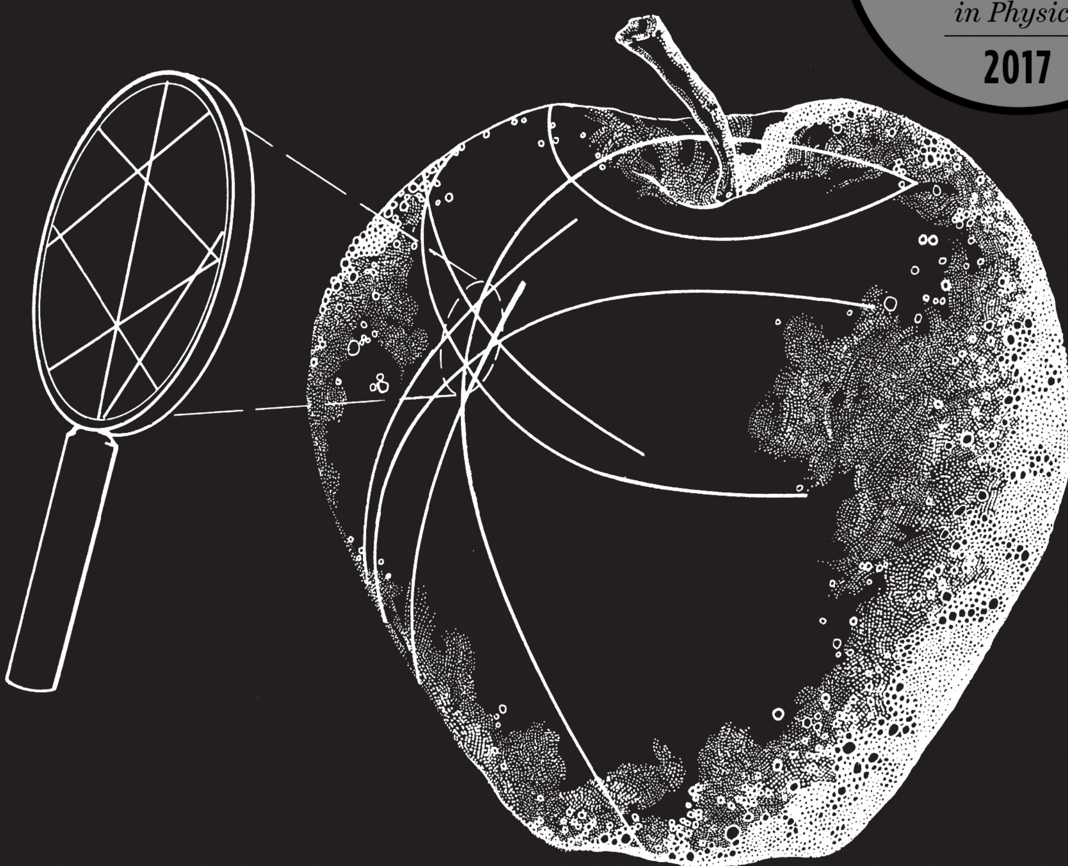
# GRAVITATION

Charles W. MISNER    Kip S. THORNE    John Archibald WHEELER

KIP THORNE
*Co-winner*
**NOBEL PRIZE**
*in Physics*
2017



WITH A NEW FOREWORD BY DAVID I. KAISER AND
A NEW PREFACE BY CHARLES W. MISNER AND KIP S. THORNE

# GRAVITATION


The Crab Nebula (NGC 1952), the remains of the supernova of July 1054, an event observed and recorded at the Sung national observatory at K'ai-feng. In the intervening 900 years, the debris from the explosion has moved out about three lightyears; i.e., with a speed about 1/300 of that of light. In 1934 Walter Baade and Fritz Zwicky predicted that neutron stars should be produced in supernova explosions. Among the first half-dozen pulsars found in 1968 was one at the center of the Crab Nebula, pulsing 30 times per second, for which there is today no acceptable explanation other than a spinning neutron star. The Chinese historical record shown here lists unusual astronomical phenomena observed during the Northern Sung dynasty. It comes from the "Journal of Astronomy," part 9, chapter 56, of the *Sung History* (*Sung Shih*), first printed in the 1340's. The photograph of that standard record used in this montage is copyright by, and may not be reproduced without permission of, the Trustees of the British Museum.

# GRAVITATION

Charles W. MISNER Kip S. THORNE John Archibald WHEELER

*With a new foreword by*
**David I. KAISER**

*and a new preface by*
**Charles W. MISNER and Kip S. THORNE**

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD

Published by Princeton University Press,
41 William Street, Princeton, New Jersey 08540

In the United Kingdom: Princeton University Press,
6 Oxford Street, Woodstock, Oxfordshire OX20 1TR

press.princeton.edu

Cover illustration by Kenneth Gwin

Errata and additional information on sources may be found at the following webpage:
http://press.princeton.edu/titles/11169.html



ISBN 978-0-691-17779-3

Library of Congress Control Number:
2017950565

Original edition published in 1973 by W. H. Freeman and Company

First Princeton University Press edition, with a new foreword by David I. Kaiser and a new preface by
Charles W. Misner and Kip S. Thorne, 2017

British Library Cataloging-in-Publication Data is available

Printed on acid-free paper. ∞

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

*We dedicate this book*
*To our fellow citizens*
*Who, for love of truth,*
*Take from their own wants*
*By taxes and gifts,*
*And now and then send forth*
*One of themselves*
*As dedicated servant,*
*To forward the search*
*Into the mysteries and marvelous simplicities*
*Of this strange and beautiful Universe,*
*Our home.*

# CONTENTS

# BOXES

# FIGURES

# FOREWORD TO THE 2017 PRINTING OF *GRAVITATION*

**DAVID I. KAISER**

A remarkable publishing event occurred in September 1973: the release of a 1,279-page book, weighing more than six pounds, with the simple title, *Gravitation*.[1] Wags were quick to remark that the book was not just about gravitation, but a significant source of it. The book acquired several nicknames, including "the phone book" (another reference to its girth) and "the big black book" (for its sleek, modern cover). Most common became "MTW," named for the authors' initials: Charles Misner, Kip Thorne, and John Wheeler.[2]

*Gravitation* focuses on the general theory of relativity, Albert Einstein's remarkable theory of gravity. Einstein completed a version of this theory, in a form we would recognize today, just over a century ago, presenting the finishing touches in a flurry of brief communications to the Prussian Academy of Sciences in November 1915. His major insight was that space and time were actors in the story of nature, not merely a fixed stage on which all other activity played out. Space and time, on Einstein's account, were dynamical—they could bend and distend in response to the distribution of matter and

Program in Science, Technology, and Society, and Department of Physics, Massachusetts Institute of Technology. Cambridge, Mass. dikaiser@mit.edu.

Portions of this essay are adapted from David Kaiser, "A Tale of Two Textbooks: Experiments in Genre," *Isis* 103 (March 2012): 126–38.

[1] The following abbreviations are used in the notes: **JAW**, John A. Wheeler papers, American Philosophical Society, Philadelphia, Pennsylvania; **KST**, Kip S. Thorne papers, in Professor Thorne's possession, California Institute of Technology, Pasadena, California.

[2] Charles W. Misner, Kip S. Thorne, and John A. Wheeler, *Gravitation* (San Francisco: W. H. Freeman, 1973). On nicknames for the book, see, e.g., "Chicago Undergraduate Physics Bibliography," available at http://www.ocf.berkeley.edu/~abhishek/chicphys.htm.

energy. That warping, in turn, would affect objects' motion, diverting them from the straight and narrow path.[3]

One year after the armistice that ended the First World War, a British team, led by Arthur Eddington, announced that they had confirmed one of Einstein's key predictions: that gravity could bend the path of starlight. The dramatic announcement propelled Einstein and his general theory to instant stardom. Yet interest in the theory waned over the 1930s. Einstein himself noted plaintively, in a preface for a colleague's textbook in 1942, "I believe that more time and effort might well be devoted to the systematic teaching of the theory of relativity than is usual at present at most universities."[4]

Years passed, but eventually some charismatic teachers began to heed Einstein's call. Among the first and most influential was John Wheeler, who began to offer Physics 570, a full-length course on general relativity, at Princeton University in 1954. He quickly attracted world-class graduate students to the subject, including Charles Misner and Kip Thorne. Fifteen years later, concerned that textbooks on general relativity had failed to keep up with modern developments, Misner, Thorne, and Wheeler teamed up to write *Gravitation*.[5] On publication, *Gravitation* joined several other new books about general relativity, including Steven Weinberg's *Gravitation and Cosmology* (1972) and Stephen Hawking's and George Ellis's *The Large Scale Structure of Space-Time* (1973).[6] Unlike those books, however, MTW defied many people's expectations for a textbook. Some just didn't know what to make of it.

[3] For succinct introductions to the early history of Einstein's work on general relativity, see Michel Janssen, "'No success like failure': Einstein's quest for general relativity," in *The Cambridge Companion to Einstein*, ed. Michel Janssen and Christoph Lehner (New York: Cambridge University Press, 2014), 167–227; Hanoch Gutfreund and Jürgen Renn, *The Road to Relativity: The History and Meaning of Einstein's "The Foundation of General Relativity"* (Princeton, NJ: Princeton University Press, 2015); and Michel Janssen and Jürgen Renn, "Arch and scaffold: How Einstein found his field equations," *Physics Today* 68 (November 2015): 30–36.

[4] Albert Einstein, "Foreword," in Peter G. Bergmann, *Introduction to the Theory of Relativity* (New York: Prentice-Hall, 1942), v. On Eddington's eclipse expedition and the early reception of general relativity, see Jean Eisenstaedt, *The Curious History of Relativity: How Einstein's Theory of Gravity was Lost and Found Again* (Princeton, NJ: Princeton University Press, 2006); Jeffrey Crelinstein, *Einstein's Jury: The Race to Test Relativity* (Princeton, NJ: Princeton University Press, 2006); Matthew Stanley, *Practical Mystic: Religion, Science, and A. S. Eddington* (Chicago: University of Chicago Press, 2007), chapter 3; and Hanoch Gutfreund and Jürgen Renn, *The Formative Years of Relativity: The History and Meaning of Einstein's Princeton Lectures* (Princeton, NJ: Princeton University Press, 2017).

[5] On the return of general relativity to physics departments' course offerings during the 1950s and 1960s, see David Kaiser, "A *psi* is just a *psi*? Pedagogy, practice, and the reconstitution of general relativity, 1942–1975," *Studies in the History and Philosophy of Modern Physics* 29 (1998): 321–338; Daniel Kennefick, *Traveling at the Speed of Thought: Einstein and the Quest for Gravitational Waves* (Princeton, NJ: Princeton University Press, 2007), chapter 6; and Alexander Blum, Roberto Lalli, and Jürgen Renn, "The reinvention of general relativity: A historiographical framework for assessing one hundred years of curved space-time," *Isis* 106, no. 3 (September 2015): 598–620. On Wheeler as an effective mentor, see Charles W. Misner, Kip S. Thorne, and Wojciech H. Zurek, "John Wheeler, relativity, and quantum information," *Physics Today* 62 (April 2009): 40–46; and Terry M. Christensen, "John Wheeler's mentorship: An enduring legacy," *Physics Today* 62 (April 2009): 55–59.

[6] Steven Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (New York: Wiley, 1972); S. W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Space-Time* (New York: Cambridge University Press, 1973).

Misner, Thorne, and Wheeler clearly intended *Gravitation* to be a textbook, pitched at advanced physics students. Wheeler's notes from an early planning meeting with his coauthors made clear that they would write the book with "the committee planning graduate courses in U. of X" in mind. While certainly thinking in terms of a textbook, however, from the start they treated the project as an experiment in the genre. The book was organized into two tracks: a core of introductory material occupying less than a third of the book, surrounded by extensions, elaborations, and applications.[7] The two tracks were not sequential; many chapters were divided, section by section, into one track or the other. Even more novel was the extensive use of "boxes" for complementary material. The boxes were set off from the main text by heavy black lines, interrupting the flow of ordinary chapter exposition, often for several pages at a time. Some of the boxes resembled the sidebars that had long been a staple of science textbooks aimed at younger students, featuring short biographies of famous physicists or brief descriptions of important experiments. But most of the boxes in *Gravitation* served a different purpose. According to Wheeler's notes, the boxes were meant to constitute "a third channel of pedagogy," beyond the two tracks. "They are distinguished from the main text by untidiness" and included "the kinds of things we would like to present in lecture hour to students who can be relied upon to learn tightly organized material and computational methods on their own from a systematic text." Their pedagogical aspirations were clear: as each author drafted a section of the book, the coauthors would "*test* a write up by asking if a student could use it to lecture from."[8]

The authors devoted spectacular attention to the physical appearance and production of the book. Thorne traded detailed letters with the artists and layout designers at the original publisher (W. H. Freeman in San Francisco), going over everything from the thickness of lines to set off the box material to arrow styles and shadings to be adopted in the hundreds of illustrations. Early on, Thorne alerted an editor at Freeman that "several features of the manuscript will require special typesetting problems." Beyond the extensive figures, tables, and boxes, the authors anticipated the need for at least six distinct typefaces, perhaps as many as eight, to properly distinguish the plethora of symbols and equations they would be treating.[9] (Before the book had even been published, Thorne worried that "the

[7] John Wheeler, handwritten notes, "Thoughts on preface, Mon., 13 July 1970," in JAW Series IV, Box F-L, Folder, "Gravitation: Notes with Charles W. Misner and Kip S. Thorne" ("committee planning graduate courses"). See also form letter from Misner, Thorne, and Wheeler to colleagues announcing forthcoming publication of the book, 13 June 1973, in KST Folder "MTW: Sample pages."

[8] John Wheeler, handwritten notes, page for insertion into draft of preface, n.d., late August 1970 ("third channel of pedagogy"); Wheeler, handwritten notes, "Plan of Book, Sat., 18 July 1970" ("*test* a write up"), both in JAW Series IV, Box F-L, Folder, "Gravitation: Notes with Charles W. Misner and Kip S. Thorne." (Emphasis in original.) On sidebars in more elementary physics textbooks, see Sharon Traweek, *Beamtimes and Lifetimes: The World of High-Energy Physicists* (Cambridge, MA: Harvard University Press, 1988), 76–81.

[9] Kip Thorne to Earl Tondreau (editor at W. H. Freeman), 14 October 1970, in KST Folder "MTW: Correspondence, 1970–May, 1973" ("Several features," typefaces). See also Thorne to Robert Ishikawa and Aidan Kelley (W. H. Freeman), 28 January 1971, in KST Folder "MTW"; and Evan Gillespie (W. H. Freeman) to Kip Thorne, 29 November 1972, in KST Folder "MTW: Publishing company, 1970–71, 1971–72."

extreme complexity of the typography" would bedevil foreign-language publishers. He recommended that they simply photograph the equations from the English edition once it became available rather than attempt to retypeset them.[10]) Given the book's unusual organization, the authors also inserted thousands of marginal comments throughout the book. Some comments summarized the material under discussion, but many others were "dependency statements": a roadmap spelling out which other sections a given discussion depended on, and which others would in turn depend on it.[11]

Having tackled every detail of composition and typesetting, imagine the authors' surprise when—two years into the process, and just three weeks before they submitted their final, edited manuscript—they learned that the publisher held a rather different conception of the book than they did. After meeting with their editor from the press, Thorne shot off a letter to his coauthors. "I was rather shocked to learn from Bruce [Armbruster, the editor] that the people at Freeman are so out-of-touch with our book that they have not been regarding it as a textbook, but rather as a technical monograph. I suppose that the enormous size of the book has something to do with it." The publisher's plan had been to produce an expensive hardcover edition, intended primarily for purchase by libraries: "Freeman had not been expecting to pick up the textbook market with this book" at all. Thorne worked hard to convince the editor that "there might be some hope of picking up student sales" as well, but that would require a complete overhaul of the publisher's printing and pricing plans.[12]

Was *Gravitation* a reference monograph for libraries or a textbook for classroom use? From that ontological difference sprang more immediate considerations. For example, how could they keep such a fabulous concoction from crumbling under its own weight? The book's unusual trim size—each of its nearly 1,300 pages was more than an inch wider and taller than standard textbooks at the time—suggested hardcover rather than paperback binding. Hardcover binding seemed all the more appropriate to the authors, for whom *Gravitation* was self-evidently a textbook, since (as Thorne explained), "it seems to me that paperback editions cannot hold up well enough with the heavy use that a student in a full year course would give the book." But hardcover binding threatened to price the book beyond the reach of a student market.[13] After assurances from the publisher that paperback binding could hold up just as ruggedly as hardcover, the authors struck a deal with W. H. Freeman: in exchange for reduced royalty rates on the paperback edition, the press would aim to keep the price of the paperback lower than the hardcover price of the recent textbook by Weinberg, *Gravitation and Cosmology*. On publication, the paperback edition of Misner, Thorne, and Wheeler's *Gravitation* sold for $19.95 (about $110 in 2017 dollars),

[10] Kip Thorne to Ya. B. Zel'dovich and I. D. Novikov, 21 June 1973, in KST Folder "MTW: Correspondence, June, 1973–."

[11] Thorne to Ishikawa and Kelley, 28 January 1971 ("dependency statements").

[12] Kip Thorne to John Wheeler and Charles Misner, with cc to Bruce Armbruster, 17 February 1972, in KST Folder, "MTW: Correspondence, 1970–May, 1973."

[13] Thorne to Wheeler and Misner with cc to Armbruster, 17 February 1972. See also Misner, Thorne, and Wheeler, form letter to colleagues, 13 June 1973, in KST Folder, "MTW: Sample pages."

and the hardcover for twice that price. With the publisher now treating the book as a textbook rather than a reference monograph, and with the compromise pricing plan in place, Thorne was confident that the book could "capture one hundred percent of the textbook market in this field—or as nearly so as possible."[14]

Like the authors and publisher, reviewers recognized the book as unusual. "A pedagogic masterpiece," announced a reviewer in *Science*; "one of the great books of science, a lamp to illuminate this Aladdin's cave of theoretical physics whose genie was Albert Einstein," crowed another in *Science Progress*. A third reviewer challenged his readers: "Imagine that three highly inventive people get together to invent a scientific book. Not just to write it, but invent the tone, the style, the methods of exposition, the format." Many reviewers lauded the rich set of illustrations and the innovative use of boxes.[15] Others complained that the two-track-plus-box organization introduced too many redundancies. "This is a difficult book to read in a linear, progressive fashion," concluded one reviewer; "there is needless repetition (indeed, almost everything is stated at least three times)," noted another. "The variety of gimmicks is bewildering—framed headings with quotations, marginal titles, 'boxes' sometimes extending over several pages, heavy type, light type, large type, small type," reported a reviewer in *Contemporary Physics*. "Clearly the book is an experiment in presentation on a grand scale."[16]

Nearly all reviewers commented on the writing style. Wheeler was already well known among physicists for his catchy slogans and engaging prose. (Among other memorable contributions, Wheeler had coined the term "black hole.") Wheeler's early planning notes for the book insisted that he and his coauthors must "make clear the idea itself. But soberly, factually, no hyperbole, no enthusiasm."[17] If that had been the intention, not all reviewers agreed on the outcome. The book featured a "prose style varying from the unusually colloquial to the unusually lyrical," wrote one reviewer. But one person's lyricism was another's doggerel. "There is a commendable attempt at informality, but this reviewer found the

[14] Thorne to Bruce Armbruster, 10 April 1973 (royalty rates, pricing vis-à-vis Weinberg's book, "capture one hundred percent"), in KST Folder "MTW: Publishing company, 1970–71, 1971–72." On pricing, see also Thorne to Richard Warrington (president), Peter Renz (science editor), and Lew Kimmick (financial manager) at W. H. Freeman, 14 February 1979, in JAW Series II, Box Fr-Gl, Folder "W. H. Freeman and Co., Publishers"; Thorne to Wheeler and Misner, 2 November 1972, in KST Folder "MTW"; Misner to Wheeler and Thorne, 18 November 1982, in KST Folder "MTW" (copy also in JAW Series II, Box Fr-Gl, Folder "W. H. Freeman and Co., Publishers"); and royalty statement from June 1993 in KST Folder "MTW: Royalty statements."

[15] Dennis Sciama, "Modern view of general relativity," *Science* 183 (March 22, 1974): 1186 ("pedagogic masterpiece"); Michael Berry, review in *Science Progress* 62, no. 246 (1975): 356–360, on 360 ("Aladdin's cave"); David Park, "Ups and downs of 'Gravitation,'" *Washington Post* (April 21, 1974): 4 ("three highly inventive people"). See also D. Allan Bromley, review in *American Scientist* (January–February 1974): 101–102.

[16] L. Resnick, review in *Physics in Canada* (June 1975), clipping in KST Folder "MTW: Reviews" ("difficult book to read"); S. Chandrasekhar, "A vast treatise on general relativity," *Physics Today* (August 1974): 47–48, on 48 ("needless repetition"); W. H. McCrea, review in *Contemporary Physics* 15, no. 4 (July 1974), clipping in KST Folder "MTW: Reviews" ("variety of gimmicks").

[17] John Wheeler, handwritten "Thoughts on preface, Mon. 13 July 1970," in JAW Series IV, Box F-L, Folder, "Gravitation: Notes with Charles W. Misner and Kip S. Thorne" ("make clear the idea"). On Wheeler's style, see also John A. Wheeler with Kenneth Ford, *Geons, Black Holes, and Quantum Foam: A Life in Physics* (New York: W. W. Norton, 1998); and Misner, Thorne, and Zurek, "John Wheeler, relativity, and quantum information."

breeziness irritating at times," came one verdict. "A 'poetical' style is understandable if one deals with such [speculative] topics as 'pregeometry.' However, 'poetical' passages in differential geometry, for example, may obstruct the understanding of an ascetic reader," concluded another.[18] One reviewer huffed that the informal writing style "comes dangerously close to being patronisingly simplistic, to the point of insulting the reader's intelligence." Another reviewer was even more scandalized by the book's tone. The intended reader, he scoffed, would be most at home with the book "if he is a regular subscriber to *Time* magazine—the writing of these authors has much in common with its breathless style."[19] Subrahmanyan Chandrasekhar, the famed astrophysicist and Nobel laureate who had grown up in India, trained in Britain, and settled in the United States, likewise noted that the book's "style fluctuates from precise mathematical rigor to evangelical rhetoric." He closed his review with the memorable observation: "There is one overriding impression this book leaves. 'It is written with the zeal of a missionary preaching to cannibals' (as J. E. Littlewood, in referring to another book, has said). But I (probably for historical reasons) have always been allergic to missionaries." (Thorne wrote to Chandrasekhar that the closing paragraph had left him "chuckling for about ten minutes.")[20]

While acknowledging the book's unusual organization, writing style, and pedagogical innovations, most reviewers treated the book as the authors had intended: a textbook primarily for graduate-level coursework in the technical details of gravitational physics. The authors had set out to corner the market for textbooks on the topic, and they largely succeeded. A few years after publication, their book was still selling between 4,000 and 5,000 copies per year, while their main competitor, Weinberg's *Gravitation and Cosmology*, had dropped to around 1,000 copies per year. Thorne noted to the original publisher—with fanfare but not much hyperbole—that by the late 1970s, "a large fraction of the physics graduate students in the Western world bought a copy of *Gravitation*."[21] The book sold 50,000 copies during its first decade, at a time when institutions in the United States graduated about 1,000 PhDs in physics per year, and no other country came close to those annual totals.[22]

[18] Sciama, "Modern view of general relativity," 1186 ("prose style"); Resnick, review in *Physics in Canada* ("commendable attempt"); J. Bicak, review in *Bulletin of the Astronomical Institute of Czechoslovakia* 26, no. 6 (1975): 377–378 ("A 'poetical' style").

[19] Alan Farmer, review in *Journal of the British Interplanetary Society* 27 (1974): 314–315, on 314 ("comes dangerously close"); Ian Roxburgh, "Geometry is all, or is it?" *New Scientist* (September 26, 1974): 828 ("a regular subscriber").

[20] Chandrasekhar, "A vast treatise on general relativity," 48; Thorne to Chandrasekhar, 21 June 1974, in KST Folder "MTW: Reviews." On Chandrasekhar's career, see K. C. Wali, *Chandra: A Biography of S. Chandrasekhar* (Chicago: University of Chicago Press, 1991); and Arthur I. Miller, *Empire of the Stars: Obsession, Friendship, and Betrayal in the Quest for Black Holes* (Boston: Houghton Mifflin, 2005).

[21] Kip Thorne to Peter Renz, 15 June 1983, in KST Folder "MTW" ("large fraction of the physics graduate students"); Thorne to Warrington, Renz, and Kimmick, 14 February 1979, on annual sales of *Gravitation* and Weinberg's textbook.

[22] Sales figures from royalty statement of June 1993 in KST Folder "MTW: Royalty statements." On PhD conferral rates, see David Kaiser, "Cold war requisitions, scientific manpower, and the production of American physicists after World War II," *Historical Studies in the Physical and Biological Sciences* 33 (2002): 131–159;

Yet from the start, some readers saw much more in *Gravitation* than a vehicle for training soon-to-be specialists. The original publisher, for one, reversed course in a dramatic way. A decade after having written off the book as merely a reference work for library purchase, editors at W. H. Freeman decided to advertise a specially reduced price on the book—nearly 25% off list price—to subscribers of the popular magazine *Scientific American*. Thorne countered that a better way to test "the elasticity in the demand" for the book would be to offer that reduced price to "that portion of the market which concerns me most": students and young academics. He urged the publisher to offer the reduced price to university bookstores rather than *Scientific American* devotees.[23]

Nonetheless, the publisher was on to something. On the book's publication, reviews had run not just in such venues as *Science* and *Physics Today*; the *Washington Post* devoted a full-page review to the book, and a daily newspaper in San Antonio, Texas, likewise recommended it. The reviewer in the *Post*, himself a physicist at Williams College, acknowledged that "perhaps it is strange to review here a textbook full of mathematics, a book, moreover, whose 6.7-pound bulk the young, the old and the infirm can scarcely lift. But," he declared, "those who read like to know what is being published and discussed." And *Gravitation* certainly warranted discussion. The book's engaging prose "awakens hope that the fuzzy and lugubrious 'style' that still spreads its gloom over so much of American science may not be in fashion forever." Moreover, the book's unusual organization seemed akin to recent trends in avant-garde filmmaking, such as the French *nouvelle vague*. "There are very few stories that should be told sequentially," the reviewer avowed. All the better that *Gravitation*, like the hip filmmakers, had discovered "strategies for breaking up a linear narrative."[24] The San Antonio reviewer likewise encouraged his readers. "I am not a mathematician, and the 200 or so pages I've read are not all that formidable," he explained. "If you're curious and have an imagination, you won't be cowed. The challenge is stiff, but fascinating." The organization of the book was "phenomenal," and the topic inspiring. He concluded, "This is a fabulous, rewarding book." Novelists could scarcely hope for a more enthusiastic review.[25]

Fan letters also streamed in to the authors from a wide assortment of readers. Many came from colleagues who reported how much they enjoyed teaching from the book in their formal classes.[26] But others came from further afield. One reader wrote from a hospital in Italy—it is not clear whether the handwritten letter came from a patient or a physician—to

---

and David Kaiser, "Booms, busts, and the world of ideas: Enrollment pressures and the challenge of specialization," *Osiris* 27 (2012): 276–302.

[23] Kip Thorne to Peter Renz, 10 August 1983, in KST Folder "MTW."

[24] Park, "Ups and downs of 'Gravitation,'" 4.

[25] Robert Pincus, "Gravity theory excites the mind," clipping in KST Folder "MTW: Reviews." The clipping does not indicate date, publication title, or page number, but advertisements on the same page as the review clearly indicate that the newspaper was based in San Antonio, Texas.

[26] See, e.g., Andrzej Trautman to Charles Misner, Kip Thorne, and John Wheeler, 10 January 1974, in KST Folder "MTW"; Heinz Pagels to Wheeler, 1 February 1974, in KST Folder "MTW Reviews"; Philip B. Burt to Wheeler, 12 November 1974, in KST Folder "MTW"; and Robert Rabinoff to Misner, Thorne, and Wheeler, 10 March 1978, in KST Folder "MTW: Reviews."

press the authors on whether their views about the cosmos had changed during the three years since the book's publication. (The letter writer had been keeping up with more recent discussions in the field by reading the Italian-language version of *Scientific American*.) He had more specific questions, too. In particular, what was the fate of life in a universe that cycled from big bang to big crunch? He was so desperate for a response that he promised $200 to anyone (the authors or their graduate students) who might take the time to answer. "Don't be offended by my proposal. Time = Money."[27]

An engineer in Brussels turned to the book for a different reason. He decided to pick up *Gravitation* to help him learn English before beginning military service. "My hopes have been completely fulfilled: *Gravitation* is worth reading to learn English because it makes enjoy Physics!" The book so inspired him that he drew seven full-page, whimsical cartoons in the style of Antoine de Saint-Exupéry's *The Little Prince* to illustrate concepts he had learned from *Gravitation*.[28]

Readers closer to home wrote to the authors as well. Especially poignant was a letter that Thorne received from a reader in Portland, Oregon. "I stumble here, fall down there, and generally make a fool of myself as I wander about your textbook," the correspondent explained, "but I am gaining a sense of balance and a few tools with which to deal with the subject." His dedication to the book was impressive:

> When friends ask me about what I am doing I have made the mistake of telling them the truth [about his attempts to read *Gravitation*]. Sometimes I think they are right, I feel as though I am on the brink of madness. I go out to have a beer and listen to someone talk about his love affairs, the clutch in his pick-up truck, the problems with his children, the plumbing, the bus service. I look at him and see him dealing with all these important issues and I ask myself why do I care if I ever understand the difference between leptons and leprosy?

