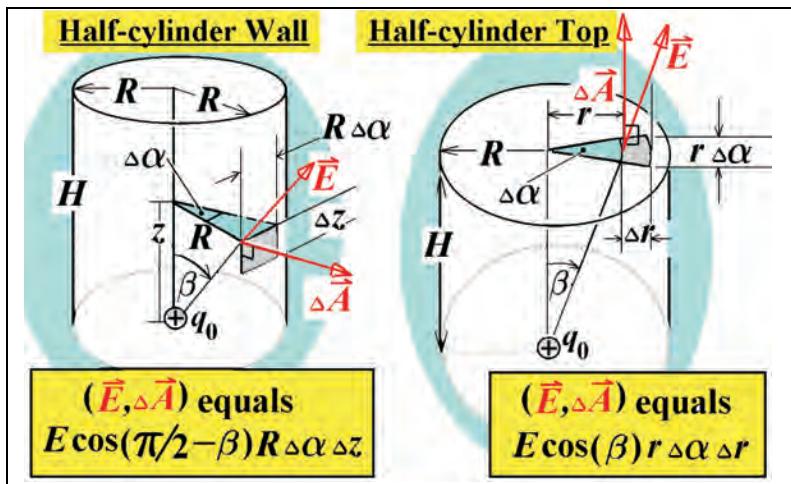
(20.5.5)
Exercise

Gauss' Law for Electricity in a Cylinder. By direct computation, verify Gauss' law for electricity over the surface of a cylinder of radius R that encloses a point charge q_0 . Specifically, show that

$$\oint \langle \vec{E}, d\vec{A} \rangle = \frac{q_0}{\epsilon_0}$$

when the enclosed charge lies H units from the top and H' units from the bottom on the radial axis. (The cylinder has total height $H + H'$.)

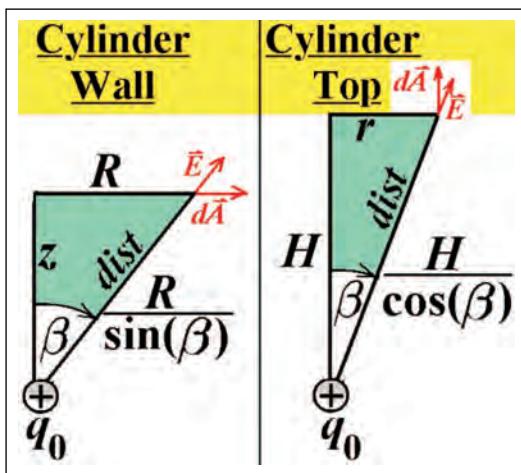
(20.5.5a) **FIGURE:** “Upper” Half-cylinder



Hints: First, compute the surface integral $\oint \langle \vec{E}, d\vec{A} \rangle$ over the “upper half surface” of Figure (20.5.5a) that consists of the top H units of the cylinder wall and the top circle. Then integrate over the “lower half surface” which is the bottom H' units of the cylinder’s wall along with its bottom circle.

Figure (20.5.5a) shows the values of the electric field vector \vec{E} , the included angle β , and the differential surface element $\Delta \vec{A}$ that we need to know before computing the closed Riemann integral $\oint \langle \vec{E}, d\vec{A} \rangle = \lim_{\Delta A \rightarrow 0} \sum_{\Delta A \rightarrow 0} \langle \vec{E}, \Delta \vec{A} \rangle$. (To recall the definition of the Riemann integral, see (18.7.4a).)

(20.5.5b) **FIGURE:** Distances from \vec{E} to q_0 .



the circular top (right side of (20.5.5b)), $dist = H/\cos(\beta)$.

- **Value of E .** Formula (19.3.5), pg. 267, tells us that the magnitude of the electric field is $E = (q_0/4\pi\epsilon_0)(1/dist^2)$ where $dist$ is the distance from the surface element to the single enclosed charge q_0 . On the cylinder wall (left triangle of (20.5.5b)), $dist = R/\sin(\beta)$ and on

• **Value of the Included Angle.** As can be read from Figures (20.5.5a) and (20.5.5b), the angle between the electric field vector \vec{E} , and the *vertical* area vector $\Delta \vec{A}$ on the top circular surface is β . But on the cylinder wall, between \vec{E} and the *horizontal* area vector $\Delta \vec{A}$ lies the included angle $(\pi/2 - \beta)$.

• **Value of ΔA .** On the cylinder wall (left cylinder in (20.5.5a)), the differential area element is $\Delta A = \Delta z \ R \Delta \alpha$. On the top circular surface (right cylinder of (20.5.5a)), the differential area element is $\Delta A = \Delta r \ r \ \Delta \alpha$.

• **The Final Integrals.** With the values of E , β , and ΔA , the calculated values for the inner products $\langle \vec{E}, \Delta \vec{A} \rangle$ in the boxes shown in Figure (20.5.5a) are justified. Now take the sum $\sum_{\Delta A} \langle \vec{E}, \Delta \vec{A} \rangle$ as $\Delta \vec{A}$ runs over the entire cylinder wall (20.5.5c), and then over the entire circular top (20.5.5d). In each case, take the limit as the maximum value of $\Delta A \rightarrow 0$ to produce the Riemann integrals $\oint \langle \vec{E}, d\vec{A} \rangle$ which assume respective forms

$$(20.5.5c) \quad \int E \cos(\pi/2 - \beta) R \, d\alpha \, dz \quad (\text{cylinder wall}), \text{ and}$$

$$(20.5.5d) \quad \int E \cos(\beta) r \, d\alpha \, dr \quad (\text{top surface}).$$

- **Eliminating the r , dr , and z , dz Variables.** The integrals above can be simplified by eliminating the dependent r and z variables and integrating only with respect to β and α . Specifically, as Figure (20.5.5b) illustrates, $r = H \tan(\beta)$ so that $dr = [H/\cos(\beta)]d\beta$. Similarly, $z = R/\tan(\beta)$ which, when differentiated with respect to β , yields

$$dz = -[R/\sin^2(\beta)]d\beta.$$

- **An Efficient Computation.** Once the variables r , dr , z , and dz are eliminated (above), the disagreeable surface integrals, (20.5.5c) and (20.5.5d), assume the simple $\int \sin(\beta)d\beta$ form, whose sum produces the total surface integral (wall + top) of the upper cylinder:

$$(20.5.5e) \quad \underbrace{\frac{q_0}{2\varepsilon_0} \int_0^{\beta_0} \sin(\beta)d\beta}_{(20.5.5d)} + \underbrace{\frac{q_0}{2\varepsilon_0} \int_{\beta_0}^{\pi/2} \sin(\beta)d\beta}_{(20.5.5c)} = \frac{q_0}{2\varepsilon_0} \int_0^{\pi/2} \sin(\beta)d\beta.$$

The angle β_0 is formed at the meeting point of the cylinder wall and the circular surface, that is, when both triangles in Figure (20.5.5b) are the same right triangle with sides $\{H, R, \sqrt{H^2 + R^2}\}$. Equivalently, β_0 is the unique angle for which $\tan(\beta_0) = R/H$.

- **(20.5.5f) How to Measure Angle β .** The angle β , formed by a line originating from the point charge q_0 , is measured from the vertical axis of the cylinder. Angle β increases as the line rotates away from the vertical as shown in Figures (20.5.5a) and (20.5.5b). Thus, a vertical line pointing up from q_0 is at an angle $\beta = 0$ and at an angle $\beta = \pi/2$ when the line points horizontally.

- **The Other Half of the Cylinder.** The surface integral above covers only the upper H units of the cylinder. Since neither the radius R nor the wall height H appears in the integral, the identical integral, $(q_0/2\varepsilon_0) \int_0^{\pi/2} \sin(\beta)d\beta$, will result when integrating over the lower H' units of the cylinder.

(21) Towards Maxwell's Equations

Preview

The Source of Magnetism.

- Magnetism is generated by moving electrical charges and (*in a quantum setting that has nothing to do with moving electrons*) by spin on particles such as the neutron and the neutrino.
- Some atoms become tiny magnets because of their moving electrons and also because of the spin of the atom's electrons, protons, and neutrons. Not all atoms have this property.

In 1819, HANS CHRISTIAN OERSTED noticed that a compass needle was deflected by a wire carrying a current. Then, in 1820, JEAN-BAPTISTE BIOT (1774–1862) and FÉLIX SAVART (1791–1841) conducted experiments that measured the force exerted on a compass by the electric current.



(21.1) Biot-Savart Law: Magnetism from Electricity

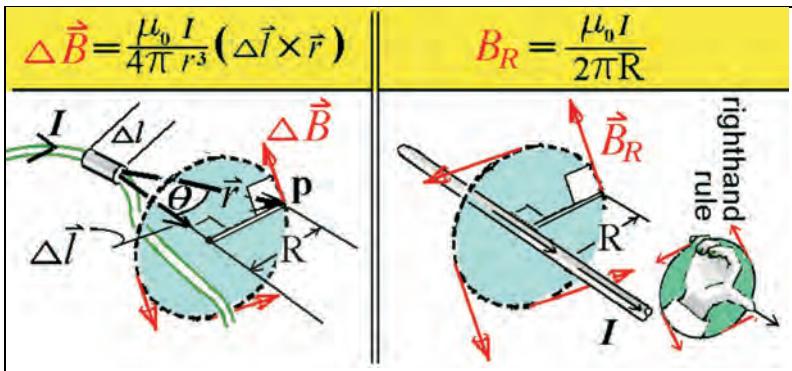
The numerical value of current, I , is measured as the number of positive coulombs, Δq , passing through a cross-section of the wire over a time period Δt . This definition is motivated by the fact that in the presence of an electric field, the stronger the field, the more electrons move per unit time through a wire. Specifically,

$$(21.1.1) \quad I = \frac{\Delta q}{\Delta t} \left(\lim_{\Delta t \rightarrow 0} \frac{\Delta q}{\Delta t} = \frac{dq}{dt} \text{ denotes the } \underline{\text{instantaneous}} \text{ rate of change} \right).$$

Biot and Savart measured the magnetic field caused by an electric current I (21.1.1). They found that the magnetic field \vec{B} at

a point \mathbf{p} is the sum of the contributions of all the differential $\Delta \vec{B}$ elements, each of which is generated by the current I passing through a differential length Δl of the wire. Each tiny $\Delta \vec{B}$ is perpendicular to both the small wire length Δl and the vector \vec{r} from Δl to \mathbf{p} . (See the left panel of Figure (21.1.2).)

(21.1.2) **FIGURE:** The Biot-Savart Law and the Right-Hand Rule



(21.1.3)

NOTE: The 2-dimensional cross-section of the magnetic field vectors shown in the right-hand panel of Figure (21.1.2) (consistent with the wheel-axle right-hand rule (18.4.5)), shows concentric circles of constant magnitude around an infinite, straight wire. In actual 3-space, the magnetic circles “add up” to form concentric cylinders of constant magnitude that sheathe the infinite wire.

(21.2) Quantitative Results for Biot-Savart

We discuss two cases for measurements of induced magnetic fields, namely, when a current I flows through

- a small, differential wire element (21.2.1), and
- a string of differential wire elements, that is, an infinite straight wire (21.2.3).

(21.2.1) Biot-Savart Law for Differential Wire Elements.

Left Panel of Figure (21.1.2): A small, differential magnetic field vector $\Delta \vec{B}$ is induced by a current I flowing through a tiny piece of wire with length Δl . The 3-dimensional vector $\Delta \vec{l}$ represents the 1-dimensional length Δl of the wire segment (see left panel of Figure

(18.3.3)). The distance vector \vec{r} runs from the wire segment to the point \mathbf{p} in space where $\Delta\vec{B}$, the magnetic field vector, is measured. The distance from the wire segment to \mathbf{p} is $\|\vec{r}\| = r$. Finally, the differential magnetic field vector $\Delta\vec{B}$ is given by

$$(21.2.1a) \quad \Delta\vec{B} = \left(\frac{\mu_0 I}{4\pi r^3}\right)(\Delta\vec{l} \times \vec{r}), \quad \text{Biot-Savart Law}$$

so that its magnitude is

$$(21.2.1b) \quad \|\Delta\vec{B}\| = \Delta B = \left(\frac{\mu_0 I}{4\pi r^2}\right)\Delta l \sin \theta, \quad \begin{cases} \text{since, from (18.6.3),} \\ \|\Delta\vec{l} \times \vec{r}\| = \Delta l (\sin \theta) r \end{cases}$$

where θ is the angle in the plane formed by the vectors $\Delta\vec{l}$ and \vec{r} . Equation (21.2.1a), a statement of the Biot-Savart law, shows that the differential magnetic field vector $\Delta\vec{B}$ is the cross product (18.6.3) of vectors $\Delta\vec{l}$ and \vec{r} . Hence, $\Delta\vec{B}$ is always perpendicular to the plane formed by the vectors $\Delta\vec{l}$ and \vec{r} .

(21.2.2)

DEFINITION: *The constant*

$$(21.2.2a) \quad \mu_0$$

of (21.2.1a) is called the magnetic permeability constant.

Permeability, μ_0 (with the 0 subscript), depends on the *vacuum* in which $\Delta\vec{B}$ of (21.2.1a) is measured. Measurements of $\Delta\vec{B}$ taken in other environments—nitrogen, air, water, etc.—produce their own permeability constants, μ .

(21.2.3)

Biot-Savart Law for an Infinite Wire.

Right Panel of Figure (21.1.2): In this panel, the Biot-Savart law (21.2.1a) is extended from a single differential wire segment with length Δl to an infinite *sum* of such wire lengths, i.e., to an infinite, straight wire carrying the same current I .

The resulting magnetic field vector, $\vec{B}_R = \sum \Delta\vec{B}$, depends on the radial distance, R , from the wire, and has magnitude

$$(21.2.3a) \quad \|\vec{B}_R\| = B_R = \frac{\mu_0 I}{2\pi R}.$$

In (21.2.3a), the vector \vec{B}_R is tangent to the circle with radius R and points clockwise or counterclockwise, in accordance with the wheel-axle orientation of the right-hand rule (18.4.5).

(21.2.4) **Note:** Equation (21.2.3a) is not trivial. Exercise (21.8.1) takes you through the integration process to prove its validity.

At all points on the circumference of the dotted circle with radius R (perpendicular to the wire segment vector, $\Delta\vec{l}$), $\|\Delta\vec{B}\| = \Delta B$, the length of the magnetic field vector, is constant—only the direction of $\Delta\vec{B}$ changes as it remains tangent to the dotted circle.

We can use this special, simple case to motivate Ampère's Law which is developed in the following section.

(21.3) Ampère's Law

We will need to relate A , a surface area, with L , the boundary, edge, or loop around that area. Closed “potato” surfaces do not have a boundary, but “pancake” surfaces, floating in space, will have an outside boundary. Our surfaces will have no holes or inside edges.

The following notation reinforces the connection between a surface and its boundary:

(21.3.1)

NOTATION: For a surface in \mathbf{R}^3 with area A , we use the symbol

$$(21.3.1a) \quad \text{bdry}(A)$$

to denote the boundary of enclosed area A .

Note that for a closed surface, like a sphere, there is no boundary. But if a 2-dimensional boundary $\text{bdry}(A)$ does exist for an area A , then A will not be a unique area—a single boundary can accommodate infinitely many areas. For example, a circle can be the boundary (edge) of a flat disk or it can be the rim of a concave bowl.

(21.3.2) **NOTE:** Similar to our dual usage of the symbol q for both a charged particle and its electrical charge (19.2.1), we use the same symbol, A , for the physical surface and for the numerical measure of its area. Similarly, $\text{bdry}(A)$ denotes both the boundary and the length of the boundary.

(21.3.3) **THEOREM: (Ampère's Law for Steady State I_{net} .)** Given a closed loop $\text{bdry}(A)$ defining area A , through which there flow several steady-state currents, I_1, I_2, \dots, I_n . Then the net current, $I_{\text{net}} = I_1 + I_2 + \dots + I_n$, flowing through the area A (enclosed by loop $\text{bdry}(A)$), induces a magnetic field \vec{B}_l whose values at a point l on $\text{bdry}(A)$ are constrained by the closed line integral

$$(21.3.3a) \quad \oint_{\text{bdry}(A)} \langle \vec{B}_l, d\vec{l} \rangle = \mu_0 I_{\text{net}}. \quad \text{Ampère's Law}$$

PROOF: We take the simplest case, which is illustrated by the right-hand panel of Figure (21.1.2). The more general non-circular boundary is beyond the scope of this book but an exposition can be found in the book *Classical Electrodynamics* [21] by J. D. Jackson. We assume the boundary loop, $\text{bdry}(A)$, is the circumference of a circular area A with radius R . Assume that there is only one infinite current-carrying wire that passes through the center of the circle A . With the symmetry of the circle, the norm (*but not the direction*) of the magnetic field vector at any point, l , on the circle can be denoted by

$$(21.3.3b) \quad B_l = B_R,$$

since B_l depends only on the radius R . Let Δl_i denote the length of the i th circumference element, $\vec{\Delta l}_i$ on the circumference, $\text{bdry}(A)$, and let l_i be a point in the interval Δl_i . Then

$$(21.3.3c) \quad \langle \vec{B}_{l_i}, \vec{\Delta l}_i \rangle = \underbrace{B_{l_i} (\sin 90^\circ) \Delta l_i}_{\text{from Def. (18.6.1)}} = B_R \Delta l_i. \quad \text{from (21.3.3b)}$$

Using (18.8.5a) to compute the left-hand side of (21.3.3a), we obtain

$$\oint_L \langle \vec{B}_R, \vec{\Delta l} \rangle = \lim_{\max \Delta l_i \rightarrow 0} \left[\sum_{\Delta l_i} B_R \Delta l_i \right] \quad \begin{array}{l} \text{Definition (18.8.5a)} \\ \text{and (21.3.3c)} \end{array}$$

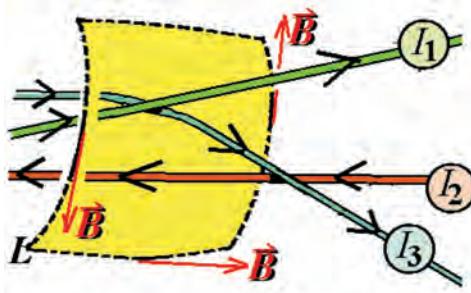
$$= B_R \cdot \left[\lim_{\substack{\max \Delta l_i \\ \rightarrow 0}} \sum \Delta l_i \right] \quad \begin{matrix} \text{factor out constant} \\ \text{number, } B_R \end{matrix}$$

$$= B_R \cdot \left[2\pi R \right]. \quad \begin{matrix} \lim \sum \\ = \text{circle circumference} \\ = 2\pi R \end{matrix}$$

Replace B_R above with its equivalent, $\mu_0 I_{net}/2\pi R$ from (21.2.3a), (which is proven in Exercise (21.8.1)), and we see that hypothesis (21.3.3a) results. This proves the theorem. ■

(21.3.4) **NOTE:** The use of (21.2.3a) in the last line of our proof tells us that the Biot-Savart Law ((21.2.3) for a wire) implies Ampère's law. The reverse implication is also true.

(21.3.5) **FIGURE:** Ampère's Law in the General Case.



Although we shall consider the case of a *single* wire passing through a region, non-trivial mathematics (involving closed Jordan curves (18.4.2)), allow Ampère's law (21.3.3a) to remain valid with more general closed loops L (18.4.3) that may surround several wires. If the k th wire carries current I_k , then the total current through the loop L is $I = \sum_k I_k$, the net (*arithmetic*) sum of all enclosed currents.

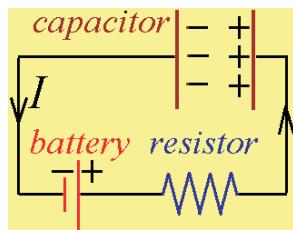
(21.4) Maxwell Adds to Ampère's Law

Around 1865, Maxwell noted a deficiency in Ampère's law,

$$\int \langle \vec{B}_l, d\vec{l} \rangle = \mu_0 I_{net} \quad (21.3.3a),$$

which says that a steady-state current, I_{net} , induces a steady-state magnetic field, \vec{B} .

- (21.4.1) **Here is the problem:** A CAPACITOR or CONDENSER consists of a pair of conducting surfaces separated by an insulator. In our diagram, a parallel plate capacitor is initially uncharged. As electrons begin to flow clockwise through the wire (which means that the current I flows *counter-clockwise*), the left-hand plate accumulates negatively-charged electrons that produce an electric field that repels the like-charged negative electrons from the right-hand plate. This “chasing away” of electrons acts as if there is a current passing through the space between the plates where an excess positive charge is created on the right-hand plate.



- (21.4.2) **There is no physical current or charge** traveling between the plates, yet a magnetic field *is* induced around the capacitor. Although Ampère's law described the effects of charges moving along a wire (that induces a magnetic field), Maxwell realized that a consistent treatment of time-varying (non steady-state) situations can be reached by introducing a term in the right-hand side of (21.3.3a) that takes into account the effects of one capacitor plate on the other across the insulating medium. He labeled the term the DISPLACEMENT CURRENT not because it represents a motion of charges, but because it plays a similar rôle to the usual current in the equations. With this deceptively simple modification, the equations no longer suffered from the problems described above, but in addition satisfied other consistency requirements such as conservation of charge. More importantly, the modified equations predicted a wealth of new phenomena whose practical applications have defined many aspects of our current electronic-dependent society.

Maxwell's idea was a momentous discovery indeed.

- (21.4.3) **What term should be added to (21.3.3a) in order to fix Ampère's steady-state law?** Maxwell's ingenious solution was to describe the induced magnetic field in terms of the physical, time-varying current, $I(t)$, flowing through the wire in the circuit. He reasoned as follows:

Gauss' law (20.3.1a), pg. 280, tells us that $\Phi_{\vec{E}}(A) = q/\epsilon_0$. Differentiating both sides of this equality with respect to time t yields

$$(21.4.3a) \quad \frac{d}{dt} \Phi_{\vec{E}}(A) = \frac{1}{\varepsilon_0} \cdot \frac{dq}{dt} = \frac{1}{\varepsilon_0} \cdot I(t). \quad \text{from (21.1.1)}$$

To match the $\mu_0 I_{net}$ form of Ampère's law (21.3.3a), we solve for $I(t)$ in (21.4.3a) to obtain

$$(21.4.3b) \quad \mu_0 I(t) = \mu_0 \varepsilon_0 \frac{d}{dt} \Phi_{\vec{E}}(A)$$

which, when added to Ampère's law, yields

$(21.4.3c) \quad \oint_{bdry(A)} \langle \vec{B}_l, d\vec{l} \rangle = \mu_0 I_{net} + \mu_0 \varepsilon_0 \frac{d}{dt} \Phi_{\vec{E}}(A).$	Ampère-Maxwell Law
---	---------------------------

The Ampère-Maxwell Law (21.4.3c), is the final (corrected) version of Ampère's law (21.3.3a) which now conserves charge. (*Recall A is an area through which the changing circuit flows, and \vec{B}_l is the induced magnetic field measured at points l on the boundary, $bdry(A)$.*)

This ingenious addition of the “virtual” current term (21.4.3b) by Maxwell into (21.4.3c), is a deceptively simple, yet consequential and robust idea.

(21.5) Faraday's Law: Electricity from Magnetism

- (21.5.1) **It was Michael Faraday** who first noticed that a moving magnet through a wire loop induced a current corresponding to a non-zero electromagnetic force.

As it turns out, the wire and the current are not necessary. The electric field exists with or without the wire. That is, a changing magnetic field $\Delta \vec{B}$ in any bounded area will induce a driving emf, or electromotive force, \vec{E}_l , at points l on any boundary of that area, whether or not that boundary is formed by a physical wire.

To quantify the relation between $d/dt (\Phi_{\vec{B}}(A))$, the change in a magnetic field through any closed area, A , and \vec{E}_l , the induced electrical pressure at point l on $bdry(A)$ (the boundary of A), Faraday devised the following closed line integral equation:

$$(21.5.1a) \oint_{bdry(A)} \langle \vec{E}_l, d\vec{l} \rangle = -\frac{d}{dt} \Phi_{\vec{B}}(A). \quad \text{Faraday's Law}$$

The minus sign in Faraday's law (21.5.1a) is the mathematical expression of Lentz's Law (21.6.1) following, which will state that the direction of the magnetic field produced is always opposite to the direction of the change in the original magnetic field.

(21.6) Lentz's Law: The Positive Side of Negativity

We know that electric fields and currents induce magnetic fields (21.1), (21.4.3c). Faraday's Law (21.5.1a) describes the converse—how changes in magnetic fields induce electric fields or currents.

In this section, we describe LENTZ'S LAW, that describes the “self-canceling” direction of the induced current. Here are the details.

- (21.6.1) **Statement of Lentz's Law:** Imagine an area A has closed boundary $bdry(A)$ that encloses a magnetic field \vec{B} . Suppose that at one time, the magnetic field changes by a small amount ΔB . Then

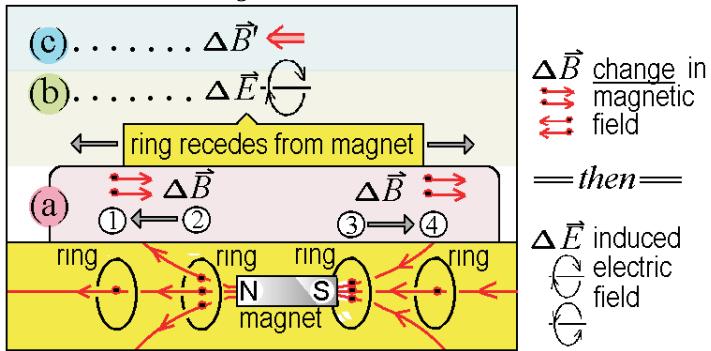
First, $\underline{\Delta \vec{B} \rightarrow \Delta \vec{E}}$. From Faraday's Law (21.5.1a), we see that the magnetic field change, $\Delta \vec{B}$ within area A , induces an electric field change, $\Delta \vec{E}$, along $bdry(A)$. If a wire loop lies on the closed $bdry(A)$, then a current ΔI is produced in the closed wire circuit due to the electromotive force generated by $\Delta \vec{E}$.

Then, $\underline{\Delta \vec{E} \rightarrow \Delta \vec{B}'}$. From the Maxwell-Ampère Law (21.4.3c), an electric field change $\Delta \vec{E}$ (or loop current I) along $bdry(A)$ induces its own magnetic field change, $\Delta \vec{B}'$. The minus sign in Faraday's Law ensures that the induced magnetic field, $\Delta \vec{B}'$, is *opposite* to the original field change $\Delta \vec{B}$. This opposition principle is **Lentz's Law**—although one might question giving a mere minus sign the title of “law.”

(21.6.2) Illustrating Lenz's Law

The following figure, which illustrates Lenz's law, analyses *changes* in the number of leftward-directed magnetic field lines enclosed by the ring as it changes positions between ① and ②, and between ③ and ④, in Figure (21.6.3). The loops in positions ① and ④, the most remote from the magnet, are shown capturing only one leftward-directed magnetic field line. (*All magnetic field lines have a direction as we saw in Fig. (19.4.2).*) The loops closest to the magnet, in Positions ② and ③, capture three leftward-directed field lines.

(21.6.3) FIGURE: Illustrating Lenz's Law



Line (a), Fig. (21.6.3). Whenever the ring recedes from the magnet, whether moving leftward, $\textcircled{1} \leftarrow \textcircled{2}$, or rightward, $\textcircled{3} \rightarrow \textcircled{4}$, there is a (rightward) change, $\Delta B = \leftrightarrow$, in the number of leftward-directed magnetic field lines that are encircled by the wire loop. In both cases, this rightward-directed change, \leftrightarrow causes a two-line *decrease* in the number of leftward-directed magnetic field lines enclosed by the wire loop — from three down to one.

Line (b), Fig. (21.6.3). This magnetic field change, $\Delta B = \leftrightarrow$, induces an electric force $\Delta \vec{E}$ in the wire loop, which leads to Line (c) following:

Line (c), Fig. (21.6.3). $\Delta \vec{E}$ induces the magnetic field, $\Delta \vec{B}' = \leftarrow$, as defined by the right-hand rule (18.4.6). This change always *opposes* the direction of the original magnetic field change, $\Delta B = \leftrightarrow$.

These statements (*along with Exercise (21.8.3)*) complete the explanation of Figure (21.6.3), which fully illustrates Lentz's law. ■

(21.7) Maxwell's Four Equations

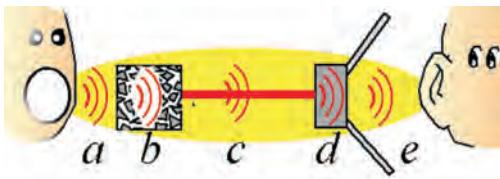
(21.7.1)



Maxwell's four equations first appeared in 1873, in his text *A Treatise on Electricity and Magnetism*. It would be three years later, in October 1876, that ALEXANDER GRAHAM BELL and THOMAS WATSON developed a TELEPHONE that employed the electromagnetic properties articulated by Maxwell. The first voice signal was carried between telephones connected by a two-mile wire.

(21.7.1a) The function of the telephone depends on the following sequence of frequency-preserving events:

- *A voice will vibrate the air (a)*
- *and induce corresponding and varying electrical conductivity of a medium, e.g., carbon particles in the mouthpiece (b),*
- *that conducts a current with a corresponding varying voltage through a transmission line (c).*
- *This fluctuating voltage induces corresponding varying magnetic forces at the receiver's end (d)*
- *that "push and pull" on a speaker cone in the receiver's earpiece (e) with corresponding frequencies that agree with the original voice frequencies. The cone vibrations are transmitted through the air to reproduce the final sound waves that finally arrive at the listener's ear.*



(21.7.2)

The esthetic and simplifying power of Maxwell's equations impelled MAX BORN to note [3], pg. 181–2:

The four symbolic formulae [of Maxwell's equations] show wonderful symmetry. Formal agreement of this kind is by

no means a matter of indifference. It exhibits the underlying simplicity of phenomena in nature, which remains hidden from direct perception because of the limitations of our senses and reveals itself only to our analytic faculty.

- (21.7.3)** **Maxwell's Four Equations:** We combine Gauss' laws of electricity and magnetism, along with Faraday's law (*magnetism induces electricity*), and the Maxwell-Amp`ere law (*electricity induces magnetism*) to obtain the following quartet of equations, known as MAXWELL'S EQUATIONS for electricity and magnetism.
-

Gauss' law for electricit [from (20.3.1a), pg. 280]. An electric charge, q_{enc} (a collection of electric "monopoles"), within a closed surface A is computed from the values of the electric field \vec{E} at points on that surface according to the closed surface integral

$$(21.7.3a) \quad \oint_A \langle \vec{E}, d\vec{A} \rangle = \frac{q_{enc}}{\varepsilon_0}.$$

Gauss' law for magnetism [from (20.4.1a), pg. 282]. (No magnetic monopoles exist) The net magnetic flux (positive field lines in, minus field lines out) through any closed surface A is zero. This quantity is computed from the values of the magnetic field \vec{B} at points on that surface according to the closed surface integral,

$$(21.7.3b) \quad \oint_A \langle \vec{B}, d\vec{A} \rangle = 0.$$

Faraday's law. [from (21.5.1a), pg. 297] A changing magnetic field, \vec{B} , passing through a surface, A , produces an electric field, \vec{E}_l , whose values at points l on $bdry(A)$ are constrained by the line integral

$$(21.7.3c) \quad \oint_{bdry(A)} \langle \vec{E}_l, d\vec{l} \rangle = -\frac{d}{dt} \Phi_{\vec{B}}(A).$$

The minus sign is the mathematical expression of Lentz's law (21.6.3) which imposes a left-hand rule (instead of the standard right-hand rule) on the orientation of the enclosing loop, $bdry(A)$.

.....

Ampère–Maxwell law [from (21.4.3c), pg. 296]. A changing electric field \vec{E} through a surface A produces a magnetic field \vec{B}_l whose values at points, l , on $\text{bdry}(A)$, the closed boundary of surface A , are constrained by the line integral

$$(21.7.3d) \quad \oint_{\text{bdry}(A)} \langle \vec{B}_l, d\vec{l} \rangle = \mu_0 I_{\text{net}} + \varepsilon_0 \mu_0 \frac{d}{dt} \Phi_{\vec{E}}(A).$$

(21.8) Exercises

(21.8.1) Exercise

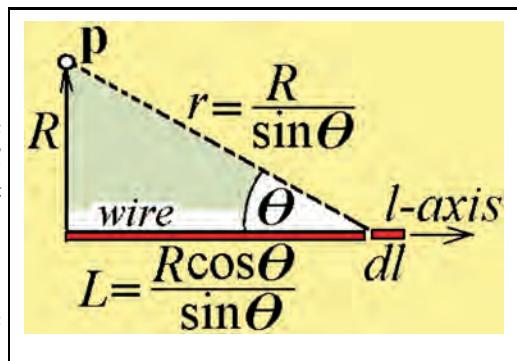
Extending Biot-Savart to an Infinite Wire. The Biot-Savart law (21.2.1a) is given for a current I passing through a differential wire segment $\Delta\vec{l}$. Extend this result by replacing segment $\Delta\vec{l}$ with a sum of such differential lengths, that is, with an infinite, thin, straight wire and show that the induced magnetic field vectors \vec{B} have constant magnitude

$$(21.8.1a) \quad \|\vec{B}\| = \frac{\mu_0 I}{2\pi R}$$

on concentric cylinders (with the wire as central axis) with radius R .

Hint:

Equation (21.2.1b) tells us that the magnitude of the magnetic field vector, induced at point \mathbf{p} , by current I through a Δl -length wire is $\Delta B = (\mu_0 I / 4\pi r^2) \Delta l \sin \theta$, where r is the distance from Δl to point \mathbf{p} .



For the purposes of integration, set $\Delta B = dB$ and $\Delta l = dl$, so that the accumulated magnetic field values, B , over half the wire, from 0 to infinity, is given by the integral

$$(*) \quad \lim_{L \rightarrow \infty} \int_{l=0}^{l=L} dB = \int_{l=0}^{l=\infty} dB = \frac{\mu_0 I}{4\pi} \int_{l=0}^{l=\infty} (\sin \theta / r^2) dl.$$

Note that the figure shows a right triangle with constant side, R , and variable sides, r and L , both of which are expressed in terms of a single variable, θ . Accordingly, convert the integral (*) to a single-variable integral, $\int_{\theta=\pi/2}^{\theta=0} \cdots d\theta$, where, necessarily, angle θ goes from $\pi/2$ to 0 as the variable L goes from 0 to ∞ . This integral over a half ray $[0, \infty)$ will give you only half of the desired result.

(21.8.2)

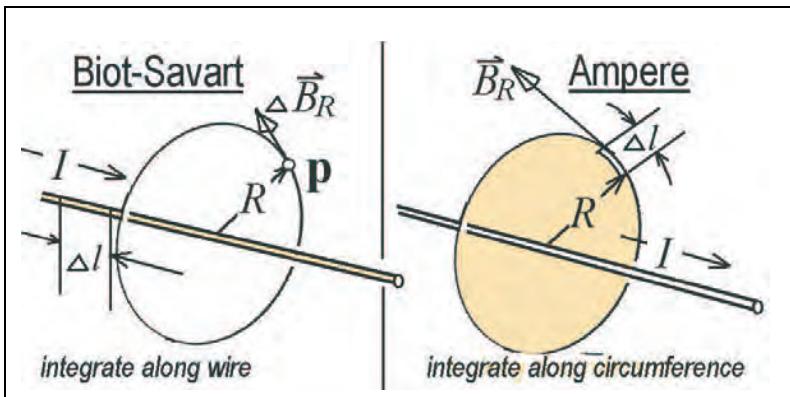
Exercise

Calc

Trig

The Biot-Savart Law Agrees with the General Ampère Law.

In the infinite wire case of Exercise (21.8.1), show that Ampère's law gives the same result, $B_R = \mu_0 I / 2\pi R$, for the magnitude of the magnetic field vector \vec{B}_R at a point R units from an infinite wire.



Hint: Left Panel: The Biot-Savart integration proceeds along the $\Delta \vec{l}$ elements of the infinite wire which hosts a current I while the point \mathbf{p} is fixed. The value of the differential magnetic vectors $\Delta \vec{B}_R$ add together as the integration advances along the wire.

Right Panel: Ampère's Law requires that a current I passes through the circular area and integration proceeds along the $\Delta \vec{l}$ elements of the circumference, which has length $2\pi R$. The length B_R (but not the direction) of the magnetic vector \vec{B}_R is constant at every point on the circle with the current passing perpendicularly through the center.

(21.8.3)

Exercise

Use Figure (21.6.3) to show how Lentz's Law applies when the ring approaches the magnet.

Hint: Reason similarly to the argument following Figure (21.6.3) which describes the case of the magnet receding from the magnet.

(22) Electromagnetism: A Qualitative View

We show how Maxwell's equations (21.7) and some accurate laboratory measurements lead to an expression for c , the speed in a vacuum of electromagnetic (*EM*) waves. Maxwell did not know that light was a subset of his electromagnetic waves. Other electromagnetic waves include infrared, and ultraviolet radiation, microwaves, radio waves, and X and gamma rays (*see Figure (22.3.3)*).

(22.1) Magnetic Waves from an Infinite Wire

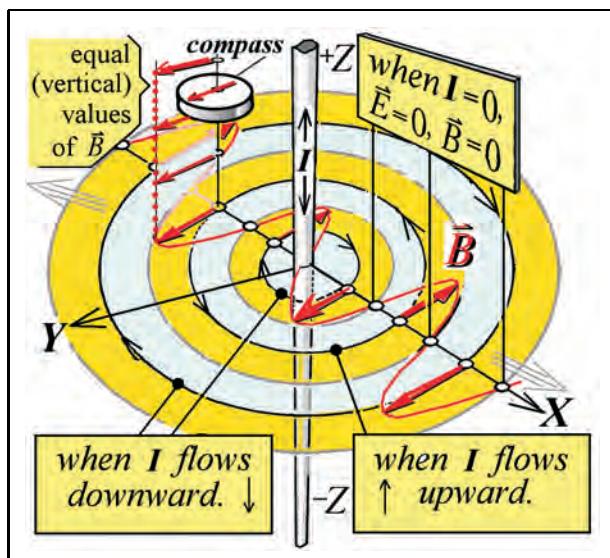
Electromagnetic waves are time-dependent. One of the simplest examples is ordinary household alternating current (AC) in a wire where electrons flow—alternately and repeatedly—in one direction, and then in the opposite direction.

Figure (22.1.1) shows a vertical wire with an alternating current. On the circle perpendicular to the wire (in the X - Y plane) with radius r measured from the wire, the magnetic field vectors are tangent to the r circle and have constant value B_r . As the notation suggests, B_r is constant on circles of radius r but changes as the radius r changes.

Since the current $I(t)$ in the “axle” repeatedly reverses its direction, the right-hand rule (18.4.6), (18.4.5), assures us that the orientation of the magnetic ripples—clockwise or counterclockwise—must also reverse in time (*see Figure (22.1.1)*).

Note: Although alternating current occurs in the wires attached to our home appliances, there is no measurable alternating magnetic field. This is because home appliances use two wires that are close enough to each other so that they may be considered to share the same “axle” while their currents run in opposite directions. Hence, their induced alternating magnetic fields cancel each other.

(22.1.1)

FIGURE: Vertical Alternating Current Propagates an Electromagnetic Wave.

The Biot-Savart law applied to an infinite wire tells us that an upward current \uparrow along the Z -axis induces a (cylindrical) magnetic ripple with field vectors oriented in the counterclockwise direction. Conversely, a downward current \downarrow along the Z -axis induces a (cylindrical) magnetic ripple with field vectors oriented in the clockwise direction (see Figure (21.1.2), pg. 290).

(22.1.1a) The compass in Figure (22.1.1) is shown during an “instantaneous snapshot.” Its south-seeking needle is tangent to the (cylindrical) ripple and is aligned counterclockwise. Note that the compass would give the same reading at *any* point on a vertical line passing through the original point since the values of the magnetic field \vec{B} are uniform on concentric cylinders surrounding the wire.

(22.1.1b) A horizontal cross-section of the uniform cylindrical magnetic fields produces circles. In this horizontal plane, as shown in Figure (22.1.1), the vectors of the magnetic field \vec{B} are plotted only on a single line, the X -axis.

At each point x on the X -axis, a vector of length $\|\vec{B}(x)\| = B(x)$ is drawn with its tail at x and tangent to the magnetic field, (i.e., perpendicular to the X -axis). The orientation of the B arrow imitates

the behavior of a compass needle (*see Figure (19.4.2)*, pg. 268) that might be placed at the point x on the X -axis.

(22.2) Wave Propagation

(22.2.1)



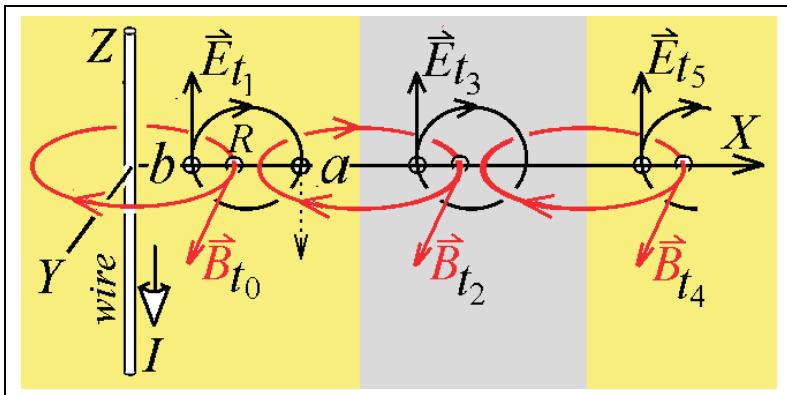
The Mechanics of EM Wave Propagation: a preview. In Figure (22.1.1) we saw how alternating current produced an outward-propagating, time-dependent, alternating *magnetic* field \vec{B} . In what follows, we shall see how a companion (and perpendicular) electric field \vec{E} is produced by a varying magnetic field \vec{B} .