Yet still he could not shake his "obsession" with Einstein's own question, "whether or not God had any choice in the creation of the Universe." He needed to know: "Could God be a traveling technician whose responsibility is to supervise gravitational collapses and big bangs?"[29]

Six years after publication, with annual sales still brisk, John Wheeler tried to assess the reasons for the book's success. Writing to his editor, Wheeler surmised that "many people buy the book who are attracted by the mystique, the boxes, the interesting illustrations, the ideas but who don't expect to and never will get deep into the mathematics." He figured

[27] Luigi Vignato to Charles Misner, Kip Thorne, and John Wheeler, 20 July 1976, in KST Folder "MTW: Correspondence, June, 1973–"; Wheeler to Vignato, 2 August 1976, in the same folder. Wheeler did not directly address Vignato's question, but he did enclose a preprint of his recent essay: John Wheeler, "Genesis and observership," in *Foundational Problems in the Special Sciences*, ed. Robert E. Butts and Jaakko Hintikka (Boston: Reidel, 1977), 3–33.

[28] Jadoul Michel to Charles Misner, Kip Thorne, and John Wheeler, August 1983, in KST Folder "MTW."

[29] Dan Foley to Kip Thorne, 7 February 1980, in KST Folder "MTW." See also Thorne to Foley, 27 February 1980, in the same folder.

about half the purchasers fell into that category—and he was eager not to lose them. In thinking about revising and updating the book, Wheeler concluded that "I think we can add a few things and take away a lot of things to keep this group 'on board.'"[30] Those plans fell through—Misner, Thorne, and Wheeler never did undertake a revision of their massive masterpiece—but Wheeler's observation nonetheless rang true. In their effort to write a specialized textbook, they had produced a hybrid work, as attractive to *Scientific American* subscribers for its "mystique" as to doctoral students struggling to enter the field.

*Gravitation* narrowly escaped the pigeonhole of library-only reference work and went on to sell tens of thousands of copies. The book received extensive analysis and review in physicists' specialist journals, even as it inspired passion—even ecstacy—among journalists and nonspecialist readers. Somehow this hulking book, stuffed to overflowing with equations so complicated they required multiple typefaces and elaborate marginal notes, excited broad, crossover appeal. To this day, Misner, Thorne, and Wheeler's book—like Einstein's elegant theory at its core—continues to inspire alluringly large questions. Why are we here? What is our place in the cosmos? Einstein helped spur those questions a century ago. *Gravitation* marks a major milestone on that continuing quest. It is a tribute to Princeton University Press that this fascinating book—with its jumble of equations, nonlinear structure, and, at times, soaring lyricism—will once again be easily available to students and seekers alike.

## ACKNOWLEDGMENTS

My thanks to Professor Kip Thorne for generously sharing his personal files with me and for granting permission to quote from them; to Charles Greifenstein and Anne Harney of the American Philosophical Society for their assistance with the John A. Wheeler papers; and to Marie Burks, Michael Gordin, Yoshiyuki Kikuchi, Roberto Lalli, Bernard Lightman, David Singerman, Alma Steingart, Marga Vicedo, Benjamin Wilson, and Aaron Wright for helpful comments on an earlier draft. I also thank Ingrid Gnerlich for her kind invitation to contribute this Foreword to the new edition of Misner, Thorne, and Wheeler's magnificent volume, *Gravitation*.

[30] John Wheeler to Peter Renz, 28 June 1979, in KST Folder "MTW"; copy also in JAW Series II, Box Fr-Gl, Folder "W. H. Freeman and Co. Publishers."

# PREFACE TO THE 2017 PRINTING OF *GRAVITATION*

**CHARLES W. MISNER AND KIP S. THORNE**

As we look back on our sixty-year love affair with Einstein's general relativity, our primary emotion is joy: joy at having participated in an amazingly fruitful era of exploration and transformation. In geologists' terminology, we have lived a blessed bit of the Anthropocene epoch from a favored perch in the world, seeing wonders, while, fortunately, avoiding personally the wars and devastations that have afflicted so many others.

There is an immense contrast in human understanding of gravity in action from the 1950s, when John Wheeler recruited us into Einstein's arena, to the present time. In the 1950s, curved spacetime was a complex though beautiful way to interpret one observational datum from each of four phenomena: the bending of light by the Sun, the perihelion motion of Mercury, the gravitational redshift from the white dwarf 40 Eridani B, and the expansion of the universe. Today we have observational data by the megabyte. The icons for these data are (1) the WMAP-based plot of the variations of temperature of the cosmic microwave radiation as a function of angular scale—the marker for the advent of precision cosmology; and (2) the "chirp" plots of LIGO's first directly observed gravitational wave, marking the advent of gravitational wave astronomy. Along with these icons, there has been a wealth of other great insights and discoveries as the general relativity community expanded from a few dozen to a few thousand during the six decades since 1952, when John Wheeler began dreaming of this textbook.

## THE CONTEXT IN WHICH WE WROTE *GRAVITATION*

General relativity had an exciting first two decades (1915–1939) and then became a two-decade backwater for physicists (1939–1958), as nuclear physics, elementary particle physics, and condensed matter physics came to the fore. In parallel, in mathematics, the field now called differential geometry was blossoming. For example, the concept of a manifold was clarified in the decades of the 1930s through the 1950s, and Milnor (1956) placed a capstone on this progress when he showed, by an example with the seven-dimensional sphere, that two manifolds that are equivalent at the level of continuous functions could be different in an essential way at the level of differentiable functions.

We were fortunate to enter relativity near the beginning of a remarkable renaissance (ca. 1958–1978), one enabled in part by the new mathematics and driven initially, in large measure, by our mentor John Wheeler, and then driven by a sequence of astronomical discoveries: the cosmic microwave background (CMB), and phenomena associated with black holes and neutron stars: quasars, pulsars, jets from galactic nuclei, compact X-ray sources, and gamma-ray bursts. It was late in this renaissance that we wrote *Gravitation*.

The relativity textbooks that preceded *Gravitation* were too old to incorporate the wonderful new observations and the new mathematical underpinnings. They treated Riemannian geometry as Einstein and then Pauli (1921) had, with almost no concept of the idea of a topological manifold that could carry properties (such as tangent vectors and 1-forms) even though it had been assigned no metric. They also, then, used no idea of points in the manifold (events in spacetime) as being conceptually superior to the various lists of coordinates used to identify them. And these texts tended to describe the physics almost entirely in terms of (old-fashioned) mathematics, with little attention paid to the heuristic but powerful tools by which modern physicists make rapid progress: physical arguments and pictures, geometric diagrams, and intuitive viewpoints. In *Gravitation*, our goal was to present relativity in physicists' physical, visual, and intuitive language, accompanied by the modern mathematics from which this language springs. The result was an advanced textbook with a far larger word-to-equation ratio than anything ever before seen in this field; a book filled with "purple prose," as John's wife, Janette, referred to it. But a book that also teaches relativity's mathematical underpinnings.

With our purple prose and pictures, we sought to transform how scientists think about relativity. And we think we succeeded, at least to some degree.

## *GRAVITATION*'S GEOMETRIC VIEWPOINT

A major part of our approach is the geometric viewpoint on general relativity that we learned from John—a viewpoint that contrasts starkly with the field-theoretic viewpoint taken by Steven Weinberg in the relativity textbook (Weinberg 1972) that he wrote in parallel with our writing *Gravitation*.

For situations where spacetime is strongly curved and where we focus on regions comparable to or larger than its radius of curvature (e.g., black holes and a closed model universe), this geometric viewpoint is essential, or at least superior. For the causal structure of spacetime (horizons, singularities, Hawking's second law of black hole mechanics), it is also essential. For most other situations, while not essential, it is powerful. And whenever field-theoretic techniques are more useful than geometry (e.g., in the evolution of structure in the early universe), one can easily descend from the heights of geometry to the nitty gritty of field theory. (OK. Our prejudice is showing. Starkly.)

After decades steeped in the geometric viewpoint, one of us (Kip) has become so enamored of it that, with Stanford astrophysicist Roger Blandford, he has crafted a much broader textbook permeated with this viewpoint: a book titled *Modern Classical Physics* (Thorne and Blandford 2017; henceforth "MCP"), which covers all the areas of classical physics that PhD physicists should be exposed to but often are not, at least in North America. That book and this reprinting of *Gravitation* are being published simultaneously by the same publisher, Princeton University Press.

## HOW USEFUL CAN *GRAVITATION* BE TODAY?

*Gravitation* was published in 1973, near enough to the end of the Relativistic Renaissance that most of that Renaissance's major theoretical insights and observational discoveries were in hand. While there have been some major additional insights and discoveries in the four decades since, they are few enough that *Gravitation* is seriously out of a date in only a moderate number of areas; primarily cosmology (Part VI), gravitational waves (Part VIII), experimental tests of general relativity (Part IX), and observations but not the theory of black holes and neutron stars (Parts V and VII).

This may account, in part, for *Gravitation*'s longevity: it continues to be used as supplemental reading in a large number of relativity courses around the world even today, 44 years after its publication. And in recent years, it has still been the primary textbook for a few courses.

## CHAPTER-BY-CHAPTER STATUS OF *GRAVITATION*

As an aid to students, teachers, and other readers as they choose a path through relativity in the modern era, we offer here a chapter-by-chapter description of what in *Gravitation* is out of date and what is not; what is missing that we think so important that we would include it in a full year, advanced course in general relativity if we were teaching one; and where readers can go to learn about the missing developments.

1. **Parts I, II, III, and IV, the fundamentals of general relativity**, have not changed significantly over the past 44 years, so Chapters 1–22 on the fundamentals are almost fully up to date. The only exceptions are the following.
   A. *Chapter 8, Differential Geometry*, should be augmented by an introduction to symbolic manipulation software (e.g., *Maple*, *Mathematica*, and *Matlab*) for computing connection coefficients and curvature tensors and performing other tensorial calculations; and *Chapter 14, Calculation of Curvature*, could be augmented by a deeper treatment of symbolic manipulation.

B. To *Part IV, Einstein's Geometric Theory of Gravity*, we would add four new topics:
   a. *Numerical relativity*, which underpins gravitational wave observations and is teaching us about the nonlinear dynamics of curved spacetime; for example, Maggiore (2017), or for far greater detail, Baumgarte and Shapiro (2010) and Shibata (2016).
   b. *Gravitational lensing*, which is based on the linearized approximation to general relativity (Section 18.1) and has become a major tool for astronomy; for example, MCP or Straumann (2013), or for far greater detail, Schneider, Ehlers, and Falco (1992).
   c. *The Einstein field equation in higher dimensions*, particularly four space dimensions and one time dimension, which is motivated by string theory's requirement for higher dimensions and by the Randall-Sundrum (1999a,b) insight that one or more of these higher dimensions could be macroscopic. This topic often goes under the name "Braneworlds." For a brief treatment see, for example, Zee (2013); for much greater detail at the level of *Gravitation*, see Maartens and Koyama (2010).
   d. *Quantum field theory in curved spacetime* (which could be added at the end of Chapter 22). This topic underpins, most importantly, Hawking radiation from black holes; see below. For a brief introduction see, for example, Carroll (2004); for more thorough treatments, see Wald (1994) and Parker and Toms (2009).

2. **Part V, Relativistic Stars**, is similarly almost fully up to date, with the following two exceptions.
   A. *Chapter 24, Pulsars and Neutron Stars; Quasars and Supermassive Stars*, is completely out of date. Observations and observation-driven astrophysical theory have transformed our understanding profoundly. See, for example, Straumann (2013) or Maggiore (2017) or, for far greater detail, Shapiro and Teukolsky (1983), which is somewhat out of date but excellent and thorough.
   B. *Chapter 25* on geodesic orbits in the Schwarzschild spacetime should be augmented by exercises on computing orbits numerically to give the reader physical insight—which is best done by numerically integrating the Hamilton equations that follow from the super-Hamiltonian (Exercise 25.2); see Levin and Perez-Giz (2008).

3. **Part VI, The Universe**, is for the most part tremendously out of date.
   A. *Chapter 27, Idealized Cosmologies*, is an exception. The fundamental ideas and equations for idealized cosmologies have not changed, but the emphasis of this chapter is archaic. John, our mentor—whose intuition and prescience

were usually superb (Misner, Thorne, and Zurek 2009)—was firmly convinced that our universe would turn out to be closed and have vanishing cosmological constant; so in *Gravitation*, the closed Friedman cosmology is given great emphasis. Since *Gravitation* was published, a rich set of cosmological observations has revealed that our universe is very nearly flat spatially and has a positive cosmological constant (or something resembling it). So this chapter should be augmented by a more detailed treatment of the material in Section 27.11, and most importantly, by an in-depth treatment of the de Sitter solution of the Einstein equation with cosmological constant—as, for example, in Hawking and Ellis (1973). As a side issue (a Box), we would add the anti–de Sitter (AdS) solution (e.g., Hawking and Ellis 1973), because of its importance today in explorations of fundamental physics (e.g., the AdS/CFT correspondence).

B. *Chapter 28, Evolution of the Universe into Its Present State*, and *Chapter 29, Present State and Future Evolution of the Universe*, are completely out of date and thus only of historic interest. During the past two decades, these subjects have been thoroughly transformed by cosmological observations and associated theory. For a fully up-to-date, pedagogical treatment, we recommend chapter 28 of MCP, or at a more elementary level, Schneider (2015). The most useful advanced textbook may be Weinberg (2008).

C. Cosmological observations over the past two decades suggest that *Chapter 30, Anisotropic and Homogeneous Cosmologies* is likely not relevant to the early evolution of our universe. However, it is of great importance for a fundamental new topic to be discussed below (see 4.D): the physical structure of singularities.

D. To this cosmological Part of *Gravitation*, we would add a major new topic (chapter): *Inflationary expansion in the very early universe*, as treated, for example, in Peacock (1999); Hobson, Efstathiou, and Lasenby (2006); Sasaki (2015); and section 28.7.1 of MCP.

4. **Part VII, Gravitational Collapse and Black Holes**, is surprisingly up to date, in large measure because it focuses on theory and says little about observations. However, a few new theoretical developments (some major) have emerged since 1973 that should be included in any year-long advanced course on general relativity.

A. To *Chapter 31, Schwarzschild Geometry*, and *Chapter 32, Gravitational Collapse*, we would add nothing.

B. To *Chapter 33, Black Holes*, we would add the following topics.

a. Exercises to explore geodesic orbits around a Kerr black hole numerically, by integrating Hamilton's equations for the super-Hamiltonian (33.27c) (Levin and Perez-Giz 2008).

b. A discussion of quasinormal modes of a Kerr black hole, motivated by Exercise 33.14; see, for example, chapter 12 of Maggiore (2017). (The first hint of these modes was found by Vishveshwara, 1970, in the form of ringdown waves like those that LIGO has detected 45 years later. By 1973, when *Gravitation* was published, the concept of quasinormal modes was fully in hand along with the equations for computing them, but the first numerical computation of their complex eigenfrequencies and eigenfunctions, by Chandrasekhar and Detweiler, 1975, was still two years in the future.)
c. Spherical accretion onto a Schwarzschild black hole and accretion disks around a Kerr black hole: at least a few exercises as, for example, in MCP. These topics are touched on in Box 33.3 of *Gravitation*, but given their great astrophysical importance today, they deserve greater and more up-to-date detail; see, for example, the brief discussion in Straumann (2013), the longer discussion in Abramovicz and Fragile (2013), or the very detailed discussion in Meier (2012).
d. The Blandford-Znajek (1977) mechanism by which magnetic fields extract spin energy from black holes to power jets; see, for example, MCP. For far greater detail, see Thorne, Price, and MacDonald (1986), which emphasizes the relativity, and McKinney, Tchekhovskoy, and Blandford (2012), which emphasizes the astrophysics.
e. Some discussion of astronomical observations of black holes and their astrophysical roles in the universe; see, for example, Narayan and McClintock (2015) or for greater detail, Meier (2012) and Schneider (2015). (What remarkable developments there have been here, since *Gravitation* was published!)
f. Hawking radiation, the associated thermal atmosphere of a black hole, and black-hole thermodynamics (all of which were developed within a year of publication of *Gravitation*, in the wake of Stephen Hawking's and others' bringing quantum field theory in curved spacetime into a sufficiently mature form; see 1.B.c above). See, for example, Carroll (2004) for a moderately brief, pedagogical treatment, and Wald (1994) for greater detail.

C. To *Chapter 34, Global Techniques, Horizons, and Singularity Theorems*, we would add a new set of topics that have been explored using global techniques since *Gravitation* was published: wormholes and topological censorship; and closed timelike curves, chronology horizons, and chronology protection. See, for example, Everitt and Roman (2012) for a not very technical discussion with references to the most important literature or Friedman and Higuchi (2008) for greater technical detail.

D. To *Part VII* we would also add the equivalent of one more chapter on *The Physical Structure of Generic Singularities and the Interiors of Black Holes*. This chapter would include the following.
  a. The material in Chapter 30 (3.C above) on the Kasner and Mixmaster solutions of Einstein's equation, plus a more detailed discussion of the Belinsky, Khalatnikov, and Lifshitz (BKL) analysis, which suggests there is a generic, spatially inhomogeneous variant of Mixmaster (p. 806 of *Gravitation*); also, a description of numerical relativity simulations (Garfinkle 2004; Lim et al. 2009) that prove this to be true and reveal some surprising twists missed by BKL.
  b. Analyses that show that the inner horizons, $r = r_-$, of a Kerr or Reissner-Nordstrøm black hole (Fig. 34.4) are highly unstable and that material or radiation falling into the hole triggers these instabilities, converting the inner horizons into generic null singularities (Poisson and Israel 1990; Marolf and Ori 2012).

5. **In Part VIII, Gravitational Waves**, the chapters on the theory of the waves and their generation are largely up to date, but the chapter on their detection is extremely out of date. More specifically:
A. *Chapter 35, Propagation of Gravitational Waves*, is essentially up to date.
B. The topics covered in *Chapter 36, Generation of Gravitational Waves*, are essentially up to date, but they need to be augmented by the following.
  a. An overview of gravitational wave sources that are likely to be observed in the next decade or two; see, for example, Buonanno and Sathyaprakash (2015); Creighton and Anderson (2011); or, for far greater detail, Maggiore (2017).
  b. A sketch of the post-Newtonian expansion of the waves from compact binary stars (higher-order corrections to this chapter's quadrupolar analysis); see, for example, Straumann (2013); and for greater detail, Poisson and Will (2014) or Blanchet (2014). Exercise 39.15 of *Gravitation* could be a starting point for this.
  c. A description of numerical relativity simulations of the inspiral and merger of black-hole binaries, and black-hole/neutron-star binaries, and their gravitational waves (e.g., Choptuik, Lehner, and Pretorius 2015; Maggiore 2017); also, the nonlinear dynamics of curved spacetime triggered by black-hole mergers (e.g., Owen et al. 2011; Scheel and Thorne 2013).
  d. A sketch of the analysis that shows that early-universe inflation parametrically amplifies gravitational vacuum fluctuations coming off the big bang, to produce a spectrum of primordial gravitational waves; see, for example, Mukhanov (2005) or Maggiore (2017).

C. *Chapter 37, Detection of Gravitational Waves*, is highly out of date. Although nothing is wrong with this chapter, and it can be of conceptual value (particularly Sections 37.1–37.3), it focuses on vibrating mechanical detectors, which have largely been abandoned. So in a modern course, we would replace Sections 37.4–37.10 by the following.
   a. An overview of the four types of detectors that are expected to open up four different gravitational-wave frequency bands in the next two decades: ground-based interferometers, such as LIGO, which have already opened the high-frequency band (10–10,000 Hz); space-based detectors, such as LISA, in which drag-free spacecraft track each other with laser beams, that are expected to open up the low-frequency band (periods of minutes to hours) in the next 15 or 20 years; pulsar timing arrays (PTAs), which are expected to open the very low frequency band (periods of a year to a few tens of years) in the coming decade; and so-called B-mode polarization patterns in the CMB, which are induced by primordial gravitational waves with periods of millions to billions of years (the extremely low frequency band) and which may be definitively measured in the next decade or so. See, for example, Berger et al. (2015) and Maggiore (2017).
   b. Detailed analyses of (idealized) ground-based interferometers, space-based detectors, and PTAs (e.g., Creighton and Anderson 2011; Saulson 2017; MCP); and analyses of the influence of gravitational waves on CMB polarization (e.g., Maggiore 2017).
   c. A summary of observations of gravitational waves, which in 2017 are solely those by LIGO, such as Abbott et al. (2016a).

D. This could be a good spot, in an advanced course, to present an overview of what is known about the nonlinear dynamics of vacuum, curved spacetime (e.g., Scheel and Thorne 2014)—most of which has already been mentioned above.
   a. The chaotic spacetime dynamics near a generic Mixmaster (BKL) singularity (4.D.a).
   b. The more gentle dynamics near a generic null singularity (4.D.b).
   c. The phase transitions, critical behavior, and scaling that show up in (nongeneric) “critical” gravitational collapse; for example, Choptuik, Lehner, and Pretorius (2015).
   d. The interacting “tidal tendices” and “frame-drag vortices” that generate the gravitational waves in black-hole collisions; for example, MCP, or for greater detail Scheel and Thorne (2014), or for still greater detail Owen et al. (2011).

e. Nonlinear, two-dimensional turbulence (energy cascades from small scales to large scales) triggered by mode–mode coupling in perturbations of a fast spinning black hole; see Yang, Zimmerman, and Lehner (2015).

We suspect that these just "scratch the surface" on nonlinear spacetime dynamics, and that a rich range of other phenomena will be discovered in the coming years.

6. **Part IX, Experimental Tests of General Relativity**, is all correct, but since *Gravitation* was published, rapidly improving technology and vigorous efforts by creative experimenters have moved the most accurate experimental tests from errors of a few percent to errors as small as one part in 100,000; so, obviously, a huge amount of updating is necessary.
   A. A modern course might simply follow the discussion of experimental tests in recent pedagogical references, such as Will (2014, 2015).
   B. Or it might do the following.
      a. Preserve the discussion of foundational tests in *Chapter 38*, augmented by an overview of the current status of those and related experiments from Will (2014, 2015).
      b. Preserve the pedagogical discussion of the post-Newtonian approximation and the parametrized post-Newtonian formalism in *Chapter 39*, augmented by the corresponding analysis for the orbital motion of compact binaries (a straightforward extension of Exercise 39.15).
      c. Preserve the analysis of solar system experiments in *Chapter 40*, augmented by an overview of the current status of those experiments as in Will (2014, 2015).
      d. Add discussion and some analyses of experimental tests in binary pulsars (e.g., Straumann 2013; Will 2014, 2015); and also experimental tests based on gravitational wave observations of binary black holes, for which expectations are discussed in Yunes and Siemens (2013) and in Gair et al. (2013), and results are just beginning to emerge from LIGO (e.g., Abbott et al. 2016b).

7. **Part X, Frontiers**, is a beautiful overview of some important ideas that occupied John Wheeler's attention in the era when we wrote this book with him.
   A. *Chapter 41, Spinors*, is an introduction to this important topic in mathematical physics—an introduction that mixes the deep mathematics with the intuitive, visual, and physical viewpoint that was John's hallmark. This chapter stands on its own, with no need for change.

B. The *Regge Calculus*, laid out so beautifully in *Chapter 42*, has played a powerful conceptual role in general relativity for decades, but has never (yet) become an effective tool for numerical computations.

C. *Superspace*, as treated in *Chapter 43*, has long been a powerful underpinning for some approaches to formulating laws of quantum gravity.

D. *Chapter 44, Beyond the End of Time*, describes prescient ideas on which John focused in the 1960s–1980s. It is of great historical import, and it contains ideas that continue to have influence.

We commend these chapters to readers, followed by a perusal of modern applications on the physics archive, https://arxiv.org.

*Gravitation* and these updates clearly constitute far more material than can be covered in a full year course, just as *Gravitation* by itself did in 1973, when first published. Today, as then, a teacher or student or reader will want to select which portions to focus on, and at what depth. But the above summary does convey what we think important and worthy of study in 2017.

## ACKNOWLEDGMENTS

Above all, we are indebted to our mentor and coauthor, John Archibald Wheeler, who enticed us into the arena of general relativity six decades ago with his optimism, enthusiasm, and eagerness for adventure.

Many colleagues, friends, students—and students of students of students—have rewarded us by embracing this heavy tome. They have our sincere thanks. We hope they continue to appreciate the beauty of the ideas described in our book: the intellectual universe that Einstein opened for humanity more than 100 years ago. And we hope and expect that they will help others, in coming decades, to see the magnificence and subtlety of Nature through this window.

We relied on textbooks, as well as on John, as we struggled to learn general relativity in the 1950s and early 1960s. We thank the authors of those texts: Peter Bergmann, Christian Møller, Richard Tolman, John Synge, and Lev Landau and Evgeny Lifshitz. Beyond these, John Wheeler encouraged Misner to look at the 1955 text by André Lichnerowicz, which does reflect the then-current differential geometry. Much of Part III in *Gravitation* reflects Misner's efforts to help Wheeler restate the differential geometry that Misner had learned from his Notre Dame mathematics mentor Arnold Ross, his Princeton mathematics advisor Donald Spencer, and fellow Princeton graduate students.

For the research that has transformed this field, financial support was needed. On behalf of our colleagues as well as ourselves, we thank the program directors who targeted that support with great wisdom over the early years, particularly Joshua Goldberg (for the U.S. Air Force), and Harry Zapolsky and then Richard Isaacson (for the National Science

Foundation), followed by many others when the necessary investments got large. The impact of gravitational wave observations will be huge over the coming decades. The chief architect on the rocky course from small-scale R&D to the massively big collaborations required for success was Barry Barish, while Joseph Weber's pioneering insights, courage, and determination from his beginnings with Wheeler in 1956 should never be overlooked.

For helping bring this textbook to fruition in 1973, we are indebted to many people; see the original Acknowledgments on page li.

Until 2015, *Gravitation* continued to sell many hundred copies a year. Then, through a series of acquisitions of publishers by publishers by publishers, it wound up in the hands of Macmillan, which took it out of print. Through persistence, finesse, and firmness, Joan Winstein succeeded in extracting all rights to *Gravitation* from Macmillan, and to her we are deeply grateful.

We were fortunate that Princeton University Press eagerly embraced the idea of producing a new hardback printing at a remarkably low purchase price. We thank the superb staff at the Press, who have worked so effectively with us to bring this printing to fruition, particularly Peter Dougherty, Ingrid Gnerlich, Arthur Werneck, Karen Carter, Lisa Black, and Jessica Massabrook.

And we thank David Kaiser for his beautiful new Foreword to this printing.

July 1, 2017

## REFERENCES

Abbott, B. P., R. Abbott, T. D. Abbott, M. R. Abernathy, et al., 2016a, "Observation of gravitational waves from a binary black hole merger," *Phys. Rev. Lett. 116,* 061102.

Abbott, B. P., R. Abbott, T. D. Abbott, M. R. Abernathy, et al., 2016b, "Tests of general relativity with GW150914," *Phys. Rev. Lett. 116,* 221101.

Abramovicz, M. A., and P. C. Fragile, 2013, "Foundations of black hole accretion disk theory," *Living Rev. Relat. 16,* 1.

Ashtekar, A., B. K. Berger, J. Isenberg, and M. MacCallum, 2015, *General Relativity and Gravitation: A Centenial Perspective,* Cambridge University Press, Cambridge.

Baumgarte, T. W., and S. Shapiro, 2010, *Numerical Relativity: Solving Einstein's Equations on the Computer,* Cambridge University Press, Cambridge.

Berger, B. K., K. Danzmann, G. Gonzalez, et al., 2015, "Receiving gravitational waves," in Ashtekar et al. (2015).

Blanchet, L., 2014, "Gravitational radiation from post-Newtonian sources and inspiraling compact binaries," *Living Rev. Relat. 17,* 2.

Blandford, R. D., and R. L. Znajek, 1977, "The electromagnetic extraction of energy from Kerr black holes," *Mon. Not. Roy. Astron. Soc. 179,* 433–456.

Buonanno, A., and B. S. Sathyaprakash, 2015, "Sources of gravitational waves: Theory and observations," in Ashtekar et al. (2015).

Carroll, S., 2004, *Spacetime and Geometry: An Introduction to General Relativity,* Pearson Education, Harlow, Essex, England.

Chandrasekhar, S., and S. Detweiler, 1975, "The quasi-normal modes of the Schwarzschild black hole," *Proc. Roy. Soc. A 344,* 441–452.

Choptuik, M. W., L. Lehner, and F. Pretorius, 2015, "Probing strong-field gravity through numerical simulations," in Ashtekar et al. (2015).

Creighton, D. E., and W. G. Anderson, 2011, *Gravitational-Wave Physics and Astronomy,* Wiley-VCH, Weinheim, Germany.

Everitt, A., and T. Roman, 2012, *Time Travel and Warp Drives,* University of Chicago Press, Chicago.

Friedman, J. L., and A. Higuchi, 2008, "Topological censorship and chronology protection," *Ann. Phys. 15,* 109–128. arXiv:0801.0735.

Gair, J. R., M. Vallisneri, S. L. Larson, and J. G. Baker, 2013, "Testing general relativity with low-frequency, space-based gravitational-wave detectors," *Living Rev. Relat. 16,* 7.

Garfinkle, D., 2004, "Numerical simulations of generic singularities," *Phys. Rev. Lett. 93,* 161101.

Hawking, S. W., and G.F.R. Ellis, 1973, *The Large Scale Structure of Space-Time,* Cambridge University Press, Cambridge.

Hobson, M. P., G. Efstathiou, and A. N. Lasenby, 2006, *General Relativity: An Introduction for Physicists,* Cambridge University Press, Cambridge.

Levin, J., and G. Perez-Giz, 2008, "A periodic table for black hole orbits," *Phys. Rev. D 77,* 103005.

Lim, W. C., L. Andersson, D. Garfinkle, and F. Pretorius, 2009, "Spikes in the mixmaster regime of $G_2$ cosmologies," *Phys. Rev. D 69,* 123526.

Maartens, R., and K. Koyama, 2010, "Brane-world gravity," *Living Rev. Relat. 13,* 5.

Maggiore, M., 2017, *Gravitational Waves, Volume 2: Astrophysics and Cosmology,* Oxford University Press, Oxford.

Marolf, D., and A. Ori, 2012, "Outgoing gravitational shock-wave at the inner horizon: The late-time limit of black hole interiors," *Phys. Rev. D 86,* 124026.

McKinney, J. C., A. Tchekhovskoy, and R. D. Blandford, 2012, "General relativistic magnetohydrodynamical simulations of magnetically choked accretion flows around black holes," *Mon. Not. Roy. Astron. Soc. 423,* 3083–3117.

Meier, D. L., 2012, *Black Hole Astrophysics: The Engine Paradigm,* Springer, Cham, Switzerland.

Milnor, J. W., 1956, "On manifolds homeomorphic to the 7-sphere," *Ann. Math. 64,* 399–405.

Misner, C. W., K. S. Thorne, and W. H. Zurek, 2009, "John Wheeler, relativity, and quantum information," *Phys. Today,* April, 40.

Mukhanov, V., 2005, *Physical Foundations of Cosmology,* Cambridge University Press, Cambridge.

Narayan, R., and J. E. McClintock, 2015, "Observational evidence for black holes," in Ashtekar et al. (2015).

Owen, R., J. Brink, Y. Chen, J. D. Kaplan, et al., 2011, "Frame-dragging vortexes and tidal tendexes attached to colliding black holes: Visualizing the curvature of spacetime," *Phys. Rev. Lett. 106,* 151101.

Parker, L., and D. Toms, 2009, *Quantum Field Theory in Curved Spacetime: Quantized Fields and Gravity,* Cambridge University Press, Cambridge.

Pauli, Wolfgang, 1921, "Relativitätstheorie," in *Encyklopadie der Mathetamischen Wissenschaften II,* 539–775, B. G. Teubner, Leipzig.

Peacock, J. A., 1999, *Cosmological Physics*, Cambridge University Press, Cambridge.

Poisson, E., and W. Israel, 1990, "Internal structure of black holes," *Phys. Rev. D 41,* 1796–1809.

Poisson, E., and C. M. Will, 2014, *Gravity: Newtonian, Post-Newtonian, Relativistic,* Cambridge University Press, Cambridge.

Randall, L., and R. Sundrum, 1999a, "An alternative to compactification," *Phys. Rev. Lett. 83,* 4690–4693.

Randall, L. and R. Sundrum, 1999b, "Large mass hierarchy from a small extra dimension," *Phys. Rev. Lett. 83,* 3370–3373.

Sasaki, M., 2015, "Inflationary cosmology," in Ashtekar et al. (2015).

Saulson, P. R., 2017, *Fundamentals of Interferometric Gravitational Wave Detectors,* 2nd ed., World Scientific, Singapore.

Scheel, M. A., and K. S. Thorne, 2014, "Geometrodynamics: The nonlinear dynamics of curved spacetime," *Phys. Uspekhi 83,* 342–351.

Schneider, P., 2015, *Extragalactic Astronomy and Cosmology,* 2nd ed., Springer, Heidelberg.

Schneider, P., J. Ehlers, and E. Falco, 1992, *Gravitational Lensing,* Springer-Verlag, Berlin.

Shapiro, S. L., and S. A. Teukolsky, 1983, *Black Holes, White Dwarfs and Neutron Stars: The Physics of Compact Objects,* Wiley, New York.

Shibata, M., 2016, *Numerical Relativity,* World Scientific, Singapore.

Straumann, N., 2013, *General Relativity,* 2nd ed., Springer, Dordrecht.

Thorne, K. S., and R. D. Blandford, 2017, *Modern Classical Physics,* Princeton University Press, Princeton, N.J.; cited in text as MCP.

Thorne, K. S., R. H. Price, and D. A. MacDonald, 1986, *Black Holes: The Membrane Paradigm,* Yale University Press, New Haven, Conn.

Vishveshwara, C. V., 1970,"Scattering of gravitational radiation by a Schwarzschild black-hole," *Nature 227,* 936–938, Fig. 3.

Wald, R. M., 1994, *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics,* University of Chicago Press, Chicago.

Weinberg, S., 1972, *Gravitation and Cosmology. Principles and Applications of the General Theory of Relativity,* Wiley, New York.

Weinberg, S., 2008, *Cosmology,* Oxford University Press, Oxford.

Will, C. M., 2014, "The confrontation between general relativity and experiment," *Living Rev. Relat. 17,* 4.

Will, C. M., 2015, "Was Einstein right? A centenary assessment," in Ashtekar et al. (2015).

Yang, H., A. Zimmerman, and L. Lehner, 2015, "Turbulent black holes," *Phys. Rev. Lett. 114,* 081101.

Yunes, N., and X. Siemens, 2013, "Gravitational-wave tests of general relativity with ground-based detectors and pulsar-timing arrays," *Living Rev. Relat. 16,* 9.

Zee, A., 2013, *Einstein Gravity in a Nutshell,* Princeton University Press, Princeton, N.J.

# PREFACE

This is a textbook on gravitation physics (Einstein's "general relativity" or "geometrodynamics"). It supplies two tracks through the subject. The first track is focused on the key physical ideas. It assumes, as mathematical prerequisite, only vector analysis and simple partial-differential equations. It is suitable for a one-semester course at the junior or senior level or in graduate school; and it constitutes—in the opinion of the authors—the indispensable core of gravitation theory that every advanced student of physics should learn. The Track-1 material is contained in those pages of the book that have a 1 outlined in gray in the upper outside corner, by which the eye of the reader can quickly pick out the Track-1 sections. In the contents, the same purpose is served by a gray bar beside the section, box, or figure number.

The rest of the text builds up Track 1 into Track 2. Readers and teachers are invited to select, as enrichment material, those portions of Track 2 that interest them most. With a few exceptions, any Track-2 chapter can be understood by readers who have studied only the earlier Track-1 material. The exceptions are spelled out explicitly in "dependency statements" located at the beginning of each Track-2 chapter, or at each transition within a chapter from Track 1 to Track 2.

The entire book (all of Track 1 plus all of Track 2) is designed for a rigorous, full-year course at the graduate level, though many teachers of a full-year course may prefer a more leisurely pace that omits some of the Track-2 material. The full book is intended to give a competence in gravitation physics comparable to that which the average Ph.D. has in electromagnetism. When the student achieves this competence, he knows the laws of physics in flat spacetime (Chapters 1–7). He can predict orders of magnitude. He can also calculate using the principal tools of modern differential geometry (Chapters 8–15), and he can predict at all relevant levels of precision. He understands Einstein's geometric framework for physics (Chapters

16–22). He knows the applications of greatest present-day interest: pulsars and neutron stars (Chapters 23–26); cosmology (Chapters 27–30); the Schwarzschild geometry and gravitational collapse (Chapters 31–34); and gravitational waves (Chapters 35–37). He has probed the experimental tests of Einstein's theory (Chapters 38–40). He will be able to read the modern mathematical literature on differential geometry, and also the latest papers in the physics and astrophysics journals about geometrodynamics and its applications. If he wishes to go beyond the field equations, the four major applications, and the tests, he will find at the end of the book (Chapters 41–44) a brief survey of several advanced topics in general relativity. Among the topics touched on here, superspace and quantum geometrodynamics receive special attention. These chapters identify some of the outstanding physical issues and lines of investigation being pursued today.

Whether the department is physics or astrophysics or mathematics, more students than ever ask for more about general relativity than mere conversation. They want to hear its principal theses clearly stated. They want to know how to "work the handles of its information pump" themselves. More universities than ever respond with a serious course in Einstein's standard 1915 geometrodynamics. What a contrast to Maxwell's standard 1864 electrodynamics! In 1897, when Einstein was a student at Zurich, this subject was not on the instructional calendar of even half the universities of Europe.[1] "We waited in vain for an exposition of Maxwell's theory," says one of Einstein's classmates. "Above all it was Einstein who was disappointed,"[2] for he rated electrodynamics as "the most fascinating subject at the time"[3]—as many students rate Einstein's theory today!

Maxwell's theory recalls Einstein's theory in the time it took to win acceptance. Even as late as 1904 a book could appear by so great an investigator as William Thomson, Lord Kelvin, with the words, "The so-called 'electromagnetic theory of light' has not helped us hitherto . . . it seems to me that it is rather a backward step . . . the one thing about it that seems intelligible to me, I do not think is admissible . . . that there should be an electric displacement perpendicular to the line of propagation."[4] Did the pioneer of the Atlantic cable in the end contribute so richly to Maxwell electrodynamics—from units, and principles of measurement, to the theory of waves guided by wires—because of his own early difficulties with the subject? Then there is hope for many who study Einstein's geometrodynamics today! By the 1920's the weight of developments, from Kelvin's cable to Marconi's wireless, from the atom of Rutherford and Bohr to the new technology of high-frequency circuits, had produced general conviction that Maxwell was right. Doubt dwindled. Confidence led to applications, and applications led to confidence.

Many were slow to take up general relativity in the beginning because it seemed to be poor in applications. Einstein's theory attracts the interest of many today because it is rich in applications. No longer is attention confined to three famous but meager tests: the gravitational red shift, the bending of light by the sun, and

---

[1] G. Holton (1965).
[2] L. Kolbros (1956).
[3] A. Einstein (1949a).
[4] W. Thomson (1904).
Citations for references will be found in the bibliography.

the precession of the perihelion of Mercury around the sun. The combination of radar ranging and general relativity is, step by step, transforming the solar-system celestial mechanics of an older generation to a new subject, with a new level of precision, new kinds of effects, and a new outlook. Pulsars, discovered in 1968, find no acceptable explanation except as the neutron stars predicted in 1934, objects with a central density so high ($\sim 10^{14}$g/cm$^3$) that the Einstein predictions of mass differ from the Newtonian predictions by 10 to 100 per cent. About further density increase and a final continued gravitational collapse, Newtonian theory is silent. In contrast, Einstein's standard 1915 geometrodynamics predicted in 1939 the properties of a completely collapsed object, a "frozen star" or "black hole." By 1966 detailed digital calculations were available describing the formation of such an object in the collapse of a star with a white-dwarf core. Today hope to discover the first black hole is not least among the forces propelling more than one research: How does rotation influence the properties of a black hole? What kind of pulse of gravitational radiation comes off when such an object is formed? What spectrum of x-rays emerges when gas from a companion star piles up on its way into a black hole?[5] All such investigations and more base themselves on Schwarzschild's standard 1916 static and spherically symmetric solution of Einstein's field equations, first really understood in the modern sense in 1960, and in 1963 generalized to a black hole endowed with angular momentum.

Beyond solar-system tests and applications of relativity, beyond pulsars, neutron stars, and black holes, beyond geometrostatics (compare electrostatics!) and stationary geometries (compare the magnetic field set up by a steady current!) lies geometrodynamics in the full sense of the word (compare electrodynamics!). Nowhere does Einstein's great conception stand out more clearly than here, that the geometry of space is a new physical entity, with degrees of freedom and a dynamics of its own. Deformations in the geometry of space, he predicted in 1918, can transport energy from place to place. Today, thanks to the initiative of Joseph Weber, detectors of such gravitational radiation have been constructed and exploited to give upper limits to the flux of energy streaming past the earth at selected frequencies. Never before has one realized from how many kinds of processes significant gravitational radiation can be anticipated. Never before has there been more interest in picking up this new kind of signal and using it to diagnose faraway events. Never before has there been such a drive in more than one laboratory to raise instrumental sensitivity until gravitational radiation becomes a workaday new window on the universe.

The expansion of the universe is the greatest of all tests of Einstein's geometrodynamics, and cosmology the greatest of all applications. Making a prediction too fantastic for its author to credit, the theory forecast the expansion years before it was observed (1929). Violating the short time-scale that Hubble gave for the expansion, and in the face of "theories" ("steady state"; "continuous creation") manufactured to welcome and utilize this short time-scale, standard general relativity resolutely persisted in the prediction of a long time-scale, decades before the astro-

[5]As of April 1973, there are significant indications that Cygnus X-1 and other compact x-ray sources may be black holes.

physical discovery (1952) that the Hubble scale of distances and times was wrong, and had to be stretched by a factor of more than five. Disagreeing by a factor of the order of thirty with the average density of mass-energy in the universe deduced from astrophysical evidence as recently as 1958, Einstein's theory now as in the past argues for the higher density, proclaims "the mystery of the missing matter," and encourages astrophysics in a continuing search that year by year turns up new indications of matter in the space between the galaxies. General relativity forecast the primordial cosmic fireball radiation, and even an approximate value for its present temperature, seventeen years before the radiation was discovered. This radiation brings information about the universe when it had a thousand times smaller linear dimensions, and a billion times smaller volume, than it does today. Quasistellar objects, discovered in 1963, supply more detailed information from a more recent era, when the universe had a quarter to half its present linear dimensions. Telling about a stage in the evolution of galaxies and the universe reachable in no other way, these objects are more than beacons to light up the far away and long ago. They put out energy at a rate unparalleled anywhere else in the universe. They eject matter with a surprising directivity. They show a puzzling variation with time, different between the microwave and the visible part of the spectrum. Quasistellar objects on a great scale, and galactic nuclei nearer at hand on a smaller scale, voice a challenge to general relativity: help clear up these mysteries!

If its wealth of applications attracts many young astrophysicists to the study of Einstein's geometrodynamics, the same attraction draws those in the world of physics who are concerned with physical cosmology, experimental general relativity, gravitational radiation, and the properties of objects made out of superdense matter. Of quite another motive for study of the subject, to contemplate Einstein's inspiring vision of geometry as the machinery of physics, we shall say nothing here because it speaks out, we hope, in every chapter of this book.

Why a new book? The new applications of general relativity, with their extraordinary physical interest, outdate excellent textbooks of an earlier era, among them even that great treatise on the subject written by Wolfgang Pauli at the age of twenty-one. In addition, differential geometry has undergone a transformation of outlook that isolates the student who is confined in his training to the traditional tensor calculus of the earlier texts. For him it is difficult or impossible either to read the writings of his up-to-date mathematical colleague or to explain the mathematical content of his physical problem to that friendly source of help. We have not seen any way to meet our responsibilities to our students at our three institutions except by a new exposition, aimed at establishing a solid competence in the subject, contemporary in its mathematics, oriented to the physical and astrophysical applications of greatest present-day interest, and animated by belief in the beauty and simplicity of nature.

*High Island*
*South Bristol, Maine*
*September 4, 1972*

*Charles W. Misner*
*Kip S. Thorne*
*John Archibald Wheeler*

# ACKNOWLEDGMENTS

Deep appreciation goes to all who made this book possible. A colleague gives us a special lecture so that we may adapt it into one of the chapters of this book. Another investigator clears up for us the tangled history of the production of matter out of the vacuum by strong tidal gravitational forces. A distant colleague telephones in references on the absence of any change in physical constants with time. One student provides a problem on the energy density of a null electromagnetic field. Another supplies curves for effective potential as a function of distance. A librarian writes abroad to get us an article in an obscure publication. A secretary who cares types the third revision of a chapter. Editor and illustrator imaginatively solve a puzzling problem of presentation. Repeat in imagination such instances of warm helpfulness and happy good colleagueship times beyond count. Then one has some impression of the immense debt we owe to over a hundred-fifty colleagues. Each face is etched in our mind, and to each our gratitude is heartfelt. Warm thanks we give also to the California Institute of Technology, the Dublin Institute for Advanced Studies, the Institute for Advanced Study at Princeton, Kyoto University, the University of Maryland, Princeton University, and the University of Texas at Austin for hospitality during the writing of this book. We are grateful to the Academy of Sciences of the U.S.S.R., to Moscow University, and to our Soviet colleagues for their hospitality and the opportunity to become better acquainted in June–July 1971 with Soviet work in gravitation physics. For assistance in the research that went into this book we thank the National Science Foundation for grants (GP27304 and 28027 to Caltech; GP17673 and GP8560 to Maryland; and GP3974 and GP7669 to Princeton); the U.S. Air Force Office of Scientific Research (grant AF49-638-1545 to Princeton); the U.S. National Aeronautics and Space Agency (grant NGR 05-002-256 to Caltech, NSG 210-002-010 to Maryland); the Alfred P. Sloan Foundation for a fellowship awarded to one of us (K.S.T.); and the John Simon Guggenheim Memorial Foundation and All Souls College, Oxford, England, for fellowships awarded to another of us (C.W.M.).

PART I

# SPACETIME PHYSICS

*Wherein the reader is led, once quickly (§1.1), then again more slowly, down the highways and a few byways of Einstein's geometrodynamics—without benefit of a good mathematical compass.*

CHAPTER 1

# GEOMETRODYNAMICS IN BRIEF

## §1.1. THE PARABLE OF THE APPLE

*One day in the year 1666 Newton had gone to the country, and seeing the fall of an apple, as his niece told me, let himself be led into a deep meditation on the cause which thus draws every object along a line whose extension would pass almost through the center of the Earth.*

VOLTAIRE (1738)

Once upon a time a student lay in a garden under an apple tree reflecting on the difference between Einstein's and Newton's views about gravity. He was startled by the fall of an apple nearby. As he looked at the apple, he noticed ants beginning to run along its surface (Figure 1.1). His curiosity aroused, he thought to investigate the principles of navigation followed by an ant. With his magnifying glass, he noted one track carefully, and, taking his knife, made a cut in the apple skin one mm above the track and another cut one mm below it. He peeled off the resulting little highway of skin and laid it out on the face of his book. The track ran as straight as a laser beam along this highway. No more economical path could the ant have found to cover the ten cm from start to end of that strip of skin. Any zigs and zags or even any smooth bend in the path on its way along the apple peel from starting point to end point would have increased its length.

"What a beautiful geodesic," the student commented.

His eye fell on two ants starting off from a common point $P$ in slightly different directions. Their routes happened to carry them through the region of the dimple at the top of the apple, one on each side of it. Each ant conscientiously pursued

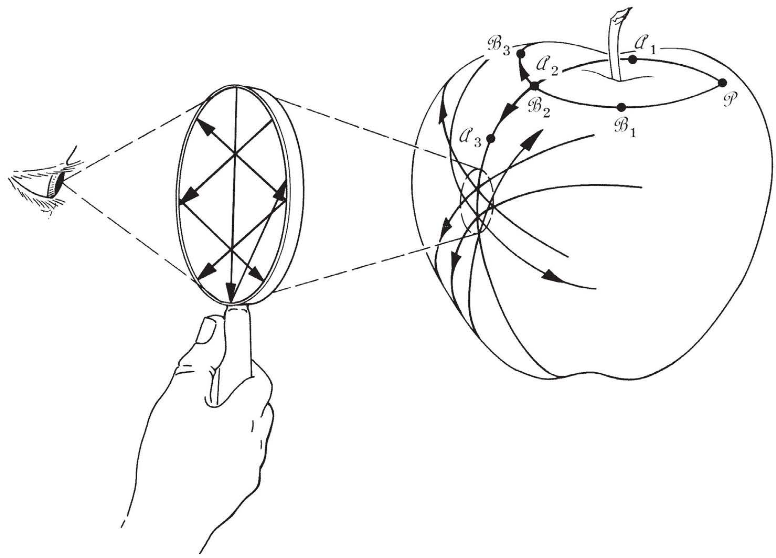



**Figure 1.1.**
The Riemannian geometry of the spacetime of general relativity is here symbolized by the two-dimensional geometry of the surface of an apple. The geodesic tracks followed by the ants on the apple's surface symbolize the world line followed through spacetime by a free particle. In any sufficiently localized region of spacetime, the geometry can be idealized as flat, as symbolized on the apple's two-dimensional surface by the straight-line course of the tracks viewed in the magnifying glass ("local Lorentz character" of geometry of spacetime). In a region of greater extension, the curvature of the manifold (four-dimensional spacetime in the case of the real physical world; curved two-dimensional geometry in the case of the apple) makes itself felt. Two tracks $\mathscr{A}$ and $\mathscr{B}$, originally diverging from a common point $\mathscr{P}$, later approach, cross, and go off in very different directions. In Newtonian theory this effect is ascribed to gravitation acting at a distance from a center of attraction, symbolized here by the stem of the apple. According to Einstein a particle gets its moving orders locally, from the geometry of spacetime right where it is. Its instructions are simple: to follow the straightest possible track (geodesic). Physics is as simple as it could be locally. Only because spacetime is curved in the large do the tracks cross. Geometrodynamics, in brief, is a double story of the effect of geometry on matter (causing originally divergent geodesics to cross) and the effect of matter on geometry (bending of spacetime initiated by concentration of mass, symbolized by effect of stem on nearby surface of apple).

his geodesic. Each went as straight on his strip of appleskin as he possibly could. Yet because of the curvature of the dimple itself, the two tracks not only crossed but emerged in very different directions.

Einstein's local view of physics contrasted with Newton's "action at a distance"

"What happier illustration of Einstein's geometric theory of gravity could one possibly ask?" murmured the student. "The ants move as if they were attracted by the apple stem. One might have believed in a Newtonian force at a distance. Yet from nowhere does an ant get his moving orders except from the local geometry along his track. This is surely Einstein's concept that all physics takes place by 'local action.' What a difference from Newton's 'action at a distance' view of physics! Now I understand better what this book means."

Physics is simple only when analyzed locally

And so saying, he opened his book and read, "Don't try to describe motion relative to faraway objects. *Physics is simple only when analyzed locally*. And locally the world line that a satellite follows [in spacetime, around the Earth] is already as straight as any world line can be. Forget all this talk about 'deflection' and 'force of gravitation.' I'm inside a spaceship. Or I'm floating outside and near it. Do I feel any 'force of gravitation'? Not at all. Does the spaceship 'feel' such a force? No. Then why talk about it? Recognize that the spaceship and I traverse a region of spacetime free of all force. Acknowledge that the motion through that region is already ideally straight."

The dinner bell was ringing, but still the student sat, musing to himself. "Let me see if I can summarize Einstein's geometric theory of gravity in three ideas: (1) locally, geodesics appear straight; (2) over more extended regions of space and time, geodesics originally receding from each other begin to approach at a rate governed by the curvature of spacetime, and this effect of geometry on matter is what we mean today by that old word 'gravitation'; (3) matter in turn warps geometry. The dimple arises in the apple because the stem is there. I think I see how to put the whole story even more briefly: *Space acts on matter, telling it how to move. In turn, matter reacts back on space, telling it how to curve.* In other words, matter here," he said, rising and picking up the apple by its stem, "curves space here. To produce a curvature in space here is to force a curvature in space there," he went on, as he watched a lingering ant busily following its geodesic a finger's breadth away from the apple's stem. "Thus matter here influences matter there. That is Einstein's explanation for 'gravitation.'"

Space tells matter how to move

Matter tells space how to curve

Then the dinner bell was quiet, and he was gone, with book, magnifying glass—and apple.

## §1.2. SPACETIME WITH AND WITHOUT COORDINATES

*Now it came to me: . . . the independence of the gravitational acceleration from the nature of the falling substance, may be expressed as follows: In a gravitational field (of small spatial extension) things behave as they do in a space free of gravitation. . . . This happened in 1908. Why were another seven years required for the construction of the general theory of relativity? The main reason lies in the fact that it is not so easy to free oneself from the idea that coordinates must have an immediate metrical meaning.*

ALBERT EINSTEIN [in Schilpp (1949), pp. 65–67.]