In fact from Maxwell's equation (21.7.3c) we see that the integral around a boundary that surrounds an electric field mathematically produces an expression that depends on magnetic flux. Similarly, from Maxwell's equation (21.7.3d), the integral around a boundary that surrounds a magnetic field produces an expression that depends on electric flux.

As hypothesized by Maxwell, and as confirmed by countless experiments, a changing electric field induces a changing magnetic field under all circumstances, and *vice versa*.

In (22.2.2), we will look at a snapshot of an alternating current $I(t) \downarrow$ as it increases downward along with the cascade of induced magnetic and electrical fields.

(22.2.2)

FIGURE: Electromagnetic Propagation Sequence

Along the intersection of the X -axis and the circles, \vec{E} vectors are always vertical, while \vec{B} vectors are always horizontal. Also, since magnetic vectors \vec{B} are changing with time, then so are the induced electrical vectors, \vec{E} , and vice versa.

Step 0. At Time t_0 , \vec{B}_{t_0} is generated. The downward electric current, $I \downarrow$ at time t_0 , induces the (red) clockwise circle of magnetic field vectors, \vec{B} , only one vector of which, \vec{B}_{t_0} , is plotted in Figure (22.2.2). The magnetic field vector, \vec{B}_{t_0} is tangential to the red circle. At the point R it:

- intersects the X -axis,
- lies in the X - Y plane,
- is perpendicular to the X -axis, and
- points in the positive Y direction.

At every point on this circle, the magnetic field has value $\|\vec{B}_{t_0}\| = B_R = \mu_0 I(t)/2\pi R$ (21.2.3a), which tells us that if the magnitude of the downward current $I(t) \downarrow$ increases, then so will the magnitude of the scalar B_R . That is,

$$(22.2.2a) \quad |I(t)| \text{ increases, implies } \Delta B / \Delta t > 0.$$

.....

Step 1. At time $t_1 > t_0$, \vec{E}_{t_1} is generated. The magnetic field \vec{B}_{t_0} at time t_1 is increasing in time (see (22.2.2a)), and, thanks to Faraday's Law (21.7.3c), induces the (black) clockwise circle of electric field vectors, \vec{E} , (where only one vector of which is displayed). The minus sign in (21.7.3c) is due to Lentz' law (21.6.3) and determines

the direction of induced electric field vector \vec{E}_{t_1} . As previously stated, magnetic vectors \vec{B} that are changing with time induce electrical vectors, \vec{E} , that are also changing in time and vice versa.

In sum, the electric field vector \vec{E}_{t_1} is tangential to the black circle centered around point R . At point b it:

- intersects the X -axis,
- lies in the X - Z plane,
- is perpendicular to the X -axis, and
- points in the positive Z direction.

It is Lentz' Law (21.6.1), pg. 297, that guarantees the *clockwise* orientation of the electric field vectors around this circle since, by the right-hand rule (18.4.5), the magnetic field it induces will oppose the increasing \vec{B}_{t_0} .

.....

Step 2. At time $t_2 > t_1 > t_0$, \vec{B}_{t_2} is generated. The vertical circle of electric field vectors at time t_1 induces the (red) clockwise magnetic field vectors, \vec{B} , only one vector of which, \vec{B}_{t_2} , is plotted.

It is the Ampère-Maxwell Law (21.7.3d), pg. 301, that guarantees existence of a magnetic field vector \vec{B} that is induced by the changing electric field vector \vec{E} . This is due to Maxwell's modification of Ampère's Law (21.3.3a) where he added the crucial \vec{B} -dependent term, $\epsilon_0\mu_0 d[\Phi_B(A)]/dt$.

(The dotted electric field vector at point a , acting like the downward current, $I \downarrow$, determines the clockwise orientation of the induced circle of magnetic field vectors.)

.....

(Steps 1 and 2 repeat.) The cycle continues when the horizontal ripple of the previous magnetic field vectors, \vec{B} , induces another vertical ring of electric field vectors \vec{B} (apply Step 1), which, in turn, induces a new horizontal loop of magnetic field vectors (apply Step 2), and so on.

Note that we get a propagating *pair* of electric and magnetic field vectors that oscillate as they move away from the (vertical) wire carrying the source alternating current. This linking of electromagnetic pairs is possible only because of Maxwell's modification of Ampère's law.

There is much we can say qualitatively, based on Figure (22.2.2).

- (22.2.3) **Summation:** The anatomy of electromagnetic wave propagation induced by an alternating current through a vertical wire:

(22.2.3a) **\vec{B} horizontal, \vec{E} vertical for every point on the X -axis.** Note that only two horizontal magnetic field vectors result from the intersection of the horizontal circle of magnetic vectors and the X -axis. Similarly, only two vertical electric field vectors are produced from the intersection of the vertical circle of electric field vectors and the X -axis. Thus, as shown in Figure (22.2.2), magnetic vectors \vec{B} are always perpendicular to the wire, while electric field vectors \vec{E} are always parallel to it. Both vectors emanate from the X -axis and are perpendicular to the X -axis.

(22.2.3b) **Direction of the velocity vector, \vec{v} .** Figure (22.2.2) also shows that the vectors \vec{E} and \vec{B} are perpendicular to each other, and the direction of the propagated electric-magnetic pair of waves is perpendicular to both \vec{E} and to \vec{B} . That is, Nature gives us a physical configuration that coincides with the mathematical cross product (18.6.3), pg. 250, call it \vec{s} , where

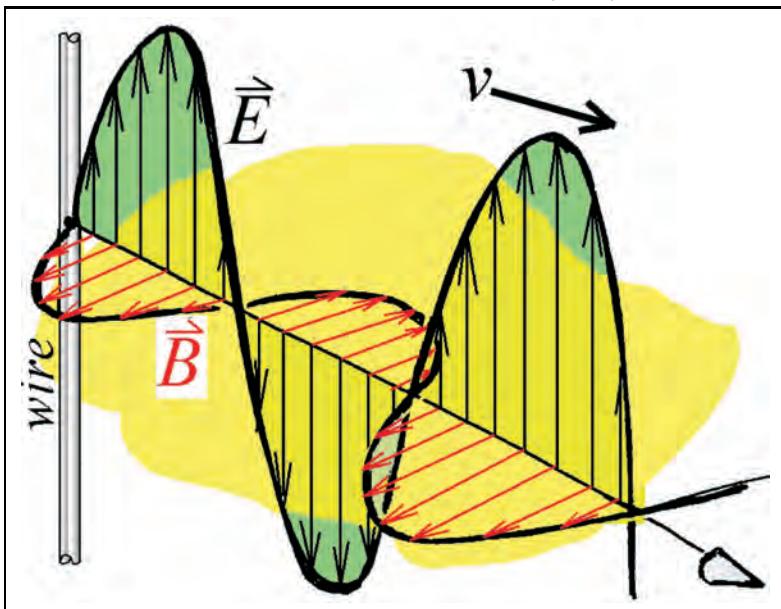
$$\vec{s} = \vec{E} \times \vec{B}.$$

(22.3) The Geometry of Electromagnetism

(22.3.1)



FIGURE: Snapshot of an electromagnetic (EM) wave



The figure takes one horizontal axis of Figure (22.1.1) and completes Figure (22.2.2). At any fixed point in space and time, the vertical electric field vectors, \vec{E} , and horizontal magnetic field vectors, \vec{B} , are always perpendicular to each other and their values propagate away from the vertical wire (the source of the alternating current) in the $\vec{E} \times \vec{B}$ direction.

If the current $I(t)$ varies in time as a sinusoidal-like function, then, as (21.2.3a) suggests, the induced magnetic field vectors, \vec{B}_t , follow along and take on the same sinusoidal pattern as $I(t)$.

Fields E and B are proportional in the sense of (18.2.1a). This will be shown in (23.3.4c), pg. 325 where we will find that $E = cB$, from which we will also conclude that the constant of proportionality is c , the speed of light. Proportionality guarantees that the graph of the electric field vectors has the same shape as the (perpendicularly drawn) graph of the magnetic field vectors, which, in the case of Figure (22.3.1), is sinusoidal.

.....

(22.3.2) Light: a member of the family.

- Intuition would dictate that waves, including electromagnetic waves, need a medium—some kind of physical “stuff” to travel through. After all, we observe water as a carrier for tidal waves and air is a carrier for sound waves. Without water or air there would be no tidal waves or sound. But electromagnetic waves are different—they do not need air or water to propagate, since they travel in the vacuum of space (we do see the stars, after all!).

Now if one does assume that electromagnetic waves are like other waves and need a medium to propagate, then that medium must have rather unique properties: it must fill all the visible cosmos, and yet be so tenuous that it generates negligible friction as the Earth moves around its orbit (otherwise the Earth would have spiraled into the Sun long ago). This hypothesized material was called the luminiferous aether (or aether, for short).

Although Maxwell himself chose to assume the existence of an aether, his equations make no reference to it. Einstein, on the other hand, assumed there was no aether necessary to carry light, a postulate (1.5.2) that greatly simplified his theory of special relativity.

- By the time Maxwell produced his equations that united electricity and magnetism, in which electromagnetic waves consist of traveling electric and magnetic fields that oscillate perpendicular to each other and to the direction of propagation, it had been well accepted (*due to the experiments of Huygens and others*) that light also behaved as a wave, and that its speed (*in a vacuum*) was c . When Maxwell predicted that the speed of electromagnetic waves was also equal to c , he hypothesized that this was no coincidence—light must be a special case.

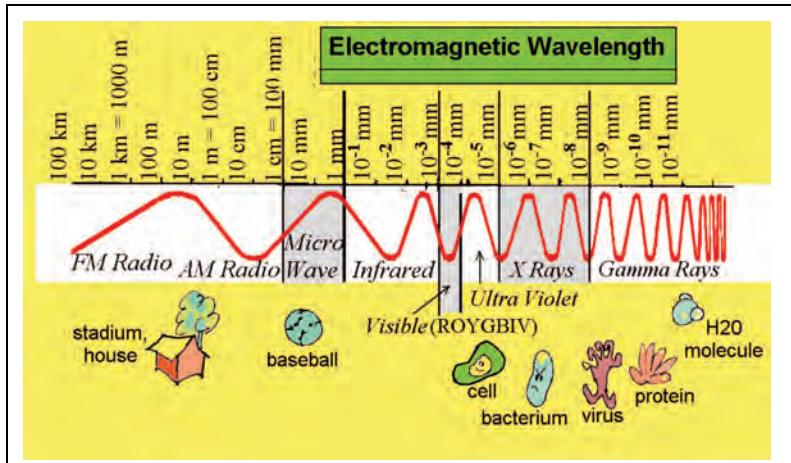
But how to prove it?

The proof came through the experiments of HEINRICH HERTZ (1857–1894) who observed that if an electromagnetic wave were to encounter a wire aligned with the waves electric field, that field would force the electrons in the wire to oscillate, generating a current that could be measured. The successful implementation of this idea confirmed all aspects of Maxwells predictions and confirmed lights character as an electromagnetic wave. Iron-

ically, Hertz did not perceive the enormous importance of his experiments which he considered to have no practical applications whatever.

The following diagram illustrates the electromagnetic spectrum and shows the various wavelengths in comparison with familiar objects.

(22.3.3)

FIGURE: Comparison of Wavelengths and Objects.

The electromagnetic spectrum includes radio waves (longest) to gamma rays (shortest). Notice the tiny fraction of the spectrum that is taken up by visible light—from about 7.5×10^{-4} mm (red) to 3.5×10^{-4} mm (violet).

Until very recently, ROY G BIV, the R(ed) O(range Y)ellow G(reen) B(lue I)ndigo V(iolet colors of the visible rainbow, represented the totality of observable reality. With devices such as compasses, radio receivers, and photographic film, we extend our five senses so that we can derive information from a wider range of the electromagnetic spectrum .



(23) Electromagnetism: A Quantitative View

The proper linking of mathematics to reality is not always clear, although the power of mathematical modeling is rarely in doubt. (*For an example of a model based on a false assumption, see (24.6), pg. 348.*) A befuddled character, Miss Withers, articulates this point of view in the 1934 book, *The Puzzle of the Silver Persian* [26], by Stuart Palmer, where we read

Miss Withers . . . sat and stared at the red coals of her fire . . . they presented innumerable, fantastic pictures, but never did they suggest the clear and definite answers to this complex puzzle which the spinster wanted to see. Finally, she took up a piece of hotel stationery. “I might do it by algebra,” she thought. “ X equals—that’s the trouble. X doesn’t equal anything. Nor, for that matter, does Y or Z .” She swept away her meaningless figures and wished wistfully that she understood relativity.

Theory Meets Observation: In this chapter, we shall see how the theoretically derived wave equation (*Lemma (16.3.7), pg. 206, and Theorem (16.4.1)*) which describes the speed v of a traveling time-dependent wave $f(x, t)$ in one dimension, describes Maxwell’s experimental observations exactly for the propagated electrical and magnetic waves, $E(x, t)$ and $B(x, t)$. We will then see how Maxwell then calculated each of their speeds to be $v = c$, the speed of light in a vacuum.

(23.1) Quantitative Preliminaries

(23.1.1)
Calculus
needed

Notations for Fields $B(x, t)$, $E(x, t)$, and Their Arithmetic Differences. We will be interested in the case where the electric and magnetic fields propagate along a direction that we will choose as the X -axis of a reference system. In this case, the magnetic and electric field values, $B(x, t)$, and $E(x, t)$, depend on the *two* variables: distance, x , and time, t . The following table lists compressed notations we will use when one of the variables is held constant (and, hence, suppressed).

(23.1.1a) Compressed Notations for $B(x,t)$, $E(x,t)$

x varies, t fixed	t varies, x fixed
$B_x \stackrel{\text{def}}{=} B(x,t)$	$B_t \stackrel{\text{def}}{=} B(x,t)$
$E_x \stackrel{\text{def}}{=} E(x,t)$	$E_t \stackrel{\text{def}}{=} E(x,t)$



(23.1.1b) **Orientation of \vec{B}_x , \vec{E}_x** : As per Figure (23.2.2), pg. 317, E_x and B_x represent the values of the electric and magnetic fields that are perpendicular to the X -axis at point x —the vector \vec{B}_x is parallel to the Y axis and vector \vec{E}_x is parallel to the Z axis. (*Sometimes E_x is used to denote the component of the electric field along the x direction, a convention we will not follow.*)

With the compressed notation of (23.1.1a), we have an economical way to write the differences between values of $B(x,t)$ and $E(x,t)$ at pairs of points, $\{x, x+\Delta x\}$ when time, t , is fixed; and between values $\{t, t+\Delta t\}$ when distance, x , is fixed. The following notation will be used for differences such as $B_{x+\Delta x} - B_x$.

(23.1.1c) Notations for ΔB , ΔE

x varies, t fixed	t varies, x fixed
$\Delta B_x \stackrel{\text{def}}{=} (B_{x+\Delta x} - B_x)$	$\Delta B_t \stackrel{\text{def}}{=} (B_{t+\Delta t} - B_t)$
$\Delta E_x \stackrel{\text{def}}{=} (E_{x+\Delta x} - E_x)$	$\Delta E_t \stackrel{\text{def}}{=} (E_{t+\Delta t} - E_t)$

Role of Partial Derivatives, $\partial/\partial x$, $\partial/\partial t$, in (23.1.1c). With sufficient (and reasonable) continuity conditions on the function B , we can always make Δx small enough to bring $\Delta B_x/\Delta x$ as close as we like to the partial derivative, $\partial B_x/\partial x$. We say this symbolically by writing

$$(23.1.1d) \quad \frac{\Delta B_x}{\Delta x} \approx \left(\frac{\partial B_x}{\partial x} \right) \quad \text{or} \quad \Delta B_x \approx \left(\frac{\partial B_x}{\partial x} \right) \Delta x. \quad \left\{ \begin{array}{l} \text{Partial} \\ \text{derivatives} \\ \text{are defined} \\ \text{in (16.3.4).} \end{array} \right.$$

A similar argument justifies the other \approx relationships of (23.1.1c) relative to ΔB_t , ΔE_x , and ΔE_t .

(23.1.2)

Mathematicians and physicists often approximate complicated objects with simple objects that are better understood.



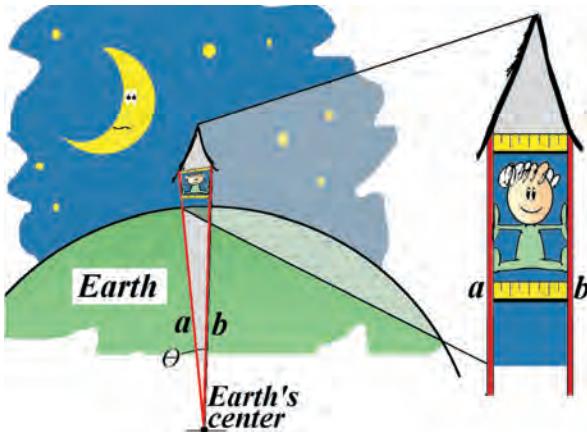
For example, oddly-shaped areas can be approximated by rectangles that are “close enough” to the original area. (*This is a key idea in calculus.*)

In general, reality is often approximated or simplified with idealized models as when, for example, we claim that a rectangular, uniformly dense wooden board with equal weights at either end, and with its pivot in the exact center (*a see-saw*), will remain in a balanced, horizontal position. In the real world, however, the wood density may vary (contrary to our assumption), and our measurements are never *exact* down to the molecular level. However, simplified models are considered valid (enough) when they work effectively.

Figure (23.1.3) illustrates similar simplifications that will be used when we apply the laws of Faraday and Ampère-Maxwell to calculate the speed of light.

(23.1.3)

FIGURE: *Approximately Parallel & Approximately Equal Lines.*



(23.1.3a) **Parallel Lines:** The vertical pie-shaped segment (with the roof on top) has its vertex at the center of the earth. Sides a and b , which form the walls of a building at the surface, are APPROXIMATELY PARALLEL if the vertex angle, θ , is small enough. In the real world,

we act as if lines a and b of Figure (23.1.3) really *are* parallel, as when we confidently use a plumb line to align “vertical” walls of a building.

(23.1.3b) Arcs are approximately straight. In our everyday experience, we pretend that the surface of a lake is horizontal and straight even though it follows the arc of the earth’s circumference. Similarly, when θ in Figure (23.1.3) is very small, it is virtually impossible to know whether the floor and ceiling are true horizontal straight lines, or whether they follow the circular arc of the earth’s surface.

But with such small θ , who can tell the difference?

(23.1.3c) Equal Lines: Often, a non-constant function f is considered to be “approximately constant” if neighboring x and x' produce values, $f(x)$ and $f(x')$, that do not differ significantly. That is, if x is close to x' , then $f(x)$ is close to $f(x')$, or $f(x) \approx f(x')$.

For example, suppose sides $a = f(x)$ and $b = f(x')$ in Figure (23.1.3) are values for some function f . Suppose side a is 20 feet shorter than side b so that from the surface of the earth, the house wall a would be a full two stories shorter than wall b . But this inequality begins to disappear when we consider the full lengths of segments a and b . In fact, if the length $a = 4,000 \text{ miles}$, and $b = 4,000 \text{ miles} + 20 \text{ feet}$, then sides a and b are APPROXIMATELY EQUAL in the sense that their lengths differ, relatively speaking, by a tiny amount—about one part in a million, or $1/10,000$ th of 1%.

(23.2) A Quantitative View of Propagation

(23.2.1) The Alternating Current Setup: We focus attention only on the X -axis of Figure (22.1.1), which is a single ray that threads its way through the magnetic field \vec{B} generated by the alternating current.

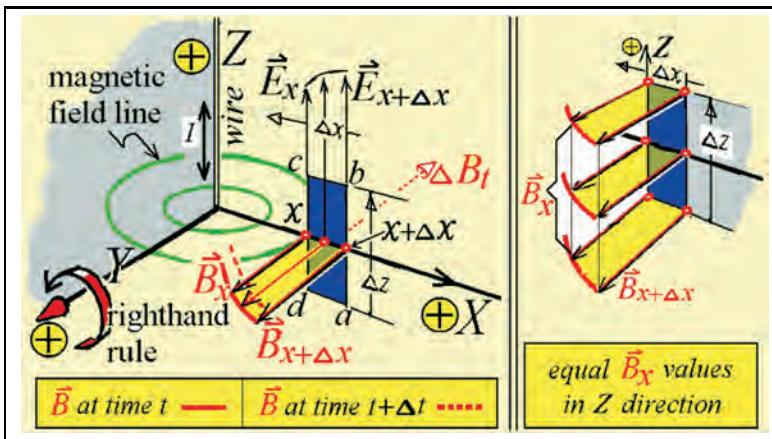
(23.2.1a) Graphing B Vectors. The visual bookkeeping method of recording values of \vec{B}_x at each point x is to imitate the needle of a south-seeking compass placed at that point.

Specifically, at each point x on the X -axis, draw a vector \vec{B}_x

- in the X - Y plane,
- perpendicular to the X -axis,
- with tail at the point x , pointing in a north→south orientation, and
- with length $\|\vec{B}_x\| = B_x$, equal to the magnitude of the magnetic field measured at point x .

To aid in the proofs of following theorems, we set down Figure (23.2.2).

(23.2.2) **FIGURE: A Changing Magnetic Field in an Area Induces an Electrical Field on its Boundary**



The magnetic \vec{B}_x vectors depend on the radial distance x from the vertical wire and are tangent to the green circles shown in the left-hand panel.

The electrical vectors \vec{E}_x also depend on the radial distance x from the vertical wire. However, due to the right-hand rule about the Y -axis, vectors \vec{E}_x are perpendicular to the X -axis, pointing in the Z direction, which is parallel to the wire. (See Note (23.1.1b) describing the configurations of \vec{E}_x and \vec{B}_x .)

(23.2.3)

Calculus
needed

THEOREM: (Faraday's Law. Magnetic Field Induces Electric Field.) Given a vertical wire with alternating current, $I(t) \downarrow$. Then the changing horizontal magnetic field vectors \vec{B} induce vertical electric field vectors, \vec{E} , such that at time t and a radial distance x from the wire, we have

$$(23.2.3a) \quad -\frac{\partial B_t}{\partial t} = \frac{\partial E_x}{\partial x}.$$

PROOF: See Figure (23.2.2). Equation (23.2.3a) is the general Faraday's Law of Induction (21.7.3c) as applied to a differential rectangular area $\Delta A = \Delta x \Delta z$ shown in the right-hand panel of Figure (23.2.2).

- **The orientation of the boundary of ΔA .** Since the positive direction of the Y -axis coincides with the positive direction of the axle (18.4.5), the right-hand rule tells us that the rectangular boundary of the area $\Delta A = \Delta x \Delta z$ in Figure (23.2.2) is oriented *counterclockwise*. That is, orientation of $bdry(\Delta A)$ is described by

$$(23.2.3b) \quad bdry(\Delta A) = a \rightarrow b \rightarrow c \rightarrow d \rightarrow a, \quad \text{in that order!}$$

- **The magnetic flux,** from (20.2.2a), is defined over the differential area $\Delta A = \Delta x \Delta z$ from Figure (23.2.2), to be

$$(23.2.3c) \quad \Phi_B(\Delta A) \approx B_t \cdot \Delta A = B_t \cdot \Delta x \Delta z.$$

Since we consider the magnetic field at a fixed distance x but at different times, we use B_t in (23.2.3c) to represent $B(x, t)$ (23.1.1c). The philosophy of (23.1.3c) allows us to assume that B is approximately constant over the sufficiently small differential area ΔA . That is, the value of $B_t = B(x, t)$ is as good a choice as $B(x', t)$ when x' is close to x .

- **The right-hand side of Faraday's law,** (21.7.3c), requires that we compute $-d/dt \Phi_{\vec{B}}(\Delta A)$, the negative of the time derivative of the magnetic flux, $\Phi_{\vec{B}}(\Delta A)$ (20.2.2a). Accordingly, when the position is fixed,

$$(23.2.3d) \quad -\frac{d}{dt} \Phi_B(\Delta A) = -\frac{d}{dt} (B_t \cdot \Delta A) \quad \text{from (23.2.3c)}$$

$$= - \left[\frac{d}{dt} B_t \right] \cdot \underbrace{\Delta x \Delta z}_{\Delta A} \quad \begin{array}{l} \text{factor out the} \\ \text{constant } \Delta A \end{array}$$

$$= -\frac{\partial B_t}{\partial t} \Delta x \Delta z. \quad \frac{d}{dt} \text{ above is actually the partial } \frac{\partial}{\partial t} \text{ with } x \text{ fixed}$$

- The left-hand side of Faraday's law, (21.7.3c), asks us to compute the closed line integral of the induced electric field, \vec{E}_l , at points l on $\text{bdry}(A)$, namely,

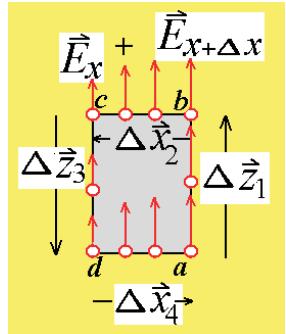
$$(23.2.3e) \quad \oint_{bdry(\Delta A)} \langle \vec{E}_l, d\vec{l} \rangle = \oint_{a \rightarrow b \rightarrow c \rightarrow d \rightarrow a} \langle \vec{E}_l, d\vec{l} \rangle \quad \begin{matrix} \text{counterclockwise,} \\ \text{from (23.2.3b)} \end{matrix}$$

Each of the integrals above is of the class given by (18.8.6a), where integration occurs along the four counterclockwise oriented chain of segments, $\Delta\vec{z}_1$, $\Delta\vec{x}_2$, $\Delta\vec{z}_3$, and $\Delta\vec{x}_4$ (*see figure at right*). Therefore, the values of integrals (1), (2), (3), and (4) of (23.2.3e), are, respectively,

(23.2.3f)

$$\langle \vec{E}_{x+\Delta x}, \Delta \vec{z}_1 \rangle_{(1)} + \langle \vec{E}_x, \Delta \vec{x}_2 \rangle_{(2)}$$

$$+ \begin{array}{c} \langle \vec{E}_x, \Delta \vec{z}_3 \rangle \\ (3) \end{array} + \begin{array}{c} \langle \vec{E}_x, \Delta \vec{x}_4 \rangle \\ (4) \end{array}.$$



Using the fact that $\|\Delta x_2\| = \|\Delta x_4\| = \Delta x$ and $\|\Delta z_1\| = \|\Delta z_3\| = \Delta z$, along with Definition (18.6.1), we obtain

$$(1) \langle \vec{E}_{x+\Delta x}, \Delta \vec{z}_1 \rangle = E_{x+\Delta x} \underbrace{\cos \theta_1}_1 \Delta z, \quad \begin{matrix} E_{x+\Delta x} \text{ constant} \\ \text{on } \Delta \vec{z}_1 \end{matrix}$$

$$(2) \langle \vec{E}_x, \Delta \vec{x}_2 \rangle = E_x \underbrace{\cos \theta_2}_{0} \Delta x, \quad \vec{E}_x \perp \Delta \vec{x}_2$$

$$(3) \langle \vec{E}_x, \Delta \vec{z}_3 \rangle = E_x \underbrace{\cos \theta_3}_{-1} \Delta z, \quad E_x \text{ constant on } \Delta \vec{z}_3$$

$$(4) \quad \langle \vec{E}_x, \Delta \vec{x}_4 \rangle = E_x \underbrace{\cos \theta_4}_{0} \Delta x, \quad \vec{E}_x \perp \Delta \vec{x}_4$$

since the angles θ_i between electric field vectors, \vec{E} , and the line segments $\Delta\vec{z}$ and $\Delta\vec{x}$ have values $\theta_1 = 0$, $\theta_2 = 90^\circ$, $\theta_3 = 180^\circ$, and $\theta_4 = 0$. Therefore, we have the sum of the inner products (23.2.3f)

$$(1) + (2) + (3) + (4) = (E_{x+\Delta x} - E_x)\Delta z$$

$$= \Delta E_x \Delta z. \quad \text{from notation of (23.1.1c)}$$

Equate this term, $\Delta E_x \Delta z$ (left-hand side of Faraday's law (21.7.3c)), to $(-\partial B_t / \partial t)\Delta x \Delta z$ of (23.2.3d), (the right-hand side of Faraday's law (21.7.3c)), and solve for $(-\partial B_t / \partial t)$ to obtain

$$(23.2.3g) \quad -\frac{\partial B_t}{\partial t} = \frac{\Delta E_x}{\Delta x}.$$

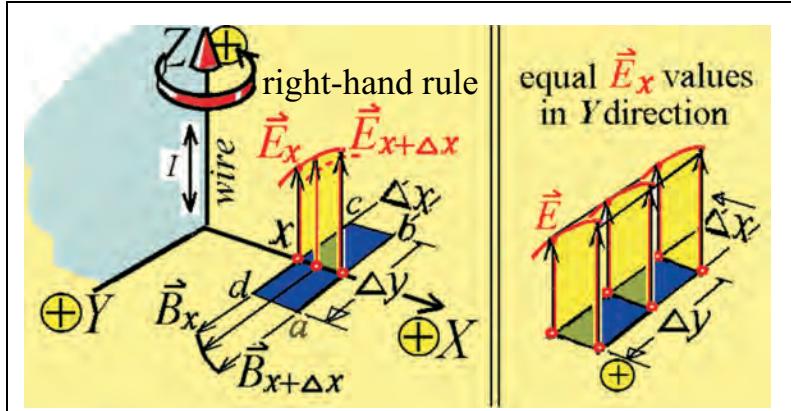
Since $\frac{\Delta E_x}{\Delta x} \rightarrow \frac{\partial E_x}{\partial x}$ as $\Delta x \rightarrow 0$, equality (23.2.3g) transforms to (23.2.3a), which proves the theorem. ■

Now that the vertical values of electric field vectors \vec{E}_x and $\vec{E}_{x+\Delta x}$ are established in (23.2.3f), we use the Ampère-Maxwell law (21.7.3d) to construct their newly-induced horizontal magnetic field vectors, \vec{B}_x , and $\vec{B}_{x+\Delta x}$.

Note: These are not the same \vec{B}_x and $\vec{B}_{x+\Delta x}$ magnetic vectors we started with—these vectors appear Δt time units later.

(23.2.4)

FIGURE: Electrical Field Change Induces Magnetic Field



(23.2.5)
Calculus
needed

THEOREM: (Ampère-Maxwell Law. Electric Field Induces Magnetic Field.) Given a vertical wire with alternating current, $I(t) \downarrow$. Then changing vertical electric field vectors \vec{E} induce horizontal magnetic field vectors \vec{B} such that at time t and at radial distance x from the wire, we have

$$(23.2.5a) \quad \mu_0 \varepsilon_0 \frac{\partial E_t}{\partial t} = -\frac{\partial B_x}{\partial x}.$$

PROOF:

- **The orientation of the boundary of ΔA .** Since the Z -axis defines the positive vertical direction, the right-hand rule (18.4.5), with the Z -axis as its axle, defines the positive orientation of the rectangular boundary with area $A = \Delta x \Delta z$ in Figure (23.2.2) to be *counterclockwise*. That is, orientation of $bdry(A)$ is described by

$$(23.2.5b) \quad bdry(A) = a \rightarrow b \rightarrow c \rightarrow d \rightarrow a, \quad \text{in that order!}$$

.....

- **The electric flux,** from (20.2.1a), is defined over the differential area $\Delta A = \Delta x \Delta z$ to be

$$(23.2.5c) \quad \Phi_B(\Delta E) \approx E_t \cdot \Delta A = E_t \cdot \Delta x \Delta y.$$

Since time t is varying while distance x is fixed, we use E_t in (23.2.5c) to represent $E(x, t)$ (23.1.1c). The philosophy of (23.1.3c) allows us to assume that E is essentially constant over the sufficiently small differential area, ΔA . That is, the value $E_t = E(x, t)$ is as good a choice as $E_{t'} = E(x, t')$ in the evaluation of E when t' is close to t .

.....

- **The right-hand side of Ampère-Maxwell law,** (21.7.3d), requires that we compute $\mu_0 \varepsilon_0 \cdot d/dt \Phi_{\vec{E}}(\Delta A)$, which is $\mu_0 \varepsilon_0$ times the time derivative of the electric flux $\Phi_{\vec{E}}(\Delta A)$ (20.2.1a). Accordingly,

$$(23.2.5d) \quad \mu_0 \varepsilon_0 \frac{d}{dt} \Phi_E(\Delta A) = \mu_0 \varepsilon_0 \cdot \frac{d}{dt} (E_t \cdot \Delta A) \quad \text{from (23.2.5c)}$$

$$= \mu_0 \varepsilon_0 \left[\frac{d}{dt} E_t \right] \cdot \underbrace{\Delta x \Delta y}_{\Delta A} \quad \begin{matrix} \text{factor out the} \\ \text{constant } \Delta A \end{matrix}$$

$$= \mu_0 \epsilon_0 \frac{\partial E_t}{\partial t} \Delta x \Delta y. \quad \left\{ \begin{array}{l} \frac{d}{dt} \text{ above is actually the} \\ \text{partial } \frac{\partial}{\partial t} \text{ with } x \text{ fixed} \end{array} \right.$$

.....

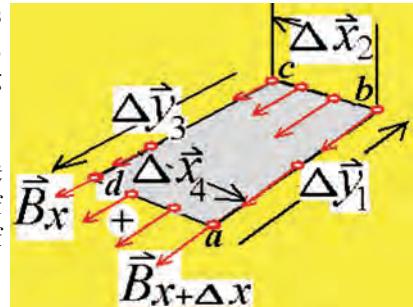
- **The left-hand side of Ampère-Maxwell's law,** (21.7.3d), asks us to compute the closed line integral of the induced magnetic field \vec{B}_l at points l on $\text{bdry}(A)$, namely,

$$(23.2.5e) \quad \oint_{\text{bdry}(\Delta A)} \langle \vec{B}_l, d\vec{l} \rangle = \oint_{a \rightarrow b \rightarrow c \rightarrow d \rightarrow a} \langle \vec{B}_l, d\vec{l} \rangle \quad \begin{array}{l} \text{counterclockwise,} \\ \text{from (23.2.3b)} \end{array}$$

$$= \int_{z=a}^{z=b} \langle \vec{B}_{x+\Delta x}, d\vec{y} \rangle + \int_{x=b}^{x=c} \langle \vec{B}_x, d\vec{x} \rangle + \int_{z=c}^{z=d} \langle \vec{B}_x, d\vec{y} \rangle + \int_{x=d}^{x=a} \langle \vec{B}_x, d\vec{x} \rangle. \quad \begin{array}{l} (1) \\ (2) \\ (3) \\ (4) \end{array}$$

.....

Each of the integrals above is of the class given by (18.8.6a), where integration occurs along the four counterclockwise oriented chain of segments, $\Delta \vec{y}_1$, $\Delta \vec{x}_2$, $\Delta \vec{y}_3$, and $\Delta \vec{x}_4$ (see figure at right). Therefore, the values of integrals (1), (2), (3), and (4) of (23.2.3e), are, respectively,



$$(23.2.5f) \quad \begin{array}{c} \langle \vec{B}_{x+\Delta x}, \Delta \vec{y}_1 \rangle + \langle \vec{B}_x, \Delta \vec{x}_2 \rangle + \langle \vec{B}_x, \Delta \vec{y}_3 \rangle + \langle \vec{B}_x, \Delta \vec{x}_4 \rangle. \\ (1) \qquad (2) \qquad (3) \qquad (4) \end{array}$$

Using the fact that $\|\Delta x_2\| = \|\Delta x_4\| = \Delta x$ and $\|\Delta y_1\| = \|\Delta y_3\| = \Delta y$, along with Definition (18.6.1), we obtain

$$(1) \quad \langle \vec{B}_{x+\Delta x}, \vec{\Delta y}_1 \rangle = B_{x+\Delta x} \underbrace{\cos \theta_1}_{-1} \Delta y,$$

$$(2) \quad \langle \vec{B}_x, \vec{\Delta x}_2 \rangle = B_x \underbrace{\cos \theta_2}_0 \Delta x,$$

$$(3) \quad \langle \vec{B}_x, \vec{\Delta y}_3 \rangle = B_x \underbrace{\cos \theta_3}_1 \Delta y,$$

$$(4) \quad \langle \vec{B}_x, \vec{\Delta x}_4 \rangle = B_x \underbrace{\cos \theta_4}_0 \Delta x$$

since the angles θ_i between the magnetic field vectors, \vec{B} , and the line segments $\Delta\vec{x}$ and $\Delta\vec{y}$ have values $\theta_1 = 180^\circ$, $\theta_2 = 90^\circ$, $\theta_3 = 0^\circ$, and $\theta_4 = 90^\circ$. Therefore, we have the sum of the inner products (23.2.5f)

$$\begin{aligned}(1) + (2) + (3) + (4) &= -(B_{x+\Delta x} - B_x)\Delta y \\ &= -\Delta B_x \Delta y. \quad \text{from notation of (23.1.1c)}\end{aligned}$$

Equate this term, $-\Delta B_x \Delta y$ (left-hand side of Ampère-Maxwell law (21.7.3d)), to $(\mu_0 \varepsilon_0 \partial B_t / \partial t) \Delta x \Delta y$ of (23.2.5d), (the right-hand side of the Ampère-Maxwell law (21.7.3d)), and solve for $\mu_0 \varepsilon_0 (\partial B_t / \partial t)$ to obtain

$$(23.2.5g) \quad \mu_0 \varepsilon_0 \frac{\partial B_t}{\partial t} = -\frac{\Delta B_x}{\Delta x}.$$

Since $\frac{\Delta B_x}{\Delta x} \rightarrow \frac{\partial B_x}{\partial x}$ as $\Delta x \rightarrow 0$, equality (23.2.5g) transforms to (23.2.5a), which proves the theorem. ■

(23.3) Theoretical Speed of Wave Propagation

We restate and summarize the results of Faraday's law and the Ampère-Maxwell law. Their differential forms, (23.2.3a) and (23.2.5a), appear in the first row of the following table as entries ① and ②. The second row, with entries ③ and ④ contains equation (16.3.7a), which relates the time derivative, $\partial/\partial t$ and the distance derivative $\partial/\partial x$, for any function traveling rightward with speed v .

To allow for all possibilities, we at first assume the two waves, \vec{B} and \vec{E} , might have different speeds, which we denote by v_B and v_E , respectively. (As we shall see in (23.3.3a), v_B necessarily equals v_E .)

(23.3.1)

$\textcircled{1} \quad -\frac{\partial B}{\partial t} = \frac{\partial E}{\partial x}$ <i>Faraday Law (23.2.3a)</i>	$\textcircled{2} \quad \mu_0 \varepsilon_0 \frac{\partial E}{\partial t} = -\frac{\partial B}{\partial x}$ <i>Ampère-Maxwell Law (23.2.5a)</i>
$\textcircled{3} \quad \frac{\partial B}{\partial x} = -\frac{1}{v_B} \frac{\partial B}{\partial t}, \quad (16.3.7a)$ <i>Mixed-Partials Property (16.3.7a)</i>	$\textcircled{4} \quad \frac{\partial E}{\partial x} = -\frac{1}{v_E} \frac{\partial E}{\partial t}, \quad (16.3.7a)$

- (23.3.2)** **Consequences of** equations ①, ②, ③, and ④ of Table (23.3.1) are further distilled and summarized as follows:

$$(23.3.2a) \quad \frac{\partial E}{\partial t} = v_E \frac{\partial B}{\partial t}, \quad \text{from ①, ④, eliminate } \frac{\partial E}{\partial x}$$

$$(23.3.2b) \quad \frac{\partial B}{\partial t} = v_B \mu_0 \varepsilon_0 \frac{\partial E}{\partial t}, \quad \text{from ②, ③, eliminate } \frac{\partial B}{\partial x}$$

$$(23.3.2c) \quad \frac{\partial E}{\partial x} = v_B \frac{\partial B}{\partial x}, \quad \text{from ①, ③, eliminate } \frac{\partial B}{\partial t}$$

$$(23.3.2d) \quad \frac{\partial B}{\partial x} = v_E \mu_0 \varepsilon_0 \frac{\partial E}{\partial x}. \quad \text{from ②, ④, eliminate } \frac{\partial E}{\partial t}$$

We now have the tools to show how Maxwell computed c , the speed of electromagnetic (EM) waves in a vacuum. Light was not known to be in the family of electromagnetic waves in Maxwell's time, although he was impressed by the "coincidental" fact that his (invisible) electromagnetic waves, and visible light, had exactly the same speed.

(23.3.3)
Calculus
needed

THEOREM: (Wave speed in terms of permeability and permittivity.) Given an electromagnetic wave with an electric field component E having speed v_E , and a magnetic field component, B , having speed v_B . If the electrical permittivity is ε_0 and magnetic permeability is μ_0 , then, necessarily,

$$(23.3.3a) \quad v_E = v_B = c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}.$$

PROOF: Since the order of partial derivatives is irrelevant, we may equate $\partial/\partial x$ of Equation (23.3.2a) with $\partial/\partial t$ of Equation (23.3.2c) to obtain $v_E = v_B$ for both the speed v_E of the magnetic wave \vec{E} and the speed v_B of the electric wave \vec{B} . Denoting this common speed by

$$(23.3.3b) \quad v_E = v_B = c$$

establishes the first two equalities of (23.3.3a). The equal speeds of (23.3.3b) are intuitively correct since, from Figure (22.2.2), the time-dependent field E generates a time dependent field B which, in its turn, generates an electric field farther from the source. In

particular, both E and B propagate together, neither lags behind the other, which implies $v_E = v_B$.