Nothing is more distressing on first contact with the idea of "curved spacetime" than the fear that every simple means of measurement has lost its power in this unfamiliar context. One thinks of oneself as confronted with the task of measuring the shape of a gigantic and fantastically sculptured iceberg as one stands with a meter stick in a tossing rowboat on the surface of a heaving ocean. Were it the rowboat itself whose shape were to be measured, the procedure would be simple enough. One would draw it up on shore, turn it upside down, and drive tacks in lightly at strategic points here and there on the surface. The measurement of distances from tack to

Problem: how to measure in curved spacetime

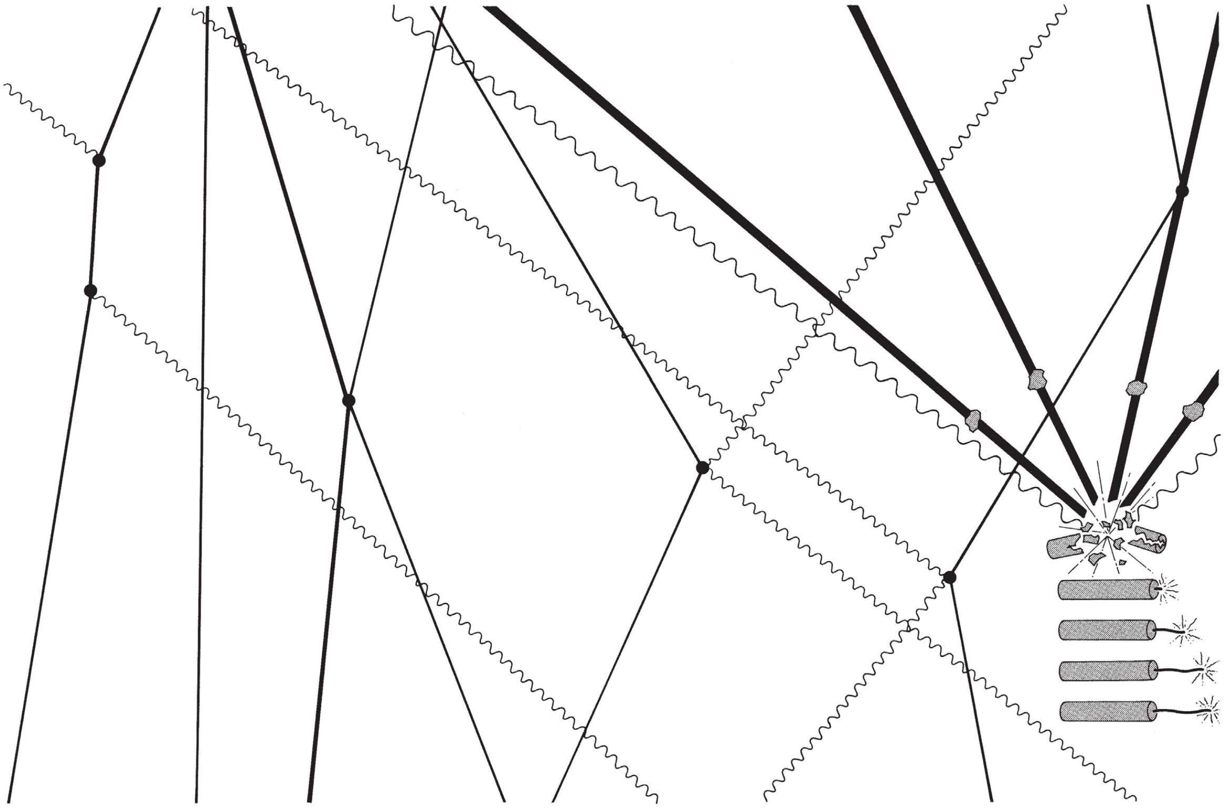


**Figure 1.2.**
The crossing of straws in a barn full of hay is a symbol for the world lines that fill up spacetime. By their crossings and bends, these world lines mark events with a uniqueness beyond all need of coordinate systems or coordinates. Typical events symbolized in the diagram, from left to right (black dots), are: absorption of a photon; reemission of a photon; collision between a particle and a particle; collision between a photon and a particle; another collision between a photon and a particle; explosion of a firecracker; and collision of a particle from outside with one of the fragments of that firecracker.

tack would record and reveal the shape of the surface. The precision could be made arbitrarily great by making the number of tacks arbitrarily large. It takes more daring to think of driving several score pitons into the towering iceberg. But with all the daring in the world, how is one to drive a nail into spacetime to mark a point? Happily, nature provides its own way to localize a point in spacetime, as Einstein was the first to emphasize. Characterize the point by what happens there! Give a point in spacetime the name "event." Where the event lies is defined as clearly and sharply as where two straws cross each other in a barn full of hay (Figure 1.2). To say that the event marks a collision of such and such a photon with such and such a particle is identification enough. The world lines of that photon and that particle are rooted in the past and stretch out into the future. They have a rich texture of connections with nearby world lines. These nearby world lines in turn are linked in a hundred ways with world lines more remote. How then does one tell the location of an event? Tell first what world lines participate in the event. Next follow each

Resolution: characterize events by what happens there

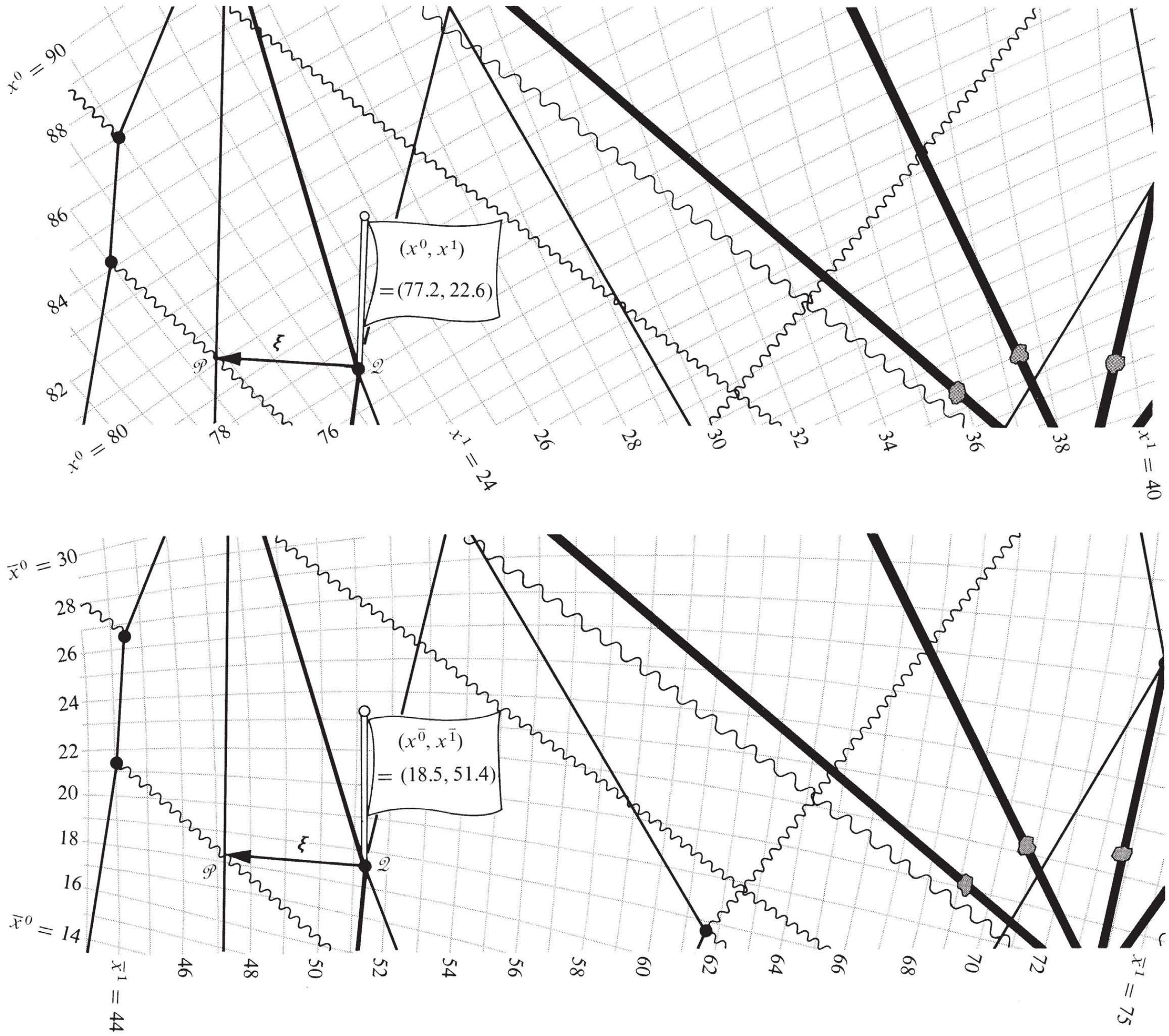



**Figure 1.3.**
Above: Assigning "telephone numbers" to events by way of a system of coordinates. To say that the coordinate system is "smooth" is to say that events which are almost in the same place have almost the same coordinates. Below: Putting the same set of events into equally good order by way of a different system of coordinates. Picked out specially here are two neighboring events: an event named "$\mathscr{Q}$" with coordinates $(x^0, x^1) = (77.2, 22.6)$ and $(x^{\bar{0}}, x^{\bar{1}}) = (18.5, 51.4)$; and an event named "$\mathscr{P}$" with coordinates $(x^0, x^1) = (79.9, 20.1)$ and $(x^{\bar{0}}, x^{\bar{1}}) = (18.4, 47.1)$. Events $\mathscr{Q}$ and $\mathscr{P}$ are connected by the separation "vector" $\boldsymbol{\xi}$. (Precise definition of a vector in a curved spacetime demands going to the mathematical limit in which the two points have an indefinitely small separation [$N$-fold reduction of the separation $\mathscr{P} - \mathscr{Q}$], and, in the resultant locally flat space, multiplying the separation up again by the factor $N$ [lim $N \to \infty$; "tangent space"; "tangent vector"]. Forego here that proper way of stating matters, and forego complete accuracy; hence the quote around the word "vector".) In each coordinate system the separation vector $\boldsymbol{\xi}$ is characterized by "components" (differences in coordinate values between $\mathscr{P}$ and $\mathscr{Q}$):

$$(\xi^0, \xi^1) = (79.9 - 77.2, 20.1 - 22.6) = (2.7, -2.5),$$

$$(\xi^{\bar{0}}, \xi^{\bar{1}}) = (18.4 - 18.5, 47.1 - 51.4) = (-0.1, -4.3).$$

See Box 1.1 for further discussion of events, coordinates, and vectors.

of these world lines. Name the additional events that they encounter. These events pick out further world lines. Eventually the whole barn of hay is catalogued. Each event is named. One can find one's way as surely to a given intersection as the city dweller can pick his path to the meeting of St. James Street and Piccadilly. No numbers. No coordinate system. No coordinates.

The name of an event can even be arbitrary

That most streets in Japan have no names, and most houses no numbers, illustrates one's ability to do without coordinates. One can abandon the names of two world lines as a means to identify the event where they intersect. Just as one could name a Japanese house after its senior occupant, so one can and often does attach arbitrary names to specific events in spacetime, as in Box 1.1.

Coordinates provide a convenient naming system

Coordinates, however, are convenient. How else from the great thick catalog of events, randomly listed, can one easily discover that along a certain world line one will first encounter event Trinity, then Baker, then Mike, then Argus—but not the same events in some permuted order?

To order events, introduce coordinates! (See Figure 1.3.) Coordinates are four indexed numbers per event in spacetime; on a sheet of paper, only two. Trinity acquires coordinates

$$(x^0, x^1, x^2, x^3) = (77, 23, 64, 11).$$

Coordinates generally do not measure length

In christening events with coordinates, one demands smoothness but foregoes every thought of mensuration. The four numbers for an event are nothing but an elaborate kind of telephone number. Compare their "telephone" numbers to discover whether two events are neighbors. But do not expect to learn how many meters separate them from the difference in their telephone numbers!

Several coordinate systems can be used at once

Nothing prevents a subscriber from being served by competing telephone systems, nor an event from being catalogued by alternative coordinate systems (Figure 1.3). Box 1.1 illustrates the relationships between one coordinate system and another, as well as the notation used to denote coordinates and their transformations.

Vectors

Choose two events, known to be neighbors by the nearness of their coordinate values in a smooth coordinate system. Draw a little arrow from one event to the other. Such an arrow is called a *vector*. (It is a well-defined concept in flat spacetime, or in curved spacetime in the limit of vanishingly small length; for finite lengths in curved spacetime, it must be refined and made precise, under the new name "tangent vector," on which see Chapter 9.) This vector, like events, can be given a name. But whether named "John" or "Charles" or "Kip," it is a unique, well-defined geometrical object. The name is a convenience, but the vector exists even without it.

Just as a quadruple of coordinates

$$(x^0, x^1, x^2, x^3) = (77, 23, 64, 11)$$

is a particularly useful name for the event "Trinity" (it can be used to identify what other events are nearby), so a quadruple of "components"

$$(\xi^0, \xi^1, \xi^2, \xi^3) = (1.2, -0.9, 0, 2.1)$$

**Box 1.1 MATHEMATICAL NOTATION FOR EVENTS, COORDINATES, AND VECTORS**

| | |
|---|---|
| **Events** are denoted by capital script, one-letter Latin names such as<br>Sometimes subscripts are used: | $\mathscr{P}, \mathscr{Q}, \mathscr{A}, \mathscr{B}$.<br>$\mathscr{P}_0, \mathscr{P}_1, \mathscr{B}_6$. |
| **Coordinates of an event** $\mathscr{P}$ are denoted by<br>or by<br><br>or more abstractly by<br>where it is understood that Greek indices can take on any value 0, 1, 2, or 3. | $t(\mathscr{P}), x(\mathscr{P}), y(\mathscr{P}), z(\mathscr{P})$,<br>$x^0(\mathscr{P}), x^1(\mathscr{P}), x^2(\mathscr{P}), x^3(\mathscr{P})$,<br><br>$x^\mu(\mathscr{P})$ or $x^\alpha(\mathscr{P})$, |
| **Time coordinate** (when one of the four is picked to play this role) | $x^0(\mathscr{P})$. |
| **Space coordinates** are<br>and are sometimes denoted by<br>It is to be understood that Latin indices take on values 1, 2, or 3. | $x^1(\mathscr{P}), x^2(\mathscr{P}), x^3(\mathscr{P})$<br>$x^j(\mathscr{P})$ or $x^k(\mathscr{P})$ or . . . . |
| **Shorthand notation:** One soon tires of writing explicitly the functional dependence of the coordinates, $x^\beta(\mathscr{P})$; so one adopts the shorthand notation for the coordinates of the event $\mathscr{P}$, and<br>for the space coordinates. One even begins to think of $x^\beta$ as representing the event $\mathscr{P}$ itself, but must remind oneself that the values of $x^0$, $x^1$, $x^2$, $x^3$ depend not only on the choice of $\mathscr{P}$ but also on the *arbitrary* choice of coordinates! | $x^\beta$<br>$x^j$ |
| **Other coordinates** for the same event $\mathscr{P}$ may be denoted<br><br>EXAMPLE: In Figure 1.3 $(x^0, x^1) = (77.2, 22.6)$ and $(x^{\bar{0}}, x^{\bar{1}}) = (18.5, 51.4)$ refer to the *same* event. The bars, primes, and hats distinguish one coordinate system from another; by putting them on the indices rather than on the $x$'s, we simplify later notation. | $x^{\bar{\alpha}}(\mathscr{P})$ or just $x^{\bar{\alpha}}$,<br>$x^{\alpha'}(\mathscr{P})$ or just $x^{\alpha'}$,<br>$x^{\hat{\alpha}}(\mathscr{P})$ or just $x^{\hat{\alpha}}$. |
| **Transformation** from one coordinate system to another is achieved by the four functions<br><br><br><br>which are denoted more succinctly | $x^{\bar{0}}(x^0, x^1, x^2, x^3)$,<br>$x^{\bar{1}}(x^0, x^1, x^2, x^3)$,<br>$x^{\bar{2}}(x^0, x^1, x^2, x^3)$,<br>$x^{\bar{3}}(x^0, x^1, x^2, x^3)$,<br>$x^{\bar{\alpha}}(x^\beta)$. |
| **Separation vector*** (little arrow) reaching from one event $\mathscr{Q}$ to neighboring event $\mathscr{P}$ can be denoted abstractly by<br>It can also be characterized by the coordinate-value differences† between $\mathscr{P}$ and $\mathscr{Q}$ (called "components" of the vector) | $\boldsymbol{u}$ or $\boldsymbol{v}$ or $\boldsymbol{\xi}$, or $\mathscr{P} - \mathscr{Q}$.<br><br>$\xi^\alpha \equiv x^\alpha(\mathscr{P}) - x^\alpha(\mathscr{Q})$,<br>$\xi^{\bar{\alpha}} \equiv x^{\bar{\alpha}}(\mathscr{P}) - x^{\bar{\alpha}}(\mathscr{Q})$. |
| **Transformation of components** of a vector from one coordinate system to another is achieved by partial derivatives of transformation equations<br>since $\xi^{\bar{\alpha}} = x^{\bar{\alpha}}(\mathscr{P}) - x^{\bar{\alpha}}(\mathscr{Q}) = (\partial x^{\bar{\alpha}}/\partial x^\beta)[x^\beta(\mathscr{P}) - x^\beta(\mathscr{Q})]$.† | $\xi^{\bar{\alpha}} = \dfrac{\partial x^{\bar{\alpha}}}{\partial x^\beta}\xi^\beta$, |
| **Einstein summation convention** is used here:<br>any index that is repeated in a product is automatically summed on | $\dfrac{\partial x^{\bar{\alpha}}}{\partial x^\beta}\xi^\beta \equiv \sum_{\beta=0}^{3} \dfrac{\partial x^{\bar{\alpha}}}{\partial x^\beta}\xi^\beta$. |

*This definition of a vector is valid only in flat spacetime. The refined definition ("tangent vector") in curved spacetime is not spelled out here (see Chapter 9), but flat-geometry ideas apply with good approximation even in a curved geometry, when the two points are sufficiently close.

†These formulas are precisely accurate only when the region of spacetime under consideration is flat and when in addition the coordinates are Lorentzian. Otherwise they are approximate—though they become arbitrarily good when the separation between points and the length of the vector become arbitrarily small.

is a convenient name for the vector "John" that reaches from

$$(x^0, x^1, x^2, x^3) = (77, 23, 64, 11)$$

to

$$(x^0, x^1, x^2, x^3) = (78.2, 22.1, 64.0, 13.1).$$

How to work with the components of a vector is explored in Box 1.1.

Coordinate singularities normally unavoidable

There are many ways in which a coordinate system can be imperfect. Figure 1.4 illustrates a coordinate singularity. For another example of a coordinate singularity, run the eye over the surface of a globe to the North Pole. Note the many meridians that meet there ("collapse of cells of egg crates to zero content"). Can't one do better? Find a single coordinate system that will cover the globe without singularity? A theorem says no. Two is the minimum number of "coordinate patches" required to cover the two-sphere without singularity (Figure 1.5). This circumstance emphasizes anew that points and events are primary, whereas coordinates are a mere bookkeeping device.

Continuity of spacetime

Figures 1.2 and 1.3 show only a few world lines and events. A more detailed diagram would show a maze of world lines and of light rays and the intersections between them. From such a picture, one can in imagination step to the idealized limit: an infinitely dense collection of light rays and of world lines of infinitesimal test particles. With this idealized physical limit, the mathematical concept of a continuous four-dimensional "manifold" (four-dimensional space with certain smoothness properties) has a one-to-one correspondence; and in this limit continuous, differentiable (i.e., smooth) coordinate systems operate. The mathematics then supplies a tool to reason about the physics.

The mathematics of manifolds applied to the physics of spacetime

Dimensionality of spacetime

A simple countdown reveals the dimensionality of the manifold. Take a point $\mathscr{P}$ in an $n$-dimensional manifold. Its neighborhood is an $n$-dimensional ball (i.e., the interior of a sphere whose surface has $n - 1$ dimensions). Choose this ball so that its boundary is a smooth manifold. The dimensionality of this manifold is $(n - 1)$. In this $(n - 1)$-dimensional manifold, pick a point $\mathscr{Q}$. Its neighborhood is an $(n - 1)$-dimensional ball. Choose this ball so that . . . , and so on. Eventually one comes by this construction to a manifold that is two-dimensional but is not yet known to be two-dimensional (two-sphere). In this two-dimensional manifold, pick a point $\mathscr{M}$. Its neighborhood is a two-dimensional ball ("disc"). Choose this disc so that its boundary is a smooth manifold (circle). In this manifold, pick a point $\mathscr{N}$. Its neighborhood is a one-dimensional ball, but is not yet known to be one-dimensional ("line segment"). The boundaries of this object are two points. This circumstance tells that the intervening manifold is one-dimensional; therefore the previous manifold was two-dimensional; and so on. The dimensionality of the original manifold is equal to the number of points employed in the construction. For spacetime, the dimensionality is 4.

This kind of mathematical reasoning about dimensionality makes good sense at the everyday scale of distances, at atomic distances ($10^{-8}$ cm), at nuclear dimensions ($10^{-13}$ cm), and even at lengths smaller by several powers of ten, if one judges by the concord between prediction and observation in quantum electrodynamics at high

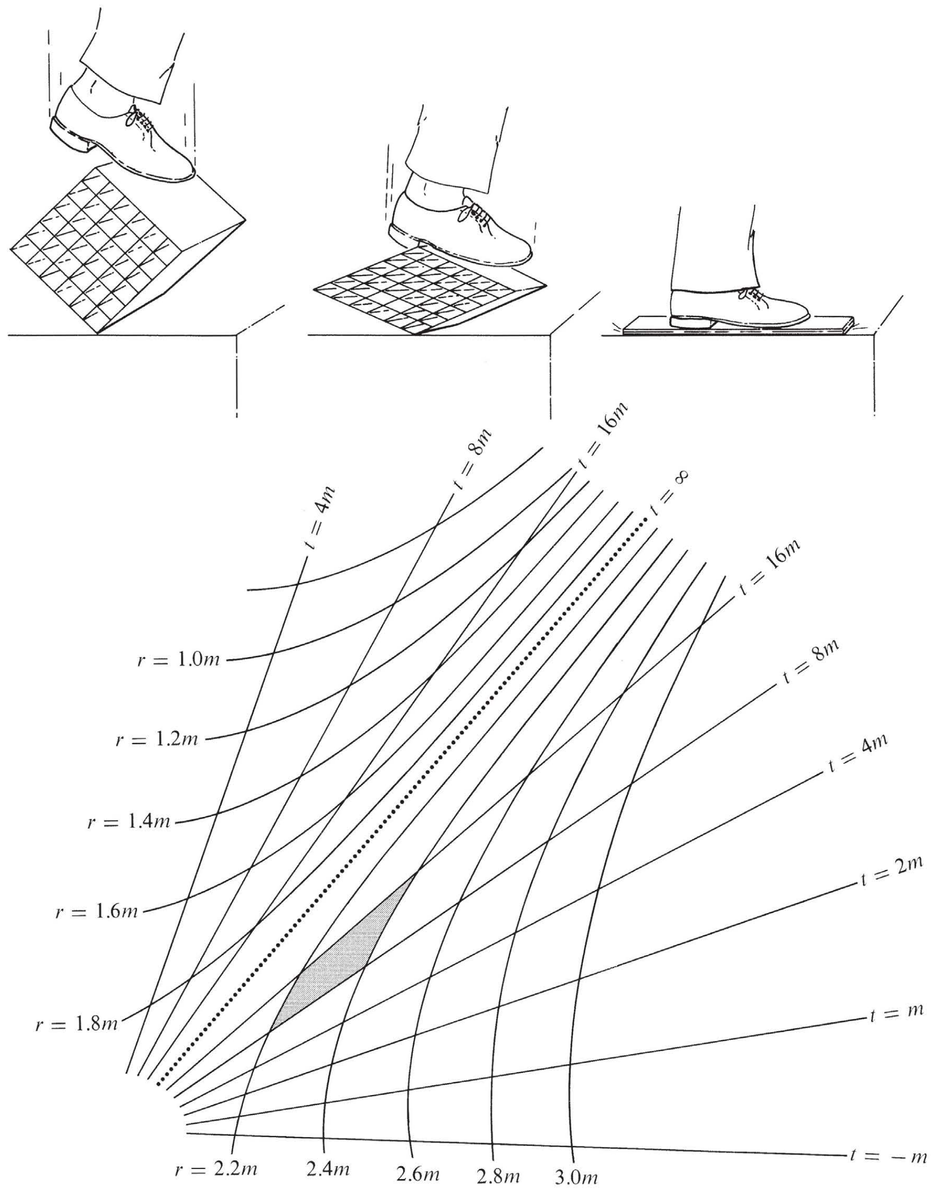



**Figure 1.4.**
How a mere coordinate singularity arises. Above: A coordinate system becomes *singular* when the "cells in the egg crate" are squashed to zero volume. Below: An example showing such a singularity in the Schwarzschild coordinates $r$, $t$ often used to describe the geometry around a black hole (Chapter 31). For simplicity the angular coordinates $\theta$, $\phi$ have been suppressed. The singularity shows itself in two ways. First, all the points along the dotted line, while quite distinct one from another, are designated by the same pair of $(r, t)$ values; namely, $r = 2m$, $t = \infty$. The coordinates provide no way to distinguish these points. Second, the "cells in the egg crate," of which one is shown grey in the diagram, collapse to zero content at the dotted line. In summary, there is nothing strange about the geometry at the dotted line; all the singularity lies in the coordinate system ("poor system of telephone numbers"). No confusion should be permitted to arise from the accidental circumstance that the $t$ coordinate attains an infinite value on the dotted line. No such infinity would occur if $t$ were replaced by the new coordinate $\bar{t}$, defined by

$$(t/2m) = \tan(\bar{t}/2m).$$

When $t = \infty$, the new coordinate $\bar{t}$ is $\bar{t} = \pi m$. The $r$, $\bar{t}$ coordinates still provide no way to distinguish the points along the dotted line. They still give "cells in the egg crate" collapsed to zero content along the dotted line.

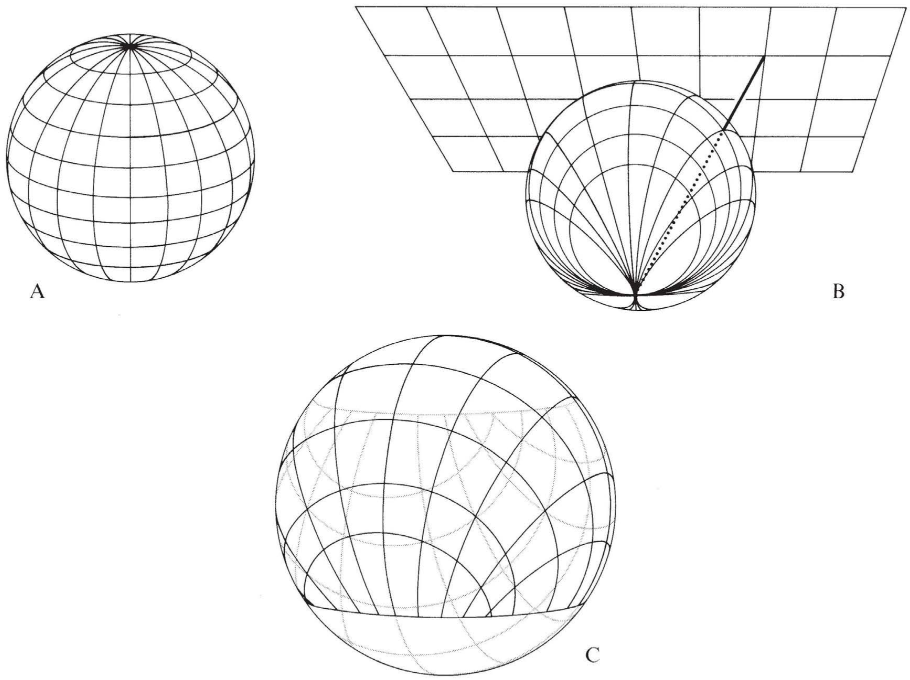



**Figure 1.5.**
Singularities in familiar coordinates on the two-sphere can be eliminated by covering the sphere with two overlapping coordinate patches. A. Spherical polar coordinates, singular at the North and South Poles, and discontinuous at the international date line. B. Projection of the Euclidean coordinates of the Euclidean two-plane, tangent at the North Pole, onto the sphere via a line running to the South Pole; coordinate singularity at the South Pole. C. Coverage of two-sphere by two overlapping coordinate patches. One, constructed as in B, covers without singularity the northern hemisphere and also the southern tropics down to the Tropic of Capricorn. The other (grey) also covers without singularity all of the tropics and the southern hemisphere besides.

energies (corresponding de Broglie wavelength $10^{-16}$ cm). Moreover, classical general relativity thinks of the spacetime manifold as a deterministic structure, completely well-defined down to arbitrarily small distances. Not so quantum general relativity or "quantum geometrodynamics." It predicts violent fluctuations in the geometry at distances on the order of the Planck length,

Breakdown in smoothness of spacetime at Planck length

$$\begin{aligned} L^* &= (\hbar G/c^3)^{1/2} \\ &= [(1.054 \times 10^{-27}\,\mathrm{g\,cm^2/sec})(6.670 \times 10^{-8}\,\mathrm{cm^3/g\,sec^2})]^{1/2} \times \\ &\qquad \times (2.998 \times 10^{10}\,\mathrm{cm/sec})^{-3/2} \qquad (1.1) \\ &= 1.616 \times 10^{-33}\,\mathrm{cm}. \end{aligned}$$

No one has found any way to escape this prediction. As nearly as one can estimate, these fluctuations give space at small distances a "multiply connected" or "foamlike" character. This lack of smoothness may well deprive even the concept of dimensionality itself of any meaning at the Planck scale of distances. The further exploration of this issue takes one to the frontiers of Einstein's theory (Chapter 44).

If spacetime at small distances is far from the mathematical model of a continuous manifold, is there not also at larger distances a wide gap between the mathematical idealization and the physical reality? The infinitely dense collection of light rays and of world lines of infinitesimal test particles that are to define all the points of the manifold: they surely are beyond practical realization. Nobody has ever found a particle that moves on timelike world lines (finite rest mass) lighter than an electron. A collection of electrons, even if endowed with zero density of charge ($e^+$ and $e^-$ world lines present in equal numbers) will have a density of mass. This density will curve the very manifold under study. Investigation in infinite detail means unlimited density, and unlimited disturbance of the geometry.

Difficulty in defining geometry even at classical distances?

However, to demand investigatability in infinite detail in the sense just described is as out of place in general relativity as it would be in electrodynamics or gas dynamics. Electrodynamics speaks of the strength of the electric and magnetic field at each point in space and at each moment of time. To measure those fields, it is willing to contemplate infinitesimal test particles scattered everywhere as densely as one pleases. However, the test particles do not have to be there at all to give the field reality. The field has everywhere a clear-cut value and goes about its deterministic dynamic evolution willy-nilly and continuously, infinitesimal test particles or no infinitesimal test particles. Similarly with the geometry of space.

No; one must accept geometry at classical distances as meaningful

In conclusion, when one deals with spacetime in the context of classical physics, one accepts (1) the notion of "infinitesimal test particle" and (2) the idealization that the totality of identifiable events forms a four-dimensional continuous manifold. Only at the end of this book will a look be taken at some of the limitations placed by the quantum principle on one's way of speaking about and analyzing spacetime.

## §1.3. WEIGHTLESSNESS

"Gravity is a great mystery. Drop a stone. See it fall. Hear it hit. No one understands why." What a misleading statement! Mystery about fall? What else should the stone do except fall? To fall is normal. The abnormality is an object standing in the way of the stone. If one wishes to pursue a "mystery," do not follow the track of the falling stone. Look instead at the impact, and ask what was the force that pushed the stone away from its natural "world line," (i.e., its natural track through spacetime). That could lead to an interesting issue of solid-state physics, but that is not the topic of concern here. Fall is. Free fall is synonymous with weightlessness: absence of any force to drive the object away from its normal track through spacetime. Travel aboard a freely falling elevator to experience weightlessness. Or travel aboard a spaceship also falling straight toward the Earth. Or, more happily, travel aboard a spaceship in that state of steady fall toward the Earth that marks a circular orbit. In each case one is following a natural track through spacetime.

Free fall is the natural state of motion

The traveler has one chemical composition, the spaceship another; yet they travel together, the traveler weightless in his moving home. Objects of such different nuclear constitution as aluminum and gold fall with accelerations that agree to better than one part in $10^{11}$, according to Roll, Krotkov, and Dicke (1964), one of the most important null experiments in all physics (see Figure 1.6). Individual molecules fall in step, too, with macroscopic objects [Estermann, Simpson, and Stern (1938)]; and so do individual neutrons [Dabbs, Harvey, Paya, and Horstmann (1965)], individual

All objects fall with the same acceleration

*(continued on page 16)*

**Figure 1.6.**
Principle of the Roll-Krotkov-Dicke experiment, which showed that the gravitational accelerations of gold and aluminum are equal to 1 part in $10^{11}$ or better (Princeton, 1964). In the upper lefthand corner, equal masses of gold and aluminum hang from a supporting bar. This bar in turn is supported at its midpoint. If both objects fall toward the sun with the same acceleration of $g = 0.59$ cm/sec$^2$, the bar does not turn. If the Au mass receives a higher acceleration, $g + \delta g$, then the gold end of the bar starts to turn toward the sun in the Earth-fixed frame. Twelve hours later the sun is on the other side, pulling the other way. The alternating torque lends itself to recognition against a background of noise because of its precise 24-hour period. Unhappily, any substantial mass nearby, such as an experimenter, located at $M$, will produce a torque that swamps the effect sought. Therefore the actual arrangement was as shown in the body of the figure. One gold weight and two aluminum weights were supported at the three corners of a horizontal equilateral triangle, 6 cm on a side (three-fold axis of symmetry, giving zero response to all the simplest nonuniformities in the gravitational field). Also, the observers performed all operations remotely to eliminate their own gravitational effects*. To detect a rotation of the torsion balance as small as $\sim 10^{-9}$ rad without disturbing the balance, Roll, Krotkov, and Dicke reflected a very weak light beam from the optically flat back face of the quartz triangle. The image of the source slit fell on a wire of about the same size as the slit image. The light transmitted past the wire fell on a photomultiplier. A separate oscillator circuit drove the wire back and forth across the image at 3,000 hertz. When the image was centered perfectly, only even harmonics of the oscillation frequency appeared in the light intensity. However, when the image was displaced slightly to one side, the fundamental frequency appeared in the light intensity. The electrical output of the photomultiplier then contained a 3,000-hertz component. The magnitude and sign of this component were determined automatically. Equally automatically a proportional D.C. voltage was applied to the electrodes shown in the diagram. It restored the torsion balance to its zero position. The D.C. voltage required to restore the balance to its zero position was recorded as a measure of the torque acting on the pendulum. This torque was Fourier-analyzed over a period of many days. The magnitude of the Fourier component of 24-hour period indicated a ratio $\delta g/g = (0.96 \pm 1.04) \times 10^{-11}$. Aluminum and gold thus fall with the same acceleration, despite their important differences summarized in the table.

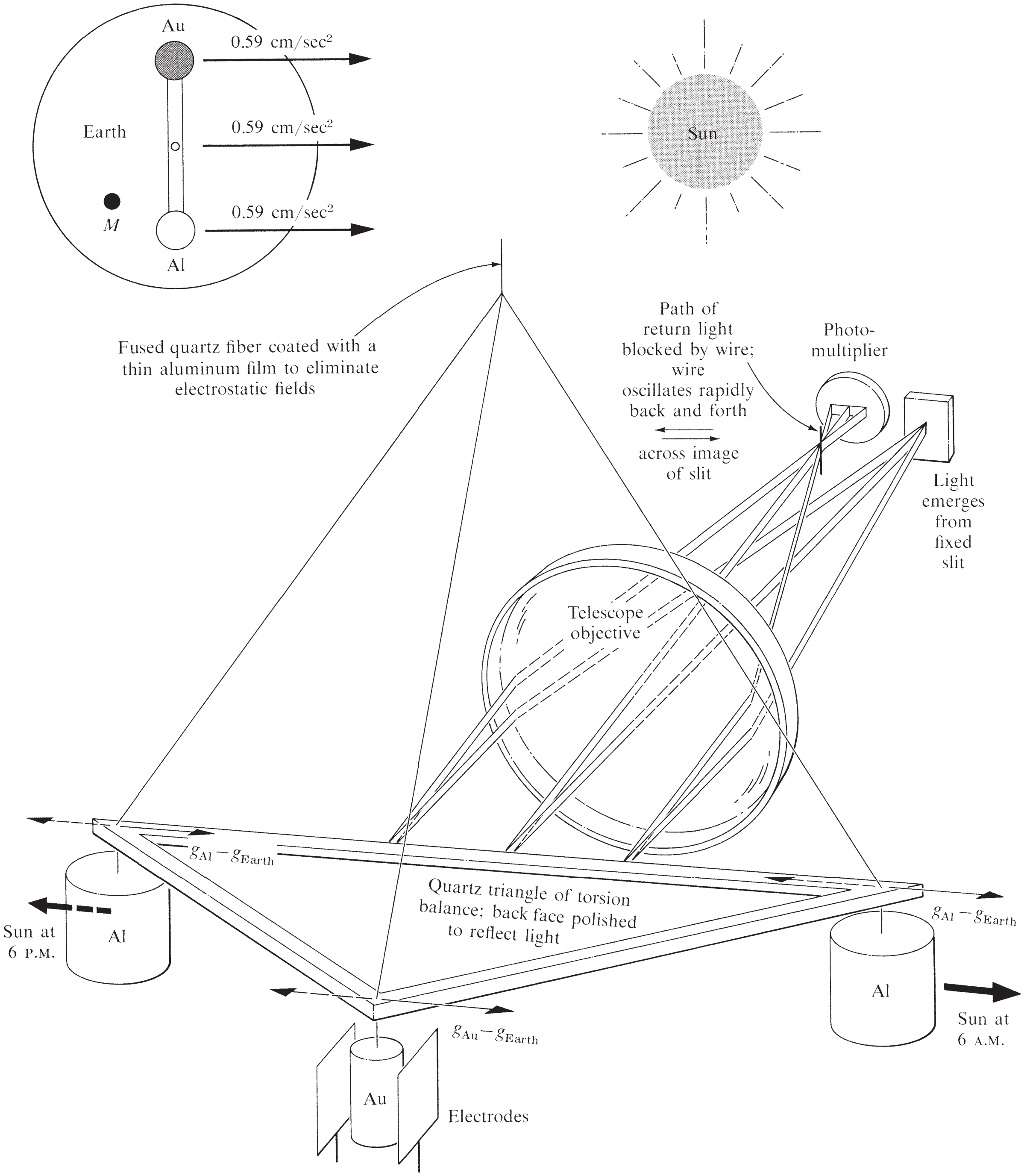
| *Ratios* | *Al* | *Au* |
|---|---|---|
| $\dfrac{\text{Number of neutrons}}{\text{Number of protons}}$ | 1.08 | 1.5 |
| $\dfrac{\text{Mass of kinetic energy of K-electron}}{\text{Rest mass of electron}}$ | 0.005 | 0.16 |
| $\dfrac{\text{Electrostatic mass-energy of nucleus}}{\text{Mass of atom}}$ | 0.001 | 0.004 |

*The theoretical implications* of this experiment will be discussed in greater detail in Chapters 16 and 38.

Braginsky and Panov (1971) at Moscow University performed an experiment identical in principle to that of Dicke-Roll-Krotkov, but with a modified experimental set-up. Comparing the accelerations of platinum and aluminum rather than of gold and aluminum, they say that

$$\delta g/g \lesssim 1 \times 10^{-12}.$$

*Other perturbations had to be, and were, guarded against. (1) A bit of iron on the torsion balance as big as $10^{-3}$ cm on a side would have contributed, in the Earth's magnetic field, a torque a hundred times greater than the measured torque. (2) The unequal pressure of radiation on the two sides of a mass would have produced an unacceptably large perturbation if the temperature difference between these two sides had exceeded $10^{-4}$ °K. (3) Gas evolution from one side of a mass would have propelled it like a rocket. If the rate of evolution were as great as $10^{-8}$ g/day, the calculated force would have been $\sim 10^{-7}$ g cm/sec$^2$, enough to affect the measurements. (4) The rotation was measured with respect to the pier that supported the equipment. As a guarantee that this pier did not itself rotate, it was anchored to bed rock. (5) Electrostatic forces were eliminated; otherwise they would have perturbed the balance.


Au
0.59 cm/sec$^2$
Earth
0.59 cm/sec$^2$
$M$
0.59 cm/sec$^2$
Al
Sun
Fused quartz fiber coated with a thin aluminum film to eliminate electrostatic fields
Path of return light blocked by wire; wire oscillates rapidly back and forth across image of slit
Photo-multiplier
Light emerges from fixed slit
Telescope objective
$g_{Al} - g_{Earth}$
Quartz triangle of torsion balance; back face polished to reflect light
Sun at 6 P.M.
Al
$g_{Al} - g_{Earth}$
Al
Sun at 6 A.M.
$g_{Au} - g_{Earth}$
Au
Electrodes

electrons [Witteborn and Fairbank (1967)] and individual mu mesons [Beall (1970)]. What is more, not one of these objects has to see out into space to know how to move.

Contemplate the interior of a spaceship, and a key, penny, nut, and pea by accident or design set free inside. Shielded from all view of the world outside by the walls of the vessel, each object stays at rest relative to the vessel. Or it moves through the room in a straight line with uniform velocity. That is the lesson which experience shouts out.

Forego talk of acceleration! That, paradoxically, is the lesson of the circumstance that "all objects fall with the same acceleration." Whose fault were those accelerations, after all? They came from allowing a groundbased observer into the act. The

**Box 1.2 MATERIALS OF THE MOST DIVERSE COMPOSITION FALL WITH THE SAME ACCELERATION ("STANDARD WORLD LINE")**

**Aristotle:** "the downward movement of a mass of gold or lead, or of any other body endowed with weight, is quicker in proportion to its size."

**Pre-Galilean literature:** metal and wood weights fall at the same rate.

**Galileo:** (1) "the variation of speed in air between balls of gold, lead, copper, porphyry, and other heavy materials is so slight that in a fall of 100 cubits [about 46 meters] a ball of gold would surely not outstrip one of copper by as much as four fingers. Having observed this, I came to the conclusion that in a medium totally void of resistance all bodies would fall with the same speed." (2) later experiments of greater precision "diluting gravity" and finding same time of descent for different objects along an inclined plane.

**Newton:** inclined plane replaced by arc of pendulum bob; "time of fall" for bodies of different composition determined by comparing time of oscillation of pendulum bobs of the two materials. Ultimate limit of precision in such experiments limited by problem of determining effective length of each pendulum: (acceleration) = $(2\pi/\text{period})^2(\text{length})$.

**Lorand von Eötvös,** Budapest, 1889 and 1922: compared on the rotating earth the vertical defined by a plumb bob of one material with the vertical defined by a plumb bob of other material. The two hanging masses, by the two unbroken threads that support them, were drawn along identical world lines through spacetime (middle of the laboratory of Eötvös!). If cut free, would they also follow identical tracks through spacetime ("normal world line of test mass")? If so, the acceleration that draws the actual world line from the normal free-fall world line will have a standard value, $\boldsymbol{a}$. The experiment of Eötvös did not try to test agreement on the magnitude of $\boldsymbol{a}$ between the two masses. Doing so would have required (1) cutting the threads and (2) following the fall of the two masses. Eötvös renounced this approach in favor of a static observation that he could make with greater precision, comparing the *direction* of $\boldsymbol{a}$ for the two masses. The direction of the supporting thread, so his argument ran, reveals the direction in which the mass is being dragged away from its normal world line of "free fall" or "weightlessness." This acceleration is the vectorial resultant of (1) an acceleration of magnitude $\boldsymbol{g}$, directed outward against so-called gravity, and (2) an acceleration directed toward the axis of rotation of the earth, of magnitude $\omega^2 R \sin\theta$ ($\omega$, angular ve-push of the ground under his feet was driving him away from a natural world line. Through that flaw in his arrangements, he became responsible for all those accelerations. Put him in space and strap rockets to his legs. No difference!* Again the responsibility for what he sees is his. Once more he notes that "all objects fall with

*"No difference" spelled out amounts to Einstein's (1911) principle of the local equivalence between a "gravitational field" and an acceleration: "*We arrive at a very satisfactory interpretation of this law of experience, if we assume that the systems* K *and* K′ *are physically exactly equivalent, that is, if we assume that we may just as well regard the system* K *as being in a space free from gravitational fields, if we then regard* K *as uniformly accelerated. This assumption of exact physical equivalence makes it impossible for us to speak of the absolute acceleration of the system of reference, just as the usual theory of relativity forbids us to talk of the absolute velocity of a system; and it makes the equal falling of all bodies in a gravitational field seem a matter of course.*"

locity; $R$, radius of earth; $\theta$, polar angle measured from North Pole to location of experiment). This centripetal acceleration has a vertical component $-\omega^2 R \sin^2\theta$ too small to come into discussion. The important component is $\omega^2 R \sin\theta\cos\theta$, directed northward and parallel to the surface of the earth. It deflects the thread by the angle

$$\frac{\text{horizontal acceleration}}{\text{vertical acceleration}} = \frac{\omega^2 R \sin\theta\cos\theta}{g} = \frac{3.4\ \text{cm/sec}^2}{980\ \text{cm/sec}^2}\sin\theta\cos\theta = 1.7\times 10^{-3}\ \text{radian at } \theta = 45^\circ$$

from the straight line connecting the center of the earth to the point of support. A difference, $\delta g$, of one part in $10^8$ between $g$ for the two hanging substances would produce a difference in angle of hang of plumb bobs equal to $1.7\times 10^{-11}$ radian at Budapest ($\theta = 42.5^\circ$). Eötvös reported $\delta g/g$ less than a few parts in $10^9$.

**Roll, Krotkov, and Dicke,** Princeton, 1964: employed as fiducial acceleration, not the 1.7 cm/sec$^2$ steady horizontal acceleration, produced by the earth's rotation at $\theta = 45^\circ$, but the daily alternating 0.59 cm/sec$^2$ produced by the sun's attraction. Reported $|g(\text{Au}) - g(\text{Al})|/g$ less than $1\times 10^{-11}$. See Figure 1.6.

**Braginsky and Panov,** Moscow, 1971: like Roll, Krotkov, and Dicke, employed Sun's attraction as fiducial acceleration. Reported $|g(\text{Pt}) - g(\text{Al})|/g$ less than $1\times 10^{-12}$.

**Beall**, 1970: particles that are deflected less by the Earth's or the sun's gravitational field than a photon would be, effectively travel faster than light. If they are charged or have other electromagnetic structure, they would then emit Čerenkov radiation, and reduce their velocity below threshold in less than a micron of travel. The threshold is at energies around $10^3\ mc^2$. Ultrarelativistic particles in cosmic-ray showers are not easily identified, but observations of $10^{13}$ eV muons show that muons are not "too light" by as much as $5\times 10^{-5}$. Conversely, a particle P bound more strongly than photons by gravity will transfer the momentum needed to make pair production $\gamma \rightarrow \text{P} + \bar{\text{P}}$ occur within a submicron decay length. The existence of photons with energies above $10^{13}$ eV shows that $e^{\pm}$ are not "too heavy" by 5 parts in $10^9$, $\mu^{\pm}$ not by 2 in $10^4$, $\Lambda$, $\Xi^-$, $\Omega^-$ not by a few per cent.

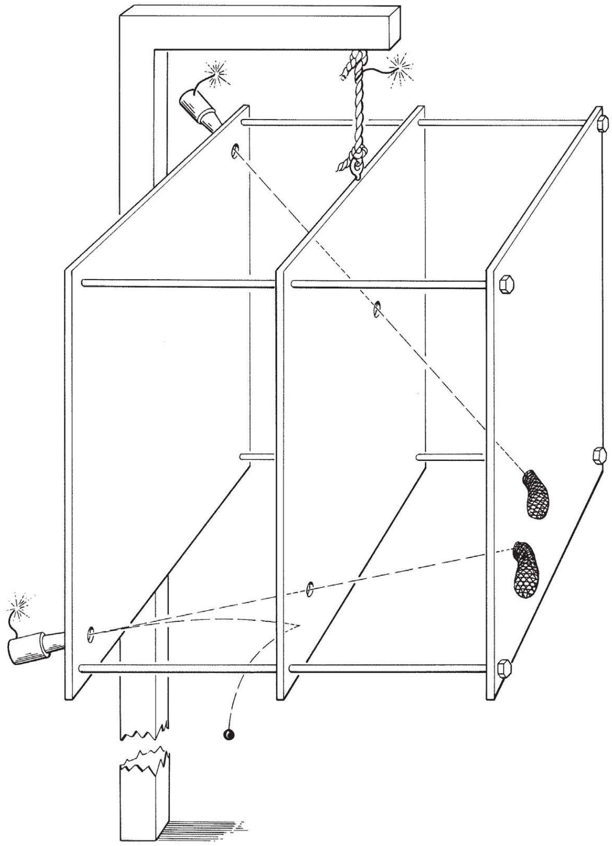


**Figure 1.7.**
"Weightlessness" as test for a local inertial frame of reference ("Lorentz frame"). Each spring-driven cannon succeeds in driving its projectile, a steel ball bearing, through the aligned holes in the sheets of lucite, and into the woven-mesh pocket, when the frame of reference is free of rotation and in free fall ("normal world line through spacetime"). A cannon would fail (curved and ricocheting trajectory at bottom of drawing) if the frame were hanging as indicated when the cannon went off ("frame drawn away by pull of rope from its normal world line through spacetime"). Harold Waage at Princeton has constructed such a model for an inertial reference frame with lucite sheets about 1 m square. The "fuses" symbolizing time delay were replaced by electric relays. Penetration fails if the frame (1) rotates, (2) accelerates, or (3) does any combination of the two. It is difficult to cite any easily realizable device that more fully illustrates the meaning of the term "local Lorentz frame."

Eliminate the acceleration by use of a local inertial frame

the same acceleration." Physics looks as complicated to the jet-driven observer as it does to the man on the ground. Rule out both observers to make physics look simple. Instead, travel aboard the freely moving spaceship. Nothing could be more natural than what one sees: every free object moves in a straight line with uniform velocity. This is the way to do physics! Work in a very special coordinate system: a coordinate frame in which one is weightless; *a local inertial frame of reference.* Or calculate how things look in such a frame. Or—if one is constrained to a ground-based frame of reference—use a particle moving so fast, and a path length so limited, that the ideal, freely falling frame of reference and the actual ground-based frame get out of alignment by an amount negligible on the scale of the experiment. [Given a 1,500-m linear accelerator, and a 1 GeV electron, time of flight $\simeq (1.5 \times 10^5 \text{ cm})/$

$(3 \times 10^{10}\ \text{cm/sec}) = 0.5 \times 10^{-5}$ sec; fall in this time $\sim\frac{1}{2}gt^2 = (490\ \text{cm/sec}^2)(0.5 \times 10^{-5}\ \text{sec})^2 \simeq 10^{-8}$ cm.]

In analyzing physics in a local inertial frame of reference, or following an ant on his little section of apple skin, one wins simplicity by foregoing every reference to what is far away. Physics is simple only when viewed locally: that is Einstein's great lesson.

Newton spoke differently: "Absolute space, in its own nature, without relation to anything external, remains always similar and immovable." But how does one give meaning to Newton's absolute space, find its cornerstones, mark out its straight lines? In the real world of gravitation, no particle ever follows one of Newton's straight lines. His ideal geometry is beyond observation. "A comet going past the sun is deviated from an ideal straight line." No. There is no pavement on which to mark out that line. The "ideal straight line" is a myth. It never happened, and it never will.

Newton's absolute space is unobservable, nonexistent

> "It required a severe struggle [for Newton] to arrive at the concept of independent and absolute space, indispensible for the development of theory. . . . Newton's decision was, in the contemporary state of science, the only possible one, and particularly the only fruitful one. But the subsequent development of the problems, proceeding in a roundabout way which no one could then possibly foresee, has shown that the resistance of Leibniz and Huygens, intuitively well-founded but supported by inadequate arguments, was actually justified. . . . It has required no less strenuous exertions subsequently to overcome this concept [of absolute space]"
>
> [A. EINSTEIN (1954)].

What is direct and simple and meaningful, according to Einstein, is the geometry in every local inertial reference frame. There every particle moves in a straight line with uniform velocity. *Define* the local inertial frame so that this simplicity occurs for the first few particles (Figure 1.7). In the frame thus defined, every other free particle is observed also to move in a straight line with uniform velocity. Collision and disintegration processes follow the laws of conservation of momentum and energy of special relativity. That all these miracles come about, as attested by tens of thousands of observations in elementary particle physics, is witness to the inner workings of the machinery of the world. The message is easy to summarize: (1) physics is always and everywhere locally Lorentzian; i.e., locally the laws of special relativity are valid; (2) this simplicity shows most clearly in a local Lorentz frame of reference ("inertial frame of reference"; Figure 1.7); and (3) to test for a local Lorentz frame, test for weightlessness!

But Einstein's local inertial frames exist, are simple

In local inertial frames, physics is Lorentzian

## §1.4. LOCAL LORENTZ GEOMETRY, WITH AND WITHOUT COORDINATES

On the surface of an apple within the space of a thumbprint, the geometry is Euclidean (Figure 1.1; the view in the magnifying glass). In spacetime, within a limited region, the geometry is Lorentzian. On the apple the distances between point and point accord with the theorems of Euclid. In spacetime the intervals ("proper distance," "proper time") between event and event satisfy the corresponding theorems of Lorentz-Minkowski geometry (Box 1.3). These theorems lend themselves

Local Lorentz geometry is the spacetime analog of local Euclidean geometry.

*(continued on page 23)*

**Box 1.3 LOCAL LORENTZ GEOMETRY AND LOCAL EUCLIDEAN GEOMETRY: WITH AND WITHOUT COORDINATES**

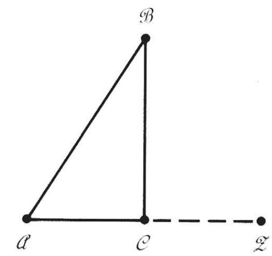
## I. Local Euclidean Geometry

What does it mean to say that the geometry of a tiny thumbprint on the apple is Euclidean?

A. *Coordinate-free language* (Euclid):
Given a line $\mathscr{AC}$. Extend it by an equal distance $\mathscr{CZ}$. Let $\mathscr{B}$ be a point not on $\mathscr{AZ}$ but equidistant from $\mathscr{A}$ and $\mathscr{Z}$. Then

$$s_{\mathscr{AB}}^2 = s_{\mathscr{AC}}^2 + s_{\mathscr{BC}}^2.$$

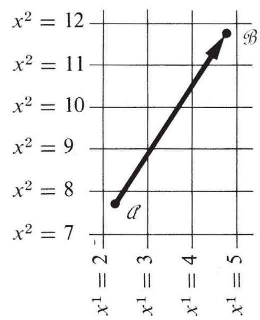
(Theorem of Pythagoras; also other theorems of Euclidean geometry.)




B. *Language of coordinates* (Descartes):
From any point $\mathscr{A}$ to any other point $\mathscr{B}$ there is a distance $s$ given in suitable (Euclidean) coordinates by

$$s_{\mathscr{AB}}^2 = [x^1(\mathscr{B}) - x^1(\mathscr{A})]^2 + [x^2(\mathscr{B}) - x^2(\mathscr{A})]^2.$$

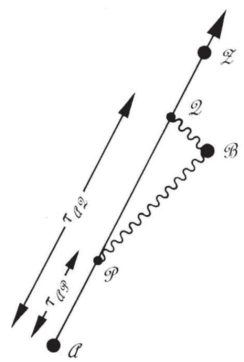
If one succeeds in finding any coordinate system where this is true for all points $\mathscr{A}$ and $\mathscr{B}$ in the thumbprint, then one is guaranteed that (i) this coordinate system is locally Euclidean, and (ii) the geometry of the apple's surface is locally Euclidean.




## II. Local Lorentz Geometry

What does it mean to say that the geometry of a sufficiently limited region of spacetime in the real physical world is Lorentzian?

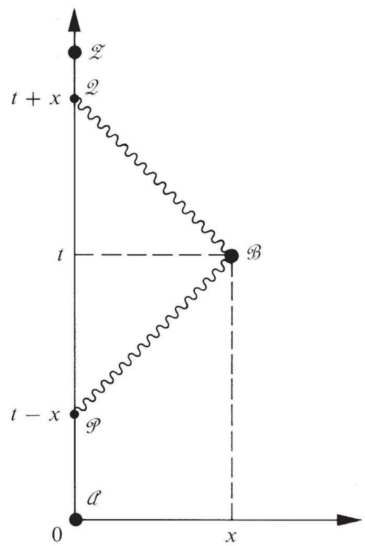
A. *Coordinate-free language* (Robb 1936):
Let $\mathscr{AZ}$ be the world line of a free particle. Let $\mathscr{B}$ be an event not on this world line. Let a light ray from $\mathscr{B}$ strike $\mathscr{AZ}$ at the event $\mathscr{Q}$. Let a light ray take off from such an earlier event $\mathscr{P}$ along $\mathscr{AZ}$ that it reaches $\mathscr{B}$. Then the proper distance $s_{\mathscr{AB}}$ (spacelike separation) or proper time $\tau_{\mathscr{AB}}$ (timelike separation) is given by




$$s_{\mathscr{AB}}^2 \equiv -\tau_{\mathscr{AB}}^2 = -\tau_{\mathscr{AQ}}\tau_{\mathscr{AP}}.$$

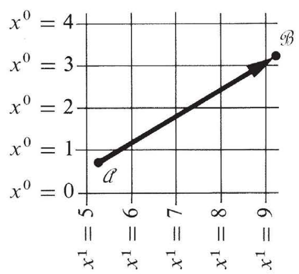
Proof of above criterion for local Lorentz geometry, using coordinate methods in the local Lorentz frame where particle remains at rest:

$$\tau_{\mathcal{AB}}{}^2 = t^2 - x^2 = (t - x)(t + x)$$
$$= \tau_{\mathcal{AP}}\tau_{\mathcal{AQ}}.$$




B. *Language of coordinates* (Lorentz, Poincaré, Minkowski, Einstein):
From any event $\mathcal{A}$ to any other nearby event $\mathcal{B}$, there is a proper distance $s_{\mathcal{AB}}$ or proper time $\tau_{\mathcal{AB}}$ given in suitable (local Lorentz) coordinates by

$$\begin{aligned} s_{\mathcal{AB}}{}^2 = -\tau_{\mathcal{AB}}{}^2 = &-[x^0(\mathcal{B}) - x^0(\mathcal{A})]^2 \\ &+ [x^1(\mathcal{B}) - x^1(\mathcal{A})]^2 \\ &+ [x^2(\mathcal{B}) - x^2(\mathcal{A})]^2 \\ &+ [x^3(\mathcal{B}) - x^3(\mathcal{A})]^2. \end{aligned}$$




If one succeeds in finding any coordinate system where this is locally true for all neighboring events $\mathcal{A}$ and $\mathcal{B}$, then one is guaranteed that (i) this coordinate system is locally Lorentzian, and (ii) the geometry of spacetime is locally Lorentzian.

### III. Statements of Fact

The geometry of an apple's surface is locally Euclidean everywhere. The geometry of spacetime is locally Lorentzian everywhere.

**Box 1.3 (*continued*)**

## IV. Local Geometry in the Language of Modern Mathematics

A. *The metric for any manifold:*
At each point on the apple, at each event of spacetime, indeed, at each point of any "Riemannian manifold," there exists a geometrical object called the *metric tensor* $\boldsymbol{g}$. It is a machine with two input slots for the insertion of two vectors:

$$\begin{array}{ccc} & \text{slot 1} & \text{slot 2} \\ & \downarrow & \downarrow \\ \boldsymbol{g}( & \quad , & \quad ). \end{array}$$

If one inserts the same vector $\boldsymbol{u}$ into both slots, one gets out the square of the length of $\boldsymbol{u}$:

$$\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{u}) = \boldsymbol{u}^2.$$

If one inserts two different vectors, $\boldsymbol{u}$ and $\boldsymbol{v}$ (it matters not in which order!), one gets out a number called the "scalar product of $\boldsymbol{u}$ on $\boldsymbol{v}$" and denoted $\boldsymbol{u} \cdot \boldsymbol{v}$:

$$\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{g}(\boldsymbol{v}, \boldsymbol{u}) = \boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}.$$

The metric is a linear machine:

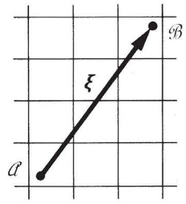
$$\begin{aligned} \boldsymbol{g}(2\boldsymbol{u} + 3\boldsymbol{w}, \boldsymbol{v}) &= 2\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) + 3\boldsymbol{g}(\boldsymbol{w}, \boldsymbol{v}), \\ \boldsymbol{g}(\boldsymbol{u}, a\boldsymbol{v} + b\boldsymbol{w}) &= a\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) + b\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{w}). \end{aligned}$$

Consequently, in a given (arbitrary) coordinate system, its operation on two vectors can be written in terms of their components as a bilinear expression:

$$\begin{aligned} \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) &= g_{\alpha\beta} u^\alpha v^\beta \\ &\qquad \text{(implied summation on } \alpha, \beta) \\ &= g_{11}u^1v^1 + g_{12}u^1v^2 + g_{21}u^2v^1 + \cdots. \end{aligned}$$

The quantities $g_{\alpha\beta} = g_{\beta\alpha}$ ($\alpha$ and $\beta$ running from 0 to 3 in spacetime, from 1 to 2 on the apple) are called the "components of $\boldsymbol{g}$ in the given coordinate system."