.....

If we eliminate any one of the four partial derivatives, $(\partial E/\partial x)$, $(\partial E/\partial t)$, $(\partial B/\partial x)$, $(\partial B/\partial t)$, from the equations, (23.3.2a)-(23.3.2d), we will obtain the formula $c = 1/\sqrt{\mu_0\varepsilon_0}$.

For example, eliminate $(\partial B/\partial t)$ from (23.3.2a) and (23.3.2b), to obtain $\partial E/\partial t = (c^2\mu_0\varepsilon_0)\partial E/\partial t$, from which it follows that $1 = c^2\mu_0\varepsilon_0$. This completes the third equality of Equation (23.3.3a) and the proof is done. ■

NOTE: *Wave equation and Maxwell's equations.* With (23.3.3a) in hand, we can show that Faraday's law and the Ampère-Maxwell law imply that each of the fields, \vec{B} and \vec{E} , satisfies the wave equation (16.4.1a). See Exercise (23.6).

Now that we know that the speed c (23.3.3a) applies to both components, \vec{B} and \vec{E} of the electromagnetic wave, we can relate their magnitudes at any point x_0 in space.

(23.3.4)

Calculus
needed

THEOREM: (Linking E and B .) Given an electromagnetic wave with electric field component \vec{E} and magnetic field component \vec{B} . Assume the null initial condition,

$$(23.3.4a) \quad E(0, 0) = B(0, 0) = 0.$$

Then the electromagnetic wave propagates with speed c (23.3.3a) in the direction

$$(23.3.4b) \quad \vec{E} \times \vec{B}$$

where the magnitudes of the electric and magnetic components, E and B , are proportional as indicated by the equation

$$(23.3.4c) \quad E(x_0, t_0) = c \cdot B(x_0, t_0)$$

for all distances x_0 and times t_0 .

PROOF: The direction (23.3.4b) of the propagated wave was established in (22.2.3b).

To validate (23.3.4c), integrate (23.3.2c) over x (holding t_0 fixed). Replacing the symbol $\partial/\partial x$ with d/dx , since the second variable, t_0 , is fixed, we have

$$\int_{x=0}^{x=x_0} \left(\frac{dE}{dx} \right) dx = \int_{x=0}^{x=x_0} c \cdot \left(\frac{dB}{dx} \right) dx$$

whose left and right sides compute to be, respectively,

$$(23.3.4d) \quad [E(x_0, t_0) - E(0, t_0)] = c \cdot [B(x_0, t_0) - B(0, t_0)].$$

Similarly, integrating (23.3.2a) over t , holding $x = 0$ fixed, we obtain

$$\int_{t=0}^{t=t_0} \left(\frac{dE}{dt} \right) dt = \int_{t=0}^{t=t_0} c \cdot \left(\frac{dB}{dt} \right) dt$$

whose left and right sides are, respectively,

$$(23.3.4e) \quad [E(0, t_0) - E(0, 0)] = c \cdot [B(0, t_0) - B(0, 0)].$$

Now once we add Equations (23.3.4d) and (23.3.4e) and use fact (23.3.4a), which holds that $B(0, 0) = E(0, 0) = 0$, we see that (23.3.4c) results. This ends the proof. ■



(23.4) Maxwell's Calculation of c

- (23.4.1) List of Units.** Here follows a brief list, in the MKS (meter-kilogram-second) system of units, that we will use:
- mtr: The **meter**, a unit of distance.
 - sec: The **second**, a unit of time.
 - C : The **coulomb**, the unit of charge on a particle q .
 - A : The **ampere** = C/sec , defined as one coulomb per second. Amperes measure the coulombs/second (an electric current) passing through a wire which is analogous to measuring gallons/second (a liquid current) of water flowing through a conducting pipe.
 - N : The **newton**, a unit of force.

From this list, we take away the following: If one ampere A equals one coulomb per second, i.e., $1A = 1C/1\text{sec.}$, then, consistent with methodology of dimensional analysis (Appendix (D)), this equality translates to the unitless quotient

$$(23.4.2) \quad 1 = \frac{A}{C/\text{sec}}.$$

Recall that permittivity (19.2.5a) ϵ_0 appears in Coulomb's Law (19.2.4b), pg. 264, and can therefore be measured in electrostatic experiments. Similarly, permeability (21.2.2) μ_0 appears in the Biot-Savart Law (21.2.1a), pg. 291, and can be measured in magnetostatic experiments. Accepted values are

$$(23.4.3) \quad \begin{aligned} \epsilon_0 &= 8.854\ 187\ 817\ 62 \times 10^{-12} \frac{C^2}{N \cdot \text{mtr}^2} \quad \text{and} \\ \mu_0 &= 1.256\ 637\ 061\ 43 \times 10^{-6} \frac{N}{A^2} \end{aligned}$$

whose product is

$$\begin{aligned} \epsilon_0 \mu_0 &\approx 11.1265 \times 10^{-18} \frac{C^2}{\text{mtr}^2 A^2} \quad \text{newtons "N" cancel.} \\ &= 11.1265 \times 10^{-18} \frac{C^2}{\text{mtr}^2 \cdot A^2} \underbrace{\left(\frac{A}{C/\text{sec}} \right)^2}_1 \quad \text{from (23.4.2)} \\ &= 11.1265 \times 10^{-18} \frac{\cancel{C}^2}{\text{mtr}^2 \cancel{A}^2} \left(\frac{\cancel{A}^2 \text{ sec}^2}{\cancel{C}^2} \right) \quad \text{unit cancellation} \\ &= 11.1265 \times 10^{-18} \frac{\text{sec}^2}{\text{mtr}^2}. \quad \text{final form} \end{aligned}$$

With $\epsilon_0 \mu_0$ in hand (*at the end of this chain of equalities*), we finally calculate $(\sqrt{\epsilon_0 \mu_0})^{-1}$, which is c , the speed of light in a vacuum. In so doing, we discover that

$$(23.4.4) \quad c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} \approx 2.99792 \times 10^8 \frac{\text{mtr}}{\text{sec}}$$

where the final units, $\frac{\text{mtr}}{\text{sec}}$ (*usually denoted* $\frac{m}{s}$ *or* m/s) are, in fact, units of speed.

(23.5) Mathematical Hits

- (23.5.1) Hit: *Unification.*** Maxwell organized the known mathematical models of electricity and magnetism and then made them consistent by adding the “virtual current” or displacement current term $\varepsilon_0\mu_0 d\Phi_E(A)/dt$ to Ampère’s law (21.4.3c). Without this correction, there would be no conservation of charge and the electrical and magnetic fields would not satisfy the wave equation.

This quartet of equations (21.7.3), commonly known as Maxwell equations, unified the theory of electricity, optics, and magnetism.

- (23.5.2) Hit: *Predicting Wave Structure.*** Maxwell’s equations predicted the perpendicularity of the electric and magnetic components of electromagnetic waves (*Figure (22.3.1)*). Moreover, Maxwell’s equations show that the electric and magnetic components fields each satisfy the wave equation (23.6.1).

Wave, what wave? This theoretical mathematical wave structure was finally confirmed experimentally by HEINRICH HERTZ (1857-1894) twenty years later, in 1888. (*See (22.3.2) for more detail.*)

- (23.5.3) Hit: *Inspiring Einstein.*** Einstein’s two basic assumptions of special relativity, especially the assumption that c is constant for all observers (1.5.2b), regardless of their inertial frame, can be traced back directly to Maxwell’s work.

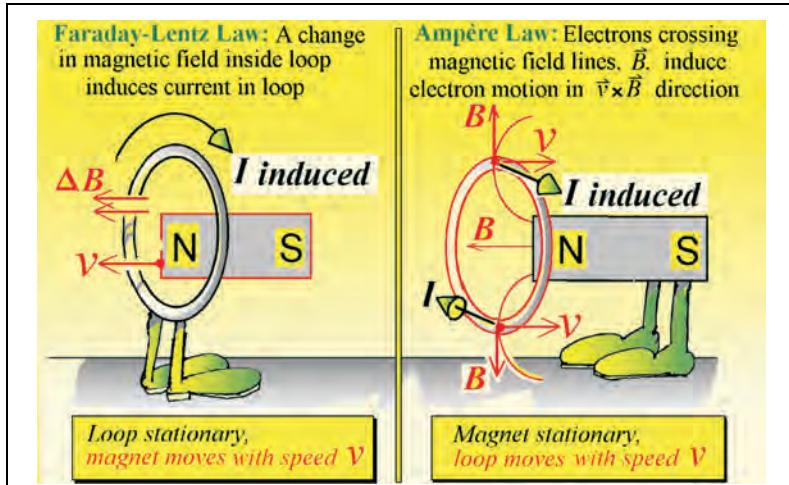
In fact, in the very opening lines of his seminal 1905 paper on special relativity [12], Einstein cites the “asymmetries” in electromagnetic theory (*See the quote in (23.5.3b).*)

The asymmetries mentioned by Einstein are these:

The current induced as a magnet moves through a stationary wire loop is determined by the Faraday-Lenz law. If, however, the magnet

is held stationary and the wire loop is moved through it, the induced current is determined by a different law, the Ampère-Maxwell law. Despite the difference in the responsible laws, the resulting current is the same and depends only on the relative motion of the wire loop (frame) and the magnet (frame).

(23.5.3a) **FIGURE:** Two Frame-dependent Explanations for Current I



- LEFT PANEL: When the wire is fixed and the magnet moves, current I is induced (in the direction shown) due to Faraday's Law (21.7.3c).
- RIGHT PANEL: When the magnet moves and the wire is stationary, then the same current I is induced, as predicted by the Ampère-Maxwell Law (21.7.3d).

To quote Einstein from *On the electrodynamics of moving bodies* [12]:

(23.5.3b) It is known that Maxwell's electrodynamics—as usually understood at the present time—when applied to moving bodies, leads to asymmetries which do not appear to be inherent in the phenomena. Take, for example, the reciprocal electrodynamic action of a magnet and a conductor. The observable phenomenon here depends on the relative motion of the conductor and the magnet, whereas the customary view draws a sharp distinction between the two cases in which either the one or the other of these bodies is in motion.

Endorsing Maxwell's work, while taking Galileo's relativity one step further (1.5.2a), Einstein postulated that all laws of physics, *including* those of electromagnetism, shall be deemed valid in all inertial frames of reference.

In addition, Einstein postulated that c , the speed of light in a vacuum, is constant for all observers (*in all frames of reference*) independent of whether the light source, or the observer, is in motion). One might argue that this postulate is unnecessary since it follows from Maxwell's equations (23.3.3a).

- (23.5.4) **Hit: Predicting relation of energy and frequency.** The wave equation defies intuition in the following sense: A rope is tied at one end as shown in Figure (16.1.3) and a wave is propagated in the rope by the vertical oscillation of its loose end which is being pulled with constant “pulling” force T . (*This perceived model is the reality phase of the analysis.*) Intuition would have it that if we hold the pulling force constant, and invest the rope with more energy by oscillating the loose end faster, then the speed of the propagated wave should increase. But it doesn’t. Let the traveling wave be modeled by Figure (16.1.6). (*This graphic represents the cartoon phase of the analysis.*) Then, analyzing the mathematical structure—vector decompositions, etc.—we arrive at Equation (16.1.6e) which tells us that as long as the pulling force T is constant, the speed v does not change. In fact, it is the frequency of the traveling wave that will change. (*This is the mathematical component of the analysis.*)

Although the propagated wave in the rope is a mechanical phenomenon, there is a parallel property with respect to light waves or electromagnetic waves—light that is invested with higher energy does not travel faster (its speed in a vacuum is always c), rather, its *frequency* increases which is evidenced in the visible spectrum by a color change from red (lower frequency) up to violet (higher frequency).



(23.6) Exercises

(23.6.1) **Exercise** **Calculus required** Maxwell equations imply the wave equation. Show that the differential forms of the Faraday and Ampère-Maxwell laws, items ① and ② of (23.3.1), imply that each of the fields, $B(x, t)$ and $E(x, t)$, satisfies the wave equation (16.4.1a).

Hint: Assume the functions $B(x, t)$ and $E(x, t)$ are sufficiently smooth so that mixed partials are equal (independent of the order of differentiation). That is,

$$\frac{\partial^2 B}{\partial x \partial t} = \frac{\partial^2 B}{\partial t \partial x} \quad \text{and} \quad \frac{\partial^2 E}{\partial x \partial t} = \frac{\partial^2 E}{\partial t \partial x}.$$

The figure shows a graph (drawn horizontally in the X - Y plane) of values of \vec{B}_x on the X -axis over a differential width of Δx .

.....

(23.6.2) Prove the following

Theorem: The function $B(x, t)$ is real-valued depending on the space variable x and time t that travels to the right as t increases. That is, in the X - Y plane, the graph of $B_x^{(t+\Delta t)}$ is to the right of the earlier $B_x^{(t)}$ graph. Then:

If $B(x, t)$ increases in x with t fixed ($\Delta B_x > 0$), then $B(x, t)$ decreases in t with x fixed ($\Delta B_t < 0$).

Conversely,

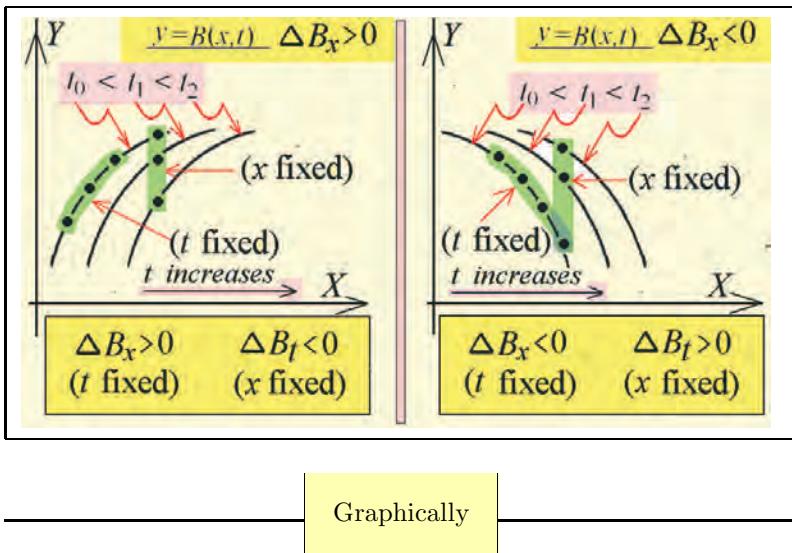
If $B(x, t)$ decreases in x with t fixed ($\Delta B_x < 0$), then $B(x, t)$ increases in t with x fixed ($\Delta B_t > 0$).

Formally, in the notation of (23.1.1a), we have, for any (x, t) ,

$$(23.6.2a) \quad \begin{aligned} \Delta B_x > 0 \ (\text{t is fixed}) &\implies \Delta B_t < 0 \ (\text{x is fixed}), \\ \Delta B_x < 0 \ (\text{t is fixed}) &\implies \Delta B_t > 0 \ (\text{x is fixed}). \end{aligned}$$

Hint: Properties (23.6.2a) can be seen in two ways — graphically and analytically: Figure (23.6.2b) shows two panels in which the function $y = B(x, t)$ moves to the right as t increases.

(23.6.2b) **FIGURE:** Function $y = B(x,t)$ Traveling to the Right as Time Increases



Left panel of (23.6.2b).

Fix t . The function $B(x,\cdot)$ in the $X-Y$ plane has positive slope ($\underline{\Delta B_x > 0}$). This is shown by the points in the slanted grey area where increasing x implies increasing $y = B(x,\cdot)$.

Fix x . The sequence of points in the vertical grey area shows that as t increases, the values of $y = B(x,t)$ decrease ($\underline{B_t < 0}$).

Right panel of (23.6.2b).

Fix t . Function $B(x,\cdot)$ in the $X-Y$ plane has negative slope ($\underline{\Delta B_x < 0}$). This is shown by the points in the slanted grey area where increasing x implies decreasing $y = B(x,\cdot)$.

Fix x . The sequence of points in the vertical grey area shows that as t increases, the values of $y = B(x,t)$ increase ($\underline{B_t > 0}$).

The underlined quantities establish (23.6.2a).

Analytically

In (16.3.7), substitute B for f to see that the signs of $\partial B_x / \partial x$ and $\partial B_t / \partial t$ are always opposite to each other. That is, (16.3.7a) tells us that B_x always increases (decreases) in x as B_t decreases (increases) in t . Briefly put,

$$\text{sign}(\Delta B_x) = -\text{sign}(\Delta B_t).$$

IX. Final Thoughts

Part (X) Final Thoughts.
Some historical context is given along with perspectives on the process of knowing, its applicability, and reliability.



(24) Epilogue: Final Thoughts

(24.1) A Coming of Age

- (24.1.1) **Old Mathematics.** Thousands of years of mathematics have produced more than mere lists of formulas and deduced results. One component of our intellectual heritage that has been defined and nourished by mathematical progress is the *process of proof*.

The theorem of PYTHAGORAS OF SAMOS (*circa* 580–500 B.C.) generalized a property of right triangles that had been known for a few special cases. (*a RIGHT TRIANGLE is any triangle in which two sides form a 90° angle.*) For example, it was known that for the (3, 4, 5), right triangle, $3^2 + 4^2 = 5^2$, and for the (5, 12, 13) triangle, we have $5^2 + 12^2 = 13^2$. Pythagoras' theorem says that for *all* right triangles with sides (a, b, c) , we have $a^2 + b^2 = c^2$. The converse is also true—if the sides (a, b, c) of some triangle satisfy the equation $a^2 + b^2 = c^2$, then it is necessarily a *right* triangle.



Today, there are dozens of independent proofs of the Pythagorean theorem¹. This theorem has come to transcend the formula. Today, mathematicians have produced generalizations that go well beyond triangles in \mathbf{R}^2 . The far-reaching nature of the Pythagorean theorem is nicely described by Robert Crease in his article, *Pythagoras* [7].

Over 2,300 years ago, EUCLID (325–265 B.C.) wrote his *Elements* [15] in which he compiled and organized a body of geometric theorems. More importantly, he provided proofs that rendered the results accessible to and testable by any reader.

Then, ARCHIMEDES (287–212 B.C.), with his use of *infinitesimals*, was on the verge of discovering calculus before he was slain by a Roman soldier at Syracuse as he was drawing figures in the sand during the Second Punic War. The PALIMPSESTS of Archimedes, documents of Archimedes' works that were copied by monks in the tenth century, are the only source of his *The Methods* and *Stomachion* (*a 14-piece version of the 7-piece tangram*) along with the only known Greek text of his work, *On Floating Bodies*.

¹ See the online animation (www.special-relativity-illustrated.com) or [8], pg. 249 for a non-computational proof.

(24.1.2)



In the twelfth century, the Palimpsest parchment would be washed and cut to be recycled into liturgical texts. The Palimpsests were rediscovered in 1906 and only in 1998, were they made available for further study when a private collector gave them over to scholars at the WALTERS ART MUSEUM in Baltimore, Maryland. Fortunately, recovery of the erasures are being made there with the use of multi-spectral imaging.

(See [www.archimedespalimpsest.org.](http://www.archimedespalimpsest.org/))

(24.1.3)

Earliest Mathematical Modeling. Early on, the non-concrete, or abstract operations of *addition*, *subtraction*, and *multiplication* were used to model real, concrete phenomena.

For example, addition and subtraction would be used to count numbers of villagers, sheep, or bushels of grain. The product, or multiplication of numbers, $a \cdot b$, would serve as the mathematical model of the area inside a rectangle with sides a and b .

Measurement of motion, however, was limited to *averages* over long time intervals — a two-hour journey of 60 cubits, say, defined an *average speed* (not *instantaneous speed*) of 30 cubits per hour.

(24.1.4)

Newer Mathematics: *The Recent Blossoming.* If four hundred years ago can be called *recent*, then mathematics has only recently been able to accurately model *instantaneous* changes in motions of objects. This achievement came almost simultaneously with the development or invention of integral and differential calculus by GOTTFRIED WILHELM LEIBNIZ (1646–1716) and ISAAC NEWTON (1642–1727), which finally allowed for the mathematical representation of speeds and accelerations at arbitrary *instantaneous points* of time, as opposed to *extended intervals* of time. (For more, see *The Triumph of Numbers* by I. B. Cohen [6].)

It was about two centuries after the invention of the calculus, when JAMES CLERK MAXWELL (1831–1879) used these tools to develop a mathematical model—four elegant equations—that united the fields of electricity, magnetism and optics and established the wave property of electromagnetic (*EM*) transmissions. As if that weren’t enough, Maxwell’s equations also predicted the frame-independent speed c (*in a vacuum*) for electromagnetic waves of all frequencies. It would be decades before *visible light* was confirmed to be a special case of Maxwell’s electromagnetic waves. (see *Chapter 21*.)

-
- (24.1.5) **The Scientific Method.** The Scientific Method (see Chapter (F)), a product of the Age of Enlightenment, requires claims of one individual to be experimentally verified and tested by others. One feature of the scientific method that makes it unique in the course of history is that truth is accepted by the persuasion of such testing and *not* by the power of authority or other coercions. Unfortunately, this methodology had not prevailed through the ages. How else can one explain ARISTOTLE'S (384–322 B.C.) apparently unchallenged claim than *men have more teeth than women do?* (See (F.6.2c), pg. 418.)
-

(24.2) Einstein's Annus Mirabilis

- (24.2.1)  **HISTORY** In 1903, WILBUR AND ORVILLE WRIGHT invent the first gasoline-powered and manned airplane. In 1905, during Einstein's miracle year, MARY ANDERSON invents windshield wipers. In this year, COCA COLA, which was marketed as a health tonic, contained caffeine along with some natural COCAINE from the kola nut. (By 1931 the cocaine was removed.)
-

- (24.2.2) **Enter Einstein.** The year is 1905, the so-called ANNUS MIRABILIS, (*miracle year*) for Albert Einstein in which he publishes not one, but *four* landmark papers, all in the journal, *Annalen der Physik*. He is a 26-year old clerk in the Swiss Patent Office in Bern and has just earned his doctoral degree in physics from the Zürich Polytechnic Institute of Physics.
-

- (24.2.3) **The Photoelectric Effect.** In March of 1905, EINSTEIN publishes his first paper. It explains the PHOTOELECTRIC EFFECT, in which a light beam directed at a metallic surface will produce the emission of electrons (electrical current). By positing that light, for all its wave properties, is, at the same time, composed of particles, or individual energy packets (PHOTONS, packets of energy, quanta), Einstein explains why higher intensity (brightness) of the light only produces more electrons (current) but does *not* produce more higher-energy electrons (voltage). This is because a higher energy of photons

corresponds to higher frequency (bluer color) and not to increased brightness, which corresponds to a larger number of photons. For this paper, which lay the foundations of quantum theory, Einstein won the 1921 NOBEL PRIZE in physics.

(24.2.4)

Brownian Motion. In his second paper of 1905, Einstein explains the random motion of small particles in “motionless” liquid solution—for example, pollen in standing water was observed to be in constant motion even though the water was apparently still.² Einstein hypothesizes that motion of the pollen particles results from their being battered by the moving molecules of the liquid. (*Yet, the molecules are too small to be seen by the naked eye.*) By applying statistical techniques to this *Brownian motion*, Einstein calculated the average distance traveled by the pollen under the assumption that this motion is generated by water molecules. His results were in excellent agreement with the observations and provided a very strong argument for the existence of molecules. At this time, prestigious physicists, such as ERNST MACH (1838–1916), would still be opposed to the idea that molecules even exist.

(24.2.5)

Special Relativity. In his third paper of 1905, *On the Electrodynamics of Moving Bodies* [12], which came to be known as his *special relativity* paper, EINSTEIN acknowledges the work of Maxwell³ as is seen from his choice of title and his opening lines (see (23.5.3b)).

The asymmetry Einstein refers to in his opening lines is between the Faraday-Lenz law (21.7.3c) and the Ampère-Maxwell law (21.7.3d), which, somehow, need two separate frames of reference to explain a single phenomenon. Yet, Maxwell’s equations (see Chapter (21)) also showed that the speed of light in a vacuum, which Maxwell calculated from a pair of laboratory measurements (see (23.4)), does *not* depend on any frame of reference.

²Some historians count this work as two papers since it draws on Einstein’s 1905 doctoral thesis in which he used statistical techniques to infer the size of sugar molecules based on the rate at which sugar dissolved in water. The paper on Brownian motion used some of these statistical techniques.

³ Ironically, the Zürich Polytechnic Institute of Physics did not offer courses or seminars in Maxwell’s theory—Einstein had to study it on his own.

Einstein responds to this “asymmetry” in Maxwell’s theory (*see Figure (23.5.3a)*) by assuming from the outset that all laws of physics are equally valid in all frames of reference and that the speed of light is the same for all observers, regardless of the relative speed of source and observer.

Einstein’s simplification eventually leads us, in spite of our dissenting intuition, to the results that moving clocks run slow and moving objects always shrink in the direction of motion.

-
- (24.2.6) **Equivalence of Energy and Mass.** EINSTEIN publishes his fourth paper of 1905 entitled *Does the Inertia of a Body Depend Upon Its Energy Content?* where the relativistic addition of speeds $v \oplus w$ (7.3.1) leads directly to $E = mc^2$, the famous equation of equivalence between rest mass m and energy E . Chapter (15) is devoted to a derivation of this equation.
-

(24.3) Comparing Relativities

- (24.3.1) **How Similar are the Galilean and Lorentzian Transformations?** For speeds v , Gaussian matrices have form $G_v = \begin{bmatrix} 1 & -v \\ 0 & 1 \end{bmatrix}$ (5.4.1a), while Lorentzian matrices take the form $L_v = \gamma_v \begin{bmatrix} 1 & -v \\ -v/c^2 & 1 \end{bmatrix}$ as given in (7.2.2).

Note that if, in the matrix L_v , we let $c \rightarrow \infty$ then the Lorentz matrix L_v approaches the Galilean matrix G_v . Recall that in Galilean physics, the speed of light is infinite while in relativistic physics, the speed of light in a vacuum is the fixed and finite quantity c .

When the speed $c \rightarrow \infty$ in L_v , the four individual entries of the matrix L_v approach the respective four entries of the Galilean matrix G_v .

Although the matrices G_v and L_v approach each other entrywise as $c \rightarrow \infty$, they are far apart from each other structurally in the sense

that G_v always has only *one* linearly independent eigenvector (A.5.1) in \mathbf{R}^2 while L_v always has *two* linearly independent eigenvectors.

(*Eigenvector structure is explored in Exercises (24.9.2) and (24.9.3).*)

Comparing Galilean and Relativistic Speeds

- (24.3.2)** **Addition-of-Speeds formulas.** Given frame \mathcal{F}^A sees frame \mathcal{F}^B moving with speed $-1 < v < 1$ and frame \mathcal{F}^B sees frame \mathcal{F}^C moving with speed $-1 < w < 1$. Then frame \mathcal{F}^A sees frame \mathcal{F}^C moving with relative speed $v + w$ in the Galilean case, and with speed $v \oplus w$ in the Lorentzian case. Symbolically,

$$(24.3.2a) \quad \mathcal{F}^A \xrightarrow{v} \mathcal{F}^B \xrightarrow{w} \mathcal{F}^C \text{ implies } \begin{cases} \mathcal{F}^A \xrightarrow{v+w} \mathcal{F}^C & \text{is Galilean,} \\ \mathcal{F}^A \xrightarrow{v \oplus w} \mathcal{F}^C & \text{is Lorentzian.} \end{cases}$$

The sums, $v + w$ and $v \oplus w$, which are virtually the same for small values, are each derived in two ways—analytically and geometrically.

(i) Analytically: In the classical, Galilean/Newtonian case, the addition of speeds $v + w$ of frame \mathcal{F}^C as seen from frame \mathcal{F}^A is calculated with the Galilean matrix (5.3.4) (*see Theorem (5.5.2), pg. 91*). In the relativistic Lorentzian case, the addition of speeds $v \oplus w$ of frame \mathcal{F}^C as seen from frame \mathcal{F}^A is calculated with the Lorentz matrix (7.1.3a) (*see Theorem (7.4.1), pg. 111*).

(ii) Geometrically: In the classical, Galilean/Newtonian case, we have the following graphic representation of relative speeds v and w in (24.3.2a): There exist certain triangles $\triangle(AB)$ and $\triangle(BC)$ with respective areas r and s (*see Theorem (5.6.2), pg. 93*), so that

$$(24.3.2b) \quad \text{Speed } v = \underbrace{\text{Area } \triangle(AB)}_r, \quad \text{Speed } w = \underbrace{\text{Area } \triangle(BC)}_s.$$

From (24.3.2a) and (24.3.2b), we then deduce

$$(24.3.2c) \quad \text{Speed } \mathcal{F}^A \rightarrow \mathcal{F}^C \text{ is } \underbrace{\text{Area}[\triangle(AB) + \triangle(BC)]}_{r+s} = v + w.$$

Following the wording of (24.3.2b) and (24.3.2c), we show the relativistic property of the addition of speeds.

In the non-classical, Lorentzian case, we have the following graphic representation of relative speeds v and w in (24.3.2a): There exist certain triangle-like regions which we loosely denote as $\Delta(AB)$ and $\Delta(BC)$, with respective areas r and s (see Figure (B.3.4), pg. 381) so that

(24.3.2d)

$$\text{Speed } v = \underbrace{\tanh(\text{Area } \Delta(AB))}_{\tanh(r)} ,$$

$$\text{Speed } w = \underbrace{\tanh(\text{Area } \Delta(BC))}_{\tanh(s)} .$$

From (24.3.2a) and (24.3.2d), we then deduce

(24.3.2e)

$$\text{Speed } \mathcal{F}^A \rightarrow \mathcal{F}^C \text{ is } \underbrace{\tanh(\text{Area } [\Delta(AB) + \Delta(BC)])}_{\tanh(r+s)} = v \oplus w.$$

The constant speed $v \oplus w$ of frame \mathcal{F}^C as seen from frame \mathcal{F}^A is obtained as the hyperbolic tangent of the addition of areas r, s of certain triangle-like regions in the Minkowski spacetime graph. (See Theorem (7.5.6), pg. 115.) That is,

(24.3.2f) speed $v \oplus w = \tanh(r + s)$.

We see in Figure (24.3.3) that for small v and w , $v + w$ and $v \oplus w$ are practically the same. When v and w are small, the T_A and T_B axes (in both Galilean and Lorentzian spacetime graphs) are virtually vertical. Hence, the hyperbola in the right-hand panel is very close to the horizontal line of the left-hand panel and the graphs come closer to agreement.

How Similar are the Area Triangles?

(24.3.2g) When Areas Conflate. Another way to force the right-hand Lorentzian panel of Figure (24.3.3) to closely approximate the left-hand Galilean panel, is to draw the Minkowski diagrams so that the photon speed increases from 1 to ∞ . (*In Galilean/Newtonian physics, light has infinite speed.*) That is, take the photon's 45° worldline (*indicating photon speed $v = 1$*) and rotate it clockwise about the origin until it approaches the horizontal X -axis (*indicating photon speed $v = \infty$*). Since the photon's speed is determined by the inverse of the slope of its worldline, a decreasing worldline slope $\downarrow 0$ implies an increasing photon speed $\uparrow \infty$.

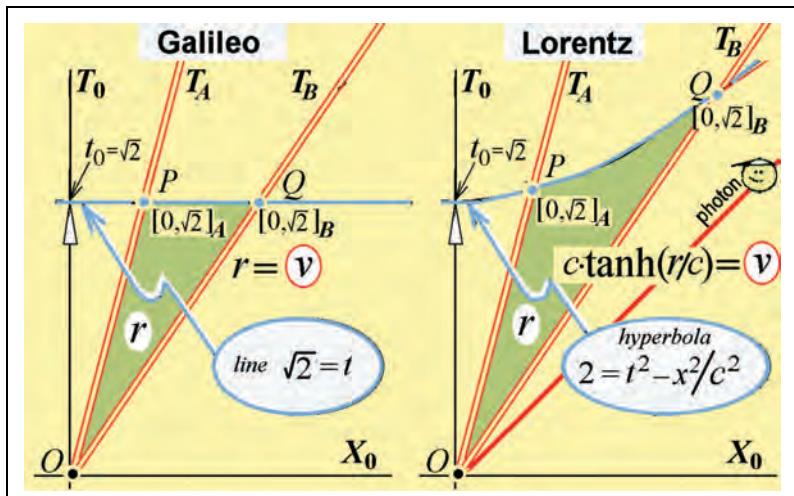
Here are the details: The speed of light is infinite in Galilean physics, but is finite in relativistic physics. This is reflected graphically in the spacetime X - T diagrams as follows:

- In Galilean physics, simultaneous events (fixed value of t) are horizontal lines.
- In relativistic physics, simultaneous events are tangent lines to the hyperbola

$$(t')^2 = t^2 - (x/c)^2 \text{ where } t' \text{ is a fixed time.}$$

(See Theorem (8.4.1), pg. 123 and the diagram in the body of the proof.) In fact, $\pm 1/c$ are the inverses of the slopes of the two asymptotes to this hyperbola.

(24.3.3) **FIGURE:** Comparing Galilean and Lorentzian Spacetime Graphs: $v = \text{the speed of } \mathcal{F}^B \text{ relative to } \mathcal{F}^A$



- LEFT PANEL OF FIGURE (24.3.3): For the classical Galilean case, when the speed of light is essentially infinite, the area $r = v$ of the triangle POQ is exactly equal to the speed v of \mathcal{F}^B as seen from \mathcal{F}^A . This was illustrated in the diagram of Theorem (5.6.2), pg. 93.
- RIGHT PANEL OF FIGURE (24.3.3): In the relativistic Lorentzian case, the area r of the triangular-like region POQ plays a role similar to that in the Galilean case. When the speed of light for all observers is c , then the speed of \mathcal{F}^B as seen from \mathcal{F}^A is $v = c(\tanh(r/c))$. Now when r/c is tiny, then $\tanh(r/c) \approx r/c$. Hence, when v is tiny, so is area r , in which case $v = c(\tanh(r/c)) \approx c(r/c) = r$, which is the Galilean case of the left panel.

(24.4) Against Conventional Wisdom

Special relativity tells us that each of us carries our own personal clock which, from each individual's point of view, never runs slower than the personal clock of any other individual.

In spite of this truth—that clocks in motion run slower—the difference in the ticking rates of the clocks is so tiny that we never perceive this difference in our day-to-day experience.

If clock rates vary and the variances are not apparent, then it is easy to behave as if clock rates do *not* vary. This is similar to saying that since the earth's curvature is not apparent (*is a lake flat or does it bulge in the middle?*) then it is easy to behave in day-to-day activities as if the earth is flat.

Hence, we believe the opposite of the relativistic truth. Experience tells us it is “obvious” that all clocks, regardless of their motion, run at the *same* rate.

(24.4.1)

This Galilean (*non-relativistic*) interpretation of physics says it is possible to construct a single theoretical reference clock by which we may all synchronize our personal watches, regardless of our relative speeds.

Here are two examples:

(24.4.2)

I. A Successful Universal Clock: *Measuring Longitude*. It was JOHN HARRISON (1693–1776) who responded to the challenge announced in 1714 by Isaac Newton, in the name of the *Royal Society of London for the Promotion of Natural Knowledge*, to develop an accurate means for ships at sea to determine their east-west distance (longitude) from England. A prize of £20,000 was offered. While other scientists thought that certain astronomical observations might lead to a solution, Harrison decided to build a *clock* of extreme accuracy that would withstand the continual lurchings of a ship at sea. (*Every 15° interval of east-west longitude corresponds to one hour’s difference between noon at each end. See Exercise (24.9.5).*)

After almost half a century of trials, Harrison’s fourth generation clock, called H4, was tested in 1764 at sea and successfully met the conditions of the Royal Society’s challenge. In spite of Harrison’s documented success, influential scientists of the Royal Society refused to authorize the award to a lower-class “tinkerer.” Only after the intervention of KING GEORGE III in 1772 did Harrison finally receive his £20,000 award. See the book *Longitude* by Dava Sobel [31] for a nice overview of these events.

In sum, the success of Harrison’s clock confirmed that it is useful and even necessary to adopt the Galilean view of time (24.4.1) where arbitrary observers can synchronize their clocks to a single and *accurate* universal clock.

(24.4.3)

II. A Successful Use of Universal Time: Train Schedules and Time Zones. Another example of effective use of the Galilean view of time (24.4.1) which re-enforced conventional wisdom that all clocks run at the same rate, is seen in the history of railroad timetables, all of which are, indeed, synchronized to a single, theoretical clock. In addition, TIME ZONES were created for the very purpose of facilitating the synchronization of railroad schedules over wide regions to the single theoretical Galilean reference clock.

Less than 200 years ago, there were no time zones. Each locality determined noon to be that time of day when the sun was at its highest point in the sky. The sun dial, with noon being marked by the shortest shadow, was an adequate timepiece—as long as coordination was not necessary with sundials of other communities.

In England, The Great Western Railway, as part of an industry-wide attempt to counteract the inconsistencies of local times, started to run on fixed London time in November of 1840. Following suit, most public clocks were on the same London time by 1855.

In the United States and Canada, it was 1883 when railroads created synchronized, regional time zones. However, the general population had not yet warmed to the notion. Local pride was a factor causing Detroit to maintain two or more time standards until 1905 (the year of Einstein's paper [12]) when they finally adopted Central Standard Time.

(24.4.4)

These examples show how EINSTEIN'S thinking in 1905, where any observer sees moving clocks running slower, went against conventional wisdom and “common sense” which held that any observer sees all clocks running at the same rate.

(24.5) Some Experimental Results

(24.5.1)

Earliest Experiments. We have already encountered experiments whose results are consistent with the theory of special relativity.

For example, light was first confirmed to have finite speed with the experiment of Rømer who inferred this property from observations of Io, a moon of Jupiter (17.2.4). (*In Galilean physics, the speed of light is infinite.*) This finite speed was measured, to reasonable accuracy, by Fizeau via his ingenious cogs-and-mirrors experiment (17.3.2).

Later, de Sitter showed that the speed of light did not depend on the speed of its source when he tracked the light from binary stars (*Figures (17.4.1) and (17.4.3)*). This is consistent with our skipping stone model of wave transmission which predicts the observed wave speed is independent of the speed of its source. (*see the Skipping Stone model (16.5) and Figure (16.5.4)*).

On the other hand, the speed of the transmitting medium *is* communicated to (added to) the speed of the transmitted wave (16.5.5). This is why the “failure” of the Michelson-Morley experiment (17.5.4) eventually convinced scientists that light, unlike any other known wave, did *not* need a transmitting medium or aether (17.1.2).

Finally, using data from laboratory experiments (*Chapters (19)–(23)*), Maxwell calculated c , the theoretical speed of an electromagnetic wave in a vacuum. Moreover, the mathematics did not require a transmitting medium for electromagnetic waves (*although Maxwell thought there must be one*) nor did Maxwell’s model require a particular frame of reference for these waves (23.4).

Confirming the Theory of Special Relativity

- (24.5.2) **Can We Observe Moving Clocks Running Slow?** A MU MESON, or MUON for short, is an elementary particle with a negative charge and a mass about 20 times that of the electron. Statistically, half of any group of stationary muons will decay after about 3 microseconds (*one microsecond is 10^{-6} , a millionth, of a second*). In any muon population, some muons live more than 3 microseconds and others live less but on average, the life span is

$$\text{life}_{\text{stationary}} = 3 \text{ microseconds.}$$

In nature, high-velocity muons with speeds v_{moving} close to the speed of light are produced when cosmic rays—subatomic charged particles that originate in outer space—hit Earth’s atmosphere. Experi-

tal observation shows that the average life span of these high-speed communities of muons is

$$\text{life}_{\text{moving}} = (\gamma_v) \cdot \text{life}_{\text{stationary}},$$

where $\gamma_v = 1/\sqrt{1 - (v_{\text{moving}}/c)^2}$ is the familiar Lorentzian factor (see (1.8.2) and (7.1.3b)) and v_{moving} is the observed muon speed.

As predicted by special relativity, the clock in the moving frame of the streaming muons runs slower than Earth clocks by exactly a factor of $\sigma_v = \sqrt{1 - (v_{\text{moving}}/c)^2} < 1$ (recall (1.8.4)).

- (24.5.3) ***GPS Satellites and Relativity.*** From the dozen or so GPS (GLOBAL POSITIONING SYSTEM) satellites circling Earth, accurate calculations of Earth positions are based on the accuracy of the onboard atomic clocks.

Special relativity tells us that all moving clocks run slow. Hence onboard clocks of GPS satellites, which are in motion relative to Earth, will run slower than Earth clocks. The slower rate is about 7 microseconds each day.

General relativity says that nearness to mass will also cause clocks to run slow. In particular, the tug of Earth's gravitational field slows the onboard GPS clocks. But when the GPS clocks are placed in orbit, further from Earth, the gravitational effect is reduced. This causes the GPS clocks to run *faster* than they did when assembled on Earth. The faster rate is about 45 microseconds each day.

As a result, onboard GPS atomic clocks, when measured on Earth, are designed to run 38 microseconds slower each day than accurate Earth clocks. Without this adjustment, accumulated clock errors would cause GPS calculations of global positions to be off by more than 10 kilometers (6 miles) each day.