B. *Components of the metric in local Lorentz and local Euclidean frames:*
To connect the metric with our previous descriptions of the local geometry, introduce local Euclidean coordinates (on apple) or local Lorentz coordinates (in spacetime).




Let $\boldsymbol{\xi}$ be the separation vector reaching from $\mathscr{A}$ to $\mathscr{B}$. Its components in the local Euclidean (Lorentz) coordinates are

$$\xi^\alpha = x^\alpha(\mathscr{B}) - x^\alpha(\mathscr{A})$$

(cf. Box 1.1). Then the squared length of $\boldsymbol{u}_{\mathscr{A}\mathscr{B}}$, which is the same as the squared distance from $\mathscr{A}$ to $\mathscr{B}$, must be (cf. I.B. and II.B. above)

$$\begin{aligned} \boldsymbol{\xi} \cdot \boldsymbol{\xi} &= \boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{\xi}) = g_{\alpha\beta}\xi^\alpha\xi^\beta \\ &= s_{\mathscr{A}\mathscr{B}}{}^2 = (\xi^1)^2 + (\xi^2)^2 \text{ on apple} \\ &= -(\xi^0)^2 + (\xi^1)^2 + (\xi^2)^2 + (\xi^3)^2 \\ &\qquad\qquad \text{in spacetime.} \end{aligned}$$

Consequently, the components of the metric are

$$\begin{aligned} & g_{11} = g_{22} = 1,\ g_{12} = g_{21} = 0; \\ & \text{i.e., } g_{\alpha\beta} = \delta_{\alpha\beta} \quad \text{on apple, in local Euclidean coordinates;} \\ & g_{00} = -1,\ g_{0k} = 0,\ g_{jk} = \delta_{jk} \\ & \qquad \text{in spacetime, in local Lorentz coordinates.} \end{aligned}$$

These special components of the metric in local Lorentz coordinates are written here and hereafter as $g_{\hat\alpha\hat\beta}$ or $\eta_{\alpha\beta}$, by analogy with the Kronecker delta $\delta_{\alpha\beta}$. In matrix notation:

$$\|g_{\hat\alpha\hat\beta}\| = \|\eta_{\alpha\beta}\| = \begin{array}{cc} & \overset{\text{--- }\beta \longrightarrow}{\begin{array}{cccc} 0 & 1 & 2 & 3 \end{array}} \\ \alpha\downarrow \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \left\| \begin{array}{cccc} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right\| \end{array}$$

to empirical test in the appropriate, very special coordinate systems: Euclidean coordinates in Euclidean geometry; the natural generalization of Euclidean coordinates (local Lorentz coordinates; local inertial frame) in the local Lorentz geometry of physics. However, the theorems rise above all coordinate systems in their content. They refer to intervals or distances. Those distances no more call on coordinates for their definition in our day than they did in the time of Euclid. Points in the great pile of hay that is spacetime; and distances between these points: that is geometry! State them in the coordinate-free language or in the language of coordinates: they are the same (Box 1.3).

## §1.5. TIME

Time is defined so that motion looks simple.

*Time is awake when all things sleep.*
*Time stands straight when all things fall.*
*Time shuts in all and will not be shut.*
*Is, was, and shall be are Time's children.*
*O Reasoning, be witness, be stable.*

VYASA, the *Mahabarata* (*ca.* A.D. 400)

Relative to a local Lorentz frame, a free particle "moves in a straight line with uniform velocity." What "straight" means is clear enough in the model inertial reference frame illustrated in Figure 1.7. But where does the "uniform velocity" come in? Or where does "velocity" show itself? There is not even one clock in the drawing!

A more fully developed model of a Lorentz reference frame will have not only holes, as in Fig. 1.7, but also clock-activated shutters over each hole. The projectile can reach its target only if it (1) travels through the correct region in space and (2) gets through that hole in the correct interval of time ("window in time"). How then is time defined? Time is defined so that motion looks simple!

The time coordinate of a local Lorentz frame is so defined that motion looks simple

No standard of time is more widely used than the day, the time from one high noon to the next. Take that as standard, however, and one will find every good clock or watch clashing with it, for a simple reason. The Earth spins on its axis and also revolves in orbit about the sun. The motion of the sun across the sky arises from neither effect alone, but from the two in combination, different in magnitude though they are. The fast angular velocity of the Earth on its axis (roughly 366.25 complete turns per year) is wonderfully uniform. Not so the apparent angular velocity of the sun about the center of the Earth (one turn per year). It is greater than average by 2 per cent when the Earth in its orbit (eccentricity 0.017) has come 1 per cent closer than average to the sun (Kepler's law) and lower by 2 per cent when the Earth is 1 per cent further than average from the sun. In the first case, the momentary rate of rotation of the sun across the sky, expressed in turns per year, is approximately

$$366.25 - (1 + 0.02);$$

in the other,

$$366.25 - (1 - 0.02).$$

Taking the "mean solar day" to contain $24 \times 3{,}600 = 86{,}400$ standard seconds, one sees that, when the Earth is 1 per cent closer to (or further from) the sun than average, then the number of standard seconds from one high noon to the next is greater (or less) than normal by

$$\frac{0.02 \text{ (drop in turns per year)}}{365.25 \text{ (turns per year on average)}} 86{,}400 \text{ sec} \sim 4.7 \text{ sec}.$$

This is the bookkeeping on time from noon to noon. No standard of time that varies so much from one month to another is acceptable. If adopted, it would make the speed of light vary from month to month!

This lack of uniformity, once recognized (and it was already recognized by the ancients), forces one to abandon the solar day as the standard of time; that day does not make motion look simple. Turn to a new standard that eliminates the motion of the Earth around the sun and concentrates on the spin of the Earth about its axis: the sidereal day, the time between one arrival of a star at the zenith and the next arrival of that star at the zenith. Good! Or good, so long as one's precision of measurement does not allow one to see changes in the intrinsic angular velocity of the Earth. What clock was so bold as first to challenge the spin of the Earth for accuracy? The machinery of the heavens.

Halley (1693) and later others, including Kant (1754), suspected something was amiss from apparent discrepancies between the paths of totality in eclipses of the sun, as predicted by Newtonian gravitation theory using the standard of time then current, and the location of the sites where ancient Greeks and Romans actually recorded an eclipse on the day in question. The moon casts a moving shadow in space. On the day of a solar eclipse, that shadow paints onto the disk of the spinning Earth a black brush stroke, often thousands of kilometers in length, but of width generally much less than a hundred kilometers. He who spins the globe upon the table and wants to make the shadow fall rightly on it must calculate back meticulously to determine two key items: (1) where the moon is relative to Earth and sun at each moment on the ancient day in question; and (2) how much angle the Earth has turned through from then until now. Take the eclipse of Jan. 14, A.D. 484, as an example (Figure 1.8), and assume the same angular velocity for the Earth in the intervening fifteen centuries as the Earth had in 1900 (astronomical reference point). One comes out wrong. The Earth has to be set back by 30° (or the moon moved from its computed position, or some combination of the two effects) to make the Athens observer fall under the black brush. To catch up those 30° (or less, if part of the effect is due to a slow change in the angular momentum of the moon), the Earth had to turn faster in the past than it does today. Assigning most of the discrepancy to terrestrial spin-down (rate of spin-down compatible with modern atomic-clock evidence), and assuming a uniform rate of slowing from then to now

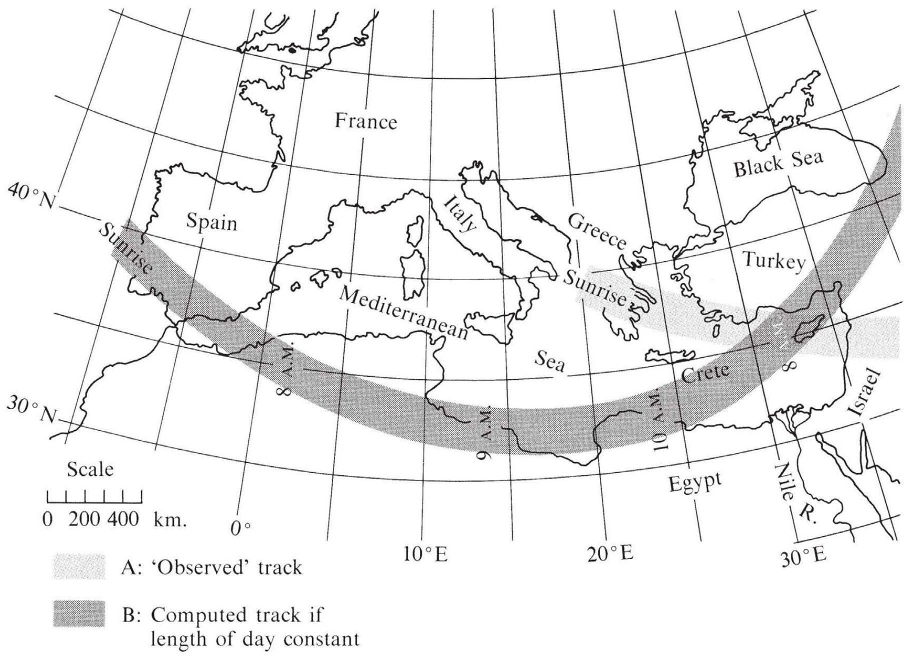



**Figure 1.8.**
Calculated path of totality for the eclipse of January 14, A.D. 484 (left; calculation based on no spin-down of Earth relative to its 1900 angular velocity) contrasted with the same path as set ahead enough to put the center of totality (at sunrise) at Athens [displacement very close to 30°; actual figure of deceleration adopted in calculations, 32.75 arc sec/(century)$^2$]. This is "undoubtedly the most reliable of all ancient European eclipses," according to Dr. F. R. Stephenson, of the Department of Geophysics and Planetary Physics of the University of Newcastle upon Tyne, who most kindly prepared this diagram especially for this book. He has also sent a passage from the original Greek biography of Proclus of Athens (died at Athens A.D. 485) by Marinus of Naples, reading, "Nor were there portents wanting in the year which preceded his death; for example, such a great eclipse of the Sun that night seemed to fall by day. For a profound darkness arose so that stars even appeared in the sky. This happened in the eastern sky when the Sun dwelt in Capricorn" [from Westermann and Boissonade (1878)].

Does this 30° for this eclipse, together with corresponding amounts for other eclipses, represent the "right" correction? "Right" is no easy word. From one total eclipse of the sun in the Mediterranean area to another is normally many years. The various provinces of the Greek and Roman worlds were far from having a uniform level of peace and settled life, and even farther from having a uniform standard of what it is to observe an eclipse and put it down for posterity. If the scores of records of the past are unhappily fragmentary, even more unhappy has been the willingness of a few uncritical "investigators" in recent times to rush in and identify this and that historical event with this and that calculated eclipse. Fortunately, by now a great literature is available on the secular deceleration of the Earth's rotation, in the highest tradition of critical scholarship, both astronomical and historical. In addition to the books of O. Neugebauer (1959) and Munk and MacDonald (1960), the paper of Curott (1966), and items cited by these workers, the following are key items. (For direction to them, we thank Professor Otto Neugebauer—no relation to the other Neugebauer cited below!) For the ancient records, and for calculations of the tracks of ancient eclipses, F. K. Ginzel (1882, 1883, 1884); for an atlas of calculated eclipse tracks, Oppolzer (1887) and Ginzel (1899); and for a critical analysis of the evidence. P. V. Neugebauer (1927, 1929, and 1930). This particular eclipse was chosen rather than any other because of the great reliability of the historical record of it.

(angular velocity correction proportional to first power of elapsed time: angle correction itself proportional to square of elapsed time), one estimates from a correction of

$$30^\circ \text{ or } 2 \text{ hours} \qquad 1{,}500 \text{ years ago}$$

the following corrections for intermediate times:

$$30^\circ/10^2, \text{ or } 1.2 \text{ min} \qquad 150 \text{ years ago,}$$

$$30^\circ/10^4, \text{ or } 0.8 \text{ sec} \qquad 15 \text{ years ago.}$$

Thus one sees the downfall of the Earth as a standard of time and its replacement by the orbital motions of the heavenly bodies as a better standard: a standard that does more to "make motion look simple." Astronomical time is itself in turn today being supplanted by atomic time as a standard of reference (see Box 1.4, "Time Today").

Good clocks make spacetime trajectories of free particles look straight

Look at a bad clock for a good view of how time is defined. Let $t$ be time on a "good" clock (time coordinate of a local inertial frame); it makes the tracks of free particles through the local region of spacetime look straight. Let $T(t)$ be the reading of the "bad" clock; it makes the world lines of free particles through the local region of spacetime look curved (Figure 1.9). The old value of the acceleration, translated into the new ("bad") time, becomes

$$0 = \frac{d^2x}{dt^2} = \frac{d}{dt}\left(\frac{dT}{dt}\frac{dx}{dT}\right) = \frac{d^2T}{dt^2}\frac{dx}{dT} + \left(\frac{dT}{dt}\right)^2 \frac{d^2x}{dT^2}.$$

To explain the apparent accelerations of the particles, the user of the new time introduces a force that one knows to be fictitious:

$$F_x = m\frac{d^2x}{dT^2} = -m\frac{\left(\dfrac{dx}{dT}\right)\left(\dfrac{d^2T}{dt^2}\right)}{\left(\dfrac{dT}{dt}\right)^2}. \tag{1.2}$$

It is clear from this example of a "bad" time that Newton thought of a "good" time when he set up the principle that "Time flows uniformly" ($d^2T/dt^2 = 0$). Time is defined to make motion look simple!

Our choice of unit for measuring time: *the geometrodynamic centimeter.*

The principle of uniformity, taken by itself, leaves free the scale of the time variable. The quantity $T = at + b$ satisfies the requirement as well as $t$ itself. The history of timekeeping discloses many choices of the unit and origin of time. Each one required some human action to give it sanction, from the fiat of a Pharaoh to the communique of a committee. In this book the amount of time it takes light to travel one centimeter is decreed to be the unit of time. Spacelike intervals and timelike intervals are measured in terms of one and the same geometric unit: the centimeter. Any other decision would complicate in analysis what is simple in nature. No other choice would live up to Minkowski's words, "Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality."

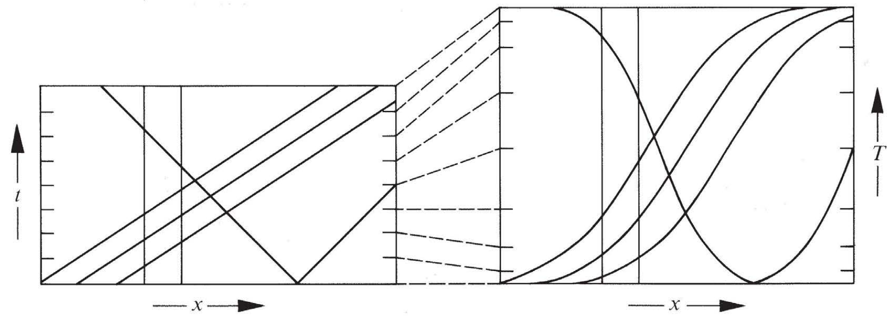



**Figure 1.9.**
Good clock (left) vs. bad clock (right) as seen in the maps they give of the same free particles moving through the same region of spacetime. The world lines as depicted at the right give the impression that a force is at work. The good definition of time eliminates such fictitious forces. The dashed lines connect corresponding instants on the two time scales.

One can measure time more accurately today than distance. Is that an argument against taking the elementary unit to be the centimeter? No, provided that this definition of the centimeter is accepted: *the geometrodynamic standard centimeter is the fraction*

$$1/(9.460546 \times 10^{17}) \tag{1.3}$$

*of the interval between the two "effective equinoxes" that bound the tropical year 1900.0.* The tropical year 1900.0 has already been recognized internationally as the fiducial interval by reason of its definiteness and the precision with which it is known. Standards committees have *defined the ephemeris second* so that 31,556,925.974 sec make up that standard interval. Were the speed of light known with perfect precision, the standards committees could have given in the same breath the number of centimeters in the standard interval. But it isn't; it is known to only six decimals. Moreover, the *international centimeter* is defined in terms of the orange-red wavelength of $Kr^{86}$ to only nine decimals (16,507.6373 wavelengths). Yet the standard second is given to 11 decimals. We match the standard second by arbitrarily defining the geometrodynamic standard centimeter so that

$$9.4605460000 \times 10^{17}$$

such centimeters are contained in the standard tropical year 1900.0. The speed of light then becomes exactly

$$\frac{9.4605460000 \times 10^{17}}{31{,}556{,}925.974} \text{ geometrodynamic cm/sec.} \tag{1.4}$$

This is compatible with the speed of light, as known in 1967, in units of "international cm/sec":

$$29{,}979{,}300{,}000 \pm 30{,}000 \text{ international cm/sec.}$$

**Box 1.4 TIME TODAY**

Prior to 1956 the second was defined as the fraction 1/86,400 of the mean solar day.

From 1956 to 1967 the "second" meant the ephemeris second, defined as the fraction 1/(31,556,925.9747) of the tropical year 00h00m00s December 31, 1899.

Since 1967 the standard second has been the SI (Système International) second, defined as 9,192,631,770 periods of the unperturbed microwave transition between the two hyperfine levels of the ground state of $Cs^{133}$.

Like the foregoing evolution of the unit for the time *interval,* the evolution of a time *coordinate* has been marked by several stages.

Universal time, UTO, is based on the count of days as they actually occurred historically; in other words, on the actual spin of the earth on its axis; historically, on mean solar time (solar position as corrected by the "equation of time"; i.e., the faster travel of the earth when near the sun than when far from the sun) as determined at Greenwich Observatory.

UT1, the "navigator's time scale," is the same time as corrected for the wobble of the earth on its axis ($\Delta t \sim 0.05$ sec).

UT2 is UT1 as corrected for the periodic fluctuations of unknown origin with periods of one-half year and one year ($\Delta t \sim 0.05$ sec; measured to 3 ms in one day).

Ephemeris Time, ET (as defined by the theory of gravitation and by astronomical observations and calculations), is essentially determined by the orbital motion of the earth around the sun. "Measurement uncertainties limit the realization of accurate ephemeris time to about 0.05 sec for a nine-year average."

Coordinated Universal Time (UTC) is broadcast on stations such as WWV. It was adopted internationally in February 1971 to become effective January 1, 1972. The clock rate is controlled by atomic clocks to be as uniform as possible for one year (atomic time is measured to $\sim$0.1 microsec in 1 min, with diffusion rates of 0.1 microsec per day for ensembles of clocks), but is changed by the infrequent addition or deletion of a second—called a "leap second"—so that UTC never differs more than 0.7 sec from the navigator's time scale, UT1.

THE TIMES
Wednesday
June 21 1972

**Time suspended for a second**

Time will stand still throughout the world for one second at midnight, June 30. All radio time signals will insert a "leap second" to bring Greenwich Mean Time into line with the earth's loss of three thousandths of a second a day.

**The signal from the Royal Greenwich Observatory to Broadcasting House at midnight GMT (1 am BST July 1) will be six short pips marking the seconds 55 to 60 inclusive, followed by a lengthened signal at the following second to mark the new minute.**

The foregoing account is abstracted from J. A. Barnes (1971). The following is extracted from a table (not official at time of receipt), kindly supplied by the Time and Frequency Division of the U.S. National Bureau of Standards in Boulder, Colorado.

Timekeeping capabilities of some familiar clocks are as follows:

Tuning fork wrist watch (1960),
1 min/mo.

Quartz crystal clock (1921–1930),
1 $\mu$sec/day,
1 sec/yr.

Quartz crystal wrist watch (1971),
0.2 sec/2 mos.,
1 sec/yr.

Cesium beam (atomic resonance, $Cs^{133}$), (1952–1955),
0.1 $\mu$sec/day,
0.5 $\mu$sec/mo.

Rubidium gas cell ($Rb^{87}$ resonance), (1957),
0.1 $\mu$sec/day,
1–5 $\mu$sec/mo.

Hydrogen maser (1960),
0.01 $\mu$sec/2 hr,
0.1 $\mu$sec/day.

Methane stabilized laser (1969),
0.01 $\mu$sec/100 sec.

Recent measurements [Evenson *et al.* (1972)] change the details of the foregoing 1967 argument, but not the principles.

## §1.6. CURVATURE

Gravitation seems to have disappeared. Everywhere the geometry of spacetime is locally Lorentzian. And in Lorentz geometry, particles move in a straight line with constant velocity. Where is any gravitational deflection to be seen in that? For answer, turn back to the apple (Figure 1.1). Inspect again the geodesic tracks of the ants on the surface of the apple. Note the reconvergence of two nearby geodesics that originally diverged from a common point. What is the analog in the real world of physics? What analogous concept fits Einstein's injunction that physics is only simple when analyzed locally? Don't look at the distance from the spaceship to the Earth. Look at the distance from the spaceship to a nearby spaceship! Or, to avoid any possible concern about attraction between the two ships, look at two nearby test particles in orbit about the Earth. To avoid distraction by the nonlocal element (the Earth) in the situation, conduct the study in the interior of a spaceship, also in orbit about the Earth. But this region has already been counted as a local inertial frame! What gravitational physics is to be seen there? None. Relative to the spaceship and therefore relative to each other, the two test particles move in a straight line with uniform velocity, to the precision of measurement that is contemplated (see Box 1.5, "Test for Flatness"). Now the key point begins to appear: precision of measurement. Increase it until one begins to discern the gradual acceleration of the test particles away from each other, if they lie along a common radius through the center of the Earth; or toward each other, if their separation lies perpendicular to that line. In Newtonian language, the source of these accelerations is the tide-producing action of the Earth. To the observer in the spaceship, however, no Earth is to be seen. And following Einstein, he knows it is important to analyze motion locally. He represents the separation of the new test particle from the fiducial test particle by the vector $\xi^k (k = 1, 2, 3$; components measured in a local Lorentz frame). For the acceleration of this separation, one knows from Newtonian physics what he will find: if the Cartesian $z$-axis is in the radial direction, then

Gravitation is manifest in relative acceleration of neighboring test particles

$$\begin{aligned}
\frac{d^2\xi^x}{dt^2} &= -\frac{Gm_{\text{conv}}}{c^2r^3}\,\xi^x; \\
\frac{d^2\xi^y}{dt^2} &= -\frac{Gm_{\text{conv}}}{c^2r^3}\,\xi^y; \\
\frac{d^2\xi^z}{dt^2} &= \frac{2Gm_{\text{conv}}}{c^2r^3}\,\xi^z.
\end{aligned} \tag{1.5}$$

Proof: In Newtonian physics the acceleration of a single particle toward the center of the Earth in conventional units of time is $Gm_{\text{conv}}/r^2$, where $G$ is the Newtonian constant of gravitation, $6.670 \times 10^{-8}\ \text{cm}^3/\text{g sec}^2$ and $m_{\text{conv}}$ is the mass of the Earth in conventional units of grams. In geometric units of time (cm of light-travel time), the acceleration is $Gm_{\text{conv}}/c^2r^2$. When the two particles are separated by a distance $\xi$ perpendicular to $r$, the one downward acceleration vector is out of line with the other by the angle $\xi/r$. Consequently one particle accelerates toward the other by the stated amount. When the separation is parallel to $r$, the relative acceleration is given by evaluating the Newtonian acceleration at $r$ and at $r + \xi$, and taking the difference ($\xi$ times $d/dr$) Q.E.D. In conclusion, the "local tide-producing acceleration" of Newtonian gravitation theory provides the local description of gravitation that Einstein bids one to seek.

Relative acceleration is caused by curvature

What has this tide-producing acceleration to do with curvature? (See Box 1.6.) Look again at the apple or, better, at a sphere of radius $a$ (Figure 1.10). The separation of nearby geodesics satisfies the "equation of geodesic deviation,"

$$d^2\xi/ds^2 + R\xi = 0. \tag{1.6}$$

Here $R = 1/a^2$ is the so-called Gaussian curvature of the surface. For the surface of the apple, the same equation applies, with the one difference that the curvature $R$ varies from place to place.

**Box 1.5 TEST FOR FLATNESS**

1. Specify the extension in space $L$ (cm or m) and extension in time $T$ (cm or m of light travel time) of the region under study.

2. Specify the precision $\delta\xi$ with which one can measure the separation of test particles in this region.

3. Follow the motion of test particles moving along initially parallel world lines through this region of spacetime.

4. When the world lines remain parallel to the precision $\delta\xi$ for all directions of travel, then one says that "in a region so limited and to a precision so specified, spacetime is flat."

EXAMPLE: Region just above the surface of the earth, 100 m $\times$ 100 m $\times$ 100 m (space extension), followed for $10^9$ m of light-travel time ($T_{\text{conv}} \sim$ 3 sec). Mass of Earth, $m_{\text{conv}} = 5.98 \times 10^{27}$ g, $m = (0.742 \times 10^{-28}$ cm/g$) \times (5.98 \times 10^{27}$ g$) =$ 0.444 cm [see eq. (1.12)]. Tide-producing acceleration $R^z{}_{0z0}$ (relative acceleration in $z$-direction of two test particles initially at rest and separated from each other by 1 cm of vertical elevation) is

$$\begin{aligned}(d/dr)(m/r^2) &= -2m/r^3\\ &= -0.888\text{ cm}/(6.37 \times 10^8\text{ cm})^3\\ &= -3.44 \times 10^{-27}\text{ cm}^{-2}\end{aligned}$$

("cm of relative displacement per cm of light-travel time per cm of light-travel time per cm of vertical separation"). Two test particles with a vertical separation $\xi^z = 10^4$ cm acquire in the time $t = 10^{11}$ cm (difference between time and proper time negligible for such slowly moving test particles) a relative displacement

$$\begin{aligned}\delta\xi^z &= -\tfrac{1}{2}R^z{}_{0z0}t^2\xi^z\\ &= 1.72 \times 10^{-27}\text{ cm}^{-2}(10^{11}\text{ cm})^2\,10^4\text{ cm}\\ &= 1.72\text{ mm.}\end{aligned}$$

(Change in relative separation less for other directions of motion). When the minimum uncertainty $\delta\xi$ attainable in a measurement over a 100 m spacing is "worse" than this figure (exceeds 1.72 mm), then to this level of precision the region of spacetime under consideration can be treated as flat. When the uncertainty in measurement is "better" (less) than 1.72 mm, then one must limit attention to a smaller region of space or a shorter interval of time or both, to find a region of spacetime that can be regarded as flat to that precision.

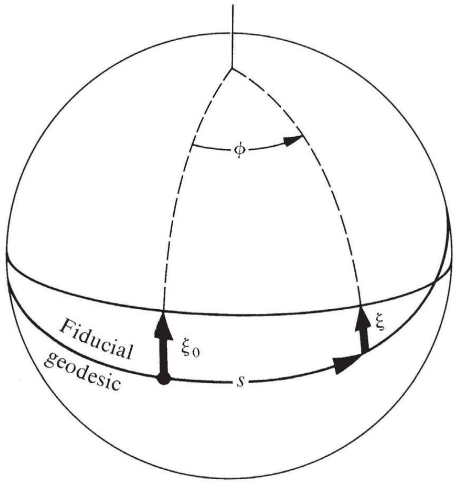



**Figure 1.10.**
Curvature as manifested in the "acceleration of the separation" of two nearby geodesics. Two geodesics, originally parallel, and separated by the distance ("geodesic deviation") $\xi_0$, are no longer parallel when followed a distance $s$. The separation is $\xi = \xi_0 \cos \phi = \xi_0 \cos (s/a)$, where $a$ is the radius of the sphere. The separation follows the equation of simple harmonic motion, $d^2\xi/ds^2 + (1/a^2)\,\xi = 0$ ("equation of geodesic deviation").

The direction of the separation vector, $\boldsymbol{\xi}$, is fixed fully by its orthogonality to the fiducial geodesic. Hence, no reference to the direction of $\boldsymbol{\xi}$ is needed or used in the equation of geodesic deviation; only the magnitude $\xi$ of $\boldsymbol{\xi}$ appears there, and only the magnitude, not direction, of the relative acceleration appears.

In a space of more than two dimensions, an equation of the same general form applies, with several differences. In two dimensions the *direction* of acceleration of one geodesic relative to a nearby, fiducial geodesic is fixed uniquely by the demand that their separation vector, $\boldsymbol{\xi}$, be perpendicular to the fiducial geodesic (see Figure 1.10). Not so in three dimensions or higher. There $\boldsymbol{\xi}$ can remain perpendicular to the fiducial geodesic but rotate about it (Figure 1.11). Thus, to specify the relative acceleration uniquely, one must give not only its magnitude, but also its direction.

The relative acceleration in three dimensions and higher, then, is a vector. Call it "$D^2\boldsymbol{\xi}/ds^2$," and call its four components "$D^2\xi^\alpha/ds^2$." Why the capital $D$? Why not "$d^2\xi^\alpha/ds^2$"? Because our coordinate system is completely arbitrary (cf. §1.2). The twisting and turning of the coordinate lines can induce changes from point to point in the components $\xi^\alpha$ of $\boldsymbol{\xi}$, even if the vector $\boldsymbol{\xi}$ is not changing at all. Consequently, the accelerations of the components $d^2\xi^\alpha/ds^2$ are generally not equal to the components $D^2\xi^\alpha/ds^2$ of the acceleration!

Curvature is characterized by Riemann tensor

How, then, in curved spacetime can one determine the components $D^2\xi^\alpha/ds^2$ of the relative acceleration? By a more complicated version of the equation of geodesic deviation (1.6). Differential geometry (Part III of this book) provides us with a geometrical object called the *Riemann curvature tensor,* "***Riemann***." ***Riemann*** is

(*continued on page 34*)

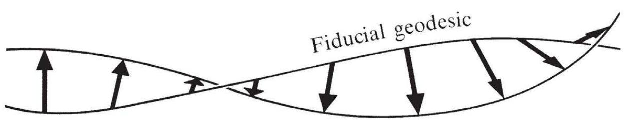
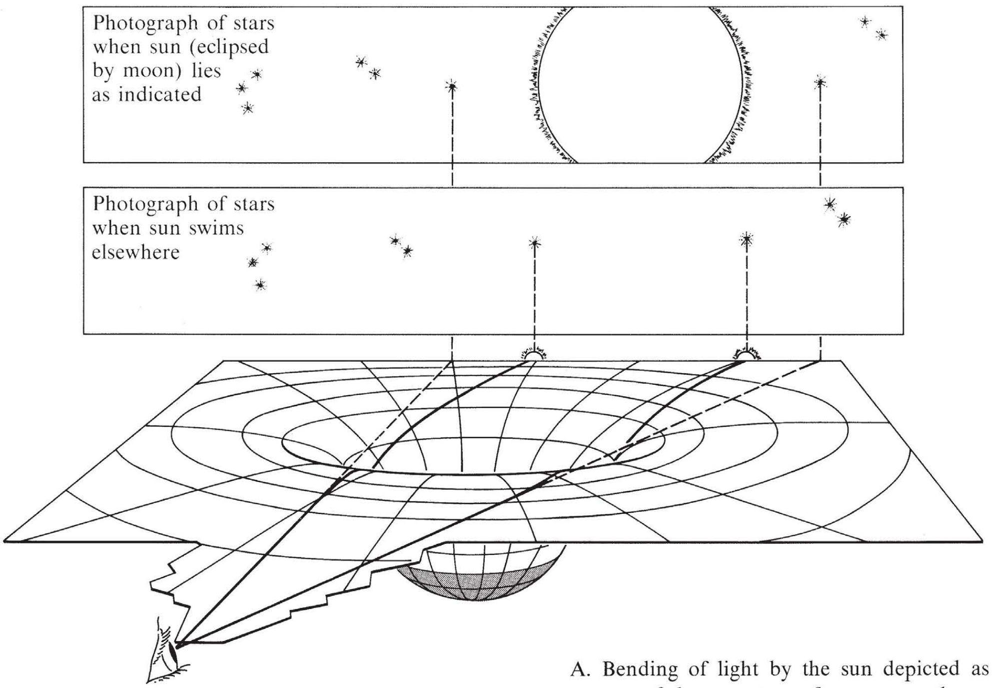



**Figure 1.11.**
The separation vector $\boldsymbol{\xi}$ between two geodesics in a curved three-dimensional manifold. Here $\boldsymbol{\xi}$ can not only change its length from point to point, but also rotate at a varying rate about the fiducial geodesic. Consequently, the relative acceleration of the geodesics must be characterized by a direction as well as a magnitude; it must be a vector, $D^2\boldsymbol{\xi}/ds^2$.

## Box 1.6 CURVATURE OF WHAT?

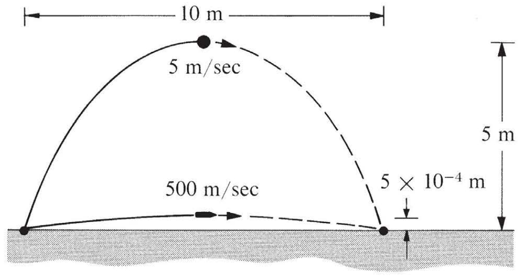
Nothing seems more attractive at first glance than the idea that gravitation is a manifestation of the curvature of space (A), and nothing more ridiculous at a second glance (B). How can the tracks of a ball and of a bullet be curved so differently if that curvature arises from the geometry of space? No wonder that great Riemann did not give the world a geometric theory of gravity. Yes, at the age of 28 (June 10, 1854) he gave the world the mathematical machinery to define and calculate curvature (metric and Riemannian geometry). Yes, he spent his dying days at 40 working to find a unified account of electricity and gravitation. But if there was one reason more than any other why he failed to make the decisive connection between gravitation and curvature, it was this, that he thought of space and the curvature of space, not of spacetime and the curvature of spacetime. To make that forward step took the forty years to special relativity (1905: time on the same footing as space) and then another ten years (1915: general relativity). Depicted in spacetime (C), the tracks of ball and bullet appear to have comparable curvature. In fact, however, neither track has any curvature at all. They both look curved in (C) only because one has forgotten that the spacetime they reside in is itself curved—curved precisely enough to make these tracks the straightest lines in existence ("geodesics").

If it is at first satisfying to see curvature, and curvature of spacetime at that, coming to the fore in so direct a way, then a little more reflection produces a renewed sense of concern. Curvature with respect to what? Not with respect to the labo-

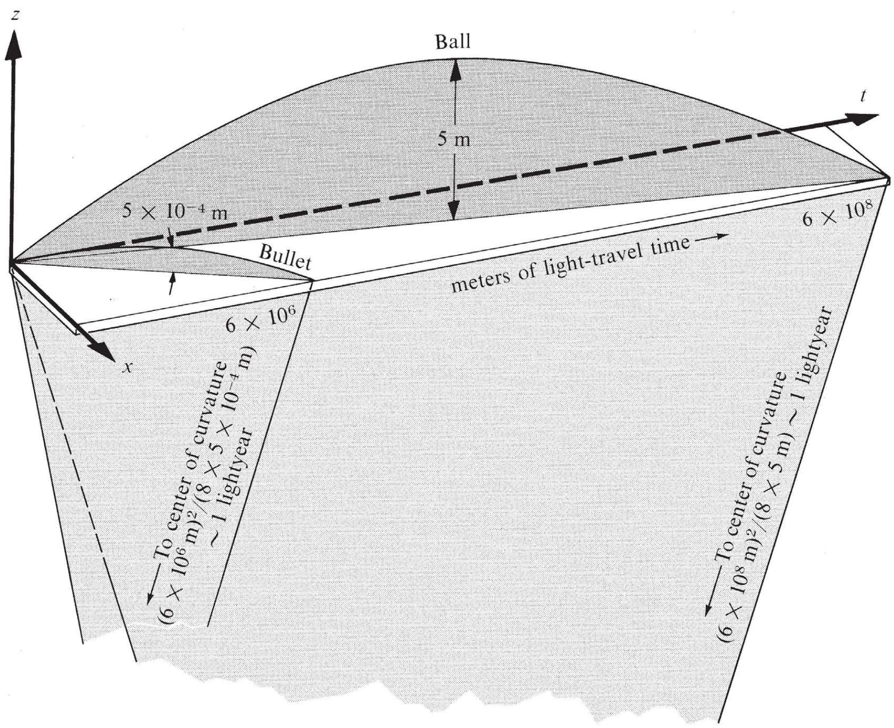



A. Bending of light by the sun depicted as a consequence of the curvature of space near the sun. Ray of light pursues geodesic, but geometry in which it travels is curved (actual travel takes place in spacetime rather than space; correct deflection is twice that given by above elementary picture). Deflection inversely proportional to angular separation between star and center of sun. See Box 40.1 for actual deflections observed at time of an eclipse.

ratory. The earth-bound laboratory has no simple status whatsoever in a proper discussion. First, it is no Lorentz frame. Second, even to mention the earth makes one think of an action-at-a-distance version of gravity (distance from center of earth to ball or bullet). In contrast, it was the whole point of Einstein that physics looks simple only when analyzed locally. To look at local physics, however, means to compare one geodesic of one test particle with geodesics of other test particles traveling (1) nearby with (2) nearly the same directions and (3) nearly the same speeds. Then one can "look at the separations between these nearby test particles and from the second time-rate of change of these separations and the 'equation of geodesic deviation' (equation 1.8) read out the curvature of spacetime."




B. Tracks of ball and bullet through space as seen in laboratory have very different curvatures.




C. Tracks of ball and bullet through spacetime, as recorded in laboratory, have comparable curvatures. Track compared to arc of circle: (radius) = (horizontal distance)$^2$/8 (rise).

the higher-dimensional analog of the Gaussian curvature $R$ of our apple's surface. ***Riemann*** is the mathematical embodiment of the bends and warps in spacetime. And ***Riemann*** is the agent by which those bends and warps (curvature of spacetime) produce the relative acceleration of geodesics.

***Riemann***, like the metric tensor $\boldsymbol{g}$ of Box 1.3, can be thought of as a family of machines, one machine residing at each event in spacetime. Each machine has three slots for the insertion of three vectors:

$$\begin{array}{cccc} & \text{slot 1} & \text{slot 2} & \text{slot 3} \\ & \downarrow & \downarrow & \downarrow \\ \textbf{\textit{Riemann}}\ ( & , & , & ). \end{array}$$

Choose a fiducial geodesic (free-particle world line) passing through an event $\mathcal{Q}$, and denote its unit tangent vector (particle 4-velocity) there by

$$\boldsymbol{u} = \boldsymbol{dx}/d\tau;\ \text{components, } u^\alpha = dx^\alpha/d\tau. \tag{1.7}$$

Choose another, neighboring geodesic, and denote by $\boldsymbol{\xi}$ its perpendicular separation from the fiducial geodesic. Then insert $\boldsymbol{u}$ into the first slot of ***Riemann*** at $\mathcal{Q}$, $\boldsymbol{\xi}$ into the second slot, and $\boldsymbol{u}$ into the third. ***Riemann*** will grind for awhile; then out will pop a new vector,

$$\textbf{\textit{Riemann}}\ (\boldsymbol{u}, \boldsymbol{\xi}, \boldsymbol{u}).$$

Riemann tensor, through equation of geodesic deviation, produces relative accelerations

The equation of geodesic deviation states that this new vector is the negative of the relative acceleration of the two geodesics:

$$D^2\boldsymbol{\xi}/d\tau^2 + \textbf{\textit{Riemann}}\ (\boldsymbol{u}, \boldsymbol{\xi}, \boldsymbol{u}) = 0. \tag{1.8}$$

The Riemann tensor, like the metric tensor (Box 1.3), and like all other tensors, is a linear machine. The vector it puts out is a linear function of each vector inserted into a slot:

$$\begin{aligned} &\textbf{\textit{Riemann}}\ (2\boldsymbol{u}, a\boldsymbol{w} + b\boldsymbol{v}, 3\boldsymbol{r}) \\ &= 2 \times a \times 3\ \textbf{\textit{Riemann}}\ (\boldsymbol{u}, \boldsymbol{w}, \boldsymbol{r}) + 2 \times b \times 3\ \textbf{\textit{Riemann}}\ (\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{r}). \end{aligned} \tag{1.9}$$

Consequently, in any coordinate system the components of the vector put out can be written as a "trilinear function" of the components of the vectors put in:

$$\boldsymbol{r} = \textbf{\textit{Riemann}}\ (\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \Longleftrightarrow r^\alpha = R^\alpha{}_{\beta\gamma\delta}\, u^\beta\, v^\gamma\, w^\delta. \tag{1.10}$$

(Here there is an implied summation on the indices $\beta$, $\gamma$, $\delta$; cf. Box 1.1.) The $4 \times 4 \times 4 \times 4 = 256$ numbers $R^\alpha{}_{\beta\gamma\delta}$ are called the "components of the Riemann tensor in the given coordinate system." In terms of components, the equation of geodesic deviation states

$$\frac{D^2\xi^\alpha}{d\tau^2} + R^\alpha{}_{\beta\gamma\delta}\frac{dx^\beta}{d\tau}\,\xi^\gamma\,\frac{dx^\delta}{d\tau} = 0. \tag{1.8$'$}$$

In Einstein's geometric theory of gravity, this equation of geodesic deviation summarizes the entire effect of geometry on matter. It does for gravitation physics what the Lorentz force equation,

Equation of geodesic deviation is analog of Lorentz force law

$$\frac{D^2x^\alpha}{d\tau^2} - \frac{e}{m}F^\alpha{}_\beta \frac{dx^\alpha}{d\tau} = 0, \tag{1.11}$$

does for electromagnetism. See Box 1.7.

The units of measurement of the curvature are $\mathrm{cm}^{-2}$ just as well in spacetime as on the surface of the apple. Nothing does so much to make these units stand out clearly as to express mass in "geometrized units":

Geometrized units

$$\begin{aligned} m(\mathrm{cm}) &= (G/c^2)m_{\mathrm{conv}}(\mathrm{g}) \\ &= (0.742 \times 10^{-28}\ \mathrm{cm/g})m_{\mathrm{conv}}(\mathrm{g}). \end{aligned} \tag{1.12}$$

**Box 1.7 EQUATION OF MOTION UNDER THE INFLUENCE OF A GRAVITATIONAL FIELD AND AN ELECTROMAGNETIC FIELD, COMPARED AND CONTRASTED**

| | *Electromagnetism* [*Lorentz force, equation (1.11)*] | *Gravitation* [*Equation of geodesic deviation (1.8′)*] |
|---|---|---|
| Acceleration is defined for one particle? | Yes | No |
| Acceleration defined how? | Actual world line compared to world line of uncharged "fiducial" test particle passing through same point with same 4-velocity. | Already an uncharged test particle, which can't accelerate relative to itself! Acceleration measured relative to a nearby test particle as fiduciary standard. |
| Acceleration depends on all four components of the 4-velocity of the particle? | Yes | Yes |
| Universal acceleration for all test particles in same locations with same 4-velocity? | No; is proportional to $e/m$ | Yes |
| Driving field | Electromagnetic field | Riemann curvature tensor |
| Ostensible number of distinct components of driving field | $4 \times 4 = 16$ | $4^4 = 256$ |
| Actual number when allowance is made for symmetries of tensor | 6 | 20 |
| Names for more familiar of these components | 3 electric<br>3 magnetic | 6 components of local Newtonian tide-producing acceleration |

This conversion from grams to centimeters by means of the ratio

$$G/c^2 = 0.742 \times 10^{-28}\ \text{cm/g}$$

is completely analogous to converting from seconds to centimeters by means of the ratio

$$c = \frac{9.4605460000 \times 10^{17}\ \text{cm}}{31{,}556{,}925.974\ \text{sec}}$$

(see end of §1.5). The sun, which in conventional units has $m_{\text{conv}} = 1.989 \times 10^{33}$ g, has in geometrized units a mass $m = 1.477$ km. Box 1.8 gives further discussion.

Components of Riemann tensor evaluated from relative accelerations of slowly moving particles

Using geometrized units, and using the Newtonian theory of gravity, one can readily evaluate nine of the most interesting components of the Riemann curvature tensor near the Earth or the sun. The method is the gravitational analog of determining the electric field strength by measuring the acceleration of a slowly moving test particle. Consider the separation between the geodesics of two nearby and slowly moving ($v \ll c$) particles at a distance $r$ from the Earth or sun. In the standard, nearly inertial coordinates of celestial mechanics, all components of the 4-velocity of the

**Box 1.8 GEOMETRIZED UNITS**

Throughout this book, we use "geometrized units," in which the speed of light $c$, Newton's gravitational constant $G$, and Boltzman's constant $k$ are all equal to unity. The following alternative ways to express the number 1.0 are of great value:

$$1.0 = c = 2.997930 \cdots \times 10^{10}\ \text{cm/sec}$$

$$1.0 = G/c^2 = 0.7425 \times 10^{-28}\ \text{cm/g};$$

$$1.0 = G/c^4 = 0.826 \times 10^{-49}\ \text{cm/erg};$$

$$1.0 = Gk/c^4 = 1.140 \times 10^{-65}\ \text{cm/K};$$

$$1.0 = c^2/G^{1/2} = 3.48 \times 10^{24}\ \text{cm/gauss}^{-1}.$$

One can multiply a factor of unity, expressed in any one of these ways, into any term in any equation without affecting the validity of the equation. Thereby one can convert one's units of measure from grams to centimeters to seconds to ergs to . . . . For example:

$$\begin{aligned}\text{Mass of sun} = M_\odot &= 1.989 \times 10^{33}\ \text{g}\\ &= (1.989 \times 10^{33}\ g) \times (G/c^2)\\ &= 1.477 \times 10^{5}\ \text{cm}\\ &= (1.989 \times 10^{33}\ g) \times (c^2)\\ &= 1.788 \times 10^{54}\ \text{ergs}.\end{aligned}$$

The standard unit, in terms of which everything is measured in this book, is centimeters. However, occasionally conventional units are used; in such cases a subscript "conv" is sometimes, but not always, appended to the quantity measured:

$$M_{\odot\text{conv}} = 1.989 \times 10^{33}\ \text{g}.$$

fiducial test particle can be neglected except $dx^0/d\tau = 1$. The space components of the equation of geodesic deviation read

$$d^2\xi^k/d\tau^2 + R^k{}_{0j0}\xi^j = 0. \tag{1.13}$$

Comparing with the conclusions of Newtonian theory, equations (1.5), we arrive at the following information about the curvature of spacetime near a center of mass:

$$\begin{Vmatrix} R^{\hat{x}}{}_{\hat{0}\hat{x}\hat{0}} & R^{\hat{y}}{}_{\hat{0}\hat{x}\hat{0}} & R^{\hat{z}}{}_{\hat{0}\hat{x}\hat{0}} \\ R^{\hat{x}}{}_{\hat{0}\hat{y}\hat{0}} & R^{\hat{y}}{}_{\hat{0}\hat{y}\hat{0}} & R^{\hat{z}}{}_{\hat{0}\hat{y}\hat{0}} \\ R^{\hat{x}}{}_{\hat{0}\hat{z}\hat{0}} & R^{\hat{y}}{}_{\hat{0}\hat{z}\hat{0}} & R^{\hat{z}}{}_{\hat{0}\hat{z}\hat{0}} \end{Vmatrix} \begin{matrix} = \\ = \\ = \end{matrix} \begin{Vmatrix} m/r^3 & 0 & 0 \\ 0 & m/r^3 & 0 \\ 0 & 0 & -2m/r^3 \end{Vmatrix} \tag{1.14}$$

(units $\text{cm}^{-2}$). Here and henceforth the caret or "hat" is used to indicate the components of a vector or tensor in a local Lorentz frame of reference ("physical components," as distinguished from components in a general coordinate system). Einstein's theory will determine the values of the other components of curvature (e.g., $R^{\hat{x}}{}_{\hat{z}\hat{x}\hat{z}} = -m/r^3$); but these nine terms are the ones of principal relevance for many applications of gravitation theory. They are analogous to the components of the electric field in the Lorentz equation of motion. Many of the terms not evaluated are analogous to magnetic field components—ordinarily weak unless the source is in rapid motion.

This ends the survey of the effect of geometry on matter ("effect of curvature of apple in causing geodesics to cross"—especially great near the dimple at the top, just as the curvature of spacetime is especially large near a center of gravitational attraction). Now for the effect of matter on geometry ("effect of stem of apple in causing dimple")!

## §1.7. EFFECT OF MATTER ON GEOMETRY

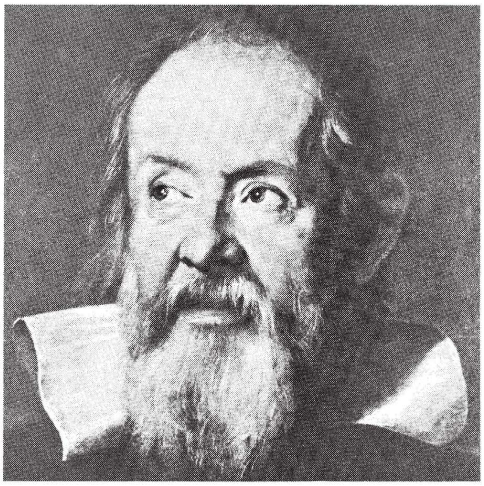
*The weight of any heavy body of known weight at a particular distance from the center of the world varies according to the variation of its distance therefrom; so that as often as it is removed from the center, it becomes heavier, and when brought near to it, is lighter. On this account, the relation of gravity to gravity is as the relation of distance to distance from the center.*

AL KHĀZINĪ (*Merv, A.D. 1115*), *Book of the Balance of Wisdom*

Figure 1.12 shows a sphere of the same density, $\rho = 5.52\ \text{g/cm}^3$, as the average density of the Earth. A hole is bored through this sphere. Two test particles, $A$ and $B$, execute simple harmonic motion in this hole, with an 84-minute period. Therefore their geodesic separation $\boldsymbol{\xi}$, however it may be oriented, undergoes a simple periodic motion with the same 84-minute period:

$$d^2\xi^j/d\tau^2 = -\left(\frac{4\pi}{3}\rho\right)\xi^j, \qquad j = x \text{ or } y \text{ or } z. \tag{1.15}$$

**Box 1.9 GALILEO GALILEI**
**Pisa, February 15, 1564—Arcetri, Florence, January 8, 1642**


Uffizi Gallery, Florence

*"In questions of science the authority of a thousand is not worth the humble reasoning of a single individual."*

GALILEO GALILEI (1632)

*"The spaces described by a body falling from rest with a uniformly accelerated motion are to each other as the squares of the time intervals employed in traversing these distances."*

GALILEO GALILEI (1638)

*"Everything that has been said before and imagined by other people* [*about the tides*] *is in my opinion completely invalid. But among the great men who have philosophised about this marvellous effect of nature the one who surprised me the most is Kepler. More than other people he was a person of independent genius, sharp, and had in his hands the motion of the earth. He later pricked up his ears and became interested in the action of the moon on the water, and in other occult phenomena, and similar childishness."*

GALILEO GALILEI (1632)

*"It is a most beautiful and delightful sight to behold* [*with the new telescope*] *the body of the Moon . . . the Moon certainly does not possess a smooth and polished surface, but one rough and uneven . . . full of vast protuberances, deep chasms and sinuosities . . . stars in myriads, which have never been seen before and which surpass the old, previously known, stars in number more than ten times. I have discovered four planets, neither known nor observed by any one of the astronomers before my time . . . got rid of disputes about the Galaxy or Milky Way, and made its nature clear to the very senses, not to say to the understanding . . . the galaxy is nothing else than a mass of luminous stars planted together in clusters . . . the number of small ones is quite beyond determination—the stars which have been called by every one of the astronomers up to this day* nebulous *are groups of small stars set thick together in a wonderful way."*

GALILEO GALILEI IN *SIDEREUS NUNCIUS* (1610)

*"So the principles which are set forth in this treatise will, when taken up by thoughtful minds, lead to many another more remarkable result; and it is to be believed that it will be so on account of the nobility of the subject, which is superior to any other in nature."*

GALILEO GALILEI (1638)

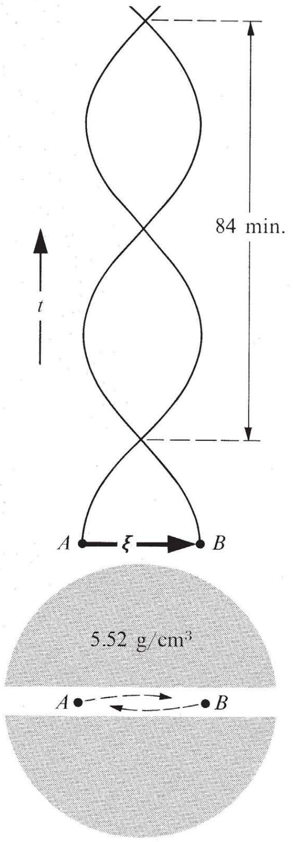



**Figure 1.12.**
Test particles $A$ and $B$ move up and down a hole bored through the Earth, idealized as of uniform density. At radius $r$, a particle feels Newtonian acceleration

$$\begin{aligned}\frac{d^2r}{d\tau^2} &= \frac{1}{c^2}\frac{d^2r}{dt_{\text{conv}}{}^2}\\ &= -\frac{G}{c^2}\frac{(\text{mass inside radius } r)}{r^2}\\ &= -\left(\frac{G}{r^2c^2}\right)\left(\frac{4\pi}{3}\rho_{\text{conv}}r^3\right)\\ &= -\omega^2 r.\end{aligned}$$

Consequently, each particle oscillates in simple harmonic motion with precisely the same angular frequency as a satellite, grazing the model Earth, traverses its circular orbit:

$$\omega^2(\text{cm}^{-2}) = \frac{4\pi}{3}\rho(\text{cm}^{-2}),$$

$$\omega^2_{\text{conv}}(\text{sec}^{-2}) = \frac{4\pi G}{3}\rho_{\text{conv}}(\text{g/cm}^3).$$

Comparing this actual motion with the equation of geodesic deviation (1.13) for slowly moving particles in a nearly inertial frame, we can read off some of the curvature components for the interior of this model Earth.

The Riemann tensor inside the Earth

$$\begin{Vmatrix} R^{\hat{x}}{}_{\hat{0}\hat{x}\hat{0}} & R^{\hat{y}}{}_{\hat{0}\hat{x}\hat{0}} & R^{\hat{z}}{}_{\hat{0}\hat{x}\hat{0}} \\ R^{\hat{x}}{}_{\hat{0}\hat{y}\hat{0}} & R^{\hat{y}}{}_{\hat{0}\hat{y}\hat{0}} & R^{\hat{z}}{}_{\hat{0}\hat{y}\hat{0}} \\ R^{\hat{x}}{}_{\hat{0}\hat{z}\hat{0}} & R^{\hat{y}}{}_{\hat{0}\hat{z}\hat{0}} & R^{\hat{z}}{}_{\hat{0}\hat{z}\hat{0}} \end{Vmatrix} = (4\pi\rho/3)\begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix} \tag{1.16}$$

This example illustrates how the curvature of spacetime is connected to the distribution of matter.

Let a gravitational wave from a supernova pass through the Earth. Idealize the Earth's matter as so nearly incompressible that its density remains practically unchanged. The wave is characterized by ripples in the curvature of spacetime, propagating with the speed of light. The ripples will show up in the components $R^j{}_{0k0}$ of the Riemann tensor, and in the relative acceleration of our two test particles. The left side of equation (1.16) will ripple; but the right side will not. Equation (1.16) will break down. No longer will the Riemann curvature be generated directly and solely by the Earth's matter.

Effect of gravitational wave on Riemann tensor

Nevertheless, Einstein tells us, a part of equation (1.16) is undisturbed by the waves: its trace

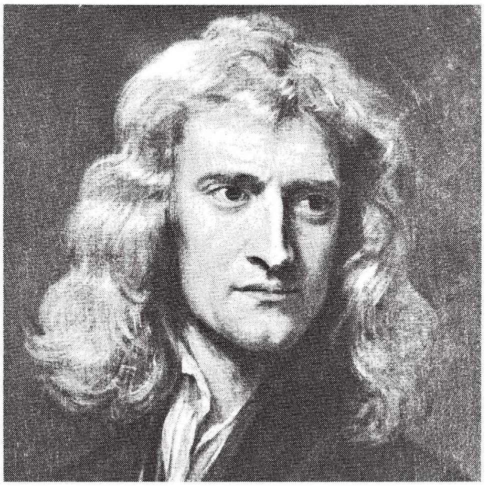
$$R_{\hat{0}\hat{0}} \equiv R^{\hat{x}}{}_{\hat{0}\hat{x}\hat{0}} + R^{\hat{y}}{}_{\hat{0}\hat{y}\hat{0}} + R^{\hat{z}}{}_{\hat{0}\hat{z}\hat{0}} = 4\pi\rho. \tag{1.17}$$

Even in the vacuum outside the Earth this is valid; there both sides vanish [cf. (1.14)].