- (24.5.4) **The formula $v \oplus w$ for the relativistic addition of speeds v and w ,** was closely approximated by measurements from the experiment of Fizeau (see Section (13.2)).
-

- (24.5.5) **NOTE:** For a comprehensive overview of experimental results that confirm properties of special relativity, see the website of John Baez at
www.math.ucr.edu/home/baez/physics/Relativity/SR/experiments.html.

(24.6) Bad Assumption, Good Result

- (24.6.1) **A remarkable dispute.** In a conversation with DAVID ELLERMAN, it was noted that the book, *The Best of All Possible Worlds* pp. 48–50 [14] by IVAR EKELAND describes the disagreement between RENÉ DESCARTES (1596–1650) and PIERRE FERMAT (1601–1665) over the following:

(24.6.1a) **Question:** Does light travel faster through air than it does through water?

Descartes answered this question in the negative, maintaining that the speed of light *increases* as it passes from air to water, while Fermat believed the opposite—that the speed of light *decreases* after entering water from air. Ironically, with opposite and contradictory hypotheses, these two men derived the valid result known as **Snell's Law**, one form of which is:

(24.6.2) **SNELL'S LAW.** A ray of light passes through air at speed v_1 with ANGLE OF INCIDENCE θ_1 (see diagram on the right). The light ray then passes through water at speed v_2 and ANGLE OF REFRACTION θ_2 . SNELL'S LAW states that there exists a constant m , (the REFRACTIVE INDEX or INDEX OF REFRACTION), that is independent of the angle θ_1 such that the following relation holds true:

$$(24.6.2a) \quad \sin \theta_1 = m \cdot \sin \theta_2.$$

Although the Dutchman WILLEBRORD SNELL discovered the relationship (24.6.2a) in 1620, he did not prove it. Descartes was the first to actually derive it.

How did Descartes do it? He assumed (*incorrectly*) that once the light ray entered water, its vertical speed component increased while the horizontal speed remained unchanged.

Fermat, on the other hand, assumed (*correctly*) that in the Figure of (24.6.2) with fixed points a and c , light would travel in the least time over all a - b - c paths as the position of entry point b varied horizontally.

Both Descartes and Fermat arrived at the same form (24.6.2a) of Snell's Law and both would agree that, in the case of air-to-water travel, the value of the refractive index⁴ of (24.6.2a) should be $m \approx 1.33$. But

$$(24.6.2b) \quad m = \begin{cases} (v_2/v_1) & \text{according to Descartes,} \\ (v_1/v_2) & \text{according to Fermat.} \end{cases}$$

Since the speed of light would not be known for some two hundred years⁵, it was impossible at this time to decide which of the two ratios of (24.6.2b) was valid. (Note: Exercise (24.9.7), will guide you through the derivations of (24.6.2a) and (24.6.2b).)

(24.7) A Limited Reality

- (24.7.1) **Reality as Source + Perceived Effect.** It is interesting to note how, in one form or another, philosophers through the ages have addressed reality in terms of *effect* and *cause*, or *perception* and *source* of the perception.

For example, PLATO (427–347 B.C.) likened our reality to that of beings imprisoned in a cave. From birth, these individuals can only perceive the shadows and sounds (*effects*) of the true and real world outside the cave (*source*). Anyone managing to find his or her way out of the cave would be met by the blinding richness of the outside reality.

Plato goes further with this metaphor and introduces a moral component—the enlightened escapee is obliged to return to the cave in order to inform the inhabitants about the outside reality.

⁴ Index of refraction for various materials is discussed in (13.2.2c).

⁵ In 1849, Fizeau devised a mechanism for measuring c in air. (See (17.3), pg. 221). In 1851, Fizeau measured the speed of light in water (see (13.2), pg. 164).

What did SOCRATES⁶ (469–399 B.C.) have to say about this moral obligation? In his work *The Republic, Book VII* [28], Plato devised an imaginary dialogue between his teacher, Socrates, and a character named Glaucon. During this discourse, Socrates—not the most optimistic person in the world—says,

Men would say of him that [ascended up to the light] . . . that it was better not even to think of ascending; and if any one tried to . . . lead [any of us] up to the light, let them only catch the offender, and . . . put him to death.

DAVID HUME (1711–1777) claimed that we determine our reality from our collection of individual perceptions (*effects*) resulting from a much larger reality (*source*), just as we perceive individual railroad cars (*effects*) of a larger train (*source*). Hume was not so sure that the full train of reality could be accurately known from the separate cars alone.

IMMANUEL KANT (1724–1804) was so positive about the reality of objects that he invented the new (*clarifying?*) term, “noumenal.” To say objects are NOUMENAL is to say they are real—outside and independent of the mind, as opposed to a phenomenon (*perceived effect*), also called a “thing-in-itself.”

- (24.7.2) **Beyond Dimension.** EDWIN A. ABBOTT (1838–1926), in his book *Flatland* [1], not only endorses the existence of an external reality, but goes further by suggesting that much, if not most, of true reality is *beyond perception*. In Abbott’s book, the universe of the Flatlander is a table-top land where there are only two space dimensions. Inhabitants of this two-dimensional table-top universe have no conception of up or down. They only know left-right and backward-forward dimensions.

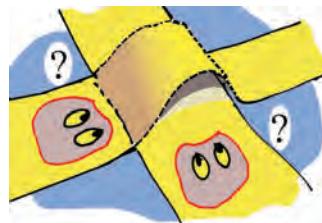
⁶There are no known writings due to Socrates. It is mostly through the writings of his student, Plato, that we know of his ideas.

(24.7.2a) In Flatland, a creature from the third dimension—a Roundlander, say—can intervene and produce results that seem downright magical to the Flatlanders. For example, when a Roundlander pokes a finger into Flatland space, as a horizontal line,



it is perceived as is every other object in Flatland. But when the finger is pulled perpendicularly upward—a direction unknown to the Flatlanders—the Flatlanders see the finger-line mysteriously shrink and then disappear into thin (*flat?*) air!

(24.7.2b) The third space dimension is as confounding to the Flatlanders as a fourth space dimension is to us. Suppose two Flatlanders traveling along separate perpendicular roads meet at the intersection. Like two cars at a four-way stop, one Flatlander must wait for the other to pass before resuming its (his/her?) own journey. Now suppose a voice from the third dimension proclaims, “Build a bridge so that no traveler has to wait for the other!” The awe-struck Flatlanders listening to this disembodied voice would have no idea of what a bridge would be in three-space, nor how it could possibly help them to avoid one traveler having to wait for the other. (*Although its practicality may be in doubt, Exercise (24.9.6) asks you to mathematically construct a bridge in four-dimensional space so that two objects at the same point (x, y, z) in \mathbf{R}^3 will be separated in \mathbf{R}^4 .*)



(24.7.2c) Our tunnel vision of reality. Abbott’s two-dimensional Flatlanders are limited in the sense that they are unaware of and cannot perceive our richer three-dimensional reality. But we are also limited since much, if not all, of what we believe to be real, lies beyond what we perceive with our five input ports (senses).

For example, we do not perceive radio waves until a radio receiver transforms them to sound. Similarly, a Geiger counter converts radioactivity (*the spontaneous emission of invisible matter or energy from the nucleus of an unstable atom*) into sound or a visible graph.

(24.7.2d) Gadgets expand our reality in the sense that instruments convert (or reduce) extra-sensory phenomena to sights, sounds, smells, tastes, or sensations that we then interpret as (effects of) “reality.”

We saw this, for example, in the experiments of Rutherford (F.12.4), pg. 429. Invisible (*to humans, at least*) alpha particles ricochet off invisible atoms of gold and hit a screen whose coating transforms each collision into a visible blip. After Rutherford applied his *intellect* and *imagination*, we expanded reality.

In another case, experimenters applied their intellect and imagination to the readings of certain physical instruments so that today we know that c , the speed of light in a vacuum, is invariant for all observers. (1.5.2b). In a remarkable *tour de force*, EINSTEIN took this single observation about c and applied *his* intellect and imagination in a grand thought experiment to produce the theory of special relativity.

(24.7.2e) We are butterflies of the Universe. The butterfly, with a life-span of only a few weeks or months, is ignorant of our realities⁷ — of mathematics, music, humor, emotion. On the other hand, as clever as we may believe ourselves to be, we suffer the anguish of acknowledging the incompatibility of three successful theories of physical reality: (i) special and (ii) general relativity (*for large scale physics*), and (iii) quantum theory (*for small scale physics*). Special and general relativity both assume continuous motion and deterministic causality



⁷For all its limitations, the Monarch butterfly does have abilities that exceed our own. Apart from knowing how to fly, every last one of them can navigate their individual migration route over the thousands of miles between North America and Mexico.

describes discontinuous motion while quantum theory that is influenced by laws of probability that are inconsistent with deterministic causality.

As long as our limitations bar us from seeing the larger reality in which all these theories are mutually consistent, we remain butterflies of the Universe.

(24.8) PIES Reality

- (24.8.1) **A Useful Oversimplification.** Reality is comprehended in stages: (i) *Physical* perception is massaged or processed by our (ii) *intellect*. The (iii) *emotional* pleasure of discovery impels further exploration after which, we may seek (iv) the *meaning* of it all.

(24.8.2) **TABLE: Four Facets of Reality**

P)physical	<i>the physical world.</i>
I)intellect	<i>thought, imagination, and logic.</i>
E)motion	<i>sensations and feelings.</i>
S)piritual	<i>the structure that provides meaning.</i>

Interrelations. These four suggested facets of Reality are not independent of each other. For example, emotions can affect our physical state and poor physical health can affect our thinking abilities.

(24.8.3) **EXAMPLES: Emotion versus Intellect.**

- Our intellect may accept the notion that moving clocks run slow, but is this idea emotionally acceptable?
- Our engineers may understand the aerodynamics of flight, but some of them might still feel uncomfortable, emotionally, with the fact that jumbo jets, weighing several tons, are being held aloft only by thin air.
- One may feel emotionally distressed walking under a ladder even though the intellect sees no cause for concern.

(24.8.4) **Many spiritualities** can be defined since there are many beliefs about our Meaning or Reason-for-Being in the Universe. Meaning will be interpreted by some, as non-existent—it is only the physical

world that truly exists. Meaning, for others, will be understood by recognizing an external force or entity, and this entity may be highly impersonal (Nature) or highly personal (God).

(24.8.5) Do scientists share a common view of Reality and Meaning?

Suffice it to quote the following interesting statements of the eminent scholars and physicists, Steven Weinberg and Freeman Dyson:

Steven Weinberg [32] proclaimed

With or without RELIGION, you would have good people doing good things and evil people doing evil things. But for good people to do evil things, that takes religion.

But Freeman Dyson [10] added

Weinberg's statement is true as far as it goes, but it is not the whole truth. To make it the whole truth, we must add an additional clause: "And for bad people to do good things—that takes religion."

(24.8.6) How Much Uncertainty Will a Scientist Tolerate? Given that scientists studying the physical universe attempt to minimize *uncertainty* (see Section (F.10)), one might assume that scientists must not tolerate uncertainty in any other discipline. Consider, then, these comments of Freeman Dyson [11]:

Both as a scientist and as a religious person, I am accustomed to living with uncertainty. Science is exciting because it is full of unsolved mysteries, and religion is exciting for the same reason. The greatest unsolved mysteries are the mysteries of our existence as conscious beings in a small corner of a vast universe. Why are we here? Does the universe have a purpose? Whence comes our knowledge of good and evil? These mysteries, and a hundred others like them, are, as yet, beyond the reach of science.

(24.8.7) Finally, this book has attempted an explanation of the processes by which we may come to understand a very small part of our Reality—the physical part. This is the beginning.

(24.9) Exercises

(24.9.1) Using Definition (A.5.1), show the following properties:

Exercise

(24.9.1a) If \vec{X} is an eigenvector for a matrix A , then so is any scalar multiple, $a\vec{X}$ for any scalar $a \neq 0$.

(24.9.1b) Suppose a matrix A has two linearly independent eigenvectors, \vec{X}_1 and \vec{X}_2 with the *same* eigenvalue, λ . Then show that the sum $\vec{X}_1 + \vec{X}_2$ is also an eigenvector with the same eigenvalue, λ .

(24.9.1c) Suppose a matrix A has two linearly independent eigenvectors, \vec{X}_1 and \vec{X}_2 with the *same* eigenvalue, λ . Then show that the linear combination $a_1\vec{X}_1 + a_2\vec{X}_2$, for scalars a_1 and a_2 , is also an eigenvector with the same eigenvalue λ .

Hint: Instead of showing this result from scratch, use the results of (24.9.1a) and (24.9.1b).

(24.9.2) Using Definition (A.5.1), find eigenvectors and eigenvalues for the following matrices:

(24.9.2a) Show that for the $n \times n$ identity matrix, I_n , *any and every* vector $\vec{X} \neq \vec{0}$ is an eigenvector. Moreover, the eigenvalue is always $\lambda = 1$.

(24.9.2b) Show that for any 2×2 triangular matrix $A_2 = \begin{bmatrix} a & b \\ 0 & c \end{bmatrix}$

with real entries, a , b , and c , the vector $\vec{X}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and all its multiples $\lambda\vec{X}_0$ is the only set of eigenvectors of A_2 . Moreover, a is the eigenvalue associated with \vec{X}_0 .

(24.9.2c) Given the Galilean matrix $G_v = \begin{bmatrix} 1 & -v \\ 0 & 1 \end{bmatrix}$. Then show

that the vector $\vec{X}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and all its multiples $\lambda\vec{X}_0$ is the only set of eigenvectors of G_v . Moreover, 1 is the eigenvalue associated with \vec{X}_0 .

Hint: Use (24.9.2b).

(24.9.3) Given the Lorentz matrix $L_v = \gamma_v \begin{bmatrix} 1 & -v \\ -v & 1 \end{bmatrix}$ (7.2.2).

(24.9.3a) Using Definition (A.5.1), find eigenvectors and eigenvalues of L_v .

Hint: From Definition (A.5.1), obtain two equations in the variables x , t , v , and eigenvalue λ . Put all terms on one side so that both equations equal zero, then show that the two eigenvalues are $\lambda = (1 \pm v)\gamma_v$. Substitute each of these eigenvalues back into the equations to obtain eigenvectors $\{\vec{X}_0, \vec{X}_1\}$, which will be mutually perpendicular (see (A.2.6b), pg. 362). Note from (24.9.1a), that once you find an eigenvector \vec{X}_i for λ , any non-zero multiple $a\vec{X}_i$ is an equivalent eigenvector for the same λ .

(24.9.4)
Exercise

Show that the geometric assertions of (24.3.2g) are valid. That is, show that the rotation of the 45° worldline of the photon forces the Lorentzian triangle-like region of the right-hand panel to approach the Galilean triangle of the left-hand panel.

(24.9.5)
Exercise

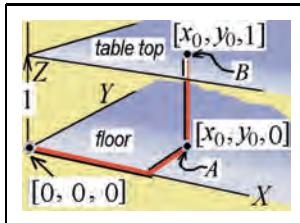
Show how an accurate clock allows for accurate calculation of longitude for ships at sea (see (24.4.2)).

Hint: An explorer carries a clock that accurately reads the time of city A where noon (sun at its maximum height) is observed at 12:00pm. If noon for the remote explorer occurs when her clock reads 11:00am, or one hour previous to noon in city A, then she knows she is exactly 15° longitude west of city A. If she sees noon at time 10:30am, or 1.5 hours previous to noon at city A, then she is $1.5 \times 15^\circ$ longitude west of city A. In general, if the explorer sees local noon occurring n hours before (after) 12:00pm, then her position is $n \times 15^\circ$ longitude west (east) of city A.

(24.9.6)
Exercise

Show how persons A and B who are approaching a collision at point $[x_0, y_0, z_0]$ in three-dimensional space \mathbf{R}^3 , can be saved when one of them veers onto a bridge in four-dimensional space \mathbf{R}^4 .

Hint: To reinforce intuition, let us see, analogously, how collision is avoided in a lower dimension. Accordingly, suppose A and B are headed for a collision at point $[x_0, y_0]$ in \mathbf{R}^2 . If we embed \mathbf{R}^2 into \mathbf{R}^3 —see the figure to the right—then A and B can each be put into different planes.



The collision point $[x_0, y_0]$ in \mathbf{R}^2 corresponds to two distinct coordinates in \mathbf{R}^3 , namely, $[x_0, y_0, 0]$ for A on the “floor” of \mathbf{R}^3 and $[x_0, y_0, 1]$ for B on the “tabletop” of \mathbf{R}^3 . Similarly, formally embed \mathbf{R}^3 into \mathbf{R}^4 —without the benefit of pictures (1.3.3)—so that the would-be collision point for A has new coordinates $[x_0, y_0, z_0, 0]$ on the “floor” of \mathbf{R}^4 and the collision point for B has new coordinates $[x_0, y_0, z_0, 1]$ on the “tabletop” of \mathbf{R}^4 .

(24.9.7)

Exercise
Calculus
needed

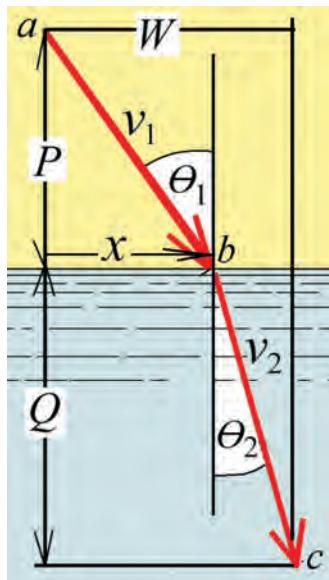
Two Derivations of Snell's Law (24.6.2a):

(24.9.7a) Derive (24.6.2a) and (24.6.2b) using Descartes' (false) hypotheses, namely, that the speed of light increases after it passes from air to water.

Hint: Equate the horizontal components of v_1 and v_2 in the Figure of (24.6.2) and the result is immediate.

(24.9.7b) Derive (24.6.2a) and (24.6.2b) using Fermat's (correct) hypothesis, namely, that in the figure to the right, light travels along path $a-b-c$ in the least time.

Hint: Write the times that light requires to travel along paths $a-b$ and $b-c$ as a function of x (see figure to the right). Find the x value that produces the minimum.



X. Appendices

Chapter A: Basics of Linear Algebra: . An overview of vector spaces, linear transformations and their matrices, eigenvectors, and eigenvalues. Linear algebraic structure plus a graphical interpretation of the Minkowski diagram lead directly to a derivation of the Lorentz transformation.

Chapter B: The hyperbolic geometry behind Minkowski diagrams is the key to explaining the paradoxes and the non-intuitive relativistic addition ($v \oplus w$) of speeds v and w .

Chapter C: Deconstruction of a Moving Train: What might be regarded as a single unit at a single moment (e.g., a train) is analyzed as an infinite sequence of points, each of which has its own clock or time. This is how we can know that the front end of a moving train is younger than its back end — albeit by an imperceptible amount of time.



(A) Linear Algebra Overview

(A.1) Mathematics as a Conduit to Reality

- (A.1.1) This book is for the reader who is familiar with calculus and matrices (linear algebra). Little or no physics is required.

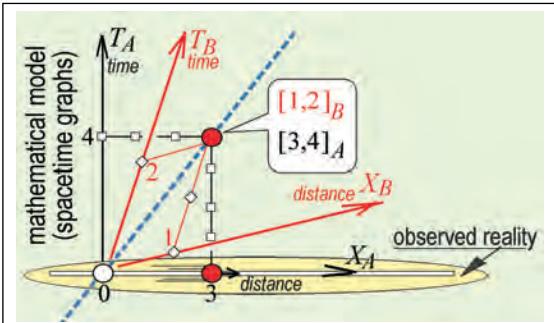
Sections (A.1.2)-(A.1.4) present a brief overview of the link between linear algebra and perceived reality. We first present a graph of a moving particle where *two* sets of (distance vs. time) axes are used:

- (A.1.2) **FIGURE:** X_A - Y_A and X_B - Y_B axes used to plot a point (dotted line) moving horizontally.

The horizontal oval shows the red dot moving horizontally as perceived in “observed reality.”

The 2-dimensional graph, the mathematical model of this reality, shows the moving dot

where the X_A - T_A coordinates are $[3, 4]_A$ (where distance = 3 units and time = 4 seconds) and the X_B - T_B coordinates of the same point are $[1, 2]_B$ (where distance = one units and time = 2 seconds).



- (A.1.3) **Linear Algebra: A Natural Language for Special Relativity** The following overview uses terminology and definitions of linear algebra whose locations in this text are indicated.

(A.1.3a) Constant Speed of Parallel Rulers. Special relativity, broadly speaking, is the study of Nature’s laws using identical rulers that move in parallel tracks at constant speeds relative to each other. Identical clocks are found at each point of each ruler.

(A.1.3b) Motion as Straight Lines. When the trajectory of a point on one of the horizontally-moving rulers is plotted on a

two-dimensional distance-versus-time graph (an X - T *spacetime* graph), the result is a straight line. This is due to the constant speed of the ruler. (See the dotted line in Figure (A.1.2) that we will later discuss in detail.)

(A.1.3c) Inherent Vector Space Structure and Matrices. Every X - T spacetime graph is actually a *vector space* (A.2.1). So when we pass from one set of coordinates to another—from $[3, 4]_A$ to $[1, 2]_B$ in Figure (A.1.2), for example—then we are talking about a *function* (A.3.2) on a *vector space* (A.2.1) that sends zero to zero and sends straight lines to straight lines. The only functions that do this are so-called *linear functions* (A.4.1) which can always be represented by *matrices* (A.4.3), (A.4.6). These matrices are the *Lorentz transformations* (7.1.3) that we discuss in detail later.

(A.1.4) **In sum**, constant motion is graphically represented by straight lines in spacetime graphs. Thanks to the vector space structure of spacetime graphs, we may pass from one coordinate system to another (from one ruler to another) using the Lorentz transformation (or Lorentz matrix) from which we gain insight into all the paradoxes of special relativity.

(A.2) Vector Spaces

To fix ideas, we present some mathematical concepts that are used in this book.

(A.2.1)

DEFINITION (informal): A VECTOR SPACE is any set of “objects” called VECTORS that can be added together (*if \vec{X} and \vec{Y} are vectors, then so is $\vec{X} + \vec{Y}$*) and multiplied by a number (*if \vec{X} is a vector and a is a number, then $a\vec{X}$ is also a vector*).

For example, the “objects” (or vectors) of Definition (A.2.1) might be 2×3 MATRICES (2×3 arrays of numbers) or FUNCTIONS (see Definition (A.3.2)) such as the set of all POLYNOMIALS of the form $a_2x^2 + a_1x + a_0$ on the real line.

(A.2.2)

DEFINITION: In the special case that the “objects” or “vectors” of (A.2.1) are lists of ordered k -tuples of real or complex numbers, then the numbers in the ordered list are called COORDINATES or COMPONENTS. These k -tuple vectors belong to a k -dimensional space which we denote \mathbf{R}^k .

(A.2.3)

NOTATION: The k -tuples of Definition (A.2.2) are written in one of two forms, depending on context, namely as a row vector

$\vec{X} = [x_1, x_2, \dots, x_k]$ or, as a column vector,

$$\vec{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}, \text{ which is also written } \text{col}[x_1, x_2, \dots, x_k].$$

(A.2.4) A SCALAR is a real or complex number.

The addition of *pairs* of vectors and multiplication of a *single* vector by a scalar are defined as follows:

For all scalars (real or complex numbers) a , and all vectors

$$(A.2.4a) \quad \vec{X} = [x_1, x_2, \dots, x_k], \quad \vec{Y} = [y_1, y_2, \dots, y_k].$$

we define the following vector operations,

VECTOR ADDITION

$$(A.2.4b) \quad [x_1, x_2, \dots, x_k] + [y_1, y_2, \dots, y_k] \stackrel{\text{def}}{=} [x_1 + y_1, x_2 + y_2, \dots, x_k + y_k] \quad \begin{array}{l} \text{entrywise} \\ \text{addition} \end{array}$$

SCALAR MULTIPLICATION

$$(A.2.4c) \quad a[x_1, x_2, \dots, x_k] \stackrel{\text{def}}{=} [ax_1, ax_2, \dots, ax_k]. \quad \begin{array}{l} \text{entrywise} \\ \text{multiplication} \end{array}$$

For more on (A.2.4c), see Exercise (3.5.1).

(A.2.5)

NOTE: Strictly speaking, there are two distinct plus signs in (A.2.4b). The boldface $+$ sign is used between pairs of k -tuples while the $+$ sign only appears between pairs of numbers. This distinction is often ignored where the same plus sign is used for both binary (E.1.1) operations.

When x_1, x_2 , and x_3 are space coordinates and $x_4 = t$ represents time (hence, all scalars must be real), then the graphs of these 4-dimensional vector spaces are *spacetime diagrams* that we shall develop in (4.1.2).

(A.2.6)

The INNER PRODUCT of two vectors, \vec{X}, \vec{Y} of the form (A.2.4a), is defined to be

(A.2.6a) $\langle \vec{X}, \vec{Y} \rangle \stackrel{\text{def}}{=} \sum_{i=1}^k x_i \bar{y}_i$ where for each variable y_i , the symbol \bar{y}_i denotes its COMPLEX CONJUGATE (A.2.13a). Vectors $\vec{X}, \vec{Y} \in \mathbf{R}^k$ are said to be PERPENDICULAR for real coordinates, and ORTHOGONAL for complex number coordinates if

$$(A.2.6b) \quad \langle \vec{X}, \vec{Y} \rangle = 0.$$



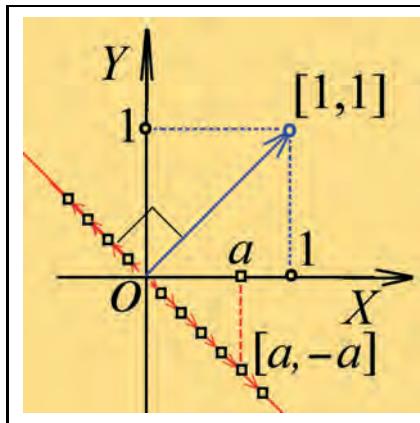
(A.2.6c) **Geometry in terms of formulas.** Note how the geometric quantity of *perpendicularity* (or 90° between pairs of vectors) is expressed in terms of the *analytic expression* (A.2.6a). That is, \vec{X} is geometrically perpendicular to \vec{Y} if and only if the expression $\langle \vec{X}, \vec{Y} \rangle$ is numerically equal to zero (A.2.6b).

(A.2.6d) **EXAMPLE: Illustrating Definition (A.2.6b)**

Definition (A.2.6b) says that vectors $[1, 1]$ and $[a, b]$ are geometrically perpendicular to each other if and only if analytically we have

$$\langle [1, 1], [a, b] \rangle = 0.$$

Definition (A.2.6a) implies that this inner product equals zero if and only if $b = -a$, for any and all a . This defines an infinity of vectors $[a, -a]$ (including $\mathbf{0} = [0, 0]$) each of which is perpendicular to $[1, 1]$.



Moreover, with (A.2.6b) we can extend the notion of perpendicularity beyond two or three dimensions. For example, even if $k > 3$, then for vectors $\vec{X}, \vec{Y} \in \mathbf{R}^k$ (A.2.3) we have a meaningful definition for \vec{X} being perpendicular to \vec{Y} even though we do not have pictures (e.g., arrows) to represent these vectors.

- (A.2.7) Any k -term sum, $r_1 + r_2 + \dots + r_k$, can be represented with the **SIGMA NOTATION** which can take any of the following forms: $\sum_{i=1}^{i=n} r_i$, $\sum_{i=1}^n r_i$, $\sum_i^n r_i$, or $\sum_i r_i$.
- (A.2.8) The **LINEAR COMBINATION** of vectors $\{\vec{X}_i\}$ in \mathbf{R}^k is any sum of the form $\sum_i a_i \vec{X}_i$ where each a_i is a scalar.
- (A.2.9) A **BASIS** for a k -dimensional vector space \mathbf{R}^k is a fixed set of k non-zero vectors $\{\vec{X}_i\}$ from which any vector \vec{X} in \mathbf{R}^k can be expressed in only one way as a linear combination
- (A.2.9a)
$$\vec{X} = \sum_i^k a_i \vec{X}_i \quad \text{for } k \text{ unique scalars } \{a_i : i = 1, 2, \dots, k\}.$$
- (A.2.10) A set of vectors $\{\vec{X}_i\}$ in \mathbf{R}^k is **LINEARLY INDEPENDENT** if there is only one way to write the zero vector as a linear combination of the vectors $\{\vec{X}_i\}$, namely, by setting each $a_i = 0$ in (A.2.9). That is,
- (A.2.10a)
$$\vec{0} = \sum_i a_i \vec{X}_i \quad \text{only if each } a_i = 0.$$

Equivalently, vectors $\{\vec{X}_i\}$ are **LINEARLY INDEPENDENT** if for any vector \vec{Y} ,

$$(\text{A.2.10b}) \quad \vec{Y} = \sum_i a_i \vec{X}_i = \sum_i b_i \vec{X}_i \quad \text{only if each } a_i = b_i.$$

That is, for any \vec{Y} , there is only one way to write \vec{Y} as a linear combination of the vectors $\{\vec{X}_i\}$.

- (A.2.11) A **STANDARD BASIS** for a vector space \mathbf{R}^k is the particular set of k vectors, $\{\vec{E}_1, \vec{E}_2, \dots, \vec{E}_k\}$,

$$(\text{A.2.11a}) \quad \underbrace{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\vec{E}_1} \quad \underbrace{\begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}}_{\vec{E}_2} \quad \cdots \quad \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}}_{\vec{E}_k}.$$

- (A.2.12) **EXAMPLE:** To illustrate (A.2.9a) when the basis is the standard basis, any vector \vec{X} in \mathbf{R}^k has the form

$$\vec{X} = \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}}_{\vec{E}_1} = x_1 \underbrace{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\vec{E}_2} + x_2 \underbrace{\begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}}_{\vec{E}_2} + \cdots + x_k \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}}_{\vec{E}_k}.$$

- (A.2.13) The Euclidean **LENGTH** or **NORM** of a k -vector of form $\vec{X} = [x_1, x_2, \dots, x_k]$ (or $\text{COL}[x_1, x_2, \dots, x_k]$) is defined by

$$(\text{A.2.13a}) \quad \underbrace{X = \|\vec{X}\|}_{\substack{\text{equivalent} \\ \text{notation}}} \stackrel{\text{def}}{=} \sqrt{x_1 \bar{x}_1 + x_2 \bar{x}_2 + \cdots + x_k \bar{x}_k} \geq 0$$

where \bar{x}_i is the **COMPLEX CONJUGATE** of x_i . That is, if for real numbers a_i and b_i we have $x_i = a_i + ib_i$, then its conjugate is $\bar{x}_i = a_i - ib_i$. Hence, for any scalar α , $\|\alpha \vec{X}\| = |\alpha| \|\vec{X}\|$ where $|\alpha|$ is the *absolute value* of α .

(A.2.14)

DEFINITION: In any vector space, a coordinatefree representation of the Straight Line L generated by two vectors \vec{X} and \vec{Y} ($\vec{X} \neq \vec{Y}$) is the set L of all (λ -dependent) points \vec{Z}_λ of the form

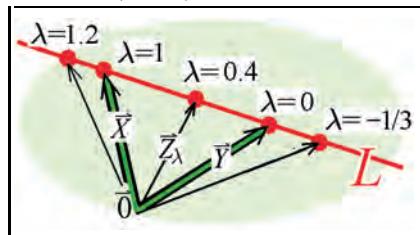
$$(A.2.14a) \quad L = \{\vec{Z}_\lambda \text{ of the form } \lambda\vec{X} + (1 - \lambda)\vec{Y} \text{ for } -\infty < \lambda < \infty\}.$$

(A.2.15)

NOTE: The equation (A.2.14a) is coordinate free and is therefore valid for vectors \vec{X} and \vec{Y} in *any* vector space as defined in (A.2.1). This means we can extend (or abstract) the geometric notion of a line of points or vectors containing the two vectors \vec{X} and \vec{Y} even though, as with vector spaces of matrices or polynomials, we do not have a picture like that of Figure (A.2.16) following.

(A.2.16) **FIGURE:** Line L equals all points $\lambda\vec{X} + (1 - \lambda)\vec{Y}$ as λ varies.

Any pair of distinct vectors on line L will suffice to generate L in accordance with (A.2.14a). The generating vector pair $\{\vec{X}, \vec{Y}\}$ in Equation (A.2.14a) and Figure on the right is not unique.



(A.2.17)

Interpreting λ in (A.2.14a) and Figure (A.2.16): The scalar (number) λ tells us where the vector $\vec{Z}_\lambda = \lambda\vec{X} + (1 - \lambda)\vec{Y}$ of (A.2.14a) lies relative to the line segment running from \vec{Y} and \vec{X} . For example,



- | | |
|------------------|--|
| $\lambda = -1/3$ | $Z_{-1/3}$ lies in front of the line segment from \vec{Y} to \vec{X} |
| $\lambda = 0$ | $Z_0 = \vec{Y}$ |
| $\lambda = 0.4$ | $Z_{0.4}$ lies almost halfway between \vec{Y} and \vec{X} |
| $\lambda = 1$ | $Z_1 = \vec{X}$ |
| $\lambda = 1.2$ | $Z_{1.2}$ lies beyond the line segment from \vec{Y} to \vec{X} . |

(A.2.18)

The LINE SEGMENT between vectors \vec{X} and \vec{Y} , denoted by \overline{XY} , is the subset of points in line L (A.2.14a) for which $0 \leq \lambda \leq 1$. The DISTANCE between \vec{X} and \vec{Y} , or the LENGTH of \overline{XY} is defined

by

$$\begin{aligned}
 \text{(A.2.18a)} \quad \text{dist}(\vec{X}, \vec{Y}) &\stackrel{\text{def}}{=} \text{length}(\overline{XY}) \\
 &= \text{length}(\vec{X} - \vec{Y}) \\
 &= \|\vec{X} - \vec{Y}\| \\
 &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_k - y_k)^2}.
 \end{aligned}$$

(A.3) Functions

Informally, a FUNCTION is a relation between a set in inputs and a set of outputs where each input is related to (associated with) only one output.

(A.3.1) **EXAMPLE:** $f(x) = x^2 + 1$: Often, but not always, the inputs of a function are *numbers* and *there is a rule or formula relating input to output*. (*Example (A.3.3a) describes a function whose inputs and outputs are not numbers, and there is no rule or formula that pairs input with output.*) When we write $f(x) = x^2 + 1$, we are stating a rule that says “to produce an output, $f(x)$, we square each input x , then add 1”. The input-output pair is $(x, f(x))$, or $(x, x^2 + 1)$.

Here follows a more formal definition which defines the function to be a list of the input-output pairings:

(A.3.2)

DEFINITION:

$f: \mathcal{X} \rightarrow \mathcal{Y}$, from set \mathcal{X} to set \mathcal{Y} is a set of pairs,

$$f \stackrel{\text{def}}{=} \{[x, y]\} \quad \text{or} \quad f \stackrel{\text{def}}{=} \{[x, f(x)]\},$$

where the first element is an input element x from \mathcal{X} , and there is only one (second) output element $y = f(x)$ in \mathcal{Y} .

If for every second element y in \mathcal{Y} , there is only one input element x in \mathcal{X} , then the function f is said to be ONE-TO-ONE, also denoted 1-1 or 1-to-1.

We say that the function f takes element x to element $y = f(x)$.

(A.3.3)

EXAMPLES: Functions. Consider the three sets,



(A.3.3a) $\mathcal{X} = \{1, \Delta, \spadesuit\}$, $\mathcal{Y} = \{5, \odot\}$, and $\mathcal{Z} = \{6, 7\}$

each of which has only finitely many elements. Use of the symbols Δ , \spadesuit , and \odot emphasizes the fact that the elements of the sets \mathcal{X} and \mathcal{Y} are not necessarily numbers.

We define the functions $f: \mathcal{X} \rightarrow \mathcal{Y}$ and $g: \mathcal{Y} \rightarrow \mathcal{Z}$ as the sets of ordered pairs (A.3.2)

(A.3.3b) $f = \{[1, \odot], [\Delta, 5], [\spadesuit, 5]\}$

$$g = \{ [5, 6], [\odot, 7] \}$$

which, in “ $f(x)$ ” notation, take the forms

(A.3.3c) $f(1) = \odot, f(\Delta) = 5, f(\spadesuit) = 5.$

$$g(5) = 6, g(\odot) = 7$$

Display (A.3.3c) shows that for each input x from \mathcal{X} (or y from \mathcal{Y}) there is only one output element $f(x)$ in \mathcal{Y} (or $g(y)$ in \mathcal{Z}).

(A.3.4)

DEFINITION: A 1-to-1 function $f: \mathcal{X} \rightarrow \mathcal{Y}$, the set of ordered pairs (x, y) , has an INVERSE FUNCTION $f^{-1}: \mathcal{Y} \rightarrow \mathcal{X}$ which is the set of ordered pairs (y, x) . Inputs of f become outputs of f^{-1} and outputs of f become inputs of f^{-1} .

(A.3.5)



EXAMPLE: An Inverse Function. Note that function f (A.3.3b) is not 1-1 because two distinct inputs Δ and \spadesuit have the same output 5. In fact, $f(\Delta) = f(\spadesuit) = 5$. The function g is 1-1 so we apply Definition (A.3.4) and *reverse the order of the ordered pairs* of function $g: \mathcal{Y} \rightarrow \mathcal{Z}$ (A.3.3b), to obtain the inverse function $g^{-1}: \mathcal{Z} \rightarrow \mathcal{Y}$. That is,

(A.3.5a) $g(5) = 6, g(\odot) = 7$

$$\text{so that } g^{-1}(6) = 5, g^{-1}(7) = \odot$$

(A.3.6) The **COMPOSITION** of two given functions,

(A.3.6a) $g : \mathcal{X} \rightarrow \mathcal{Y}$ and $f : \mathcal{Y} \rightarrow \mathcal{Z}$,

is a *third* function, denoted by

(A.3.6b) $(f \circ g) : \mathcal{X} \rightarrow \mathcal{Z}$,

that is defined by the separate actions of f and g in the following way:

(A.3.6c) $(f \circ g)(x) \stackrel{\text{def}}{=} f(g(x))$ in \mathcal{Z}

for all x in \mathcal{X} . In the terminology of (A.3.2), the function $f \circ g$ sends element $x \in \mathcal{X}$ to element $f(g(x)) \in \mathcal{Z}$.

(A.3.7)  **EXAMPLE: A Composite Function.** Consider functions $f : \mathcal{X} \rightarrow \mathcal{Y}$ and $g : \mathcal{Y} \rightarrow \mathcal{Z}$ defined in (A.3.3b). Then Definition (A.3.6) tells us that the composite function $g \circ f$ is described as follows:

$$\begin{aligned} \text{(A.3.7a)} \quad g \circ f(1) &= g(f(1)) = g(\heartsuit) = 7 \\ g \circ f(\triangle) &= g(f(\triangle)) = g(5) = 6 \\ g \circ f(\spadesuit) &= g(f(\spadesuit)) = g(5) = 6 \end{aligned}$$

which has ordered pair notation $g \circ f = \{[1, 7], [\triangle, 6], [\spadesuit, 6]\}$.

(A.4) Linear Functions and Matrices

(A.4.1) A **LINEAR FUNCTION** $T : \mathcal{X} \rightarrow \mathcal{Y}$ for vector spaces \mathcal{X} and \mathcal{Y} is defined by three properties that (i) $T(\vec{0}) = \vec{0}$, (ii) T preserves proportion, and (iii) T takes straight lines (A.2.14a) to straight lines (*proved in Theorem (4.5.6), pg. 78*). i.e.,

$$\text{(A.4.1a)} \quad T(\lambda_1 \vec{P}_1 + \lambda_2 \vec{P}_2) \stackrel{\text{def}}{=} \lambda_1 T(\vec{P}_1) + \lambda_2 T(\vec{P}_2)$$

for all vectors, \vec{P}_1, \vec{P}_2 in \mathcal{X} and all scalars, λ_1, λ_2 .

Using mathematical induction, the 2-term expression (A.4.1a) that defines the linearity of T can be extended to the n -term expression

$$(A.4.1b) \quad T\left(\sum_{i=1}^{i=n} \lambda_i \vec{P}_i\right) = \sum_{i=1}^{i=n} \lambda_i T(\vec{P}_i)$$

for all scalars λ_i and all vectors $\vec{P}_i, i = 1, 2, \dots, n$.

A useful property of any linear function T is that once its values $T(x)$ are known for a small set of input x s, (*a basis*), the values $T(x)$ are known for all possible input vectors x . Here are the details.

(A.4.2)

THEOREM: All the values of the linear function $T: \mathcal{X} \rightarrow \mathcal{Y}$ for vector spaces \mathcal{X} and \mathcal{Y} are uniquely determined just by the values of T on any basis (A.2.9) of \mathcal{X} .

PROOF: Let \vec{X} be any vector. Let $\{\vec{X}_i, i = 1, 2, \dots, n\}$ form a basis (A.2.9) which then guarantees the existence of unique scalars, $\{\lambda_i, i = 1, 2, \dots, n\}$, such that

$$(A.4.2a) \quad \vec{X} = \sum_{i=1}^{i=n} \lambda_i \vec{X}_i \qquad \text{from (A.2.9).}$$

Then

$$(A.4.2b) \quad T(\vec{X}) = T\left(\sum_{i=1}^{i=n} \lambda_i \vec{X}_i\right) = \sum_{i=1}^{i=n} \lambda_i T(\vec{X}_i) \qquad \text{from (A.4.1b).}$$

Since each λ_i is known, and each $T(\vec{X}_i)$ is known, (*the value of T on the basis vector \vec{X}_i*), it follows that the last expression of (A.4.2b) is known, which means the first expression, $T(\vec{X})$, is completely determined by its values on the n basis vectors $\{\vec{X}_i\}$. ■

MATRICES

(A.4.3) The $m \times k$ MATRIX A (*m rows, k columns*) defined by the mk scalars $a_{ij}, i = 1, 2, \dots, m, j = 1, 2, \dots, k$ is written as the rectangular array

$$(A.4.3a) \quad A = (a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \\ a_{m1} & a_{m2} & \dots & a_{mk} \end{bmatrix} \begin{array}{l} \leftarrow r_1(A) \\ \leftarrow r_2(A) \\ \vdots \\ \leftarrow r_m(A) \end{array}$$

$$\begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ c_1(A) & c_2(A) & c_k(A) \end{array}$$

where

(A.4.3b)

- $r_i(A) = [a_{i,1}, a_{i,2}, \dots, a_{i,k}]$ is the i th row of $m \times k$ matrix A and
- $c_j(A) = \text{col}[a_{1,j}, a_{2,j}, \dots, a_{m,j}]$ denotes its j th column.

With (A.4.3b), we have an alternate notation for (A.4.3a), namely,

(A.4.3c)

$$A = (a_{ij}) = \begin{bmatrix} r_1(A) \\ r_2(A) \\ \vdots \\ r_m(A) \end{bmatrix} = [c_1(A) \quad c_2(A) \quad \dots \quad c_k(A)].$$

Let the $k \times n$ matrix B (k rows, n columns) defined by the kn scalars $j = 1, 2, \dots, k$, $l = 1, 2, \dots, n$, be written

$$(A.4.3d) \quad B = (b_{jl}) \stackrel{\text{def}}{=} \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \dots & b_{kn} \end{bmatrix}.$$

Then the MATRIX PRODUCT, $A * B$ is the $m \times n$ matrix

$$(A.4.3e) \quad A * B = D = (d_{il}) = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{m1} & d_{m2} & \dots & d_{mn} \end{bmatrix}$$

(m rows, n columns) which is defined by the mn scalars c_{il} where $i = 1, 2, \dots, m$, $l = 1, 2, \dots, n$, and

$$(A.4.3f) \quad d_{il} = \langle r_i(A), \overline{d_l(B)} \rangle = \sum_{s=1}^{s=k} a_{is} \overline{d_{sl}}. \quad \text{see } (A.2.6) \text{ and } (A.4.3b)$$

where $\overline{b_{sl}}$ is the complex conjugate (A.2.13a) of b_{sl} in case b_{sl} is complex.

(A.4.4)



EXAMPLES: Matrix Products: When the matrix A is a 2×2 square matrix and the right-hand factor is a column vector matrix \vec{X} or a square 2×2 matrix, then (A.4.3e), (A.4.3f) give the entries of the products as follows:

$$(A.4.4a) \quad \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_A * \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\vec{X}} = \underbrace{\begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}}_{D = A * \vec{X}}$$

$$(A.4.4b) \quad \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_A * \underbrace{\begin{bmatrix} p & q \\ r & s \end{bmatrix}}_B = \underbrace{\begin{bmatrix} ap + br & aq + bs \\ cp + dr & cq + ds \end{bmatrix}}_{A * B}.$$

From (A.4.4b), there is a 2×2 matrix I_2 that is **Feeble**¹ in the sense that after multiplication, I_2 leaves any factor unchanged. Specifically, for any 2×2 matrix A , the products, $I_2 * A$ and $A * I_2$ are still equal to A . Here is the calculation:

$$(A.4.4c) \quad \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{I_2} * \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_A = \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_A = \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_A * \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{I_2}.$$

It also follows from (A.4.4b) that any 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, where $ad - bc \neq 0$, has a unique **Multiplicative Inverse**

$$(A.4.4d) \quad A^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \text{ see footnote}^2$$

which brings us back to the identity I_2 in the sense that

$$(A.4.4e) \quad A^{-1} * \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} * A^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2.$$

We show that *any* linear function (or transformation) T can always be effected by left-hand multiplication by a certain $m \times k$ matrix $\text{MX}(T)$ which we now define. (*The proof that $\text{MX}(T)$ is the desired matrix follows in Theorem (A.4.6).*)

¹Similarly, 0, the number zero is “feeble” with respect to addition since its combination with any other number x leaves it intact. That is, for all x , $0 + x = x + 0 = x$. Formally, feeble elements are called **identities**.

²Scalar multiplication by a matrix (*a vector*) is defined in (A.2.4c).

(A.4.5) **NOTATION:** Suppose $T : \mathbf{R}^k \rightarrow \mathbf{R}^m$ is a linear transformation. Let $\{\vec{E}_1, \vec{E}_2, \dots, \vec{E}_k\}$ be the k standard basis vectors of \mathbf{R}^k , each of which is written as a $k \times 1$ column. Then we define the $m \times k$ matrix whose j th column equals $T(\vec{E}_j)$ in \mathbf{R}^m where $j = 1, 2, \dots, k$. That is,

$$(A.4.5a) \quad \text{MX}(T) \stackrel{\text{def}}{=} \begin{bmatrix} T(\vec{E}_1) & T(\vec{E}_2) & \dots & T(\vec{E}_j) & \dots & T(\vec{E}_k) \end{bmatrix} \quad \begin{array}{cccccc} \uparrow & \uparrow & & \uparrow & & \uparrow \\ \text{Col 1} & \text{Col 2} & & \text{Col } j & & \text{Col } k \end{array}$$

(A.4.6)

THEOREM: Matrix multiplication characterizes all linear transformations. For any linear $T : \mathbf{R}^k \rightarrow \mathbf{R}^m$ as defined in (A.4.1a), $m \times k$ matrix $\text{MX}(T)$ given by (A.4.5a) has the property that for all \vec{X} in \mathbf{R}^k ,

$$(A.4.6a) \quad \underbrace{T(\vec{X})}_{T \text{ of } X} = \underbrace{\text{MX}(T) * \vec{X}}_{\text{matrix MX}(T) \text{ times vector } \vec{X}}.$$

PROOF: Choose any \vec{X} in \mathbf{R}^m . Then \vec{X} is represented by the linear combination (A.2.9a)

$$(A.4.6b) \quad \vec{X} = a_1 \vec{E}_1 + a_2 \vec{E}_1 + \dots + a_j \vec{E}_j + \dots + a_k \vec{E}_k = \sum_{j=1}^k a_j \vec{E}_j.$$

for unique scalars $\{a_j\}$ and the standard basis vectors (A.2.11) \vec{E}_j in \mathbf{R}^m where $j = 1, 2, \dots, k$.

A useful observation is the fact that from any $m \times k$ matrix, the j th column, an $m \times 1$ sub-matrix, can be isolated by simply multiplying on the right by the j th standard basis vector \vec{E}_j . In the notation of (A.4.3b), for any $k \times m$ matrix A , $c_j(A) = A * \vec{E}_j$. In our case, set $A = \text{MX}(T)$, a matrix which, by definition (A.4.5a), guarantees that its j th column is $T(\vec{E}_j)$. Then

$$(A.4.6c) \quad \text{The } j\text{th column of } \text{MX}(T) = \text{MX}(T) * \vec{E}_j = T(\vec{E}_j).$$

Since $\vec{X} = \sum_{j=1}^k a_j \vec{E}_j$ as given in (A.4.6b),

$$(A.4.6d) \quad T(\vec{X}) = T\left(\sum_{j=1}^k a_j \vec{E}_j\right) = \sum_{j=1}^k a_j T(\vec{E}_j) \quad \begin{cases} \text{linearity} \\ \text{of } T \text{ (A.4.1b)} \end{cases}$$

$$\begin{aligned}
&= \sum_{j=1}^k a_j \left(\text{MX}(T) * \vec{E}_j \right) \quad \text{from (A.4.6c), } T(\vec{E}_j) = \left(\text{MX}(T) * \vec{E}_j \right) \\
&= \text{MX}(T) * \left(\sum_{j=1}^k a_j \vec{E}_j \right) \quad \text{factor out MX}(T) \\
&= \text{MX}(T) * \vec{X} \quad \text{since } \vec{X} = \sum_{j=1}^k a_j \vec{E}_j \text{ from (A.4.6b).}
\end{aligned}$$

In this string of equalities, equate the first and last terms to show that (A.4.6a) is true. This ends the proof. ■

(A.4.7) **EXAMPLE:** The 2×2 $\text{MX}(T) = [T(\mathbf{E}_1) \ T(\mathbf{E}_2)]$ for $T: \mathbf{R}^2 \rightarrow \mathbf{R}^2$.

For all $\vec{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ in \mathbf{R}^2 , and arbitrary numbers a, b, c, d , we define

$$\begin{aligned}
\text{(A.4.7a)} \quad &T \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = \begin{bmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{bmatrix} \quad \text{linear function } T: \mathbf{R}^2 \rightarrow \mathbf{R}^2 \\
&= \begin{bmatrix} a & b \\ c & d \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \text{from (A.4.4a)} \\
&= \left[T \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \ T \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \right] * \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&\quad \text{from (A.4.7a), } T \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) = \begin{bmatrix} a \\ c \end{bmatrix} \text{ and } T \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} b \\ d \end{bmatrix} \\
&= \underbrace{\left[T \left(\vec{E}_1 \right) \ T \left(\vec{E}_2 \right) \right]}_{\text{MX}(T)} * \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \left\{ \vec{E}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \vec{E}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}
\end{aligned}$$

which confirms $\text{MX}(T) = [T(\vec{E}_1) \ T(\vec{E}_2)]$ (A.4.6a) when $m = k = 2$.

(A.4.8)

NOTATION: It will often be convenient to replace the symbol $\text{MX}(T)$ for the $m \times n$ matrix of T with the simpler symbol T . That is,

$$\text{MX}(T) \equiv T.$$

(A.5) Eigenvectors and Eigenvalues

Intuitively, an *eigenvector* (sometimes called a *proper vector*) of a given matrix is a vector that is stretched or contracted by a certain factor known as the *eigenvalue* (also called a *proper value*) under the action of (*multiplication by*) that matrix. Here are the formal definitions.

(A.5.1)

DEFINITION: Given an $n \times n$ matrix T . Then an $n \times 1$ vector $\vec{X} \neq 0$ in \mathbf{R}^n is called an EIGENVECTOR or PROPER VECTOR of T if

$$(A.5.1a) \quad T(\vec{X}) = T * \vec{X} = \lambda \vec{X}.$$

The “stretching” scalar λ , which may be real or complex, is called an EIGENVALUE or PROPER VALUE of the matrix T .

(A.5.2)



There is always at least one eigenvector for any matrix T . A deep theorem, equivalent to the *Fundamental Theorem of Algebra*,³ guarantees that every square matrix has at least one eigenvector which comes with an associated eigenvalue. Also note that Definition (A.5.1) requires that the eigenvector $\vec{X} \neq 0$ while the eigenvalue $\lambda = 0$ is allowed.

(A.5.3)



There are at most n linearly independent (A.2.10) eigenvectors for any $n \times n$ matrix T . For example, the matrix $\begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix}$ has only one eigenvalue, namely a , and its only eigenvector is $\vec{X} = \text{col}[1, 0]$ (or any multiple $b\vec{X}$ of \vec{X} where $b \neq 0$). The matrix $aI_2 = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$ has only one eigenvalue, a , yet *every* non-zero vector is an eigenvector.

Finding Matrix of T When $T(\mathbf{X}_i) = \lambda_i \mathbf{X}_i$ for Known Eigenvectors \mathbf{X}_i and Eigenvalues λ_i .

The following theorem is essential in the proof of Theorem (7.1.3).

³ The Fundamental Theorem of Algebra says that for any n th degree polynomial, $p(z) = \sum_{k=0}^n a_k z^k$, where the coefficients a_k are complex numbers, there are exactly n complex roots $\{z_i\}$ where $p(z_i) = 0$ for each $i = 1, 2, \dots, n$.

(A.5.4)

THEOREM: Given an $n \times n$ matrix T that has eigenvectors that also form a basis for \mathbf{R}^n . That is,

$$(A.5.4a) \quad T(\vec{X}_i) = T * \vec{X}_i = \lambda_i \vec{X}_i \quad \text{where } i = 1, 2, \dots, n,$$

for scalars $\{\lambda_i\}$ and n vectors $\{\vec{X}_i\}$ that are also linearly independent in \mathbf{R}^n . Then using the $n \times n$ matrices

(A.5.4b)

$$S = \begin{bmatrix} \vec{X}_1 & \vec{X}_2 & \dots & \vec{X}_n \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} \lambda_1 & 0 & & 0 \\ 0 & \lambda_2 & & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & & \lambda_n \end{bmatrix},$$

$\uparrow \quad \uparrow \quad \uparrow$
 $Col 1 \quad Col 2 \quad Col n$

we have

$$(A.5.4c) \quad T = S * D * S^{-1}.$$

PROOF: Note that S is the matrix whose i th column is the eigenvector \vec{X}_i . From (A.4.6a), the i th column of S is $\vec{X}_i = S(\vec{E}_i) = S * \vec{E}_i$ for all standard basis vectors $\{\vec{E}_i\}$ (A.2.11) of \mathbf{R}^n . That is,

$$(A.5.4d) \quad S * \vec{E}_i = \vec{X}_i. \quad \text{Equivalently, } S^{-1} * \vec{X}_i = \vec{E}_i.$$

(A.5.4e) Now for all i , $i = 1, 2, \dots, n$,

$$\begin{aligned} (S * D * S^{-1}) * \vec{X}_i &= S * D * (S^{-1} * \vec{X}_i) && * \text{ multiplication associativity} \\ &= S * (D * \vec{E}_i) && S^{-1} * \vec{X}_i = \vec{E}_i \text{ from (A.5.4d)} \\ &= \lambda_i S * (\vec{E}_i) && D * \vec{E}_i = \lambda_i \vec{E}_i \text{ from (A.5.4b)} \\ &= \lambda_i \vec{X}_i && S * (\vec{E}_i) = \vec{X}_i \text{ from (A.5.4d)} \\ &= T * \vec{X}_i. && \text{Hypothesis (A.5.4a)} \end{aligned}$$

From the chain of equalities (A.5.4e), $(S * D * S^{-1}) * \vec{X}_i = T * \vec{X}_i$ (also denoted by $T(\vec{X}_i)$) for all basis vectors $\{X_i\}$, $i = 1, 2, \dots, n$. It follows that $(S * D * S^{-1}) * \vec{X} = T * \vec{X}$ for all \vec{X} in \mathbf{R}^n and not just for basis vectors (Theorem (A.4.2)). With (A.5.4c) being established, the proof is done. ■

(B) Hyperbolic Functions

(B.1) Overview

We come to the hyperbolic functions, $\cosh x$, $\sinh x$, and $\tanh x$ (pronounced “coe-sh”, “sinsh”, and “tansh”, respectively). The functions $\cosh x$ and $\sinh x$ are the even and the odd parts of the function e^x where x is a real number. *For more on even and odd functions, see Exercise (B.4.1), pg. 383.* Here are the definitions:

(B.1.1)

DEFINITION: The HYPERBOLIC SINE, COSINE, AND TANGENT, \sinh , \cosh , and \tanh , are defined for all real x as follows:

(B.1.1a)

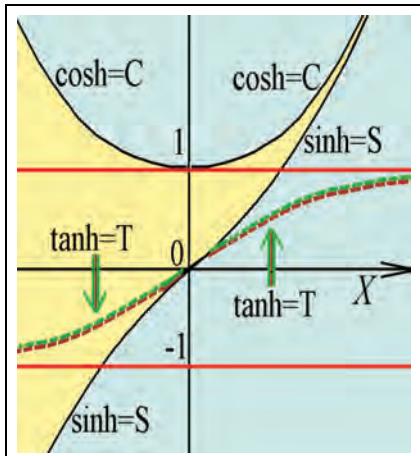
$$\sinh(x) = S_x \stackrel{\text{def}}{=} \frac{e^x - e^{-x}}{2},$$

(B.1.1b)

$$\cosh(x) = C_x \stackrel{\text{def}}{=} \frac{e^x + e^{-x}}{2},$$

(B.1.1c)

$$\begin{aligned} \tanh(x) &= T_x \stackrel{\text{def}}{=} \frac{e^x - e^{-x}}{e^x + e^{-x}} \\ &= \frac{S_x}{C_x} \stackrel{\text{def}}{=} \frac{\sinh(x)}{\cosh(x)}. \end{aligned}$$



(B.1.2) **Ranges of Hyperbolic Functions.** Note that for all real x ,

$$-1 < \tanh(x) = T_x < 1,$$

$$1 \leq \cosh(x) = C_x < \infty,$$

$$-\infty < \sinh(x) = S_x < \infty.$$

(B.1.3) **Some Hyperbolic Identities:** Lines (B.1.3a), (B.1.3c), and (B.1.3d), flow directly from the core definitions (B.1.1).

$$(B.1.3a) \quad C_x^2 - S_x^2 = 1. \quad \begin{matrix} \text{Divide this line by } C_x^2 \\ \text{to obtain} \end{matrix}$$

$$(B.1.3b) \quad 1 - T_x^2 = 1/C_x^2.$$

$$(B.1.3c) \quad C_{x+y} = C_x C_y + S_x S_y. \quad \text{Use } x+y \text{ in Def. (B.1.1b)}$$

$$(B.1.3d) \quad S_{x+y} = S_x C_y + C_x S_y. \quad \begin{matrix} \text{Use } x+y \text{ in Def. (B.1.1a).} \\ \text{Then divide this line by} \\ (B.1.3c) \text{ to obtain} \end{matrix}$$

$$(B.1.3e) \quad T_{x+y} = \frac{T_x + T_y}{1 + T_x T_y}. \quad \begin{matrix} \text{Divide (B.1.3d) by (B.1.3c).} \\ \text{Then use Def. (B.1.1c)} \end{matrix}$$

(B.2) Even and Odd Functions

(B.2.1)

DEFINITION: An EVEN FUNCTION f_E acts like x^n where n is even. That is, an even function, like x^n , will always absorb the minus sign in the sense that

$$f_E(-x) = f_E(x) \quad \text{for all } x.$$

(B.2.2)

DEFINITION: An ODD FUNCTION f_O , behaves like x^n where n is odd. Specifically, any odd function, like x^n , allows minus signs to factor out in the sense that

$$f_O(-x) = -f_O(x) \quad \text{for all } x.$$

Any function $f : R \rightarrow R$ decomposes uniquely into an even part f_1 and an odd part f_2 . The interesting thing is that the forms for f_1 and f_2 are easily obtained. In fact, we have the

(B.2.3)

THEOREM: For any function $f : R \rightarrow R$, define f_1 and f_2 by

$$(B.2.3a) \quad f_1(x) = \frac{f(x) + f(-x)}{2} \text{ and } f_2(x) = \frac{f(x) - f(-x)}{2}.$$

Then

$$(B.2.3b) \quad f = f_1 + f_2 \text{ where } f_1 \text{ is even and } f_2 \text{ is odd.}$$

Moreover, decomposition (B.2.3b) is unique in the sense that if there are other functions, g_1 and g_2 that are even and odd respectively, such that

$$(B.2.3c) \quad f = f_1 + f_2 = g_1 + g_2,$$

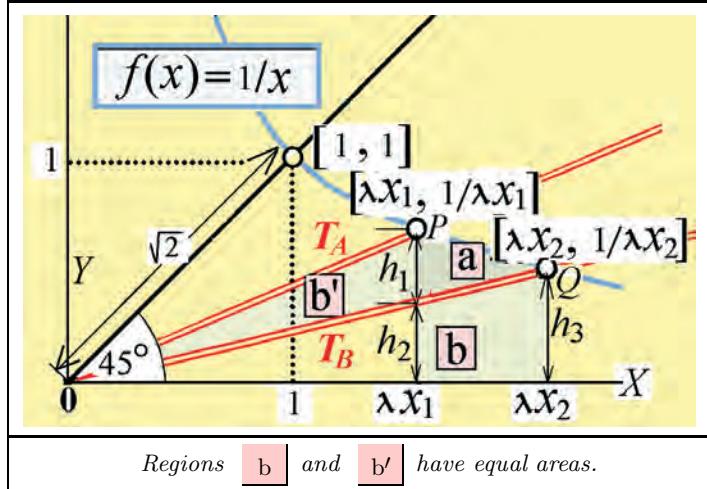
then, necessarily, $f_1 = g_1$ and $f_2 = g_2$.

PROOF: Exercise (B.4.1), part (3).

(B.3) Invariant Areas of Transformed Hyperbolas

(B.3.1)

FIGURE: Triangular and Trapezoidal Areas are Equal



In Lemma (B.3.2) and Theorem (B.3.3) following, we will show that

for fixed x_1 and x_2 on the X -axis in Figure (B.3.1), the areas of all subtended triangular areas \overline{POQ} are constant and independent of the “sliding” scalar, λ .

First, we prove

(B.3.2)

LEMMA: (Area Equality) *In Figure (B.3.1), the perpendicular X -axis and Y -axis are identically scaled. The points $0 < x_1 < x_2$ are arbitrary but fixed on the X -axis. We also have axes T_A , T_B , hyperbola $f(x) = 1/x$, and arbitrary scalar $\lambda > 0$. Then*

$$(B.3.2a) \quad \text{Area of triangle } [b'] = \text{Area of trapezoid } [b].$$

Moreover, this value is independent of λ .

PROOF: In Figure (B.3.1), the left-hand and right-hand sides of the trapezoid $[b]$ are $h_2 = (x_1/(\lambda x_2^2))$ and $h_3 = (1/(\lambda x_2))$, respectively. The horizontal base of the trapezoid has length $(\lambda x_2 - \lambda x_1)$. By taking the product, *average height \times width* $= \frac{1}{2}(h_2 + h_3)(\lambda x_2 - \lambda x_1)$, we obtain

$$(B.3.2b) \quad \text{Trapezoid Area } [b] = \frac{1}{2} \left(1 - \left(\frac{x_1}{x_2} \right)^2 \right). \quad \begin{matrix} \text{independent} \\ \text{of } \lambda \end{matrix}$$

Now triangle $[b']$ has (vertical) base $h_1 = (1/\lambda)(1/x_1 - x_1/x_2^2)$ and (horizontal) altitude λx_1 . Hence, $1/2 \text{ base } \times \text{ altitude}$ has value

$$(B.3.2c) \quad \text{Triangle Area } [b'] = \frac{1}{2} \left(1 - \left(\frac{x_1}{x_2} \right)^2 \right), \quad \begin{matrix} \text{independent} \\ \text{of } \lambda \end{matrix}$$

which proves (B.3.2a). The theorem is done. ■

We come to the theorem that quantifies the triangular area \overline{POQ} in Figure (B.3.1) bounded by the hyperbola $f(x) = 1/x$ and the T_A and T_B axes.

(B.3.3)

THEOREM: (*Calculation of Area of Fig. (B.3.1)*) In Figure (B.3.1), we are given two areas:

- Area($|a| + |b|$) of the subtended “triangle” bounded by the T_A and T_B axes and the hyperbola $f(x) = 1/x$,
- Area($|a| + |b'|$) of the subtended “trapezoids”.

Then both areas are invariant under λ and have common value

$$(B.3.3a) \quad r_{A,B} = \ln\left(\frac{x_2}{x_1}\right). \quad \text{area } r \text{ independent of } \lambda.$$

Equivalently,

$$(B.3.3b) \quad e^{r_{A,B}} = \frac{x_2}{x_1} \quad \text{exponential of both sides of (B.3.3a).}$$

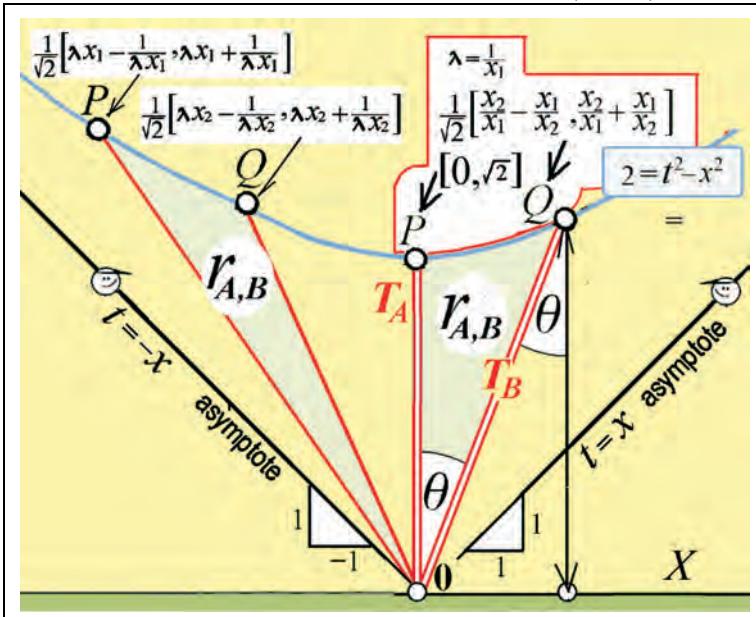
PROOF: Since (B.3.2a) guarantees that the area of $|b| = \text{area of } |b'|$, it follows from Figure (B.3.1) that the area of triangle $|a| + |b'|$ = the area of trapezoid $|a| + |b|$, which computes to be

$$(B.3.3c) \quad \int_{\lambda x_1}^{\lambda x_2} \left(\frac{dx}{x} \right) = \ln(\lambda x_2) - \ln(\lambda x_1) = \ln\left(\frac{x_2}{x_1}\right) = r_{A,B},$$

a quantity that is *independent of λ* which we have denoted $r_{A,B}$. The very last equality of (B.3.3c) is (B.3.3a). Taking the exponential of both sides of (B.3.3a) gives us (B.3.3b). Having established (B.3.3a) and (B.3.3b), the theorem is proved. ■

The following theorem uses a counterclockwise 45° rotation of Figure (B.3.1) to relate the area $r_{A,B}$, the hyperbolic tangent \tanh , and the trigonometric tangent \tan .

(B.3.4)

FIGURE: Area-Preserving Rotation of Figure (B.3.1).

(B.3.5)

THEOREM: (Relating Hyperbolic Area to Speed): In the diagram (B.3.4) the T_A and T_B axes intersect the hyperbola $2 = t^2 - x^2$ at points P and Q that are arbitrary except for the requirement that

- Area $r_{A,B}$ is fixed for triangular regions $\overline{P0Q}$.

If the line segment \overline{OP} (or T_A -axis) is perpendicular to the X -axis, forming angle θ with the T_B -axes, then

$$(B.3.5a) \quad \tan(\theta) = \tanh(r_{A,B}).$$

If $\tan(\theta) > 0$ (if $\tan(\theta) < 0$), then Q falls to the right (to the left) of P .

PROOF: We rotate Figure (B.3.1) 45° counterclockwise to produce Figure (B.3.4). Then

- $f(x) = 1/x$ rotates into the hyperbola $2 = t^2 - x^2$ of (B.3.4): Counterclockwise rotation through an angle θ of any point $\text{col}[x, y]$ in

\mathbf{R}^2 can be produced by a left-hand multiplication by the orthogonal transformation

$$(B.3.5b) \quad \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \quad \text{when } \theta = 45^\circ.$$

Using the matrix (B.3.5b) on a typical point $\text{col}[x, 1/x]$ of the hyperbola $f(x) = 1/x$ in Figure (B.3.1), we obtain

(B.3.5c)

$$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} x \\ 1/x \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} x - 1/x \\ x + 1/x \end{bmatrix} = \begin{bmatrix} \tilde{x} \\ \tilde{t} \end{bmatrix}. \quad \begin{cases} \tilde{x} \stackrel{\text{def}}{=} (x - 1/x)/\sqrt{2} \\ \tilde{t} \stackrel{\text{def}}{=} (x + 1/x)/\sqrt{2} \end{cases}$$

The output coordinates, $\tilde{x} = (x - 1/x)/\sqrt{2}$ and $\tilde{t} = (x + 1/x)/\sqrt{2}$, have the property $2 = \tilde{t}^2 - \tilde{x}^2$ which defines the hyperbola in the rotated figure (B.3.4).

- Choose $\lambda = 1/x_1$. As $\lambda > 0$ varies and points P and Q slide along the hyperbola $2 = t^2 - x^2$, the resulting triangular areas $\overline{P0Q}$ have constant area $r_{A,B}$ (B.3.3a). For $\lambda = 1/x_1$, the T_A -axis of triangular region $\overline{P0Q}$ is vertical. The T_B -axis of frame \mathcal{F}^B forms an angle θ , say, which tells us that

$$(B.3.5d) \quad \begin{aligned} \tan(\theta) &= \frac{(x_2/x_1 - x_1/x_2)}{(x_2/x_1 + x_1/x_2)} && \text{from Coordinates of } Q \text{ in (B.3.4)} \\ &= \frac{(e^{r_{A,B}} - e^{-r_{A,B}})}{(e^{r_{A,B}} + e^{-r_{A,B}})} && \text{from (B.3.3b)} \\ &= \tanh(r_{A,B}). && \text{from Def. (B.1.1c)} \end{aligned}$$

This establishes (B.3.5a) for all constant-area regions $\overline{P0Q}$ of Figure (B.3.4). Now the speed $v < 0$

- if and only if $\tanh(r_{A,B}) < 0$ from (B.3.5d)
- if and only if $r_{A,B} < 0$ from Figure (B.1.1)
- if and only if $\ln(x_2/x_1) < 0$ from (B.3.3c)
- if and only if $x_2 < x_1$ $\ln(x_2/x_1) < 0$ means $(x_2/x_1) < 1$
- if and only if axis T_B is left of T_A see axes in (B.3.1)

The proof is done. ■

(B.4) Exercises**(B.4.1)**
Exercise

Even and Odd Functions. For the functions f_1 and f_2 given by (B.2.3a), prove that (B.2.3b) is true. That is, show

$$(1) \quad f = f_1 + f_2 \quad \text{and}$$

$$(2) \quad f_1 \text{ and } f_2 \text{ of (B.2.3a) are even and odd, respectively.}$$

.....

(3) Prove (B.2.3c) is valid — there is only one way to decompose a function f into an even and an odd part.

Hint: Assume f_1, g_1 are even and f_2, g_2 are odd where $f = f_1 + f_2 = g_1 + g_2$. Take the difference to get $(f_1 - g_1) + (f_2 - g_2) = 0$. Since the zero function is the only function that is both even and odd, argue that, necessarily, $f_1 = g_1$ and $f_2 = g_2$.

(B.4.2)
Exercise

Cosh occurs in Nature. The curve of the hyperbolic cosine, $\cosh(x) = C(x)$, shown in the figure of (B.1.1), pg. 376, appears in nature and is called the **catenary curve**. Hanging strings or cables of uniform linear density (ρ pounds per linear foot, say) fall into a catenary pattern. Examples include suspension bridge cables and fine, flexible necklaces.



Question: Analyze the Diagram (B.4.2b) to show that the forces produce a catenary.

Hint: Consider Steps (B.4.2a)–(B.4.2f) following:

(B.4.2a) Calculus Preliminaries: hyperbolic derivatives

$$\text{If } y(x) = A \cosh(Bx),$$

$$\text{then } y'(x) = AB \sinh(Bx)$$

$$\text{and } y''(x) = AB^2 \cosh(Bx).$$

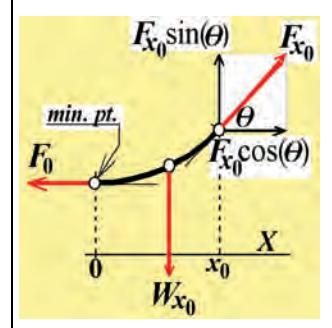
(B.4.2b) Three Forces on a Cable Segment: W_{x_0} , F_0 , F_{x_0}

The cable segment in the diagram on the right is stationary. Hence, the sum of the following three forces acting on the cable equals zero.

$$W_{x_0} = \text{Weight}$$

$$= \rho \times \text{arc length}, x = 0 \text{ to } x = x_0$$

$$= \rho \times \int_{x=0}^{x=x_0} \sqrt{1 + [y'(x)]^2} dx$$



F_0 = internal horizontal force at cable minimum,

F_{x_0} = internal force at $x = x_0$.

(B.4.2c) A Hanging Cable Coincides with the Catenary $\cosh(x)$.

Step 1: Note that Diagram (B.4.2b) decomposes F_{x_0} into its horizontal and vertical components. Set the sum of all horizontal forces to zero. Then set the sum of all vertical forces to zero. From these two equations, deduce the unitless expression

$$(B.4.2d) \quad y'(x_0) = \frac{\rho}{F_0} \int_{x=0}^{x=x_0} \sqrt{1 + [y'(x)]^2} dx.$$

Step 2: After differentiating (B.4.2d) with respect to x , show that

$$(B.4.2e) \quad \left[\frac{F_0}{\rho} y''(x) \right]^2 - \left[y'(x) \right]^2 = 1.$$

Step 3: Use the differentiation properties (B.4.2a) of $A \cosh(Bx)$ and the identity (B.1.3a) to show that (B.4.2e) implies

$$(B.4.2f) \quad y(x) = \frac{F_0}{\rho} \cosh\left(\frac{\rho}{F_0} x\right) + \text{const.}$$

(C) Deconstructing a Moving Train

(C.1) Motion Alters Age

(C.1.1)



Consider triplets of identical age, each having been born at *precisely* the same moment. (*We are inventing a biology in which triplets can be born simultaneously instead of sequentially.*) The triplets are seated inside a railroad car with two of them sitting at opposite ends while the third triplet sits in the middle row (see Figure (C.2.1)).

If the railroad car is traveling in the rightward direction, then, when each triplet is exactly opposite an observer on the platform, the triplets' various ages, as measured by the observers will all be different. This is illustrated in Figure (C.2.1) where the time is **6.15** on the platform, and the (*formerly same-age*) triplets in the train get older as we pass from the front (right) to the rear (left) of the railroad car.

Conversely, on a leftward traveling train, the triplets get younger as we pass from right to left. (*Exercise (C.3.1).*)

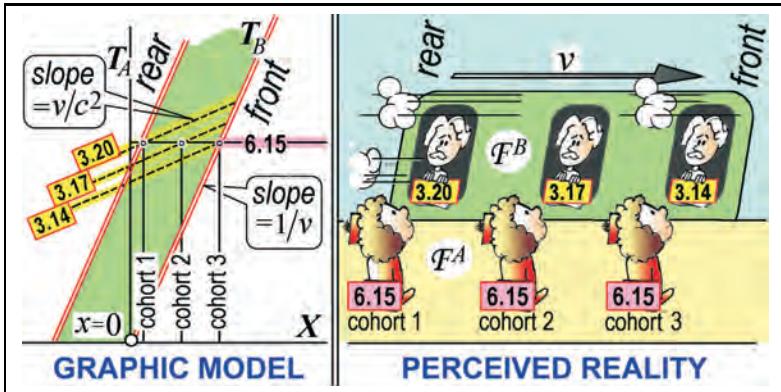
(C.1.1a) **To resolve this dilemma**, we must leap beyond our intuition and imagine the train (or any connected rod, or ruler) as a continuously connected collection of individual “slices” like beads on a string, where each slice, or bead, possesses its own synchronized clock.

In the following discussion, the train of Figure (C.2.1) can be thought of as three individual “slices,” each with its own observer and clock.

(C.2) Minkowski Diagram for a Moving Train

In Figure (C.2.1) following, we consider a train as a set of infinitely many individual points. This is why the left hand panel of Figure (C.2.1) represents all of the train’s coordinates as an *area* or *band* (of coordinates) on the spacetime graph.

(C.2.1)

FIGURE: A Moving Train in Several Time Slices

- LEFT PANEL OF (C.2.1) (*Graphic Model*): The stationary frame of the train platform has a vertical T -axis which intersects the horizontal X -axis at $x = 0$.

WORLDLINES.

- ¬ The three platform cohorts have worldlines that are vertical lines parallel to the T -axis.
- ¬ The front (rightmost) end of the train at point $x = 0$ in the moving frame has relative speed v . Its worldline radiates from the origin with slope $(1/v)$, i.e., at an angle θ from the vertical which implies $v = \tan \theta$. This worldline coincides with the T' -axis of the moving frame.
- ¬ The rear (leftmost) end of the train has its worldline to the left of, and parallel to, the T' -axis. The shaded region between the left and right worldlines represents the infinitely many parallel spacetime worldlines of the points of the train in between.

LINES of SIMULTANEOUS EVENTS.

In the stationary frame, lines of simultaneous events are horizontal (→). The graph shows simultaneous events at time 6.15. In the moving frame of the train, each line of simultaneous events (↗) has slope (v/c^2) (*Theorem (8.4.1)*). The graph shows parallel (slanted) lines of simultaneous events at respective times, 3.14, 3.17, and 3.20.

- **RIGHT PANEL OF (C.2.1)(*Perceived Reality*):** Three cohorts (observers) are fixed at equal distances on a railroad platform. The individual cohort clocks are synchronized with each other. A train passes along the platform moving from left to right. When the platform clocks all read **6.15**, each cohort observes the moving clock (on the train) at the other end of her individual line-of-sight corridor.

Although all clocks on the train are synchronized within the train's frame, the readings of the train's clocks from the platform at time **6.15** are (reading from left to right) **3.20**, **3.17**, and **3.14**.

In sum, at time **6.15** on the platform, the rear (leftmost part) of the train is seen to be older than the front (rightmost part) of the train.



(C.3) Exercises

- (C.3.1) Exercise** **Shifting Age of Train Segments.** Modify Figure (C.2.1) to show that for a leftward moving train, the left end will be younger than the right end as measured by synchronized clocks on the platform.



XI. Supplemental Material Online

Supplemental chapters C, D, E, and F along with interactive animations appear online at

www.special-relativity-illustrated.com.

Chapter D: Dimensional Analysis outlines the process of dimensional analysis that is especially popular with engineers for checking consistency in calculations.

Chapter E: Rings of Functions and Matrices isolates the underlying axiomatic structure behind matrix calculations. As a bonus, we use the same structure to see exactly why a minus times a minus must be a plus (Exercise (E.4.5)).



Chapter F: The process of the scientific method is described.

Chapter G: Mathematical logic and the anatomy of human thought are discussed.

(D) Dimensional Analysis

Here is an engineer’s device (algorithm, recipe) for a relatively straight-forward conversion of one set of units to another set of units. Consider the question

$$\text{pounds/quart} = \text{how many grams/cup?}$$

To resolve this question, the engineer’s technique requires fraction cancellation to hold true for unit *labels*—quarts, inches, pounds—as well as for ordinary *numbers*. Let us see how this works.

(D.1) Unitless Quotients of Dimensions

If we agree that 1 yard = 3 feet, then we agree that both sides of the equal sign are identical except in name. This means that their quotient is numerically equal to 1 and is unitless. For example,

$$(D.1.1) \quad 1 = \frac{1 \text{ yd}}{3 \text{ ft}} = \frac{3 \text{ ft}}{1 \text{ yd}}$$

(D.1.1a) Note: Looking at (D.1.1), you might be tempted to divide the numbers in these “quotients”. But division within these quotients makes sense only when the quotients like (D.1.1) are multiplied against other quantities with units.

Since these quotients are symbolic representations of the number 1, multiplication of any number x by these quotients does not change its value. After multiplication, the *form*, or the *units* of x may change, but not the *value* of x .

Examples (D.2.1) and (D.2.3) show how a quantity’s units, but not its values, are changed when multiplied by 1 in this special way.

.....

(D.1.2) Some Quantities with Quotient = 1:

Expanding on (D.1.1) we present a partial list of well-known equal quantities (with different names) whose algebraic quotients equal 1:

Distances	1 mile	=	5,280 feet
	1 yard	=	3 feet
	1 circumference	=	$2\pi \times \text{radius}$
	1 centimeter	=	2.54 inches
	1 meter	=	39.34 inches
Volumes	1 quart	=	4 cups
	1 gallon	=	4 quarts
	1 liter	=	1.057 quarts
Time	1 minute	=	60 seconds
	1 hour	=	60 minutes
	1 day	=	24 hours
	1 week	=	7 days
Weight	1 pound	=	454 grams
	1 kilogram	=	2.2 pounds

(D.2) Dimensions in Fractions

(D.2.1) EXAMPLE: Converting mph to fps

Suppose we want to convert 30 miles/hour (mph) to X feet/second (fps). The way we do this is to multiply the expression “30 mi/hour” by 1 in such a way as to cancel out unwanted dimensions (miles, hours) and introduce wanted dimensions (feet, seconds).

Let’s see the details of how (D.2.1) is computed.

Question: $\frac{30 \text{ miles}}{\text{hr}} = \text{how many } \frac{\text{feet}}{\text{sec}}$?

Answer:

$$\begin{aligned}
 \text{(a)} \quad \frac{30 \text{ mi}}{\text{hr}} &= \frac{30 \text{ mi}}{\text{hr}} \times \underbrace{\frac{5,280 \text{ ft}}{1 \text{ mi}}}_{1} \times \underbrace{\frac{1 \text{ hr}}{3,600 \text{ sec}}}_{1} \\
 \text{(b)} \quad &= \frac{30 \text{ mi}}{\text{hr}} \times \frac{5,280 \text{ ft}}{1 \text{ mi}} \times \frac{1 \text{ hr}}{3,600 \text{ sec}} \\
 \text{(c)} \quad &= \frac{30 \times 5,280 \text{ ft}}{3,600 \text{ sec}} = \frac{44 \text{ ft}}{\text{sec}}
 \end{aligned}$$

Hence, $\frac{30 \text{ mi}}{\text{hr}} = \frac{44 \text{ ft}}{\text{sec}}$.

(D.2.2)



NOTE: Letters are as Easy as Numbers. As is often the case, the general case is as easy to deal with as the special case. For example, the sequence of steps in (D.2.1) converts 30 mi/hour to 44 ft/sec. But with the same amount of work, by replacing “30” with “ S ” everywhere in (D.2.1), we just as easily answer the more general

Question: $[S] \frac{\text{mi}}{\text{hr}} = [\text{how many}] \frac{\text{ft}}{\text{sec}}$?

Following the steps of (D.2.1), we arrive at the

Answer: $[S] \frac{\text{mi}}{\text{hr}} = \left[\frac{22}{15} S \right] \frac{\text{ft}}{\text{sec}}$ for any speed $[S] \frac{\text{mi}}{\text{hr}}$.

The next example is consistent with the notion that the computation of sequences like (D.2.1) is equally easy whether the computations are made with symbols or specific numbers.

(D.2.3) **EXAMPLE: Converting rpm to mph.**

We consider a wheel of radius R spinning at ω revolutions/minute (rpm). (See how computations are *not* made easier if the symbol R were to be replaced by a single number like 3.830476 feet.) We will find how fast, in miles/hour (mph), a point on the rim is moving.

Here are the details:

A wheel with radius R (feet) is turning at ω revolutions per minute (rev/min, rpm). For an arbitrary point on the circumference,

Question: $\frac{\omega \text{ rev}}{\text{min}} =$ how many $\frac{\text{mi}}{\text{hour}}$?

Answer:

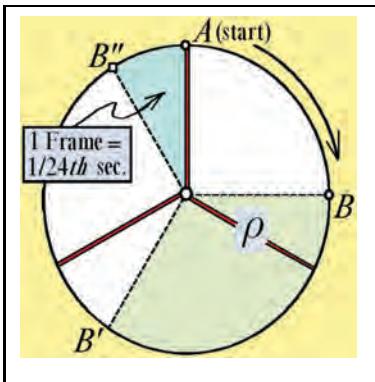
$$\begin{aligned}
 \text{(a)} \quad \frac{\omega \text{ rev}}{\text{min}} &= \frac{\omega \text{ rev}}{\text{min}} \times \underbrace{\frac{60 \text{ min}}{\text{hour}}}_1 \times \underbrace{\frac{2\pi R \text{ feet}}{\text{rev}}}_1 \\
 \text{(b)} \quad &= \frac{\omega \cancel{\text{rev}}}{\cancel{\text{min}}} \times \frac{60 \cancel{\text{min}}}{\cancel{\text{hour}}} \times \frac{2\pi R \text{ feet}}{\cancel{\text{rev}}} \\
 \text{(c)} \quad &= \frac{(\omega \times 60 \times 2\pi)R \text{ feet}}{\text{hour}} \\
 \text{(d)} \quad &= \frac{\omega (120\pi)R \text{ feet}}{\text{hour}} \times \underbrace{\frac{1 \text{ mile}}{5,280 \text{ feet}}}_1 \\
 \text{(e)} \quad &= \frac{\omega (120\pi)R \text{ feet}}{\text{hour}} \times \frac{1 \text{ mile}}{5,280 \text{ feet}} \\
 \text{(f)} \quad &= \frac{\omega \pi R \text{ mile}}{44 \text{ hour}}
 \end{aligned}$$

Hence, $\omega \frac{\text{rev}}{\text{min}} = \omega \left(\frac{\pi R}{44} \right) \frac{\text{mile}}{\text{hour}}$.

(D.3) Exercises

(D.3.1) Obtaining the True Speed of a Car from an Optical Illusion.

A car is seen in a film to be traveling from left to right. The wheels each have N spokes (our diagram shows $N = 3$ spokes) and, in the film, the spokes appear to rotate slowly in the “wrong” direction, i.e., counterclockwise. It is known that the radius of each wheel is ρ feet and the film is running at 24 frames per second.



(D.3.1a) If it takes T seconds for an (apparent) spoke to make one complete counterclockwise rotation, then show that for the car in the film, the possible true speeds are

$$2\pi\rho \left(\frac{24k}{N} - \frac{1}{T} \right) \frac{\text{ft.}}{\text{sec.}}, \quad \text{where } k = 1, 2, 3, \dots$$

Hint: The double lines of the diagram show the initial position of the wheel’s spokes as it appears in one frame. In the next frame, $1/24$ th second later, the diagram shows the new position of a spoke that starts at point A and apparently moves counterclockwise to point B'' . In fact, in the time of one frame, the spoke actually moves clockwise, either to B , to B' , or to B'' , etc.

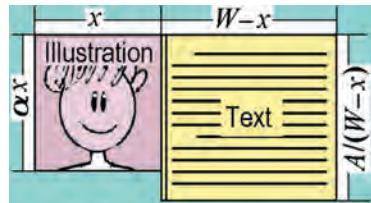
(D.3.1b) What is the car’s speed in miles per hour?

(D.3.1c) How does the expression of (D.3.1a) change if the optical illusion shows a spoke slowly rotating in the “correct” clockwise direction?

A units problem from the writing of this book. The form of Problem (D.3.1) illustrates an ideal situation—both the illustration and the text have the same height, and hence look good together. The following problem, which uses units to check accuracy, asks you to find when this ideal can and cannot be achieved. That is, when is it possible to adjust the widths of the illustration (and hence of the text) to produce equal column heights?

(D.3.2)
Exercise

A page of total width W is made up of two columns. The left column, with an illustration, has width x . The right column, containing text, therefore has width $W - x$. Since the text has a fixed number of words, it is reasonable to assume that for all column widths $W - x$ the text consumes the same area A . To say that the illustration, whether large or small, must always retain the same proportion is to say the ratio of height to width—the **aspect**—is a constant, $\alpha > 0$.



(D.3.2a) Question: Show that the heights of the text and illustration are equal if and only if left-hand column width x has the value

$$x = \left(\frac{W}{2}\right)^2 \pm \sqrt{\left(\frac{W}{2}\right)^2 - \frac{A}{\alpha}}.$$

Note: The figure above shows a right-hand column that has both a width $(W - x)$ and a height $A/(W - x)$ that are greater than the width and height of the left-hand column.

(D.3.2b) Question: What are the units of $\sqrt{(W/2)^2 - (A/\alpha)}$ above? Do the terms, $(W/2)^2$ and A/α , have the same units?

(D.3.2c) Question: Show that a solution, x , given in (D.3.2a), exists if and only if the photo is deeper (taller) than the text when the columns are evenly divided, that is, when each column has width $W/2$.

Hint: Since the square root of a real number exists only for positive real numbers, solution x , given by (D.3.2a), exists if and only if $(W/2)^2 - (A/\alpha) > 0$. Show that this algebraic statement, an inequality, is equivalent to the geometric (visual) statement of (D.3.2c), namely that the photo height is greater than the text height when the two columns are of equal length.

Note: A real-world procedure derived from algebra. An equivalent interpretation of Question (D.3.2c) is that we can easily test in

advance whether or not an equal-height solution is possible. Slide the column divider—the double vertical line of Figure (D.3.2a)—to the mid-point. If, at that test point, the illustration is taller than the text, then no solution is possible. If, on the other hand, the illustration is shorter than, or the same height as the text, then a solution is achieved by sliding the double-line column divider either to the left or to the right until equal heights for the illustration and the text result.



(E) Rings of Functions and Square Matrices

Preview

(E.1) Associative, Binary Operations

- (E.1.1) Given elements x and y . Then a BINARY OPERATION, $*$, acting between pairs of elements x and y produces the BINARY PRODUCT, $x*y$. Sets of objects, like the set of real numbers $\mathbf{R}=\mathbf{R}^1$, have binary operations + (addition) and $*$ (multiplication) that display the following properties:

The ASSOCIATIVITY OF ADDITION AND MULTIPLICATION says that pairwise operations may be associated, or paired off, in any order without changing the resulting sum. That is, for all triples, x, y, z ,

$$\begin{aligned} (\text{E.1.1a}) \quad x + (y + z) &= (x + y) + z, && \text{additive associativity} \\ x * (y * z) &= (x * y) * z. && \text{multiplicative associativity} \end{aligned}$$

IDENTITIES are “feeble” elements for addition and multiplication that leave every element intact. That is, 0 and I are additive and multiplicative identities, respectively, if for all x ,

$$\begin{aligned} (\text{E.1.1b}) \quad 0 + x &= x + 0 = x, && \text{additive identity} \\ I * x &= x * I = x. && \text{multiplicative identity} \end{aligned}$$

INVERSES for a particular x are elements for addition and multiplication that “return” us to the identities. That is, for a given x , elements $\sim x$ and x' are additive and multiplicative inverses, respectively, if

$$\begin{aligned} (\text{E.1.1c}) \quad x + \sim x &= \sim x + x = 0, && \text{additive inverse} \\ x * x' &= x' * x = I. && \text{multiplicative inverse} \end{aligned}$$

(E.1.2)

DEFINITION: A GROUP \mathcal{G} is a set of objects endowed with an associative binary operation \diamond (E.1.1a) so that for every x and y , the element $x \diamond y$ is defined in \mathcal{G} . Moreover, there is a \diamond -identity I (E.1.1b) and every element x has a \diamond -inverse (E.1.1c) x' .

Relating + and *: The distributive property,

The DISTRIBUTIVE LAW is the one rule that relates the two binary operations + and *. This law says that for all x, y, z ,

$$(E.1.2a) \quad x * (y + z) = (x * y) + (x * z).$$

(E.1.2b) **Note:** The symbol * is often denoted with a dot • or a times sign × or nothing at all, so that $(x * y) = (x \cdot y) = (x \times y) = (xy)$.

(E.1.3)



Only two objects can be combined at a time. The addition of numbers is defined only between pairs of objects. Therefore, a triple sum like $x + y + z$ has no meaning until the objects being added are paired off as in (E.1.1a).

(E.1.4)



One advantage of establishing general rules (E.1.2) for a group is that we can answer important questions about uniqueness without having to deal with particular structure of the ring elements.

For example, the multiplicative identity, I , is unique—there is no other element, I' say, of the group, where $I' * A = A = A * I'$ for all A .

Similarly, if element A has a multiplicative inverse A^{-1} as is the case with $n \times n$ matrices, then there is no other matrix in the ring, \hat{A} say, that behaves like A^{-1} —it is not the case that $\hat{A} * A = 1 = A * \hat{A}$ unless $A = \hat{A}$ to begin with.

The following theorems verify these assertions.

(E.1.5)

THEOREM: Any identity I in a ring is unique.

PROOF: Assume a different identity I' exists. Then

$$\begin{aligned} I &= I * I' && \text{from (E.2.1d), } I' \text{ is a "feeble" right identity} \\ &= I' && \text{from (E.2.1d), } I \text{ is a "feeble" left identity.} \end{aligned}$$

Since $I = I'$, there is no alternate identity. This ends the proof. ■

(E.1.6)

THEOREM: If A in a ring has a multiplicative inverse A^{-1} then that inverse is unique.

PROOF: Suppose some A in the ring has two apparently different inverses, A^{-1} and \hat{A} , where

$$(E.1.6a) \quad A^{-1} * A = I = A * \hat{A}.$$

$$\begin{aligned} \text{Then } A^{-1} &= A^{-1} * I && \text{feeble property of } I \text{ (E.2.1d)} \\ &= A^{-1} * (A * \hat{A}) && \text{hypothesis (E.1.6a)} \\ &= (A^{-1} * A) * \hat{A} && \text{associativity (E.2.1c)} \\ &= (I) * \hat{A} && \text{hypothesis (E.1.6a)} \\ &= \hat{A} && \text{feeble property of } I \text{ (E.2.1d).} \end{aligned}$$

Since $A^{-1} = \hat{A}$, there is no alternate inverse. This ends the proof. ■

But there are other sets of objects with these “number-like” properties, one of which is the class of sets called *rings*. We define this class of sets in the next section.

(E.2) Rings over the Real Numbers

(E.2.1)

DEFINITION: A RING over the real numbers is defined to be any set, \mathcal{M} say, of objects, A, B, C, \dots , with the following properties:

(E.2.1a) The set \mathcal{M} is a vector space in accordance with Definition (A.2.1), pg. 360 over the real numbers.

There is a binary multiplication operation $*$ with the following four properties:

(E.2.1b) *-CLOSURE: If A and B are in \mathcal{M} , then so is $A * B$.

(E.2.1c) *-ASSOCIATIVITY: For all A, B , and C in a vector space \mathcal{M} ,

$$(A * B) * C = A * (B * C) \quad \text{as per Def. (E.1.1a)}$$

$$(\alpha \cdot A) * B = \alpha \cdot (A * B). \quad \text{for all scalars } \alpha,$$

(E.2.1d) *-IDENTITY: In accordance with Definition (E.1.1b), there is a “feeble” 2-sided multiplicative identity, I : such that

$$A * I = A = I * A \text{ for all } A \in \mathcal{M}.$$

(E.2.1e) DISTRIBUTIVITY (connecting $+$ and $*$.). For all A, B , and C in the vector space \mathcal{M} ,

$$A * (B + C) = (A * B) + (A * C) \quad \text{and}$$

$$(A + B) * C = (A * C) + (B * C).$$

(E.2.1f) *INVERSES: Some (but not all) elements A in a ring have a two-sided multiplicative identity I as per Definition (E.1.1c). That is,

$$A^{-1} * A = I = A * A^{-1},$$

where $A^{-1} = A'$, the two-sided multiplicative inverse of element A .

(E.3) The Ring of Matrices

Multiplication is possible in \mathcal{M}_{m*n} only when the matrices are square, i.e., only when $m = n$. This follows directly from Definition (A.4.3e) since the product $A_{m*k_1} * B_{k_2*n} = C_{m*n}$ exists only when the “inner” dimensions, $k_1 = n$ and $k_2 = m$, are equal.

Once we have the multiplication, it is straightforward to test $\mathcal{M}_{n\times n}$ against the requirements of Definition (E.2.1) to prove the following theorem:

.....
(E.3.1)

THEOREM: *The set $\mathcal{M}_{n\times n}$ is a ring.*

PROOF: Exercise (E.4.7)

When the identity 1 in $\mathcal{M}_{1\times 1}$ (*which is the same as \mathbf{R}*) is generalized to $\mathcal{M}_{2\times 2}$, the following square matrix emerges:

(E.3.2)

PROPOSITION: *(An identity matrix in $\mathcal{M}_{2\times 2}$) The matrix*

$$(E.3.2a) \quad I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

is a multiplicative identity matrix in the ring $\mathcal{M}_{2\times 2}$.

PROOF: For an arbitrary 2×2 matrix A , use (A.4.3e) to directly verify the equalities,

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{I_2} * \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_A = \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_A = \underbrace{\begin{bmatrix} a & b \\ c & b \end{bmatrix}}_A * \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{I_2},$$

which, according to (E.2.1d), show that I_2 is a multiplicative identity in $\mathcal{M}_{2\times 2}$. ■

$$(E.3.3) \quad \left\| \begin{array}{l} \text{NOTE: The } n \times n \text{ diagonal matrix, } I_n \text{ (to} \\ \text{the right), generalizes } I_2 \text{ (E.3.2a) from} \\ \mathcal{M}_{2 \times 2} \text{ to } \mathcal{M}_{n \times n}. \text{ This is discussed in Ex-} \\ \text{ercise (E.4.3).} \end{array} \right\| I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

(E.3.4)

PROPOSITION: (2×2 Matrix Inverses)

$$(E.3.4a) \quad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{implies} \quad A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

whenever $ad - bc \neq 0$.

PROOF: With matrices, A and A^{-1} given in (E.3.4a), use (A.4.3e), the definition of matrix multiplication, to verify that $A^{-1} * A = I_2 = A * A^{-1}$. The validity of these equalities says that A^{-1} of (E.3.4a) passes the test (E.1.1c) and is therefore a genuine inverse of A . ■

(E.4) Exercises

(E.4.1)

Associativity of Function Composition. Show that the composition of functions operation, \circ , is associative (E.1.1a). That is, for the three functions $f : \mathcal{U} \rightarrow \mathcal{V}$, $g : \mathcal{V} \rightarrow \mathcal{W}$, and $h : \mathcal{W} \rightarrow \mathcal{Z}$, we have $(h \circ g) \circ f = h \circ (g \circ f)$.

Hint: Consider the triple of ordered pairs, (u, v) , (v, w) , and (w, z) where (u, v) , (v, w) , and (w, z) are in functions f , g , and h , respectively. Moreover, the inner entries, v and w are equal. Note that function composition can be interpreted as a “fusing” of inner entries. Thus, $f \circ (g \circ h)$ is constructed by first fusing, or collapsing, the ordered pairs of the function $g \circ h$, then fusing those ordered pairs with the ordered pairs of the function f .

(E.4.2)

Associativity of Vector Addition. Use Definition (A.2.4b) to show that the associative property of numbers (scalars), $x + (y + z) =$

$(x + y) + z$, implies a similar associativity for any triplet of vectors, i.e., \vec{X} , \vec{Y} , and \vec{Z} , in \mathbf{R}^n . That is, show

$$(E.4.2a) \quad \vec{X} + (\vec{Y} + \vec{Z}) = (\vec{X} + \vec{Y}) + \vec{Z}.$$

**(E.4.3)
Exercise**

The $n \times n$ Identity matrix I_n . Use (E.2.1d) to show that the $n \times n$ matrix I_n (E.3.3) is multiplicatively “feeble”. If I_n is an identity, then, from (E.1.5), I_n is the identity.

Hint: Use Theorem (A.4.3e) directly.

**(E.4.4)
Exercise**

A Matrix Without an Inverse. Set $F = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = (f_{i,j})$, for $i, j = 1, 2$. Show that the multiplicative inverse, F^{-1} (E.1.1c), does not exist.

Hint: (i) Take a typical 2×2 matrix, $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, as a candidate for F^{-1} , and use (A.4.3e) directly to show that the products, $F * A$ and $A * F$, can never equal I_2 (E.3.2). Alternately: (ii) Use Proposition (E.3.4) which requires $(f_{1,1}f_{2,2} - f_{1,2}f_{2,1}) \neq 0$.

**(E.4.5)
Exercise**

Why Minus \times Minus = Plus. Given any ring \mathcal{M} with additive identity 0 and multiplicative identity 1. This exercise outlines five steps, (E.4.5b)-(E.4.5f), that lead to the final result, namely for any x and y in \mathcal{M} , we always have

$$(E.4.5a)$$

$$(\sim x) * (\sim y) = x * y.$$

Accordingly, the first step is to show that for any x in \mathcal{M} ,

$$(E.4.5b)$$

$$\sim\sim x = \sim(\sim x) = x.$$

Hint: The definition of additive inverse (E.1.1c) says

$$(i) (x) + (\sim x) = 0.$$

In this equation, replace x with $\sim x$ to get

$$(ii) (\sim x) + (\sim\sim x) = (\sim\sim x) + (\sim x) = 0.$$

The term, $(\sim x)$, appears in both equations, (i) and (ii), from which it would seem that $(\sim x)$ has two (apparently different) inverses, namely, (x) and $(\sim\sim x)$. Use uniqueness (E.4.10) to show that (E.4.5b) is true.

.....

Secondly, prove that zero times anything is zero. That is, for all x in \mathcal{M} , show

(E.4.5c)

$$0 * x = 0 = x * 0.$$

Hint: The distributive law connecting $+$ and $*$ is crucial to this argument. Write

$$0 = \underbrace{1 * x}_{x, \text{ from}} + (\sim x) = (0 + 1) * x + (\sim x) = 0 * x + \underbrace{1 * x + (\sim x)}_0,$$

(E.2.1d)

from which we conclude that (E.4.5c) is valid.

.....

For the third step, prove that the minus sign, \sim , gets “absorbed into” (or “factors out of”) any x in the ring \mathcal{M} . Specifically, show

(E.4.5d)

$$(\sim 1) * x = \sim x.$$

$$\begin{aligned} \text{Hint: Write } (\sim 1) * x &= (\sim 1) * x + \underbrace{x + (\sim x)}_0 \\ &= (\sim 1) * x + 1 * x + (\sim x) \\ &= \underbrace{(\sim 1 + 1)}_0 * x + (\sim x) \end{aligned}$$

which, from (E.4.5c), proves (E.4.5d).

.....

The fourth step is to prove

(E.4.5e)

$$(\sim 1) * (\sim 1) = 1.$$

Hint: This is a special case of (E.4.5d) with $x = \sim 1$. (You will need to use (E.4.5b).)

For the fifth and final step, show that for all x, y , in the ring \mathcal{M} ,

(E.4.5f)

$$(\sim x) * (\sim y) = x * y.$$

Hint: On the left-hand side of (E.4.5f), use (E.4.5d) to replace $(\sim x)$ with its equivalent $(\sim 1) * x$ and, similarly, replace $(\sim y)$ with its equivalent $(\sim 1) * y$. Then, use (E.4.5e).

**(E.4.6)
Exercise**

Why 0^{-1} cannot exist. Show that if 0 is an additive identity (E.1.1b) in a ring, then the multiplicative inverse, 0^{-1} (E.1.1c), cannot exist.

Hint: Show that the equation $0 * 0^{-1} = I$ leads to a contradiction by using (E.4.5c).

**(E.4.7)
Exercise**

The Ring of Square Matrices. Show that for all $n \geq 1$, $\mathcal{M}_{n \times n}$ is a ring in accordance with the axioms given in Definition (E.2.1).

**(E.4.8)
Exercise**

Existence of Additive Inverse Vectors. Show that every vector \vec{X} in \mathbf{R}^n has an additive inverse, $\sim \vec{X}$, where $\sim \vec{X}$ “brings us back” to the identity vector, $\vec{0}$. That is,

$$(E.4.8a) \quad (\sim \vec{X}) + \vec{X} = \vec{0} = \vec{X} + (\sim \vec{X}).$$

Note: We know that an additive inverse, $\sim x$, exists for each number x . That is, for every number x we have a number $\sim x$ such that $(\sim x) + x = 0 = x + (\sim x)$. A similar inverse under multiplication is x^{-1} or $1/x$ for non-zero numbers x . That is, $x^{-1} \cdot x = 1 = x \cdot x^{-1}$ for all non-zero x .

Hint: Use the fact that an additive inverse $\sim x$ already exists for non-zero numbers. Then show that the inverse property (E.4.8a) must hold

for all vectors \vec{X} in \mathbf{R}^n if you define the vector $\sim\vec{X}$ in \mathbf{R}^n properly.

**(E.4.9)
Exercise**

Uniqueness of Additive Identity, $\vec{0}$. Show there is only one additive identity 0.

Hint: First, assume there are two “feeble” elements, $\vec{0}_1$ and $\vec{0}_2$, and consider the sum $\vec{0}_1 + \vec{0}_2$ which, from (E.1.1b), will imply $\vec{0}_1 = \vec{0}_2$.

**(E.4.10)
Exercise**

Uniqueness of Additive Inverses, $\sim\vec{X}$. Show there is only one additive inverse $\sim\vec{X}$. (Existence is established in Exercise (E.4.8). This exercise asks you to establish uniqueness.)

Hint: Suppose there are two (apparently different) additive inverses, $\sim\vec{X}$ and $\sim\vec{X}_0$. Then, consider the sum

$$\sim\vec{X} + \vec{X} + \sim\vec{X}_0.$$

From (E.1.1a), (E.1.1b), and (E.1.1c), it necessarily follows that $\sim\vec{X} = \sim\vec{X}_0$.

(E.4.11)



Note: The same proof for uniqueness of the identity element (or zero) $\vec{0}$, and of the additive inverse vector, $\sim\vec{X}$, applies for many mathematical structures, such as matrices, column vectors, functions, integers, etc. In the proof, we only use the common associative property (E.1.1a), the identity (E.1.1b), and the definition of inverse (E.1.1c).

This is but one example in mathematics where apparently different structures are better understood by studying common properties as opposed to their specific, diverse forms.

(F) The Scientific Method

(F.1) Reality of the Unseen

- (F.1.1) **Reality Revealed by its Effects.** The scientist seeks to understand objects through the effects produced by the hidden hand of Nature. When we see a tree bending “by itself,” we assert that an invisible wind exists.



Atomic Structure Through Indirect Effects

No one has seen an ELECTRON, yet, from observations of its effects alone, we acknowledge its existence. (*For details, see (19.6.3).*) No one has seen an ATOM, but after 1911, when Rutherford offered his explanation of why ALPHA PARTICLES (He^{++} , helium atoms without their electrons) unexpectedly ricochet off a sheet of gold foil, we have accepted as fact that every atom is mostly empty space with electrons orbiting around a small heavy core. (*For more about this experiment, see Exercise (F.12.4).*)

Planet Discovery Through Varying Light Intensity

Even with an optical telescope, no one has seen a PLANET orbiting around a star different from our own sun. Yet, in 2004 [2], a team of researchers, using humble 4-inch telescopes, was able to find a Jupiter-size planet, with a three-day “year,” orbiting around a star that is 500 light years away.

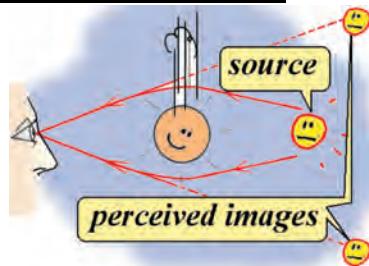
Here are the effects leading to this discovery.

A decrease of 2% in the star’s intensity every three days was measured on photographic plates. From this periodic diminishing of the

light, the existence of the planet, TrES1¹, was inferred. (*This is similar to our inferring the existence of an unseen airplane overhead when we see only its shadow passing over us*).

Planet Discovery Through General Relativity

- (F.1.2) In 2006, ANDREW GOULD [18], working with an international team, used the MICROLENSING effect that was predicted by Einstein's GENERAL THEORY OF RELATIVITY to discover an icy Neptune-sized planet orbiting around a red dwarf star.



Microlensing explains the apparent magnification of one star (*right star in the diagram*) when a space-distorting star (*the large center star in the diagram*) passes in front of it causing the source light to bend. Particular properties observed in the magnified image could only have resulted from the existence of an unseen in-orbit planet.

Planet Discovery Through Star Wobble

- (F.1.3) **Earth-like Planets.** In 2006, a team of European astronomers detected a wobble in a star named HD 69830. From this motion, they inferred the existence of three planets and an asteroid belt circling a Sun-like star situated about 42 light years from Earth in the constellation PUPPIUS. The two innermost planets are probably rocky like Earth, but they live too close to their sun to support life as we know it. (*Could there exist life as we don't know it?*) The third planet is about 18 times the size of Earth and orbits about 60 million miles from its star and is therefore in a theoretical "habitable zone" where temperatures allow existence of liquid water—Earth is about 93 million miles from the Sun. (*See the New York Times article [5] for more detail.*)

¹TrES is an acronym for Transatlantic Exoplanet Survey, and TrES1 is the name of the first planet beyond our solar system that was discovered using small, home-size telescopes with a relatively wideangle. The distance of TrES1 to its home star is about 4 million miles, or about 1/20th the distance between Earth and the sun.

The Economy Through Statistics

(F.1.4)

The Economy: Well-Being and the Invisible. It can be argued that the necessities of well-being—food, shelter, and clothing—are sufficiently concrete so as to be devoid of any abstract component. We can, after all, *taste and smell* our food, *hear* a knock on a wall, and *feel* the texture of cloth. No abstraction there. Yet, the movement and exchange of these concrete goods are somehow governed by background forces of the invisible—the forces we call the ECONOMY.

How do we understand this invisible, over-arching thing called “the economy?” Its reality is expressed only through various measurements of its *effects* — the Gross Domestic Product, the unemployment rate, the number of housing starts, and so on. These numbers are footprints of the economy.

To grasp the reality of our economy is to correctly interpret its numerical footprints—its effects. Plenty of abstraction there.

Geological Events Through Ice Analysis

(F.1.5)

An Ice History. Effects of geological events can be captured by one of nature’s blandest substances—ice. In particular, layers in cylindrical ice cores extracted from Greenlandic or Antarctic ice sheets and the patterns of bubbles captured in these cores are among the natural history markers to be found in the ice. MARIANA GOSNELL, in her book *Ice* [17] describes how the art of scientific inference is applied to these ice cores:



In black ice, bubbles tend to be tubes, not spheres. . . . The more prolonged the cold the longer the tubes will be, and the greater the cold the bigger around the tubes will be. . . . If temperatures should rise while the ice is growing, a tube may narrow or become pinched off. . . . If you sawed pieces of ice out of the core at regular intervals and examined the bubbles in it, you could read the history of weather over a winter at that spot.

(F.1.6) Are Inferences Reliable? In the previous examples, (F.1.1)–(F.1.5), our descriptions (hypotheses) of “reality” come not from direct observations of the reality itself, but indirectly—through inferences we draw from *effects of* these realities. Yet experience seems to confirm that this method *is* effective and reliable. (See (F.10.5), pg. 424, concerning a satirical view of scholarly inference.)

For example, we are extremely confident that Newton’s laws of motion and the sun-centered geometry of the solar system are True. To date, every predicted time and place of a rocket sent into space, every predicted eclipse, and every predicted position of the planets, has confirmed (agreed with) Newton’s theories.

Although confidence increases as more experiments confirm predictions of a hypothesis, confirmations almost never give us 100% certainty.

Question: If confirming experiments only increase our confidence, are all our hypotheses doomed to be tinged with uncertainty?

Answer: Yes, there is always uncertainty whenever you declare a hypothesis to be True, according to KARL POPPER. (*We will see more on Karl Popper later in Section (F.11).*)

However, there is 100% certainty only when a contradictory result occurs, in which case the contradicted hypothesis is certainly False.

(F.1.7) The Scientific Method in a Nutshell. To restate the previous ideas: The SCIENTIFIC METHOD is a process by which we consider a list of observed properties from which we **predict** other properties (or effects), $Q_0, Q_1, Q_2, \dots, Q_i, \dots$; we then infer condition P (the “cause”) from which these properties follow, i.e.,

(F.1.8) (If P then Q_0), (If P then Q_1), \dots , (If P then Q_i), \dots

Confidence in the truth of the hypothesis P increases with each confirming **test via experiment and observation**. But the hypothesis P is rejected with the first contradictory experiment or observation.

(F.2) If-then Sentences

As (F.1.8) suggests, the if-then sentence is an important component of the scientific method. Accordingly, we first set down three equivalent notations for if-then sentences.

(F.2.1) **Notation:** If P then Q , $P \Rightarrow Q$, P implies Q .

(F.2.2)  **It should not be surprising** that there are three notations in (F.2.1) for the if-then sentence. After all, the art of reasoning has historically developed in many areas of thought. Terminology for the same concept has grown to be diverse and non-standard.

Table (F.2.3) following can be regarded as a Rosetta stone that lists various equivalent notations for the if-then sentence.

(F.2.3)

TABLE: If- P -then- Q . (terminology and synonyms).

P	\Rightarrow	Q
inference		implication
cause		effect
assumption	guarantees	result
induction	implies	deduction
premise	predicts	conclusion
hypothesis	requires	consequence
general case		particular case
sufficient		necessary

(F.2.4)

 “Cause” is a tricky substitute for “Sufficient.”: It is often intuitive and effective to think of ($P \Rightarrow Q$) in terms of (Cause \Rightarrow Effect). (*This is the case in Example (F.9.3), pg. 421, and Exercise (F.12.2), pg. 428.*) But cause-and-effect depends on time—first comes the cause, then comes the effect.

Here is an example where “sufficient” does not correspond to “cause.” Consider sentences P and Q given by

P = Life (as we know it) exists at area X .

Q = Liquid water exists at area X .

Then we have the implication,

$$P \Rightarrow Q,$$

which, in natural language, says *Existence of life is sufficient to guarantee that liquid water is necessarily present.*

NOTE: The converse implication, $Q \Rightarrow P$, is False. It is not True that the existence of liquid water is a sufficient guarantee that life necessarily exists.

To sum up: Although existence of life is a sufficient condition to imply the existence of liquid water, the existence of life does not cause the existence of water (although it may cause us to believe in the existence of water—but that is another discussion.).

We conclude this section with definitions of implications related to the form $P \Rightarrow Q$.

(F.2.5)

DEFINITION: Given component sentences P and Q . Then

- $P \Rightarrow Q$ is the **(logical) implication**, P implies Q .
- $Q \Rightarrow P$ is the **converse implication**.
- $\sim Q \Rightarrow \sim P$ is the **contrapositive implication**.

where each sentence, $\sim P$ and $\sim Q$, is the **negative** of sentences, P and Q , respectively.

EXAMPLE: The sentence, P = Today is Tuesday, has for its negative, either of the sentences,

- (F.2.5a) $\sim P$ = Today is not Tuesday, or
 $\sim P$ = It is not the case that today is Tuesday.



(F.3) Property Lists

- (F.3.1) Consider the following bit of folk wisdom, ornamented with the flourishes of natural language.

(F.3.1a)

If *it walks like a duck and talks like a duck,*
then, *as far as I'm concerned, it is a duck.*



Statement (F.3.1a) suggests that it will be useful to catalog sequences of properties, Q_i , (such as walking or talking) that necessarily follow from a tentative hypothesis P (so that $P \Rightarrow Q_i$). This is the motivation for the *property list* that we now define:

|| **NOTE:** The logic and validity of sentence (F.3.1a) is discussed further in (F.7.1)

- (F.3.2)

DEFINITION: Given an object or phenomenon P with properties $Q_0, Q_1, Q_2, Q_3, \dots$. Then the sequence,

(F.3.2a) $[P | Q_0, Q_1, Q_2, Q_3, \dots]$,
(sometimes written as a vertical column),

is a PROPERTY LIST for X .

To say that the property list is **valid** is to say that every implication, $P \Rightarrow Q_i, i = 0, 1, 2, 3, \dots$, is True.

EXAMPLE: Every property list can be interpreted as a list of logical implications, as shown in the following specific case:

(F.3.3)

<i>Property List</i>			<i>Implication Form</i>
If	P	Duck	
then	Q_0	waddles	$P \Rightarrow Q_0$
"	Q_1	quacks	$P \Rightarrow Q_1$
"	Q_2	has feathers	$P \Rightarrow Q_2$
"	Q_3	weighs < 20 lbs.	$P \Rightarrow Q_3$
"	Q_4	weighs > 2 lbs.	$P \Rightarrow Q_4$
"	Q_5	can fly	$P \Rightarrow Q_5$
:	:	:	:

With the structure and notations of the previous sections in hand, we are ready to define and illustrate the scientific method.

(F.4) The Four-Step Scientific Method

In the following, P represents a hypothesis that implies properties Q_0, Q_1, Q_2, \dots . In other words, for $i = 0, 1, 2, \dots$, $(P \Rightarrow Q_i)$ {in equivalent notation, $(P$ implies $Q_i)$ or $(\text{If } P \text{ then } Q_i)$ }.

(F.4.1)

Chart of the Scientific Method

(F.4.1a) Gather Data.....

[(1) OBSERVE: Q_0 , a phenomenon that is judged to be True.

[(2) INFER: (hypothesize, guess) a condition P , from which Q_0 necessarily follows, i.e., $P \Rightarrow Q_0$ is True.

Test and Evaluate hypothesis P as follows:

(F.4.1b) Confirming

[(3) TEST: $P \Rightarrow Q_1$ is True. Experiment reveals Q_1 is True.

[(4) EVALUATE: Tentatively accept P as True. Continue testing.

$\left[\begin{array}{ll} \text{(3) TEST: } & P \Rightarrow Q_2 \text{ is True. Experiment reveals} \\ & Q_2 \text{ is True.} \\ \text{(4) EVALUATE: } & \text{\underline{Tentatively accept } } P \text{ as True. Continue testing.} \end{array} \right.$

⋮ ⋮ ⋮ ⋮

$\left[\begin{array}{ll} \text{(3) TEST: } & P \Rightarrow Q_k \text{ is True. Experiment reveals} \\ & Q_k \text{ is True.} \\ \text{(4) EVALUATE: } & \text{\underline{Tentatively accept } } P \text{ as True. Continue testing.} \end{array} \right.$

(F.4.1c) A FALSIFICATION Occurs

$\left[\begin{array}{ll} \text{(3) TEST: } & P \Rightarrow Q_{k+1} \text{ is True. But experiment reveals} \\ & Q_{k+1} \text{ is False.} \\ \text{(4) EVALUATE: } & \text{\underline{Reject or modify the hypothesis. } } P \text{ cannot be} \\ & \text{True since it implies a false } Q_{k+1}. — \text{we cannot} \\ & \text{have a (True } \Rightarrow \text{ False) situation.} \end{array} \right.$

(F.5) Is X a Duck? Applying the Scientific Method.

(F.5.1) **EXAMPLE:** With (F.3.3) as a resource, we apply the scientific method to the hypothesis,

(F.5.2) $P = X$ is a duck,

which might be formed after a reported sighting of X waddling.

(F.5.2a) Gathering Data

$\left[\text{(1) OBSERVE: } Q_0 = X \text{ waddles. } (Q_0 \text{ is a True statement.}) \right.$

$\left[\begin{array}{ll} \text{(2) INFERENCE: } & \text{Hypothesize } P = X \text{ is a duck. Waddling, } Q_0, \\ & \text{necessarily follows. } P \Rightarrow Q_0 \text{ is a True implication.} \end{array} \right.$

Test and evaluate hypothesis P (F.5.2) as follows:

(F.5.2b) **Confirming**

$\left[\begin{array}{ll} (3)_1 \text{TEST:} & \text{Encourage } X \text{ to vocalize. You hear a quack} \\ & \text{which says } Q_1 \text{ is True. Thus, } P \Rightarrow Q_1 \text{ is True.} \\ (4)_1 \text{EVALUATE:} & \underline{\text{Tentatively accept }} P \text{ as True. Continue testing.} \end{array} \right]$

$\left[\begin{array}{ll} (3)_2 \text{TEST:} & \text{Inspect } X. \text{ You see feathers, so } Q_2 \text{ is True.} \\ & \text{Thus, } P \Rightarrow Q_2 \text{ is True.} \\ (4)_2 \text{EVALUATE:} & \underline{\text{Tentatively accept }} P \text{ as True. Continue testing.} \end{array} \right]$

(F.5.2c) **A Falsification Occurs**

$\left[\begin{array}{ll} (3)_3 \text{TEST:} & \text{Weigh } X. \text{ This experiment shows } X \text{ comes in} \\ & \text{at 203 lbs! Without doubt, } Q_3 \text{ is False. Yet,} \\ & P \Rightarrow Q_3 \text{ is True. (Falsifiability, a notion articu-} \\ & \text{lulated by Karl Popper, is discussed in Section} \\ & (\text{F.11}).) \end{array} \right]$

$\left[\begin{array}{ll} (4)_3 \text{EVALUATE:} & \underline{\text{Reject }} \text{hypothesis } P. \text{ We are now 100\% certain} \\ & \text{that } P \text{ of (F.5.2) cannot be True since, in a} \\ & \text{True and valid implication, } P \Rightarrow Q_3, \text{ where } Q_3 \\ & \text{is False, we cannot have a (True} \Rightarrow \text{False) situation. The duck property list (F.3.3) must be} \\ & \text{abandoned or modified.} \end{array} \right]$



After falsification (F.5.2c), we must abandon or modify the duck property list (F.3.3) and go to step (2) of (F.5.2a) with an alternate working hypothesis P' . For example, we might (re)conjecture

(F.5.3) $P' = X$ is an imposter in a duck suit,

which generates, or is associated with, the new property list:

(F.5.4)

		<i>Property List</i>	<i>Implication Form</i>
If	P'	Duck Impersonator	
then	Q'_0	waddles	$P' \Rightarrow Q'_0$
"	Q'_1	quacks	$P' \Rightarrow Q'_1$
"	Q'_2	has feathers	$P' \Rightarrow Q'_2$
"	Q'_3	weighs ≥ 20 lbs.	$P' \Rightarrow Q'_3$
"	Q'_4	cannot fly	$P' \Rightarrow Q'_4$
:	:	:	:

Note: Properties Q'_0 , Q'_1 , Q'_2 , and Q'_3 , the first four entries of this property list, have already been observed to be True in (F.5.2a) and (F.5.2b).

This ends the description of the scientific method in testing the hypothesis $P = X$ is a duck.

(F.5.5)



Scientists have recently weighed in on this very hypothesis (F.5.2) as we see from these opening lines of an article that appeared in the *New York Times* [25]:

It looked like a duck. It swam like a duck. It is not known if it quacked like a duck.

But it definitely was not a duck.

Filling a gap in the evolution of birds, scientists have dug up fossils of a bird that lived 110 million years ago and looked remarkably like a small modern-day waterfowl.

Since these fossils, discovered near the city of Yumen in the Chinese province of Gansu, come in at the size of a modern-day robin, they fail to meet (among others) the minimum size property, which is on the property list (F.3.3) that certifies duckness.



(F.6) Whence the Scientific Method?

(F.6.1)



How did we come to this place? Since the scientific method (F.4.1) has proved to be so useful, effective and reliable, one might think it has been around for a long time. Not so. Although technology (*how things work*) has been around at least since the discovery of fire, the scientific enterprise as we know it today (*why things work*) began with sixteenth and seventeenth century figures, including SIR FRANCIS BACON (1561–1626) and GALILEO GALILEI (1564–1642).

Francis Bacon was one of the first to advocate the cycle of collecting facts (Step 1 of (F.4.1)), guessing the pattern (Step 2 of (F.4.1)), and experiment (Steps 3,4 of (F.4.1)). Galileo also relied on observation and fact-gathering. However,

(F.6.1a)

RENÉ DESCARTES (1596–1650) had a different view. Instead of starting with observations, he thought one should start with mathematical axioms and our knowledge of God from which the laws of nature can be deduced (F.2.3). (See the book, Isaac Newton by James Gleick [16].)

It is interesting to note that the word “science” did not exist in the times of the ancient Greeks. In fact, in the preface of his treatise on early Greek science, E. R. Lloyd [23] writes:

Science is a modern category, not an ancient one: there is no one term that is exactly equivalent to our ‘science’ in Greek.... But although [Greek terminology] may be used to refer to certain intellectual disciplines which we should think of as scientific, each of them means something quite different from our own term “science.”

The ancient Greeks were struggling to understand the form and substance of scientific inquiry. Lloyd continues:

Indeed, the study of early Greek science is as much a study of the development and interaction of opinions concerning the nature of the inquiry as of the content of the theories that were put forward.

(F.6.2)

Truth by Authority. In place of the scientific method, whereby assertions are tested, the Greeks seemed to rely, to a great extent, on truth by authority.

ARISTOTLE (384–322 B.C.) had been so potent an authority, that (some would say) he retarded the advancement of scientific research for centuries. For example, among his “authoritative,” unchallenged scientific statements are:

- (F.6.2a) Bodies in motion tend to slow down to their natural static state.
- (F.6.2b) A heavy object falls to the ground faster than a lighter one.
- (F.6.2c) Women have fewer teeth than men do.

(See E. R. Lloyd's book, Early Greek Science [23] and [8], pg. 90, item [172] for the source of assertion (c).)

(F.6.2d) The Freedom to Think and to Test. The interplay of social forces goes well beyond the scope of this book, but it is interesting to ponder the question: *What brings a society's leaders to the point of encouraging independent ideas that will almost certainly contradict their own?*

A partial answer may lie in the following comments of JAMES Q. WILSON, in his review of the book *How Religious Tolerance Came to the West* [34].

... material and cultural progress of a nation depends on the creation and maintenance of human freedom and that in turn depends on religious freedom. If you think the state should enforce on others your views about God and his demands it is most unlikely that you will allow the state to recognize freedom of speech or the press or an unfettered right to scientific inquiry.

The primacy of religious freedom is made evident by the fact that in the West, the struggle for religious freedom predicated by about three centuries any discussion of free speech or a free press. Until the first issue was settled, the second would have little meaning. How can people argue over who is to rule if they must first grant the right to rule to people of a chosen faith?

(F.6.3) What does Chart (F.4.1) tell us about the anatomy of the scientific method?

- The process is built on the scaffold of if-then sentences (the logical implications, $P \Rightarrow Q_i$, of (F.2.1)).
- Therefore, to understand Nature through the scientific method, is to understand the (logic of) the if-then sentence.

- To understand the if-then sentence is to understand, among other things, the distinction between *necessary* and *sufficient* which, in turn, often allows us to recognize counterfeit arguments in natural language. (*Examples follow in (F.9.1) and (F.9.2).*)

(F.7) The Logical Implication

(F.7.1) **Reducing Natural language to Symbolic Form:** The first step in understanding sentences in natural language, like (F.3.1a), is to trim the language so as to produce a more diagrammatic, symbolic form. Accordingly, replace the pronoun “it” with the symbolic variable for the unknown, “ X ,” and delete the ornamental phrase, “as far as I’m concerned.”

|| *For better or for worse, we often eliminate ornaments and flourishes of natural language in order to extract the essential meaning of phrases, and to reveal the syntactical structure.*

Sentence (F.3.1a) is made up of component sentences, P and Q , which are symbolized

(F.7.2) $Q = X$ walks and talks like a duck.
 $P = X$ is a duck.

(F.8) Induction vs. Deduction

(F.8.1)  **The table** of (F.2.3) can help us to distinguish between sound-alike questions such as “What are you inferring?” and “What are you implying?” The table can also illuminate the difference in meaning between “necessary” and “sufficient.” In fact, these terms lie at the root of the confusion between “induction” and “deduction,” which we now define.

(F.8.2)

DEFINITION:

INDUCTION or INDUCTIVE REASONING, is the logical process of passing from necessary condition Q to (at least one) sufficient condition P , which is to say

(F.8.2a)

Given Q , find P so that $P \Rightarrow Q$ is True.

DEDUCTION or DEDUCTIVE REASONING, is the logical process of passing from a sufficient condition P to a necessary condition Q , which is to say

(F.8.2b)

Given P , find Q so that $P \Rightarrow Q$ is True.

(F.9) Necessary vs. Sufficient

We expand on the correlation in Table (F.2.3) between “necessary” and “sufficient.” Clarity is the victim when confusion between these two terms becomes so severe that two sentences with different meanings are thought to have the same meaning! Here are examples:

(F.9.1)

EXAMPLE 1: Necessary Confused with Sufficient. For sentences P and Q of (F.7.2), we compare the implication, $P \Rightarrow Q$ and its contrapositive, $Q \Rightarrow P$.

(F.9.1a)	<u>Sufficient</u>	\Rightarrow	<u>Necessary</u>
	(P) X is a duck	\Rightarrow	(Q) X talks, walks like a duck.
	(Q) X talks, walks like a duck	\Rightarrow	(P) X is a duck.

Implication $(P \Rightarrow Q)$ of (F.9.1a) is *always* True. On the other hand, implication $(Q \Rightarrow P)$ is *not always* True—“ X is a duck,” is not a necessary consequence of statement Q , which says X waddles like a duck and quacks like a duck. (*See also* (F.2.3)).

Implications $(P \Rightarrow Q)$ and $(Q \Rightarrow P)$ reverse the roles of *sufficient* and *necessary*. They carry very different messages, and hence are

distinctly different implications.

The advocate of our original statement, $Q \Rightarrow P$ of (F.3.1a), is assuming, erroneously, that what is necessary (sentence Q) is also sufficient (sentence P).

(F.9.2) **EXAMPLE 2: Necessary Confused with Sufficient.**

Spoken Language.

In political discussions, we often hear arguments like the following, between party (A) and party (B):

- (F.9.2a) | (A) We need funding if the schools are to do better.
 | (B) You people always have the same response to
 | a problem — just throw money at it.

Symbolic Language.

If we paraphrase—replace the statement, “the schools do better”, with “problem gets solved,”—and if we distinguish between *sufficient* and *necessary* conditions, then statements (F.9.2a) become

- (F.9.2b) | **Sufficient** \Rightarrow **Necessary**
 | (A) Problem gets solved \Rightarrow Money is spent.
 | (B) Money is spent. \Rightarrow Problem gets solved

(F.9.2c) **Analysis of Counterfeit Disagreement (F.9.2a):**

Party (A) is saying “I believe money is *necessary*,” while party (B) is saying “I don’t believe money is *sufficient*.” Both parties believe they are arguing for or against a single proposition when, in fact, they are stating different propositions. Ironically, they may *both be right*—it could well be that money is *necessary and not sufficient* to solve the problem.

(F.9.3) **EXAMPLE: Diagnosis is Almost Always Induction.** As you try to understand why your car will not start, you may hypothesize either

- (F.9.3a) | P = the car’s battery is dead, or
 | P' = the fuel tank is empty, or
 | P'' = an ignition wire has been cut.

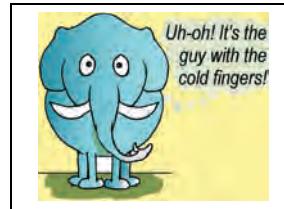
(F.9.4)



Induction and its uncertainty—a prickly friend. Inductive reasoning (see (F.8.2a)) almost never gives us 100% certainty. As we see in example (F.9.3a), we infer not one, but three plausible hypotheses.

For another example, consider the poem by JOHN GODFREY SAXE (1816–18887), inspired by a classic Indian folk tale, in which six blind men of Indostan

... went to see the Elephant
(Though all of them were blind),
That each by observation
Might satisfy his mind.



After feeling different parts of the elephant, the blind men disagree about their findings. One man feels the elephant's *side* and infers that the elephant is a *wall*. Feeling the *tusk*, another says the elephant is a *spear*. Feeling the trunk, the third man infers a *snake*. From the knee, the inference of a *tree* results. The fifth man feels the ear, and the elephant is declared to be a *fan*. Feeling the tail, the inference of the sixth man is that the elephant is a *rope*. Saxe's poem concludes,

... [the disputants]
Rail on in utter ignorance
Of what each other mean,
And prate about an Elephant
Not one of them has seen.

(F.10) Uncertainty, Popper, and Derrida.

Built-in uncertainty troubled some philosophers.

(F.10.1)



EXAMPLE: Certainty only through contradiction. KARL POPPER, for example, asserted that induction, with its inability to provide complete confidence in a hypothesis, can not be a legitimate tool of scientific inquiry. (*Non-scientific questions are another matter—many scientists acknowledge that scientific inquiry is not the only route to understanding the human condition.*) Note that complete certainty occurs in step (4) of (F.4.1c), when, without doubt, a negative result invalidates (or contradicts) the hypothesis. For more on Popper's ideas on *falsifiability* and *demarcation*, see (F.11.1) and (F.11.2).

.....

It could be worse. Unavoidable uncertainty can reenforce the belief among some that, since we don't know everything with 100% confidence, we, therefore, know nothing.

(F.10.2)

EXAMPLE: Can we know anything? JACQUES DERRIDA (1930–2004) was the father of the *deconstruction* method of inquiry which asserts that language is so replete with confusion and contradiction that Derrida himself could not clearly articulate and explain his *own ideas*.

As if to prove his point, Derrida replied, when asked by the New York Times reporter, DINITIA SMITH, to define *deconstruction*, “It is impossible to respond. I can only do something which will leave me unsatisfied.”[22].

(F.10.3)



EXAMPLE: The Marbleized-Bowling-Ball Pattern of Clouds.



After thousands of daily weather observations collected over the centuries, we might expect Earthbound mortals to have great success in inferring how cloud patterns would appear to an observer at a remote point in space.

If so, then, why is it, that only *after* NASA's astronauts took photographs of Earth from space, do we see *for the first time*, clouds in their now-familiar marbleized bowling ball pattern? (*On Christmas of 1968, APOLLO 8 crew members, FRANK BORMAN, JAMES*

LOVELL, JR., and WILLIAM ANDERS, captured the first image of Earth as a full globe in space.)

Before the NASA photos from space were released, all man-made illustrations of the Earth's clouds had shown them like puffs of cotton, in neat, fluffy patterns (*left panel of (F.10.3)*). In popular culture, clouds were never before shown as swirling wisps of vapor (*right panel of (F.10.3)*).

In fact, you can see how Earth clouds were misrepresented in the January 26, 1967 (Stardate 3113.2) STAR TREK episode, *Tomorrow is Yesterday*.

(F.10.4)



EXAMPLE: Celestial Over-Inference: If we put too little into our inferred pictures of the daytime sky, as in Example (F.10.3), then it seems that we put too much into our inferred pictures of the night time sky.

THE BIG DIPPER is seen at night as a collection of stars strewn on a flat, two-dimensional canvas, where each of the seven stars is a point of light that is equidistant to our position on Earth. Yet, the stars of the Big Dipper have distances from Earth that range from 68 light years to 210 light years.

ORION is a constellation that is interpreted as a hunter wearing his well-known and well-defined belt along with a pair of magnificently hand-tooled leather boots. How disappointed and *traumatized* might a curious child be after looking up at the night sky—only to find a few sparsely placed dots of light instead of the anticipated, richly-detailed image of Orion in all his sartorial splendor.



(F.10.5)

Archaeology and Inference. Hypothetical examples of over-inference are presented in the satirical book, *Motel of the Mysteries*, by

DAVID MACAULAY [24] in which archaeologists of the distant future (mis)interpret the buried ruins of a contemporary American motel. In one example, a telephone receiver is declared by the scholars to be a ritualistic percussion instrument used to create god-pleasing rhythms from the bell inside the larger half which rings whenever the worshiper bangs the two “instrument” halves together.

(F.11) Implications and Falsifiability of Karl Popper

(F.11.1) Falsifiability and Karl Popper



The logical structure of $P \Rightarrow Q$, the implication (F.2.5), is the basis of what philosopher Karl Popper (1902–1994) called FALSIFIABILITY which, he asserts, is the DEMARCATON boundary between science and non-science.

.....

Here are the details. When we test a hypothesis P in a scientific theory, we write the implication

(F.11.2) $P \Rightarrow Q$ that we know is True.

(F.11.2a) Case 1: An experiment produces phenomenon Q . But can we be sure that it is hypothesis P and no other hypothesis that produces phenomenon Q ? Popper says that such confirming experiments cannot give us certainty, although overconfidence that P is True increases with each confirming experiment.

(F.11.2b) Case 2: An experiment produces $\sim Q$, where (F.11.2) predicts Q . We have a contradiction that gives us, with certainty, that hypothesis P is undoubtedly not True. This is the FALSIFIABILITY of Popper.

.....

(F.11.3) Popper’s book, *Conjecture and Refutation: The Growth of Scientific Knowledge* [29], at the very outset, in Chapter I, pg. 47–8, phrases these ideas as follows:

- (1) It is easy to obtain confirmations, or verifications, for nearly every theory—if we look for confirmations.

- (2) Confirmations should count only if they are the result of risky predictions, that is to say, if unenlightened by the theory in question, we should have expected an event which was incompatible with the theory—an event which would have refuted the theory.
- (3) Every “good” scientific theory is a prohibition: it forbids certain things to happen. The more a theory forbids, the better it is.
- (4) A theory which is not refutable by any conceivable event is non-scientific. Irrefutability is not a virtue of theory (as people often think) but a vice.
- (5) Every genuine test of a theory is an attempt to falsify it, or to refute it. Testability is falsifiability; but there are degrees of testability; some theories are more testable, more exposed to refutation, than others; they take, as it were, greater risks.
- (6) Confirming evidence should not count except when it is the result of a genuine test of the theory; and this means that it can be presented as a serious but unsuccessful attempt to falsify the theory. (I now speak in such cases of “corroborating evidence.”)
- (7) Some genuinely testable theories, when found to be false, are still upheld by their admirers—for example by introducing ad hoc some auxiliary assumption, or by re-interpreting theory ad hoc in such a way that it escapes refutation.

One can sum up all this by saying that the criterion of the scientific status of a theory is its falsifiability, or refutability, or testability.

(F.11.4) **EXAMPLE: Falsification.** You make the observation

(F.11.4a) $Q =$ A bowling ball always falls faster to Earth than a feather does.

and you create the hypothesis

(F.11.4b) $P =$ Heavy objects always fall to Earth faster than lighter ones.

Observation Q is experimentally True, but is the more general hypothesis P True? With only one confirmation, we can't be certain. We might measure the descent rate of another pair of objects (a piece

of paper against a heavy rock, say) and, with a second confirmation, our confidence increases.

But if we experiment with these same objects *in a vacuum*, we observe something different:

$$Q' = \text{In a vacuum, a feather and a bowling ball fall to Earth at the same rate}$$

which is a falsifying contradiction to the property (F.11.4a) predicted by hypothesis (F.11.4b) (*see Figure (1.1.2) for a visual explanation of why Q' is True.*)

At this point, we see with our own eyes, that in a vacuum, Q of (F.11.4b) is False. This brings us to invoke (F.11.2b), Popper's rule of *falsifiability*, which guarantees that hypothesis P of (F.11.4b) is False.

(F.12) Exercises

(F.12.1) Falsifying the Flat Earth Theory. We observe

Q_0 = the surface of any lake is flat,
so we form (guess) the hypothesis

P = the entire Earth is flat.

This inference about the Earth gives us the partial property list

[$P : Q_0$] where P = The Earth is flat.

Q_0 = Any lake is flat.

(See (F.3.2) for the definition of a property list.)

We write the property list in equivalent symbolic form according to the format of section (F.5) as follows:

$$\left| \begin{array}{ll} Q_0 & \text{is True (lake } X \text{ is observed to be flat) and} \\ P \Rightarrow Q_0 & \text{is True (if all of Earth is flat, then any part of its} \\ & \text{surface (e.g., a lake) is flat.)} \end{array} \right. \dots \dots \dots$$

(F.12.1a) Question: Assume you are burdened with the inaccuracy of measurements that existed centuries ago. Then describe ex-

periments that produce observations Q_1 , and Q_2 , etc., each of which (seemingly) confirms the hypothesis P . That is, generate the property list $[P : Q_0, Q_1, Q_2, \dots]$ so that $P \Rightarrow Q_i$, $i = 0, 1, 2, 3, \dots$, are True implications (since observations Q_i , $i = 0, 1, 2, 3, \dots$, are apparently True).

Hint: Consider measuring, as best your instruments can, the “curve” of the horizon, or the distance between tall vertical buildings at the top point and at the ground.

(F.12.1b) Question: Even with inaccurate measurements, what experiment would produce an observation Q that contradicts (falsifies) hypothesis P ?

Hint: See Exercise (F.12.7).



This exercise highlights another problem in determining the certainty of our results—the accuracy of our measurements and the reliability of our five senses. It may seem that all lakes have flat surfaces, but they don’t. It may seem that the night sky is a huge black shell with pinpoint holes (stars), but it isn’t.

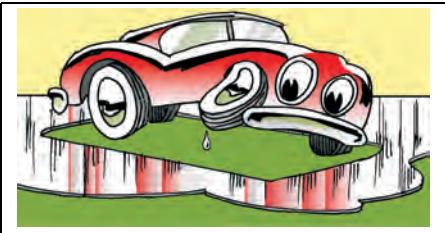
(F.12.2)

Exercise

A Dry Spot in Search of a Cause.

Observation: It is sunny when you park your car to dine at an indoor restaurant. Returning to your car, you notice

Q_1 = There is a dry spot under your car.



There are two hypotheses, namely,

P = Someone with a hose came by and sprayed your car.

P' = It rained.

.....

(F.12.2a) Question: (Refute hypothesis P) What experiment would produce an observation Q that would falsify hypothesis P ?

.....

(F.12.2b) Question: (Confirm hypothesis P') You want to increase confidence in the truth of the alternative hypothesis

$P' =$ A rainstorm caused the dry spot.

What observable effects Q_1 , Q_2 , etc., of a city-wide rainstorm can you predict?

Hint: Consider experiments that reveal facts that would be observable after a rainstorm but not with someone secretly hosing down your car.

(F.12.3) Aristotle's Method. Consider the three assertions of (F.6.2).

Exercise

(F.12.3a) Question: What steps of the scientific method (F.4.1) did Aristotle not employ when he arrived at these three assertions? (*To put it another way, what steps of (F.4.1), had Aristotle used them, would have saved him from making these erroneous statements?*)

(F.12.4) The Nucleus of the Atom.

In 1911, ERNEST RUTHERFORD announced results of his experiments that showed the atom was composed of mostly empty space in which electrons orbited around a small heavy core (nucleus). Previously, it was thought that each atom was like a solid raisin pudding with the electrons being like the embedded raisins.

NOTE: The raisin pudding model is a reminder of how false models can be consistent with observation. For example, Descartes assumed, falsely, that light increased its speed as it passed from air to water — and yet, with this false assumption, he derived Snell's law (see (24.6)).

The Experiment: Here is a description of Rutherford's experiment whose results would contradict the raisin pudding model of the atom.

Rutherford aimed a stream of alpha particles (He^{++} , helium atoms without their electrons) at a sheet of gold foil. Most of the alpha particles passed right through, which was not a surprise. But a small, significant number of alpha particles bounced back, which *was* a surprise. In Rutherford's words, it was like a naval shell being reflected back by a piece of tissue paper.



(More details of the experiment are given in (F.12.4d). In his book, *A Force of Nature* [30], Richard Reeves details his experiences as he reproduces Rutherford's experiment. Also, see Cathcart's book, *The Fly in the Cathedral: How a Group of Cambridge Scientists Won the International Race to Split the Atom* [4].)

.....
Given the accepted hypothesis of the 1900s, namely,

(*) $P =$ The gold atom is a uniformly dense “pudding” with embedded electrons having tiny mass.

(F.12.4a) Question: If P is True, then when gold foil is bombarded with alpha particles, what deductions, Q_i , would belong to the property list [$P : Q_1, Q_2, Q_3, \dots$]? That is, what predictions, Q_i , would describe what occurs after alpha particles are sprayed into gold foil that has “pudding” atoms?

.....
Given the laboratory observation

(**) $Q_0 =$ Most alpha particles pass through the gold foil undeflected, but a tiny number of particles always bounce back at various angles.

(F.12.4b) Question: Since observation Q_0 of (**) is True, is the implication $P \Rightarrow Q_0$ True with the hypothesis P given by (*)?

.....

(F.12.4c) Question: Since, as Rutherford observed, Q_0 is True, then what other phenomena or properties, Q_1, Q_2, Q_3, \dots , are consistent with the “non-pudding” Rutherford model of the atom?

.....



(F.12.4d) Historical Note: Since 1908, Rutherford had been shooting alpha particles through sheets of various materials. Fruitful results occurred with gold foil, having a thickness of about 1/50,000th of an inch, or about 2,000 gold atoms. Most alpha particles passed through the foil and registered on a photographic plate behind, suggesting there was a lot of space in the gold. But some alpha particles (about 1 in 8,000) bounced off at an angle and even ricocheted backwards. Using inductive reasoning (F.8.2a), Rutherford inferred that the atom was not uniform but that most of its mass was concentrated in a very small space.

It took until 1911 for Rutherford to develop sufficient statistical analysis to justify his non-pudding, nucleus theory of the atom. He calculated the ratio of the size of the nucleus to the gold atom to be 100,000:1, or about the size of a football field to the size of the football.

(F.12.5) Exercise

Popper’s Swans: Given observations Q_i from which hypothesis P is inferred, where

P = All swans are white.

Q_i = The i th observed swan is white, $i = 1, 2, 3, \dots$

(F.12.5a) Question: Is the property list $[P : Q_1, Q_2, Q_3, \dots]$ valid in the sense that every implication $P \Rightarrow Q_i$ is True? (*If so, then the valid property list would tentatively confirm hypothesis P .*)

.....

(F.12.5b) Question: In the context of Popper’s falsifiability (F.11.1), what single observation would falsify hypothesis P ?

(F.12.6) A Weakness in Popper's Argument? (We use Exercise (F.12.5) as an example.) Popper says, that in science,

- (1) Hypothesis P (e.g., all swans are white) is only *tentatively* True in spite of successive confirming observations, and
- (2) Hypothesis P is certainly False in the event of a single contradiction (e.g., one grey swan is observed).

(F.12.6a) Question: Isn't there a contradiction to Popper's assertion that no hypothesis P can be declared True with 100% certainty, yet its negative, $\sim P$, can be declared False with 100% certainty? If hypothesis P is certainly False, then isn't $\sim P$ (not every swan is white) a hypothesis that is certainly True?

Hint: You may find justification for Popper's "weakness" among the seven points of his own summary in (F.11.3). Also, think of a valid and viable hypothesis in terms of predictions. That is, how many "interesting" predictions do you get from the hypothesis (not every swan is white.)? It is a philosophical and subjective point of view to take, namely, that a "good" hypothesis cannot be a one-trick pony. (For example, $P = \text{This page is white with black words printed all over it.}$)

(F.12.6b) Question: In light of the previous question and the hint, state some "bad" hypotheses and some "good" hypotheses.

(F.12.6c) Question: In your view, do Newton's laws of motion represent "good" or "bad" hypotheses?

(F.12.7) Is the Earth Cylindrical or Spherical? A person of height h observes a ship leaving port. The ship has height S above the waterline.

(F.12.7a) Question: How far must the ship be from shore at the moment a person sees the mast disappear below the horizon? (If you prefer numbers instead of letters, set the person's height $h = 6$ feet, the ship's height $S = 40$ feet, and the radius of the Earth to be $R = 8,000$ miles.)

.....

(F.12.7b) Question: If observations Q_i , $i = 1, 2, 3, \dots$, lead to the inferred hypothesis P where

Q_i = Every departing northbound ship S_i eventually drops below the horizon.

P = The Earth is cylindrical around an East-West axis.

then is the property list $[P : Q_1, Q_2, Q_3, \dots]$ for hypotheses P valid in the sense that every implication, $P \Rightarrow Q_i$ is True? (Recall, all observed departing ships are *northbound* only.)

.....

(F.12.7c) **Question:** What observation (or experiment) would falsify hypothesis P of (F.12.7b)?

.....

(F.12.7d) **Question:** Using sightings of departing ships only, what set of observations, Q'_i , $i = 1, 2, 3, \dots$, would confirm (give a high level of confidence to) the new hypothesis

P' = The Earth is round (spherical)?

(F.12.8) Exercise

Sherlock Holmes. One day, Sherlock Holmes and Dr. Watson received a surprise visitor. Before the visitor could introduce himself, Holmes immediately declared, “Aha, my good man. You took the 9:42 train from Bath this morning and arrived at Victoria Station. From there, you took one of Mr. Tattersworth’s carriages before arriving here at Baker Street.”

When Dr. Watson and the astonished visitor asked Holmes to explain how he knew these facts, Holmes replied,

“I see by the mud on your shoes that you were in Bath last night. Only Bath has that colour of mud and I know it rained there heavily just last night. There is only one train from Bath that would bring you to London at this time — the 9:42, and that train arrives at Victoria Station.”

“In addition, the coach of Mr. Tattersworth is the most likely one you would



have taken from Victoria Station since he has far more coaches in supply there than does any one else.”

“After a moment, Holmes responded to Watson’s open-mouthed admiration with, “Elementary deduction, my dear Watson — elementary deduction.”

(F.12.8a) Question: Was Sherlock Holmes correct in describing his reasoning as “deduction?”

(G) Logic of the Scientific Method

(G.1) Implications Built from P, Q

This section focuses on the True or False values (TRUTH VALUES) of individual sentences and the resulting Truth values of compound sentences composed from these smaller sentences.

For example, the compound sentence P implies Q (*written* $P \Rightarrow Q$) is built from the two component sentences, P and Q . Given all four combination of Truth values for the pair of sentences, P , Q , we list the four corresponding Truth values for the compound sentence $P \Rightarrow Q$.

(G.1.1) Truth Table for $P \Rightarrow Q$.

This table shows that the only instance for which an implication $P \Rightarrow Q$ is False is when a True P implies a False Q (line (b)). The other three implications of lines (a), (c), (d) are all True, which, as we shall now see, reflects the way we assign Truth in natural language.

	P	Q	$P \Rightarrow Q$
(a)	T	T	T
(b)	T	F	F
(c)	F	T	T
(d)	F	F	T

input pair implication

(G.1.2) How Truth Tables Reflect Language: We illustrate how the truth values of (G.1.1) coincide with the way we assign Truth to $P \Rightarrow Q$ in natural language for the special case of sentences P and Q following:

(G.1.3)

- $P = X$ is a horse.
 $Q = X$ has four legs.
 $\sim P = X$ is not a horse.
 $\sim Q = X$ does not have four legs.



(G.1.4) Examples that Confirm Truth Values of Table (G.1.1).

Line (a): (True \Rightarrow True) is True. For the *same X* in (G.1.3), if P is True, then Q is always True. Implication $P \Rightarrow Q$, in natural language, becomes, “If X is a horse, then X has four legs.” Therefore, the implication $(\text{True} \Rightarrow \text{True})$, is always True. The truth value triple T,T,T, of Line (a) in (G.1.1), is justified.



(G.1.4a) Contrast this with the converse implication, $Q \Rightarrow P$. For the same X in both sentences, P and Q (as when X is a fox), P can be False and Q can be True. The converse implication, $Q \Rightarrow P$, does not always provide a $(\text{True} \Rightarrow \text{True})$ situation.

.....

Line (b): (True \Rightarrow False) is False. If X is a horse (P is True), then we cannot claim that X does not have four legs, i.e., we cannot claim that Q is False. To make such a claim, that any horse will *not* have four legs, is faulty. That is, the implication itself, $P \Rightarrow Q$, is False. In general, $(\text{True} \Rightarrow \text{False})$ is False. The truth value triple T,F,F, of line (b) in (G.1.1) is justified.

.....

**(G.1.4b)**

Line (c) $(\text{False} \Rightarrow \text{True})$; Line (d) $(\text{False} \Rightarrow \text{False})$: In these two cases of (G.1.1), we start with a premise \overline{P} whose truth value is False. From a False premise however, sound logic can produce any conclusion Q , be it True or False. For example,

- If You are the King of France (a False statement), then I can lift 400 pounds (could be either True or False). The logic of this if-then sentence is valid.
- From computer folklore, where the logic of the computer is valid, we hear, “Garbage in, [implies] garbage out”.

In other words, $(\text{False} \Rightarrow \text{True})$ and $(\text{False} \Rightarrow \text{False})$ are both True implications, even though some component sentences are False.

.....

(G.1.5)



We have seen two ways in which the truth values of component sentences P and Q determine the truth value of the implication, $P \Rightarrow Q$, namely,

- In symbolic language, by constructing the truth table (G.1.1) or
- In natural language, as we have done in (G.1.3).

We shall expand upon the structure of (G.1.1) to see how the “mechanical” use of the truth tables can give us insight into the relationship and equivalence of sentences in natural language that might not be clear at first. For example, as Theorem (G.2.1) will tell us, $P \Rightarrow Q$ carries the same meaning as $\sim Q \Rightarrow \sim P$.

(G.1.6)

Implications Built from P and Q . Consider the following truth table (G.1.7) for the three implications, $(P \Rightarrow Q)$, $(Q \Rightarrow P)$, and $(\sim Q \Rightarrow \sim P)$ that were defined in (F.2.5).

There are truth tables for sentences other than sentences of the form $(P \Rightarrow Q)$. Compound sentences such as $(P \vee Q)$ (P or Q) and $(P \wedge Q)$ (P and Q) are discussed in Exercise (G.4.1).

(G.1.7)

Truth Table for $P \Rightarrow Q$, $\sim Q \Rightarrow \sim P$, and $Q \Rightarrow P$.

	P	Q	$\sim Q$	$\sim P$	$P \Rightarrow Q$	$\sim Q \Rightarrow \sim P$	$Q \Rightarrow P$
(a)	T	T	F	F	T	T	T
(b)	T	F	T	F	F	F	T
(c)	F	T	F	T	T	T	F
(d)	F	F	T	T	T	T	T
input pair		(1)		(2)		(3)	
implication		contra positive		converse			

The truth values in columns (1), (2), and (3) follow the simple rule expressed in the commentary of (G.1.1), namely, that an implication is always True except for the one case where True \Rightarrow False.

(G.2) Equivalence of Implications

The following theorem says the two implications $P \Rightarrow Q$ and $\sim Q \Rightarrow \sim P$, although different in form, are equivalent in substance. That is, they share the same truth values and meanings.

(G.2.1)

THEOREM: *For any sentences P and Q ,*

$$(P \Rightarrow Q) \text{ equals } (\sim Q \Rightarrow \sim P)$$

in the sense that they are different forms of the same statement.

It has already been noted in (F.9.1a) and (F.9.2b) that implication $(P \Rightarrow Q)$ and its converse, $(Q \Rightarrow P)$, are not always equal.

PROOF:

The Truth Table Argument. Columns (1) and (2) of Table (G.1.7) tell us that the truth values for implications $P \Rightarrow Q$ and contrapositive $\sim Q \Rightarrow \sim P$ agree.

But if truth values agree (for the same truth values of P and Q), then how do we argue that the *meanings* of these two implications agree?

The Verbal Argument: Showing that the meanings of $P \Rightarrow Q$ and $\sim Q \Rightarrow \sim P$ are the same, will depend on our use of natural language. The most direct (albeit non-rigorous) way to argue this is to recognize the pattern

$$\begin{array}{ll} \text{(G.2.1a)} & \frac{\text{Sufficient} \Rightarrow \text{Necessary}}{P \Rightarrow Q} \end{array}$$

where P on the left-hand side of the implication is a *sufficient* condition with a *necessary* consequence Q on the right-hand side. (*The many labels in natural language for the roles of P and Q are summarized in Table (F.2.3).*) But to say condition Q is *necessary* is to say Q is essential, or indispensable. In other words: without Q , there is no P , which in symbolic language, is expressed as

$$\begin{array}{ccc} \text{(G.2.1b)} & \sim Q & \Rightarrow & \sim P \\ & \text{without } Q & \Rightarrow & \text{we don't have } P. \end{array}$$

Based on our understanding of the word “necessary,” in natural language, we have shown that the meanings of (G.2.1a) and (G.2.1a) are the same. The proof is done. ■

- (G.2.2) **EXAMPLE: Equivalent Implications.** As an application of Theorem (G.2.1), consider the special case where sentences P and Q are given by (G.1.3). Then the following implication (a) and its contrapositive (b) carry the same message:

- (a) $(P \Rightarrow Q)$: X is a horse $\Rightarrow X$ has four legs.
 - (b) $(\sim Q \Rightarrow \sim P)$: X does not have four legs. $\Rightarrow X$ is not a horse.
-

- (G.2.3) **EXAMPLE: Inequivalent Implications.** To see that the meaning of $P \Rightarrow Q$ is not necessarily the same as that of the converse, $Q \Rightarrow P$, take P and Q from (G.1.3) to write

- (a) $(P \Rightarrow Q)$: X is a horse $\Rightarrow X$ has four legs.
- (b) $(Q \Rightarrow P)$: X has four legs $\Rightarrow X$ is a horse.

Implication (a) is always True no matter what X is chosen, even though component P may be False. Implication (b), on the other hand, is False only when it becomes a ($\text{True} \Rightarrow \text{False}$) implication—and this occurs whenever X is chosen to be any four-legged creature that is not a horse. (See (G.1.3).) These two implications, (a) and (b), do not carry the same meaning.

- (G.2.4) **EXAMPLE: Tautologies.** When we have

- (G.2.4a) When $P \Rightarrow Q$ is equivalent to its converse, $Q \Rightarrow P$, we write $P \Leftrightarrow Q$.

Display (G.2.4a) says that both implications have the same truth value—it does not say that the individual component sentences, P

and Q are each True. However, the component sentences, P and Q , are equivalent in the sense that they share the same truth value—they are either *both* True or *both* False.

This agreement (G.2.4a) is called a **tautology**. In the situation of tautologies (G.2.4a), we use the phrase,

(G.2.4b) “ P is True **if and only if** Q is True.”

To see why (G.2.4a) says the same thing as (G.2.4b), see Exercise (G.4.2b).

.....

Unlike example (G.2.3), tautologies do not allow implications $P \Rightarrow Q$ where P is True while Q is False. (See Exercise (G.4.2).) Here is a specific example of a tautology:

Let $P = X$ was elected mayor.

$Q = X$ received at least 50% of the votes.

- (a) $(P \Rightarrow Q)$: X was elected mayor.
 $\Rightarrow X$ received at least 50% of the votes.
 - (b) $(Q \Rightarrow P)$: X received at least 50% of the votes.
 $\Rightarrow X$ was elected mayor.
-

(G.3) Proof by Contradiction

(G.3.1) **Proof by Contradiction** is a technique in which one constructs a hypothesis P from which consequence Q should result. But, through experiment or observation, it is seen that the *opposite* occurs (*call it $\sim Q$*). This falsification tells us that the original hypothesis P is False ($\sim P$ is True).

(G.3.2) **A Formal Outline of Proof by Contradiction:**

(G.3.2a) Pose hypothesis P (*Truth value unknown*)

(G.3.2b) Find conclusion Q so that ($P \Rightarrow Q$) is True.

(G.3.2c) Observe contradiction $\sim Q \dots (\sim Q)$ is True.

(G.3.2d) Observation $\sim Q$ contradicts the expected conclusion, Q . Therefore, hypothesis P is falsified or negated, i.e., \dots necessarily, (P) is False, or $(\sim P)$ is True.

Proposition (G.3.3) following justifies (G.3.2d).

(G.3.3)

PROPOSITION: If $(P \Rightarrow Q)$ is True and $(\sim Q)$ is True, then necessarily,

(G.3.3a) $(\sim P)$ is True (equivalently, (P) is False).

PROOF: From the fourth and second columns of the master truth table (G.1.7), pg. 437, we see, respectively, that

$(P \Rightarrow Q)$ is True (T) in rows { (a), (c), (d) }, and
 $(\sim Q)$ is True (T) in rows { (b), (d) }.

Row (d) is the only row in which both $(P \rightarrow Q)$ and $(\sim Q)$ are True as required by the assumptions of this proposition. Therefore, it is only from this common row (d) that we need to read the truth value of hypothesis (P) , which is True ($(\sim P)$ is False). This justifies (G.3.3a) and the proposition is proved. ■

(G.3.4)

To illustrate the four steps of template (G.3.2), imagine Sherlock Holmes and Chief Inspector Twiddleham walk into a deserted warehouse in pursuit of a criminal. The dubious Twiddleham pessimistically tells Holmes that the search will be in vain. In fact, Twiddleham has a hypothesis called P (for *pessimism?*) where

(G.3.4a) $P =$ No one has been in this deserted warehouse for several months.....(Truth value unknown).

Using Twiddleham's hypothesis, Holmes considers a consequence Q of Twiddleham's hypothesis, thereby creating the True implication $P \Rightarrow Q$ where

(G.3.4b) $Q =$ The pull-chain on every desk lamp in this warehouse is stationary. ($P \Rightarrow Q$) is True.

Holmes and Twiddleham enter a dark office of the warehouse where a desk lamp with a pull-chain rests on the executive desk.

“Aha, my good fellow,” Holmes exclaims to Twiddleham, “Your hypothesis P (*no one is here*) assuredly implies consequence Q (*the pull-chain should be stationary*). Yet, what you see with your own eyes, my dear fellow, is $\sim Q$ (*the pull-chain is swinging*) which is a contradiction to Q . Specifically,



(G.3.4c) $\sim Q =$ The pull-chain of the desk lamp is swinging, it is *not* stationary.” ($\sim Q$) is True

Holmes continues,

(G.3.4d) “Twiddleham, my observed phenomenon, $\sim Q$, contradicts or falsifies the conclusion Q that is expected from your hypothesis P , which, in turn, falsifies your hypothesis P . That is, $\sim P$ is True, which is to say, Someone has been here very recently.”

..... ($\sim P$) is True

Using Proposition (G.3.3), Holmes finally convinces Twiddleham that his original pessimistic hypothesis, P (*no one is here*), is False.



(G.4) Exercises(G.4.1)
Exercise**Truth Tables for Conjunction \wedge and Disjunction \vee**

Conjunction		Disjunction	
P	Q	$P \wedge Q$	$P \vee Q$
T	T	T	T
T	F	F	T
F	T	F	T
F	F	F	F
input pair		$(P \text{ and } Q)$	
		input pair	
		$(P \text{ or } Q)$	

The truth tables tell us:

$P \wedge Q$ The one time the conjunction, $P \wedge Q$, is True, is when both component sentences, P , and Q , are True.

Example: “Cecil is from Mars(F) and Today is Monday(T or F).” is a False assertion. But “Cecil is not from Mars(T) and Cecil is human(T)” is a True sentence.

$P \vee Q$ The only time the disjunction, $P \vee Q$, is False, is when both component sentences, P , and Q , are False.

Example: “Cecil is human(T) or Cecil can fly(F).” is True. But “Cecil is from Mars(F) or Cecil can fly(F).” is a False statement.

(G.4.1a) Question: Show that the implication $P \Rightarrow Q$ and the disjunction $(\sim P) \vee (Q)$ have the same four-row truth table and are, therefore, equivalent.

(G.4.1b) Construct examples of sentences P and Q and reason in natural language so that (G.4.1a) is valid.

(G.4.2) **The Tautology.** The definition of a tautology ([G.2.4a](#)) says that $(P \Rightarrow Q)$ and $(Q \Rightarrow P)$ must both be True once the truth values of P and Q are fixed. Use the truth table ([G.1.1](#)) for the following:

(G.4.2a) Question: Is it possible for both implications, $(P \Rightarrow Q)$ and $(Q \Rightarrow P)$, to be False simultaneously, i.e., for the same truth values of P and Q ? Why or why not?

.....

(G.4.2b) When the implications $(P \Rightarrow Q)$ and $(Q \Rightarrow P)$ are simultaneously True (as is required in a tautology), then show that this occurs only when P and Q are both True or both False.

This resolves the question of why [\(G.2.4a\)](#) and [\(G.2.4b\)](#) have the same meaning.

(G.4.3) **The Algebra of Thought.** Recall that the distributive law in algebra says that for any numbers x , y , and z , we have

$$(i) \quad x \times (y + z) = (x \times y) + (x \times z).$$

This law of algebra is hard-wired into our brains in the sense that it provides a template for the thought patterns we all use.

To see this, replace x , y , and z in **(i)** with sentences P , Q , and R respectively, replace “ \times ” with “and,” and replace “ $+$ ” with “or.” With these substitutions, the distributive law transforms into the following statement in natural language:

$$(ii) \quad P \text{ and } (Q \text{ or } R) = (P \text{ and } Q) \text{ or } (P \text{ and } R),$$

which, in symbolic terms, is

$$(iii) \quad P \wedge (Q \vee R) = (P \wedge Q) \vee (P \wedge R).$$

(G.4.3a) Question: Are the equal signs in **(ii)** and **(iii)** justified? Fill in the blank columns of the following truth table and show that for each of the eight input rows, the truth values for $P \wedge (Q \vee R)$ and $(P \vee Q) \wedge (P \vee R)$ are identical.

P	Q	R	$P \wedge (Q \vee R)$	$(P \wedge Q) \vee (P \wedge R)$
T	T	T	?	?
T	T	F	?	?
T	F	T	?	?
T	F	F	?	?
F	T	T	?	?
F	T	F	?	?
F	F	T	?	?
F	F	F	?	?
input triple			P and $(Q$ or $R)$	$(P$ and $Q)$ or $(P$ and $R)$

Hint: Like the arithmetic operations $+$ and \times , the symbols \vee and \wedge act only between pairs of sentences. Therefore, before filling in the “?” columns of the truth table, find the eight truth values for each the intermediate pairs of sentences, $(Q \vee R)$, $(P \wedge Q)$, and $(P \wedge R)$.

(G.4.3b) Question: Show that the “distributive law of thought,” (iii), is True when symbols \vee and \wedge are interchanged, that is, construct an eight-row truth table to show

$$(iv) \quad P \vee (Q \wedge R) = (P \vee Q) \wedge (P \vee R).$$

(G.4.4) de Morgan’s Laws. Use four-row truth tables for the following:

Exercise **(G.4.4a) Show** $\sim (P \vee Q) = \sim P \wedge \sim Q$.

(G.4.4b) Show $\sim (P \wedge Q) = \sim P \vee \sim Q$.

(G.4.4c) Craft specific sentences, P and Q , in natural language, that illustrate the validity of (G.4.4a) and (G.4.4b).

Hint: Consider related sentences like “I will go to the movies.” and “I will paint the house.”

(G.4.5)
Exercise

Proof by Contradiction. In this exercise, prove by contradiction that Sherlock Holmes is correct, as is done in Example (G.3.4):

Entering the warehouse, Inspector Twiddleham hypothesizes P , that the warehouse electricity has been uninterrupted for the last week. Holmes, looking at the electric clock on the wall, notices that it is running, but nevertheless concludes that Twiddleham's hypothesis is False.

- (G.4.5a) What consequence (or clock property) Q would render the implication $P \Rightarrow Q$ to be True?
- (G.4.5b) What observation $\sim Q$ would Holmes make so as to contradict, negate, or falsify Twiddleham's hypothesis P ?

Hint: *The white faces of electric clocks often have a red/white indicator that must be set manually to white so that a small electromagnet can hold it in place as long as the electric current is running. As soon as the current is interrupted, however, the electromagnetic field is lost and the indicator shows red. When the power is restored, the indicator remains red.*

Bibliography

- [1] Abbott, Edwin A. *Flatland: A Romance of Many Dimensions.* Dover (Thrift Edition), 96 pp., 1992.
- [2] Alonso, R., et al. TrES-1: The Transiting Planet of a Bright K0V Star. *ArXiv Astrophysics e-prints astro-ph/0408421*, August 2004.
- [3] Born, Max. *Einstein's Theory of Relativity.* Dover, New York, New York, 1965.
- [4] Cathcart, Brian. *The Fly in the Cathedral: How a Group of Cambridge Scientists Won the International Race to Split the Atom.* Farrar, Strauss and Giroux, 2005.
- [5] Chang, Kenneth. Planet Group Similar to the Solar System is Found. *New York Times, Late Edition - Final, Section A, Page 26, Column 1*, May 18 2006.
- [6] Cohen, I. B. *The Triumph of Numbers.* W. W. Norton & Co., 2005.
- [7] Crease, Robert. Pythagoras. *Physics World*, January 2006.
- [8] de Pillis, John. *777 Mathematical Conversation Starters.* Math. Assoc. of Amer., Washington, D. C., 2002.
- [9] Drake, Stillman (trans.). *Galileo Galilei: Dialogue Concerning the Two Chief World Systems—Ptolemaic and Copernican.* University of California, Berkeley, 1957.
- [10] Dyson, Freeman. Religion from the outside. *The New York Review of Books*, pp. 4–8, LII(11), June 22 2006.
- [11] Dyson, Freeman. *Progress in Religion.* Acceptance speech for the Templeton Prize for Progress in Religion, Washington National Cathedral, May 16, 2000.
- [12] Einstein, Albert. Zur elektrodynamik bewegter körper (on the electrodynamics of moving bodies). *Annalen der Physik*, 17:891, 1905.
- [13] Einstein, Albert. *Sidelights on Relativity.* Dover, New York, New York, October 1983.

- [14] Ekeland, Ivar. *The Best of All Possible Worlds*. University of Chicago Press, 2007.
- [15] Euclid (introduction by Thomas Heath). *The Thirteen Books of Euclid's Elements, Books 1 and 2 (paperback)*. Dover, New York, 1956.
- [16] Gleick, James. *Isaac Newton*. Pantheon, New York, 2004.
- [17] Gosnell, Mariana. *Ice: The Nature, the History, and the Uses of an Astonishing Substance*. Knopf, November 2005.
- [18] Gould, Andrew, *et al.* Einstein Effect Reveals Icy Exoplanet. *Scientific American*, March 15, 2006.
- [19] Hamming, R. W. On the unreasonable effectiveness of mathematics. *Amer. Math. Monthly*, 87(2), 1980.
- [20] Hawking, Stephen W. *A Brief History of Time: From the Big Bang to Black Holes*. Bantam; Rei edition, March 1, 1988.
- [21] Jackson, J. D. *Classical Electrodynamics*. Wiley, 3rd ed., 1998.
- [22] Kandell, Jonathan. *Jaques Derrida, Abstruse Theorist, Dies at 74*. New York Times, October 10, 2004.
- [23] Lloyd, G. E. R. *Early Greek Science: Thales to Aristotle*. W. W. Norton, New York · London, 1970.
- [24] Macauley, David. *Motel of the Mysteries*. Houghton Mifflin Company, 1979.
- [25] Overbye, Dennis. Duck Look-Alike Reveals Birds' Evolution. *New York Times*, June 16 2006.
- [26] Palmer, Stuart. *The Puzzle of the Silver Persian*. Bantam Books, ISBN: 0-553-25934-2, 1934.
- [27] Penrose, Roger. *The Road to Reality*. Knopf, 2004.
- [28] Plato (Kamtekar and Lee). *The Republic*. Penguin Classics, 2003.
- [29] Popper, Karl. *Conjecture and Refutations: The Growth of Scientific Knowledge*. Routledge Classics, April 1992.
- [30] Reeves, Richard. *A Force of Nature*. Discovery Series of Atlas Books, a division of W. W. Norton, 2007.

- [31] Sobel, Dava. *Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time.* Debate, 1997.
- [32] Weinberg, Steven. (interview). *New York Times*, April 20, 1999.
- [33] Wigner, Eugene P. On the unreasonable effectiveness of mathematics in the natural sciences. *Comm. on Pure and Appl. Mathematics*, 13(1), 1960.
- [34] Zagorin, Perez. *How the Idea of Religious Toleration Came to the West.* Princeton University Press, New York, 2003.

Index

Symbols

- * matrix product (A.4.3e), 370
 $\langle \vec{X}, \vec{Y} \rangle$ inner product (A.2.6a), 362
 G_v line-of-sight fcn (3.4.6a), 62
 I current (21.1.1), 289
 L_v Lorentz matrix (7.1.3), 106
 PQ line between vectors
 \vec{P} and \vec{Q} (A.2.18a), 366
 $\Delta f_x, \Delta f_t$ (16.3.8), 208
 $\|PQ\|$ length of line between
 \vec{P} and \vec{Q} (A.2.18a), 366
 \Leftrightarrow if and only if (G.2.4), 440
 $\Phi(r)$ potential (11.5.1a), 147
 $\Phi_B(A)$ mag'c. flux (20.2.2), 280
 $\Phi_E(A)$ elec. flux (20.2.1), 279
 $\Pi_{\mathcal{U}, \Delta u_i}$ partition of \mathcal{U} (18.7.1a), 252
 \Rightarrow implies (F.2.1), 410
 γ_v Lorentz factor (7.1.3b), 106
 $\text{col}[x_i]$ (A.2.3), 361
 $\vec{A} \times \vec{B}$ cross product (18.6.3), 251
 \vec{B} magnetic field (19.4.1), 268
 \vec{E} electric field (19.3.5), 267
 \vec{E}_i Std basis (A.2.11), 364
 \vec{E}_{p_0} electric field at p_0 (19.3.5), 267
 $\vec{X} + \vec{Y}$ vector add'n (A.2.4b), 361
 $\vec{e}_{0,1}$ unit vector (19.2.4c), 264
 \mathbf{R}^k k -space (A.2.2), 361
 μ_0 permeability (21.2.2), 291
 $\partial/\partial x, \partial/\partial t$ partial derivs. (16.3.4), 206
- grad** gradient (11.5.1a), 147
 σ_v Lorentz contraction factor (1.8.2), 20
 σ_v Lorentz contraction (7.1.3b), 106
 \sum notation (A.2.7), 363
 \times cross product (18.6.3), 251
 ϵ_0 permittivity (19.2.5a), 265
 $a\vec{X}$ scalar mult'n (A.2.4c), 361
 $\text{bdry}(A)$ boundary of area A (??), 292
 c speed of light (17.1.1), 217
 $c_j(A), r_i(A)$ mx row, col (A.4.3a), 370
 $dist(A, B)$ distance from A to B (A.2.18a), 366
 $r_i(A), c_j(A)$ mx row, col (A.4.3a), 370
 $r_{A,B}$ area between axes and hyperbola (B.3.3a), 380
 $v \oplus w$ Lorentz sum (7.3.1), 110
 \mathcal{F} frame of reference (1.2.1), 13
 $\mathcal{F}^A, \mathcal{F}^B$ frames of reference (3.4.3), 59
 \mathcal{F}_v moving frame (8.2.1), 120
 \mathbf{f} Fresnel drag (13.2.7b), 167
 $x \text{ MX}(T$ (A.4.6a), 372
 $\sim P, \sim Q$ negative sentences (F.2.5), 411
-
- A**
-
- Abbott, Edwin A.
Flatland (24.7.2), 350
Accommodating Universe (1.9.2), 23
& S. Hawking (9.3.2), 130
Addition

- Lorentzian $v \oplus w$ (1.13.3), 39
of vectors (A.2.4b), 361
- Addition $u \oplus v$, *see* Speeds, Lorentzian
addition of
- Addition of $u \oplus v$, *see* Speeds, Lorentzian
addition of
- Additive associativity
(E.1.1a), 396
- Additive, multiplicative identities 0,I
(E.1.1b), 396
- inverse
(E.1.1c), 396
- Aether, luminescent
(17.1.3), 218
(17.5.1), 228
(22.3.2), 311
- Age of a point (2.2.6), 47
- Algebra, Fundamental
Theorem of (A.5.2), 374
- Alpha particles
(F.1.1), 406
- Ampère, André M.
(19.1), 262
law (21.3.3), 293
law, with Maxwell
(21.4.3c), 296
- Ampère, André M.
induction (21.7.3d), 301
- Analysis, numerical
(18.7.5), 254
- Anders, William: NASA
astronaut (F.10.3), 424
- Anderson, Mary
windshield wipers (24.2.1), 337
- Angle of Incidence
(24.6.2), 348
- Annus mirabilis (24.2.2), 337
- Apollo 8 moon mission
(F.10.3), 423
- Approximately
parallel (23.1.3a), 316
straight (23.1.3b), 316
- Aquinas, Thomas (1.1.1), 11
- Archimedes' palimpsests
- (24.1.1), 335
- Area
as 3D vector (18.3.3), 242
Area addition of $u \oplus v$, *see* Speeds,
Lorentzian addition of
- Aristotle
& motion (1.1.1), 11
& motion (1.1.2), 12
& testing (24.1.5), 337
Truth by authority
(F.6.2), 418
& women's teeth
(F.6.2c), 418
- Aristotle v. Galileo (1.1.1), 11
- Armchair calculation
of c by Maxwell (23.4), 326
- Arrows
as area, length (18.3.3), 242
- Aspect of a photo
(D.3.2), 394
- Associativity of addition &
multiplication
(E.1.1a), 396
- Asymmetry in the paradoxes
(10.1), 132
- Atom, structure of
(F.1.1), 406
- Axle, *see* Wheel-axle configuration
-
- B**
-
- Bacon, Sir Francis
(F.6.1), 417
- Baez, John (17.1.1), 217
- Basis
(A.2.9), 363
standard (A.2.11), 364
- Bell, Alexander G. (21.7.1), 299
- Big Dipper (F.10.4), 424
- Biot, Jean-Baptiste (21), 289
- Biot-Savart law (21.2.1a), 291
- Blind men and the elephant
(F.9.4), 422
- Borman, Frank: NASA
astronaut (F.10.3), 423

- Born, Max (21.7.2), 299
 Brownian motion (24.2.4), 338
 Bug-Rivet paradox (14.1), 171
-
- C**
- Callisto, and c (17.2.5), 220
 Capacitor (21.4.1), 295
 Catenary curve (B.4.2), 383
 Celeritas, and c (17.1.1), 217
 Clock
 cohort (2.2.3), 46
 origin (2.2.3), 46
 Clock rate
 measuring (2.4.2), 51
 Clock, photon (2.1.1), 43
 Clocks
 Origin (3.4.3), 59
 identical (2.2.1), 45
 measuring longitude (24.4.2), 344
 synchronized (2.2.5), 47
 Closed
 integral (18.7.4), 254
 line (18.4.3), 245
 surface (18.4.3), 245
 Clouds misrepresented
 in Star Trek (F.10.3), 424
 Coca Cola & cocaine (24.2.1), 337
 Cocaine & Coca Cola (24.2.1), 337
 Coefficient
 Fresnel drag (13.2.7b), 167
 Cohort
 clock (2.2.3), 46
 observer (3.4.3), 59
 Combination, linear (A.2.8), 363
 Complex conjugate (A.2.13a), 364
 Component
 projection (18.3.5), 243
 vector (A.2.2), 361
 Composition of
 functions (A.3.6), 368
 Condenser (21.4.1), 295
 Conjugate, complex (A.2.13a), 364
 Conserving electrical charge in Ampère-Maxwell Law (21.4.3c), 296
 Constant of proportionality (18.2.3), 240
 Constant velocity, speed (4.2.1), 69
 in spacetime (4.2.3), 69
 Constellations, and induction (F.10.4), 424
 Continuous sum (18.7.3a), 253
 Contraction factor σ_v (1.8.2), 20
 (7.1.3b), 106
 Contradiction, proof by (G.4.5), 440
 Contrapositive implication (F.2.5), 411
 Conventional wisdom opposed by Einstein (24.4.4), 345
 Converse implication (F.2.5), 411
 Coordinate-free inner product (18.6.1), 249
 Coordinates vector (A.2.2), 361
 Corridor
 line-of-sight (3.4.4a), 60
 skateboarder (1.9.1), 22
 Cosh, hyperbolic (B.1.1b), 376
 Coulomb's law (19.2.4b), 264
 Coulomb, Charles-Augustin de (19.1), 262
 Critical thinking (0.1.8), 4
 an example (1.9.7), 25
 Cross product \times (18.6.3), 251
 Current displacement (21.4.2), 295
 electric (21.1.1), 289
-
- D**
- de Morgan laws (G.4.3b), 445
 de Sitter ghosts (17.4.3), 227
 de Sitter, Willem (17.4), 225
 Decomposition of vectors

- (18.5.1a), 247
 Deconstruction (F.10.2), 423
 Deduction, logical (F.8.2), 420
 Definite integral (18.7.4), 254
 Definition of one second (2.1.4), 44
 Demarcation of Karl Popper (F.11.1), 425
 Derivative, partial (16.3.4), 206
 Derrida, Jacques (F.10.2), 423
 Descartes, René dispute w. Fermat (24.6.1a), 348 rôle of axioms (F.6.1a), 417
 Diagram Galilean (5.2.1), 86 Minkowski (6.1.1), 97 spacetime (4.1.2), 68
 Disappearing sun (14.3.2), 181
 Displacement current (21.4.2), 295
 Dispute between Descartes and Fermat (24.6.1a), 348
 Distance, length in vector space (A.2.18a), 366
 Distributive law general (E.1.2a), 397 in logic (G.4.3), 444
 Doppler effect for waves (16.6.3), 214
 Drag, Fresnel coefficient (13.2.7b), 167
 Duck property list (F.3.3), 413
 Dyson, Freeman (24.8.5), 354
-
- E**
- Economy reality of (F.1.4), 408
 Effectiveness of mathematics (0.1.2), 3
 Eigenvalue (A.5.1), 374
 Eigenvector (A.5.1), 374
 Einstein, Albert Brownian Motion (24.2.4), 338 Nobel prize (24.2.3), 338 Photoelectric effect (24.2.3), 337 and $E = mc^2$ (24.2.6), 339 and Maxwell's asymmetry (24.2.5), 338 and thought experiments (24.7.2d), 352 opposing conventional wisdom (24.4.4), 345
 Ekeland, Ivar (24.6.1a), 348
 Electric current (21.1.1), 289 field (19.3.5), 267 flux (20.2.1), 279 force (19.2.4), 264 permittivity ϵ_0 (19.2.5a), 265
 Electrical charge conserved in Ampère-Maxwell Law (21.4.3c), 296
 Electricity Gauss' law of (20.3.1), 280
 Electromagnetic spectrum (22.3.3), 312 and light (19.1), 262
 Electron, discovery of (19.6), 272
 Electron, discovery of (F.1.1), 406
 Elephant and blind men (F.9.4), 422
 Ellerman, David (14.3), 179 (24.6.1a), 348
 Energy, *see* Kinetic energy
 Equation (s) of Maxwell (21.7.3), 300 of straight line (A.2.14a), 365
 Equidistance and simultaneity (6.2.2), 99
Euclid, Elements (24.1.1), 335
 Europa, and c (17.2.5), 220

- Even function (B.2.1), 377
 Event
 in spacetime (4.1.2), 68
 static, stationary (4.4.1b), 73
 Experiment
 (13.2), 164
 J. J. Thomson (19.6), 272
 Michelson-Morley (17.5), 228
-
- F**
- Factor
 γ_v of Lorentz (7.1.3b), 106
 σ_v of contraction (1.8.2), 20
 False Hypothesis of Descartes (24.6.1a), 348
 Falsifiability
 Karl Popper
 (F.11.1), 425
 definition of
 (F.11.2), 425
 (F.5.2c), 415
 Falsification (F.4.1c), 414
 Faraday, Michael
 (19.1), 262
 law of induction
 (21.5.1a), 297
 (21.7.3c), 300
 (23.2.3), 318
 Feeble elements (A.4.4b), 371
 Fermat, Pierre, dispute with Descartes (24.6.1a), 348
 Field
 electric (19.3.5), 267
 magnetic (19.4.1), 268
 vector (18.5.3), 248
 Filmstrip as spacetime graph (4.1.1), 67
 Fitzgerald, George (17.5.8), 233
 Fizeau, Hippolyte Louis
 speed of light (17.3.1), 222
 Fizeau, Hippolyte Louis
 c in a medium (13.2), 164
 Flatland (24.7.2), 350
 Flux
 electric (20.2.1), 279
 magnetic (20.2.2), 280
 Flux (20.1.3), 277
 Force
 Lorentz (19.5.3a), 272
 electric (19.2.4), 264
 field (19.2.2), 263
 line of (19.2.2), 263
 magnetic (19.5.2a), 271
 Formulas, Stephen Hawking (*footnote*), 2
 Frame \mathcal{F}_v , moving (8.2.1), 120
 Frames of reference
 $\mathcal{F}^A, \mathcal{F}^B$ (3.4.3), 59
 6 rules for (3.4), 58
 Cartesian (1.2.1), 13, 54
 as spacetime graph
 (4.1.3), 68
 as vector space (3.2.2), 55
 home (3.4.2), 58
 inertial (1.5.1), 16
 line of sight (3.4.4), 59
 synchronized pairs of
 (3.4.5a), 62
 Franklin, Benjamin
 (19.2), 263
 Frequency, wave (16.1.2), 200
 Fresnel drag (13.2.7b), 167
 Function
 (A.3.2), 366
 composition of (A.3.6), 368
 even (B.2.1), 377
 hyperbolic (B.1.1), 376
 inverse of (A.3.4), 367
 line-of-sight
 (3.4.6a), 62
 (3.4.4a), 60
 linear (A.4.1), 368
 odd (B.2.2), 377
 of symmetry (3.4.6), 62
 one-to-one (A.3.2), 366
 Fundamental Theorem
 of Algebra (A.5.2), 374
 Future
 affects the present (14.2.4), 177
 travel to (1.11.1), 29

G _____

Galilean
 Transformation (5.3.1), 87
 diagram (5.2.1), 86
 matrix pattern (5.4.1a), 90
 matrix (5.3.4a), 88
 relativity principle (1.4.2), 15
 Galileo v. Aristotle (1.1.1), 11
 Galileo, Galilei
 (1.4.1), 14
 and motion (1.1.1), 11
 (F.6.1), 417
 Ganymede, and c (17.2.5), 220
 Garbage and logic
 (G.1.4b), 436
 Gauss, Karl Friedrich
 (19.1), 262
 law of
 electricity (21.7.3a), 300
 magnetism (21.7.3b), 300
 General relativity
 and GPS clocks (24.5.3), 347
 confirming the
 twin paradox (11.2.1), 142
 and microlensing
 (F.1.2), 407
 George III, King (24.4.2), 344
 Ghosts, of de Sitter (17.4.3), 227
 Gibbs, Philip (17.1.1), 217
 Gosnell, Mariana
 analyzing ice
 (F.1.5), 408
 Gould, Andrew
 and microlensing
 (F.1.2), 407
 GPS and Relativity
 (24.5.3), 347
 GPS clocks
 and Special relativity (24.5.3), 347
 $\times \text{ grad}$ (11.5.1a), 147
 Graph
 Galilean (5.2.1), 86
 spacetime (4.1.2), 68
 Graphical addition of $u \oplus v$, , see

Speeds, Lorentzian addition of

Gravitational potential Φ (11.5.1a), 147

Greece, early science
 (F.6.1a), 417

Group, definition of
 (E.1.2), 397

H _____

Harrison, John (24.4.2), 344

Hawking, Stephen
 planetary travel (9.3.2), 130
 on formulas (*footnote*), 2

Hertz, Heinrich, Light as
 electromagnetism
 (23.5.2), 328
 (19.1), 262

electromagnetism (22.3.2), 311

Hits, mathematical
 (23.5), 328

Holmes, Sherlock
 and deduction
 (F.12.8), 433
 proof by contradiction
 (G.3.4), 441

Home frame of reference
 (3.4.2), 58

Hume, David (24.7.1), 350

Hyperbolic
 functions (B.1.1), 376
 identities (B.1.3), 377

Hypotheses, false
 true results from (24.6.1a), 348

I _____

Ice core history
 (F.1.5), 408

Identical clocks (2.2.1), 45

Identity
 hyperbolic (B.1.3), 377
 0,I of reals (E.1.1b), 396
 multiplicative
 (E.2.1d), 399

- If and only if (G.2.4), 440
 If-then sentence (F.2.1), 410
 Implication
 contrapositive (F.2.5), 411
 converse (F.2.5), 411
 logical (F.2.1), 410
 Implication (F.2.5), 411
 Imposter duck, property list (F.5.4), 416
 Incidence, angle (24.6.2), 348
 Independence, linear (A.2.10), 364
 Index, refractive (13.2.2c), 165
 (24.6.2), 348
 Induction
 electrical (21.7.3d), 301
 law of (21.5.1a), 297
 & constellations (F.10.4), 424
 logical (F.8.2), 420
 Inertial
 frame of reference (1.5.1), 16
 Inference, logical (F.8.2), 420
 Inherited ruler (4.4.3), 73
 Inner product (A.2.6a), 362
 coordinate-free (18.6.1), 249
 integrals of (18.8.5), 257
 Integral
 Riemann (18.7.4), 254
 closed (18.7.4), 254
 of $h\vec{A}\vec{V}, d\vec{A}\vec{B}$ (18.8.5a), 257
 of $h\vec{A}\vec{V}_l, d\vec{l}\vec{B}$ (18.8.5a), 257
 Intellectual round trip (0.1.3), 3
 Inter-planetary travel
 of Hawking (9.3.2), 130
 Interferometer (17.5.3), 230
 Invariance
 Lorentz (8.1.2), 119
 Inverse
 of 2×2 matrix (A.4.4d), 371
 of a function (A.3.4), 367
 square law (18.2.4b), 241
 additive, multiplicative (E.1.1c), 396
 additive (E.4.8), 404
 Io, and c (17.2.5), 220
-
- J**
- Jordan curve (18.4.2), 245
 Jordan, M. E. Camille (18.4.4), 245
-
- K**
- Kaboogle (1.13.5a), 39
 Kant, Immanuel (24.7.1), 350
 Kinetic energy
 Newtonian (15.3.1d), 191
 relativistic (15.3.1d), 191
 King George III (24.4.2), 344
-
- L**
- Law
 Ampère's (21.3.3), 293
 Ampère-Maxwell (21.4.3c), 296
 Biot-Savart (21.2.1a), 291
 Faraday's (21.5.1a), 297
 Gauss'
 electricity (20.3.1), 280
 magnetism (20.4.1), 282
 Lentz's (21.6), 297
 Lorentz elect. (19.5.3), 272
 Leibniz, Gottfried W.
 co-invents calculus (24.1.4), 336
- Length
 Minkowski (8.1.1), 119
 as 3D vector (18.3.3), 242
 measuring in motion (2.4.3), 52
 of 2D vector (A.2.13a), 364
- Length, distance
 in vector space (A.2.18a), 366

- Lentz's law (21.6), 297
 Life, and water
 (F.2.4), 410
 Light
 & electromagnetism
 (22.3.2), 311
 (19.1), 262
 speed measured with
 microwave oven (17.6.6d), 237
 speed= c (15.4), 17
 why use c (17.1.1), 217
 year (1.9.3), 23
 Lightning-Tunnel paradox
 (12.1), 153
 Line
 closed (18.4.3), 245
 of force (19.2.2), 263
 of simultaneous events
 (4.4.1a), 73
 straight (A.2.14a), 365
 Line-of-sight
 (3.4.4), 59
 and Minkowski
 invariance (8.1.2), 119
 corridor (3.4.4a), 60
 function
 (3.4.6a), 62
 (3.4.4a), 60
 Linear
 combination (A.2.8), 363
 independence (A.2.10), 364
 Linear function
 (A.4.1), 368
 Logical
 deduction (F.8.2), 420
 garbage (G.1.4b), 436
 implication (F.2.1), 410
 induction (F.8.2), 420
 inference (F.8.2), 420
 tautology (G.2.4), 440
 Lorentz
 Hendrik A. (19.1), 262
 addition $v \oplus w$ (7.3.1), 110
 contraction σ_v
 (1.8.2), 20
 (7.1.3b), 106
 factor γ_v (7.1.3b), 106
 force (19.5.3), 272
 invariance (8.1.2), 119
 matrix L_v (7.1.3), 106
 matrix pattern (7.2.2), 109
 sum $v \oplus w$
 (1.13.3), 39
 (7.3.1), 110
 (7.5.6b), 115
 transformation (7.1.3), 106
 Lorentz speeds, , see Speeds, Lorentzian addition of
 Lovell, Jr., James
 NASA astronaut
 (F.10.3), 424
 Luminescent aether
 (17.1.3), 218
 (17.5.1), 228
 (22.3.2), 311

M
 Macauley, David, *Motel of the Mysteries*
 (F.10.5), 425
 Mach, Ernst (24.2.4), 338
 Magnetic
 field (19.4.1), 268
 flux (20.2.2), 280
 force (19.5.2a), 271
 monopole (19.4.4), 269
 permeability μ_0 (21.2.2), 291
 Magnetism
 Gauss' law (20.4.1), 282
 Mass
 increasing with speed
 (15.2), 186
 symmetry principle (15.2.3b), 187
 Mathematical
 hits (23.5), 328
 unreasonable effectiveness (0.1.2),
 3
 Matrices as vector space
 (A.2.1), 360
 Matrix

- 2×2 inverse (A.4.4d), 371
 G_v of Galileo (5.3.4a), 88
 L_v of Lorentz (7.1.3), 106
 definition of (A.4.3), 369
 pattern
 of Galileo (5.4.1a), 90
 of Lorentz (7.2.2), 109
 product $A * B$ (A.4.3e), 370
 Matrix addition of $u \oplus v$, , see Speeds
 Lorentzian addition of
 Maxwell's equations
 (21.7.3), 300
 and Max Born (21.7.2), 299
 the added term (21.4.3), 295
 Maxwell, James Clerk
 (24.1.4), 336
 (19.1), 262
 armchair calculation
 of c (23.4), 326
 Measuring
 c in a medium (13.2), 164
 moving clock rate (2.4.2), 51
 moving lengths (2.4.3), 52
 object speed (2.4.2), 51
 MHz=Megahertz
 $=10^6$ cycles/sec (17.6.6d), 237
 Michelson-Morley
 experiment (17.5), 228
 Microlensing and planet
 discovery (F.1.2), 407
 Microwave oven used
 in measuring c (17.6.6d), 237
 Minkowski
 diagram (6.1.1), 97
 length (8.1.1), 118
 Minkowski length
 line-of-sight invariance
 (8.1.2), 119
 Minus \times a minus
 (E.4.5), 402
 Miracle year (24.2.2), 337
 MKS units (19.2.7a), 265
 Models
 early math'l (24.1.3), 336
 simplifying (23.1.2), 315
 Momentum
 particle (15.2.3), 187
 system (15.2.3a), 187
 Monopole, magnetic
 (19.4.4), 269
 Motel of the Mysteries
 (F.10.5), 425
 Motion, Brownian (24.2.4), 338
 Motivation for
 Time-stamp Origins Thm.
 (8.3), 121
 age of a point (2.2.6), 47
 constant wave speed
 (16.5.3), 211
 deSitter diagram (17.4.1), 226
 electric field \vec{E} (19.3.3a), 267
 event (spacetime) (4.1.2), 68
 proportionality (18.2.1), 240
 vector field (18.5.2), 248
 Moving
 clocks run slow (2.3.3), 49
 rods shrink (2.3.3), 49
 Moving clock rate
 measuring (2.4.2), 51
 Mu Meson (Muon) (24.5.2), 346
 Multiplication
 $A * B$ of matrices (A.4.3e), 370
 as early model (24.1.3), 336
 Multiplicative
 & additive identities 0,I
 (E.1.1b), 396
 & additive inverse
 (E.1.1c), 396
 associativity
 (E.1.1a), 396
 identity (E.2.1d), 399
-
- N**
-
- NASA Earth photos
 (F.10.3), 423
 Nature
 catenary curve in (B.4.2), 383
 as hidden hand
 (F.1.1), 406
 as set of properties

- (F.3.2), 412
- Necessary *vs.* sufficient
(F.9), 420
- Negative
× another negative
(E.4.5), 402
- sentence (F.2.5), 411
- Newton, Isaac
co-invents calculus
(24.1.4), 336
- laws confirmed
(F.1.6), 409
- Nobel prize
Einstein (24.2.3), 338
- Node of standing wave (17.6.6a), 236
- Norm of n -vector
(A.2.13a), 364
- Normal vector (18.3.3), 242
- Noumenal (of Kant)
(24.7.1), 350
- Numerical analysis
(18.7.5), 254
-
- ## O
- Observer
Origin (3.4.3), 59
- Odd function (B.2.2), 377
- Oersted, Hans (21), 289
- One second of time
definition of (2.1.4), 44
- One-to-one function (A.3.2), 366
- Origin
clock (3.4.3), 59
clock (2.2.3), 46
observer (3.4.3), 59
time-stamped (8.2.1), 120
- Orion, the hunter
(F.10.4), 424
- Orthogonal vectors
(A.2.6b), 362
-
- ## P
- Palimpsest of Archimedes
(24.1.1), 335
- Paradox
Accommodating Universe
(1.9.2), 23
- Asymmetry (10.1), 132
- Bug-Rivet
(14.1), 171
(1.14), 40
- Definition of (0.1.7), 4
- Pea-Shooter
(13.1), 163
(1.13), 36
- Time-Distance (1.10), 26
- Train-Tunnel
(1.12), 33
(12.1), 153
- Twin
(11.1), 141
(1.11), 29
- Parallel, approximately (23.1.3a), 316
- Partial derivative (16.3.4), 206
- Particle momentum (15.2.3), 187
- Partition $\Pi_{\mathcal{U}, \Delta u_i}$ of
arc or area (18.7.1a), 252
- Past/present, affected by
the future (14.2.4), 177
- Pattern of
Galilean matrix (5.4.1a), 90
Lorentz matrix (7.2.2), 109
- Pea Shooter paradox (13.1), 163
- Period, wave (16.1.2), 200
- Permeability μ_0 (21.2.2), 291
- Permittivity ϵ_0 (19.2.5a), 265
- Perpendicular vectors
(A.2.6b), 362
- Petri dish equation
(*footnote*), 186
- Photoelectric effect (24.2.3), 337
- Photon clock (2.1.1), 43
- Photon (24.2.3), 337
- Planets, extra-solar
how discovered
(F.1.1), 407
- in Puppis constellation
(F.1.3), 407
- w. microlensing
(F.1.2), 407

- Plato (24.7.1), 349
 Polynomials as vector space (A.2.1), 360
 Popper, Karl
 demarcation (F.11.1), 425
 falsifiability (F.11.2), 425
 induction (F.10.1), 423
 uncertainty (F.1.6), 409
 Positive
 surface direction (18.4.7), 246
 Postulates of
 special relativity (1.5.2), 16
 Potential Φ , gravitational (11.5.1a), 147
 Preferred observer (3.4.5b), 62
 Present/past, affected by the future (14.2.4), 177
 Prime Mover
 of Aristotle (1.1.1), 11
 Product
 $A * B$ of matrices (A.4.3e), 370
 as early model (24.1.3), 336
 cross \times (18.6.3), 251
 inner (A.2.6a), 362
 of two negatives (E.4.5), 402
 Projection component (18.3.5), 243
 Proof by contradiction (G.4.5), 440
 Proper
 time (2.1.4), 44
 value (A.5.1), 374
 vector (A.5.1), 374
 Property list
 of a duck (F.3.3), 413
 of duck imposter (F.5.4), 416
 Proportional (18.2.1), 240
 Puppis constellation (F.1.3), 407
 Pythagorean Theorem
 in history (24.1.1), 335
-
- Q** _____
- Qualitative *vs.* quantitative (0.1.1), 2
 Quintessence (1.1.1), 11
-
- R** _____
- Rømer, Ole (Olaus)
 speed of light (17.2.1), 219
 Rainbow, colors of (22.3.3), 312
 Reality
 and the economy (F.1.4), 408
 Reference, frame of (1.2.1), 13
 Refractive index
 (13.2.2c), 165
 (24.6.2), 348
 Relativistic
 sum $v \oplus w$ (7.3.1), 110
 Relativity
 and GPS clocks (24.5.3), 347
 general (11.2.1), 142
 incompatible with
 rigidity (14.2.3), 177
 postulates of (1.5.2), 16
 principle (1.4.2), 15
 special (1.5), 16
 general & planets (F.1.2), 407
 Religion (24.8.5), 354
 Riemann
 integral of hA, B (18.8.5), 257
 integral (18.7.4), 254
 sum (18.7.2a), 253
 Right triangles (24.1.1), 335
 Right-Hand rule for
 closed loops (18.4.6), 246
 cross products (18.6.3), 250
 wheel-axle (18.4.6), 246
 Rigidity incompatible with relativity (14.2.3), 177
 Ring
 Definition (E.2.1), 399

- Rivet-Bug paradox (14.1), 171
 Rosetta stone
 if-then sentences
 (F.2.2), 410
 Round trip, intellectual (0.1.3), 3
 ROY-G-BIV rainbow
 colors (22.3.3), 312
 Ruler, inherited (4.4.3), 73
 Rutherford, Ernest
 (F.12.4), 429
 (F.1.1), 406
-
- S**
- Savart, Félix (21), 289
 Saxe, John G, *Blind men and the elephant* poem
 (F.9.4), 422
 Scalar multiplication
 of vectors (A.2.4c), 361
 Scalar (A.2.4), 361
 Science
 in Ancient Greece
 (F.6.1a), 417
 when it thrives
 (F.6.2d), 418
 Scientific method
 overview (F.1.7), 409
 Second of time
 definition of (2.1.4), 44
 Sherlock Holmes, *see* Holmes, Sherlock
 Sigma notation \sum (A.2.7), 363
 Simplifying models (23.1.2), 315
 Simultaneity
 and equidistance (6.2.2), 99
 as subset of X -axis (4.4.3), 73
 line of (4.4.1a), 73
 Sinh, hyperbolic (B.1.1a), 376
 Six rules for frames (3.4), 58
 Skateboarder corridor
 (1.9.1), 22
 Skipping stone model
 of a wave (16.5.4), 211
 Slinky Connection (14.3), 179
 Smith, Dinitia, and J. Derrida
 (F.10.2), 423
 Snell, Willebrord
 Snell's Law (24.6.2), 348
 Socrates (24.7.1), 350
 Space of vectors (A.2.1), 360
 Spacetime
 as vector space (3.2.2), 55
 coordinates (4.1.2), 68
 Spacetime graph
 (F.1.2), 68
 & constant speed (4.2.3), 69
 and waves (16.3.2), 205
 as filmstrip (4.1.1), 67
 as frame of reference
 (F.4.3), 68
 Special relativity
 and GPS clocks (24.5.3), 347
 and rigidity (14.2.3), 177
 introduction (1.5), 16
 postulates of (1.5.2), 16
 Spectrum, electromagnetic
 (22.3.3), 312
 Speed
 how to measure (2.4.2), 51
 increases mass (15.2), 186
 of rope wave (16.1.4), 200
 Speed of light
 Maxwell's calculation
 (23.4), 326
 measured by Louis Fizeau
 (17.3.1), 222
 measured by Ole Rømer
 (17.2.1), 219
 measured with
 microwave oven (17.6.6d), 237
 value of c (1.5.4), 17
 why symbol c (17.1.1), 217
 Speed, of a system (15.2.3), 187
 Speeds, Lorentzian addition of
 (7.3), 110
 via areas (7.5), 112
 via matrices (7.4), 111
 Standard basis (A.2.11), 364
 Standing wave (17.6.6a), 236
 Star Trek misrepresents
 clouds (F.10.3), 424

- Static, stationary event (4.4.1b), 73
- Straight approximately (23.1.3b), 316
line (A.2.14a), 365
- Sufficient *vs.* necessary (F.9), 420
- Sum
- Lorentz $v \oplus w$
as \tanh (7.5.6b), 115
 - Riemann (18.7.2a), 253
 - as early model (24.1.3), 336
 - continuous (18.7.3a), 253
 - relativistic $v \oplus w$ (7.3.1), 110
 - \sum notation (A.2.7), 363
- Sun, disappearing (14.3.2), 181
- Surface
- closed (18.4.3), 245
- Swiveled Line Thm. (6.4.1), 102
- Symmetry principle (1.6.1), 18
- and functions (3.4.6), 62
 - for mass (15.2.3b), 187
- Synchronized
- clocks (2.2.5), 47
 - frames (3.4.5a), 62
- System
- momentum (15.2.3a), 187
 - speed v_{sys} (15.2.3), 187
-
- T**
- Tanh, hyperbolic (B.1.1c), 376
- Tautology (G.2.4), 440
- Teeth of women, fewer than men's (F.6.2c), 418
- Telephone, invention of (21.7.1), 299
- Term, added by Maxwell (21.4.3), 295
- Test charge q_0 (19.3.5), 267
- Theorem, Swiveled Line (6.4.1), 102
- Thinking, critical (0.1.8), 4
- Thomson, J. J.
discovers electron (19.6), 272
- Time zones, history of (24.4.3), 345
- Time, proper (2.1.4), 44
- Time-stamped
- origin theorem (8.2.3), 120
 - origin (8.2.1), 120
- Train-Tunnel paradox (12.1), 153
- Transformation
- G_v of Galileo (5.3.1), 86
 - L_v of Lorentz (7.1.3), 106
- Travel
- Hawking's interplanetary (9.3.2), 130
 - into the future (1.11.1), 29
- TrES1
- extra-solar planet (F.1.1), 407
- Triangles
- right (24.1.1), 335
- Truth table
- (G.1.7), 437
 - (G.1.1), 435
- Truth values (G.1), 435
- Tunnel-Train paradox (12.1), 153
- Twiddleham, Inspector (G.3.4), 441
- Twin paradox (11.1), 141
-
- V**
- Units, MKS (19.2.7a), 265
- Universe, Accommodating (1.9.2), 23
- Unreasonable effectiveness of mathematics (0.1.2), 3
-
- V**
- Vector
- addition (A.2.4b), 361
 - as area (18.3.3), 242
 - as length (18.3.3), 242
 - component (A.2.2), 361

- coordinate (A.2.2), 361
 decomposition (18.5.1a), 247
 electric field (19.3.5), 267
 field (18.5.3), 248
 inner product (A.2.6a), 362
 length, norm (A.2.13a), 364
 orthogonality (A.2.6b), 362
 perpendicularity (A.2.6b), 362
 space basis (A.2.9), 363
 space (A.2.1), 360
 unit $\vec{e}_{0,1}$ (19.2.4c), 264
- Vector space
 as spacetime (3.2.2), 55
- Velocity
 constant (4.2.1), 69
-
- W**
- Walters Art Museum
 Archimedes scrolls
 (24.1.2), 336
- Water, and life
 (F.2.4), 410
- Watson, Thomas (21.7.1), 299
- Wave
 equation, 1-dimensional
 (16.4.1a), 209
- frequency of (16.1.2), 200
 in a rope (16.1.4), 200
 in spacetime (16.3.2), 205
 period of (16.1.2), 200
 skipping stone model
 (16.5.4), 211
 standing (17.6.6a), 236
- Weinberg, Steven (24.8.5), 354
- Wheel-axle configuration
 orienting loops (18.4.5), 245
- Wilson, James. Q., when
 science thrives
 (F.6.2d), 418
- Windshield wipers
 invention of (24.2.1), 337
- Worldline (4.1.2), 68
- Wright, Orville & Wilbur
 (24.2.1), 337
-
- X**
- X-axis and Simultaneity (4.4.3), 73
-
- Z**
- Zeeman, Pieter
 (19.1), 262