Einstein tensor introduced

More generally, a certain piece of the Riemann tensor, called the *Einstein tensor* and denoted ***Einstein*** or ***G***, is always generated directly by the local distribution of matter. ***Einstein*** is the geometric object that generalizes $R_{\hat{0}\hat{0}}$, the lefthand side

**Box 1.10 ISAAC NEWTON**
**Woolsthorpe, Lincolnshire, England, December 25, 1642—**
**Kensington, London, March 20, 1726**



*"The description of right lines and circles, upon which geometry is founded, belongs to mechanics. Geometry does not teach us to draw these lines, but requires them to be drawn."*

[FROM P. 1 OF NEWTON'S PREFACE TO THE FIRST (1687) EDITION OF THE *PRINCIPIA*]

*"Absolute space, in its own nature, without relation to anything external, remains always similar and immovable*
*"Absolute, true, and mathematical time, of itself, and from its own nature, flows equably without relation to anything external."*

[FROM THE *SCHOLIUM* IN THE *PRINCIPIA*]

*"I have not been able to discover the cause of those properties of gravity from phenomena, and I frame no hypotheses; for whatever is not reduced from the phenomena is to be called an hypothesis; and hypotheses . . . have no place in experimental philosophy. . . . And to us it is enough that gravity does really exist, and act according to the laws which we have explained, and abundantly serves to account for all the motions of the celestial bodies, and of our sea."*

[FROM THE *GENERAL SCHOLIUM* ADDED AT THE END OF THE THIRD BOOK OF THE *PRINCIPIA* IN THE SECOND EDITION OF 1713; ESPECIALLY FAMOUS FOR THE PHRASE OFTEN QUOTED FROM NEWTON'S ORIGINAL LATIN, *"HYPOTHESES NON FINGO."*]

*"And the same year [1665 or 1666] I began to think of gravity extending to the orb of the Moon, and having found out. . . . All this was in the two plague years of 1665 and 1666, for in those days I was in the prime of my age for invention, and minded Mathematicks and Philosophy more than at any time since."*

[FROM MEMORANDUM IN NEWTON'S HANDWRITING ABOUT HIS DISCOVERIES ON FLUXIONS, THE BINOMIAL THEOREM, OPTICS, DYNAMICS, AND GRAVITY, BELIEVED TO HAVE BEEN WRITTEN ABOUT 1714, AND FOUND BY ADAMS ABOUT 1887 IN THE "PORTSMOUTH COLLECTION" OF NEWTON PAPERS] of equation (1.17). Like $R_{\hat{0}\hat{0}}$, ***Einstein*** is a sort of average of ***Riemann*** over all directions. Generating ***Einstein*** and generalizing the righthand side of (1.16) is a geometric object called the *stress-energy tensor* of the matter. It is denoted $\boldsymbol{T}$. No coordinates are need to define ***Einstein***, and none to define $\boldsymbol{T}$; like the Riemann tensor, ***Riemann***, and the metric tensor, $\boldsymbol{g}$, they exist in the complete absence of coordinates. Moreover, in nature they are always equal, aside from a factor of $8\pi$:

Stress-energy tensor introduced

$$\textbf{\textit{Einstein}} \equiv \boldsymbol{G} = 8\pi \boldsymbol{T}. \tag{1.18}$$

*"For hypotheses ought . . . to explain the properties of things and not attempt to predetermine them except in so far as they can be an aid to experiments."*

[FROM LETTER OF NEWTON TO I. M. PARDIES, 1672, AS QUOTED IN THE CAJORI NOTES AT THE END OF NEWTON (1687), P. 673]

*"That one body may act upon another at a distance through a vacuum, without the mediation of any thing else, by and through which their action and force may be conveyed from one to another, is to me so great an absurdity, that I believe no man, who has in philosophical matters a competent faculty of thinking, can ever fall into it."*

[PASSAGE OFTEN QUOTED BY MICHAEL FARADAY FROM LETTERS OF NEWTON TO RICHARD BENTLY, 1692–1693, AS QUOTED IN THE NOTES OF THE CAJORI EDITION OF NEWTON (1687), P. 643]

*"The attractions of gravity, magnetism, and electricity, reach to very sensible distances, and so have been observed . . . ; and there may be others which reach to so small distances as hitherto escape observation; . . . some force, which in immediate contract is exceeding strong, at small distances performs the chemical operations above-mentioned, and reaches not far from the particles with any sensible effect."*

[FROM QUERY 31 AT THE END OF NEWTON'S *OPTICKS* (1730)]

*"What is there in places almost empty of matter, and whence is it that the sun and planets gravitate towards one another, without dense matter between them? Whence is it that nature doth nothing in vain; and whence arises all that order and beauty which we see in the world? To what end are comets, and whence is it that planets move all one and the same way in orbs concentrick, while comets move all manner of ways in orbs very excentrick; and what hinders the fixed stars from falling upon one another?"*

[FROM QUERY 28]

*"He is not eternity or infinity, but eternal and infinite; He is not duration or space, but He endures and is present. He endures forever, and is everywhere present; and by existing always and everywhere, He constitutes duration and space. . . . And thus much concerning God; to discourse of whom from the appearances of things, does certainly belong to natural philosophy."*

[FROM THE *GENERAL SCHOLIUM* AT THE END OF THE *PRINCIPIA* (1687)]

Einstein field equation: how matter generates curvature

This *Einstein field equation,* rewritten in terms of components in an arbitrary coordinate system, reads

$$G_{\alpha\beta} = 8\pi T_{\alpha\beta}. \tag{1.19}$$

The Einstein field equation is elegant and rich. No equation of physics can be written more simply. And none contains such a treasure of applications and consequences.

Consequences of Einstein field equation

The field equation shows how the stress-energy of matter generates an average curvature (***Einstein*** $\equiv \boldsymbol{G}$) in its neighborhood. Simultaneously, the field equation is a propagation equation for the remaining, anisotropic part of the curvature: it governs the external spacetime curvature of a static source (Earth); it governs the generation of gravitational waves (ripples in curvature of spacetime) by stress-energy in motion; and it governs the propagation of those waves through the universe. The field equation even contains within itself the equations of motion ("Force =

**Box 1.11**
**ALBERT EINSTEIN**
**Ulm, Germany,**
**March 14, 1879—**
**Princeton, New Jersey,**
**April 18, 1955**

Library of E. T. Hochschule, Zürich

Académie des Sciences, Paris

Archives of California Institute of Technology

SEAL: Courtesy of the Lewis and Rosa Strauss Foundation and Princeton University Press mass $\times$ acceleration") for the matter whose stress-energy generates the curvature.

Those were some consequences of $\boldsymbol{G} = 8\pi\boldsymbol{T}$. Now for some applications.

Applications of Einstein field equation

The field equation governs the motion of the planets in the solar system; it governs the deflection of light by the sun; it governs the collapse of a star to form a black hole; it determines uniquely the external spacetime geometry of a black hole ("a black hole has no hair"); it governs the evolution of spacetime singularities at the end point of collapse; it governs the expansion and recontraction of the universe. And more; much more.

In order to understand how the simple equation $\boldsymbol{G} = 8\pi\boldsymbol{T}$ can be so all powerful, it is desirable to backtrack, and spend a few chapters rebuilding the entire picture of spacetime, of its curvature, and of its laws, this time with greater care, detail, and mathematics.

Thus ends this survey of the effect of geometry on matter, and the reaction of matter back on geometry, rounding out the parable of the apple.

*"What really interests me is whether God had any choice in the creation of the world"*

EINSTEIN TO AN ASSISTANT, AS QUOTED BY G. HOLTON (1971), P. 20

*"But the years of anxious searching in the dark, with their intense longing, their alternations of confidence and exhaustion, and the final emergence into the light—only those who have experienced it can understand that"*

EINSTEIN, AS QUOTED BY M. KLEIN (1971), P. 1315

*"Of all the communities available to us there is not one I would want to devote myself to, except for the society of the true searchers, which has very few living members at any time. . ."*

EINSTEIN LETTER TO BORN, QUOTED BY BORN (1971), P. 82

*"I am studying your great works and—when I get stuck anywhere—now have the pleasure of seeing your friendly young face before me smiling and explaining"*

EINSTEIN, LETTER OF MAY 2, 1920, AFTER MEETING NIELS BOHR

*"As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality."*

EINSTEIN (1921), P. 28

*"The most incomprehensible thing about the world is that it is comprehensible."*

EINSTEIN, IN SCHILPP (1949), P. 112

## EXERCISES

**Exercise 1.1. CURVATURE OF A CYLINDER**

Show that the Gaussian curvature $R$ of the surface of a cylinder is zero by showing that geodesics on that surface (unroll!) suffer no geodesic deviation. Give an independent argument for the same conclusion by employing the formula $R = 1/\rho_1\rho_2$, where $\rho_1$ and $\rho_2$ are the principal radii of curvature at the point in question with respect to the enveloping Euclidean three-dimensional space.

**Exercise 1.2. SPRING TIDE VS. NEAP TIDE**

Evaluate (1) in conventional units and (2) in geometrized units the magnitude of the Newtonian tide-producing acceleration $R^m{}_{0n0}(m, n = 1, 2, 3)$ generated at the Earth by (1) the moon ($m_{\text{conv}} = 7.35 \times 10^{25}$ g, $r = 3.84 \times 10^{10}$ cm) and (2) the sun ($m_{\text{conv}} = 1.989 \times 10^{33}$ g, $r = 1.496 \times 10^{13}$ cm). By what factor do you expect spring tides to exceed neap tides?

**Exercise 1.3. KEPLER ENCAPSULATED**

A small satellite has a circular frequency $\omega(\text{cm}^{-1})$ in an orbit of radius $r$ about a central object of mass $m$(cm). From the known value of $\omega$, show that it is possible to determine neither $r$ nor $m$ individually, but only the effective "Kepler density" of the object as averaged over a sphere of the same radius as the orbit. Give the formula for $\omega^2$ in terms of this Kepler density.

It is a reminder of the continuity of history that Kepler and Galileo (Box 1.9) wrote back and forth, and that the year that witnessed the death of Galileo saw the birth of Newton (Box 1.10). After Newton the first dramatically new synthesis of the laws of gravitation came from Einstein (Box 1.11).

*And what the dead had no speech for, when living,*
*They can tell you, being dead; the communication*
*Of the dead is tongued with fire beyond*
*the language of the living.*

T. S. ELIOT, in *LITTLE GIDDING* (1942)

*I measured the skies*
*Now the shadows I measure*
*Skybound was the mind*
*Earthbound the body rests*

JOHANNES KEPLER, d. November 15, 1630.
He wrote his epitaph in Latin;
it is translated by Coleman (1967), p. 109.

*Ubi materia, ibi geometria.*

JOHANNES KEPLER

PART II

# PHYSICS IN FLAT SPACETIME

*Wherein the reader meets an old friend, Special Relativity, outfitted in new, mod attire, and becomes more intimately acquainted with her charms.*

CHAPTER 2

# FOUNDATIONS OF SPECIAL RELATIVITY

*In geometric and physical applications, it always turns out that a quantity is characterized not only by its tensor order, but also by symmetry.*

HERMAN WEYL (1925)

*Undoubtedly the most striking development of geometry during the last 2,000 years is the continual expansion of the concept ''geometric object.'' This concept began by comprising only the few curves and surfaces of Greek synthetic geometry; it was stretched, during the Renaissance, to cover the whole domain of those objects defined by analytic geometry; more recently, it has been extended to cover the boundless universe treated by point-set theory.*

KARL MENGER, IN SCHILPP (1949), P. 466.

## §2.1. OVERVIEW

Curvature in geometry manifests itself as gravitation. Gravitation works on the separation of nearby particle world lines. In turn, particles and other sources of mass-energy cause curvature in the geometry. How does one break into this closed loop of the action of geometry on matter and the reaction of matter on geometry? One can begin no better than by analyzing the motion of particles and the dynamics of fields in a region of spacetime so limited that it can be regarded as flat. (See "Test for Flatness," Box 1.5).

Background assumed of reader

Chapters 2–6 develop this flat-spacetime viewpoint (special relativity). The reader, it is assumed, is already somewhat familiar with special relativity:* 4-vectors in general; the energy-momentum 4-vector; elementary Lorentz transformations; the Lorentz law for the force on a charged particle; at least one look at one equation

*For example, see Goldstein (1959), Leighton (1959), Jackson (1962), or, for the physical perspective presented geometrically, Taylor and Wheeler (1966).

in one book that refers to the electromagnetic field tensor $F_{\mu\nu}$; and the qualitative features of spacetime diagrams, including such points as (1) future and past light cones, (2) causal relationships ("past of," "future of," "neutral," or "in a spacelike relationship to"), (3) Lorentz contraction, (4) time dilation, (5) absence of a universal concept of simultaneity, and (6) the fact that the $\bar{t}$ and $\bar{z}$ axes in Box 2.4 are orthogonal even though they do not look so. If the reader finds anything new in these chapters, it will be: (i) a new viewpoint on special relativity, one emphasizing coordinate-free concepts and notation that generalize readily to curved spacetime ("geometric objects," tensors viewed as machines—treated in Chapters 2–4); or (ii) unfamiliar topics in special relativity, topics crucial to the later exposition of gravitation theory ("stress-energy tensor and conservation laws," Chapter 5; "accelerated observers," Chapter 6).

## §2.2. GEOMETRIC OBJECTS

Every physical quantity can be described by a geometric object

Everything that goes on in spacetime has its geometric description, and almost every one of these descriptions lends itself to ready generalization from flat spacetime to curved spacetime. The greatest of the differences between one geometric object and another is its scope: the individual object (vector) for the momentum of a certain particle at a certain phase in its history, as contrasted to the extended geometric object that describes an electromagnetic field defined throughout space and time ("antisymmetric second-rank tensor field" or, more briefly, "field of 2-forms"). The idea that every physical quantity must be describable by a geometric object, and that the laws of physics must all be expressible as geometric relationships between these geometric objects, had its intellectual beginnings in the Erlanger program of Felix Klein (1872), came closer to physics in Einstein's "principle of general covariance" and in the writings of Hermann Weyl (1925), seems to have first been formulated clearly by Veblen and Whitehead (1932), and today pervades relativity theory, both special and general.

All laws of physics can be expressed geometrically

A. Nijenhuis (1952) and S.-S. Chern (1960, 1966, 1971) have expounded the mathematical theory of geometric objects. But to understand or do research in geometrodynamics, one need not master this elegant and beautiful subject. One need only know that geometric objects in spacetime are entities that exist independently of coordinate systems or reference frames. A point in spacetime ("*event*") is a geometric object. The arrow linking two neighboring events ("*vector*") is a geometric object in flat spacetime, and its generalization, the "*tangent vector*," is a geometric object even when spacetime is curved. The "*metric*" (machine for producing the squared length of any vector; see Box 1.3) is a geometric object. No coordinates are needed to define any of these concepts.

The next few sections will introduce several geometric objects, and show the roles they play as representatives of physical quantities in flat spacetime.

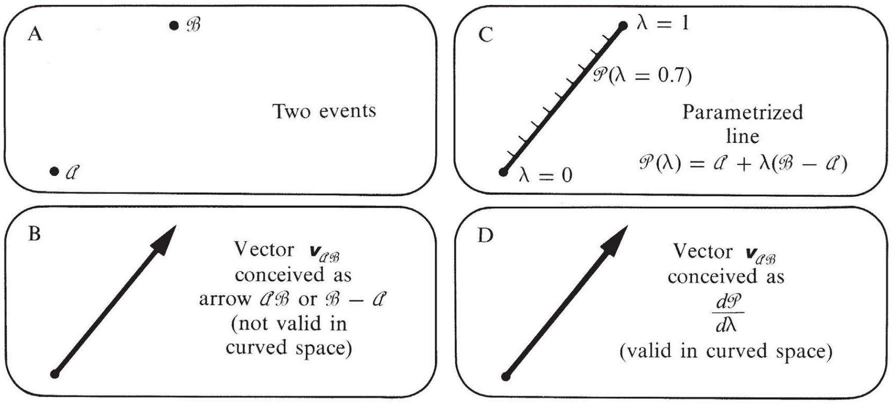



**Figure 2.1.**
From vector as connector of two points to vector as derivative ("tangent vector"; a local rather than a bilocal concept).

## §2.3. VECTORS

Ways of defining vector:

As arrow

Begin with the simplest idea of a vector (Figure 2.1B): an arrow extending from one spacetime event $\mathcal{A}$ ("tail") to another event $\mathcal{B}$ ("tip"). Write this vector as

$$\boldsymbol{v}_{\mathcal{A}\mathcal{B}} = \mathcal{B} - \mathcal{A} \text{ (or } \mathcal{A}\mathcal{B}).$$

As parametrized straight line

For many purposes (including later generalization to curved spacetime) other completely equivalent ways to think of this vector are more convenient. Represent the arrow by the parametrized straight line $\mathcal{P}(\lambda) = \mathcal{A} + \lambda(\mathcal{B} - \mathcal{A})$, with $\lambda = 0$ the tail of the arrow, and $\lambda = 1$ its tip. Form the derivative of this simple linear expression for $\mathcal{P}(\lambda)$:

$$(d/d\lambda)[\mathcal{A} + \lambda(\mathcal{B} - \mathcal{A})] = \mathcal{B} - \mathcal{A} = \mathcal{P}(1) - \mathcal{P}(0) \equiv \text{(tip)} - \text{(tail)} \equiv \boldsymbol{v}_{\mathcal{A}\mathcal{B}}.$$

This result allows one to replace the idea of a vector as a 2-point object ("bilocal") by the concept of a vector as a 1-point object ("tangent vector"; local):

As derivative of point along curve

$$\boldsymbol{v}_{\mathcal{A}\mathcal{B}} = (d\mathcal{P}/d\lambda)_{\lambda=0}. \tag{2.1}$$

*Example:* if $\mathcal{P}(\tau)$ is the straight world line of a free particle, parametrized by its proper time, then the displacement that occurs in a proper time interval of one second gives an arrow $\boldsymbol{u} = \mathcal{P}(1) - \mathcal{P}(0)$. This arrow is easily drawn on a spacetime diagram. It accurately shows the 4-velocity of the particle. However, the derivative formula $\boldsymbol{u} = d\mathcal{P}/d\tau$ for computing the same displacement (1) is more suggestive of the velocity concept and (2) lends itself to the case of accelerated motion. Thus, given a world line $\mathcal{P}(\tau)$ that is not straight, as in Figure 2.2, one must first form $d\mathcal{P}/d\tau$, and only thereafter draw the straight line $\mathcal{P}(0) + \lambda(d\mathcal{P}/d\tau)_0$ of the arrow $\boldsymbol{u} = d\mathcal{P}/d\tau$ to display the 4-velocity $\boldsymbol{u}$.

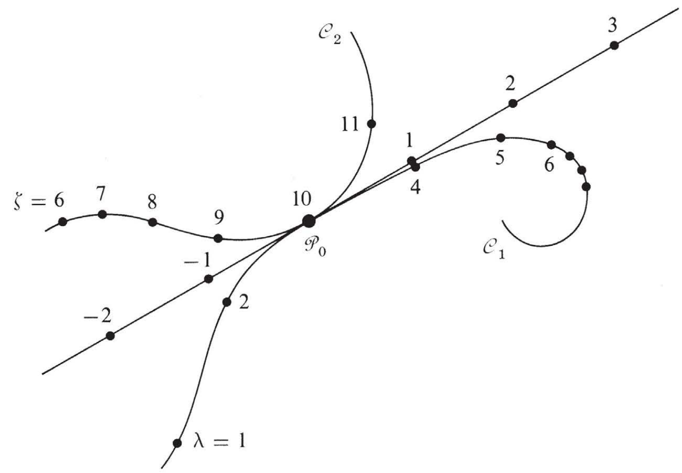



**Figure 2.2.**
Same tangent vector derived from two very different curves. That parametrized straight line is also drawn which best fits the two curves at $\mathcal{P}_0$. The tangent vector reaches from 0 to 1 on this straight line.

Components of a vector

The reader may be unfamiliar with this viewpoint. More familiar may be the components of the 4-velocity in a specific Lorentz reference frame:

$$u^0 = \frac{dt}{d\tau} = \frac{1}{\sqrt{1 - \boldsymbol{v}^2}}, \qquad u^j = \frac{dx^j}{d\tau} = \frac{v^j}{\sqrt{1 - \boldsymbol{v}^2}}, \tag{2.2}$$

where

$$v^j = dx^j/dt = \text{components of "ordinary velocity,"}$$
$$\boldsymbol{v}^2 = (v^x)^2 + (v^y)^2 + (v^z)^2.$$

Even the components (2.2) of 4-velocity may seem slightly unfamiliar if the reader is accustomed to having the fourth component of a vector be multiplied by a factor $i = \sqrt{-1}$. If so, he must adjust himself to new notation. (See "Farewell to '*ict*,'" Box 2.1.)

More fundamental than the components of a vector is the vector itself. It is a geometric object with a meaning independent of all coordinates. Thus a particle has a world line $\mathcal{P}(\tau)$, and a 4-velocity $\boldsymbol{u} = d\mathcal{P}/d\tau$, that have nothing to do with any coordinates. Coordinates enter the picture when analysis on a computer is required (rejects vectors; accepts numbers). For this purpose one adopts a Lorentz frame with orthonormal basis vectors (Figure 2.3) $\boldsymbol{e}_0$, $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, and $\boldsymbol{e}_3$. Relative to the origin $\mathcal{O}$ of this frame, the world line has a coordinate description

Basis vectors

$$\mathcal{P}(\tau) - \mathcal{O} = x^0(\tau)\boldsymbol{e}_0 + x^1(\tau)\boldsymbol{e}_1 + x^2(\tau)\boldsymbol{e}_2 + x^3(\tau)\boldsymbol{e}_3 = x^\mu(\tau)\boldsymbol{e}_\mu.$$

Expressed relative to the same Lorentz frame, the 4-velocity of the particle is

$$\boldsymbol{u} = d\mathcal{P}/d\tau = (dx^\mu/d\tau)\boldsymbol{e}_\mu = u^0\boldsymbol{e}_0 + u^1\boldsymbol{e}_1 + u^2\boldsymbol{e}_2 + u^3\boldsymbol{e}_3. \tag{2.3}$$

**Box 2.1 FAREWELL TO *"ict"***

One sometime participant in special relativity will have to be put to the sword: "$x^4 = ict$." This imaginary coordinate was invented to make the geometry of spacetime look formally as little different as possible from the geometry of Euclidean space; to make a Lorentz transformation look on paper like a rotation; and to spare one the distinction that one otherwise is forced to make between quantities with upper indices (such as the components $p^\mu$ of the energy-momentum vector) and quantities with lower indices (such as the components $p_\mu$ of the energy-momentum 1-form). However, it is no kindness to be spared this latter distinction. Without it, one cannot know whether a vector (§2.3) is meant or the very different geometric object that is a 1-form (§2.5). Moreover, there is a significant difference between an angle on which everything depends periodically (a rotation) and a parameter the increase of which gives rise to ever-growing momentum differences (the "velocity parameter" of a Lorentz transformation; Box 2.4). If the imaginary time-coordinate hides from view the character of the geometric object being dealt with and the nature of the parameter in a transformation, it also does something even more serious: it hides the completely different metric structure (§2.4) of $+++$ geometry and $-+++$ geometry. In Euclidean geometry, when the distance between two points is zero, the two points must be the same point. In Lorentz-Minkowski geometry, when the interval between two events is zero, one event may be on Earth and the other on a supernova in the galaxy M31, but their separation must be a null ray (piece of a light cone). The backward-pointing light cone at a given event contains all the events by which that event can be influenced. The forward-pointing light cone contains all events that it can influence. The multitude of double light cones taking off from all the events of spacetime forms an interlocking causal structure. This structure makes the machinery of the physical world function as it does (further comments on this structure in Wheeler and Feynman 1945 and 1949 and in Zeeman 1964). If in a region where spacetime is flat, one can hide this structure from view by writing

$$(\Delta s)^2 = (\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 + (\Delta x^4)^2,$$

with $x^4 = ict$, no one has discovered a way to make an imaginary coordinate work in the general curved spacetime manifold. If "$x^4 = ict$" cannot be used there, it will not be used here. In this chapter and hereafter, as throughout the literature of general relativity, a real time coordinate is used, $x^0 = t = ct_{\text{conv}}$ (superscript 0 rather than 4 to avoid any possibility of confusion with the imaginary time coordinate).

The components $w^\alpha$ of any other vector $\boldsymbol{w}$ in this frame are similarly defined as the coefficients in such an expansion,

Expansion of vector in terms of basis

$$\boldsymbol{w} = w^\alpha \boldsymbol{e}_\alpha. \tag{2.4}$$

Notice: the subscript $\alpha$ on $\boldsymbol{e}_\alpha$ tells which vector, not which component!

## §2.4. THE METRIC TENSOR

The metric tensor, one recalls from part IV of Box 1.3, is a machine for calculating the squared length of a single vector, or the scalar product of two different vectors.

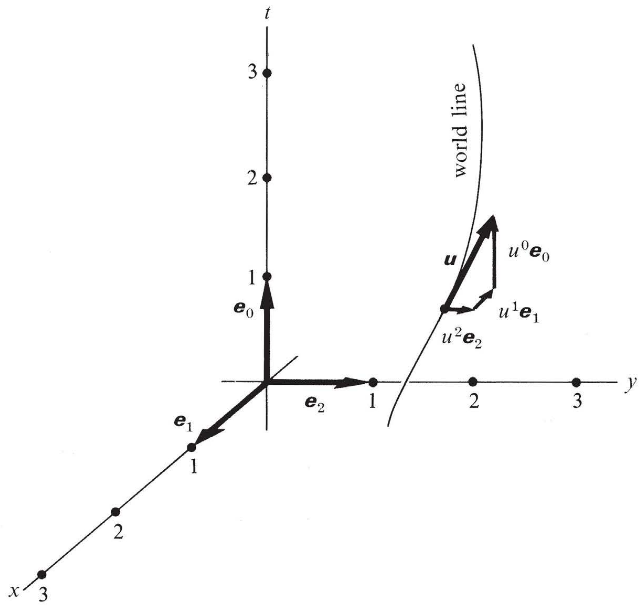



**Figure 2.3.**
The 4-velocity of a particle in flat spacetime. The 4-velocity $\boldsymbol{u}$ is the unit vector (arrow) tangent to the particle's world line—one tangent vector for each event on the world line. In a specific Lorentz coordinate system, there are basis vectors of unit length, which point along the four coordinate axes: $\boldsymbol{e}_0$, $\boldsymbol{e}_1$,$\boldsymbol{e}_2$, $\boldsymbol{e}_3$. The 4-velocity, like any vector, can be expressed as a sum of components along the basis vectors:

$$\boldsymbol{u} = u^0\boldsymbol{e}_0 + u^1\boldsymbol{e}_1 + u^2\boldsymbol{e}_2 + u^3\boldsymbol{e}_3 = u^\alpha\boldsymbol{e}_\alpha.$$

Metric defined as machine for computing scalar products of vectors

More precisely, the metric tensor $\boldsymbol{g}$ is a machine with two slots for inserting vectors

$$\begin{array}{ccc} & \text{slot 1} & \text{slot 2} \\ & \downarrow & \downarrow \\ \boldsymbol{g}( & \quad , & \quad ). \end{array} \tag{2.5}$$

Upon insertion, the machine spews out a real number:

$$\begin{aligned} \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) &= \text{“scalar product of } \boldsymbol{u} \text{ and } \boldsymbol{v}\text{,” also denoted } \boldsymbol{u} \cdot \boldsymbol{v}. \\ \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{u}) &= \text{“squared length of } \boldsymbol{u}\text{,” also denoted } \boldsymbol{u}^2. \end{aligned} \tag{2.6}$$

Moreover, this number is independent of the order in which the vectors are inserted (“symmetry of metric tensor”),

$$\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{g}(\boldsymbol{v}, \boldsymbol{u}); \tag{2.7}$$

and it is linear in the vectors inserted

$$\boldsymbol{g}(a\boldsymbol{u} + b\boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{g}(\boldsymbol{w}, a\boldsymbol{u} + b\boldsymbol{v}) = a\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{w}) + b\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{w}). \tag{2.8}$$

Because the metric “machine” is linear, one can calculate its output, for any input, as follows, if one knows only what it does to the basis vectors $\boldsymbol{e}_\alpha$ of a Lorentz frame. (1) Define the symbols ("metric coefficients") $\eta_{\alpha\beta}$ by

Metric coefficients

$$\eta_{\alpha\beta} \equiv \boldsymbol{g}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta) = \boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta. \tag{2.9}$$

(2) Calculate their numerical values from the known squared length of the separation vector $\boldsymbol{\xi} = \Delta x^\alpha \boldsymbol{e}_\alpha$ between two events:

$$\begin{aligned}(\Delta s)^2 &= -(\Delta x^0)^2 + (\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 \\ &= \boldsymbol{g}(\Delta x^\alpha \boldsymbol{e}_\alpha, \Delta x^\beta \boldsymbol{e}_\beta) = \Delta x^\alpha \Delta x^\beta \boldsymbol{g}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta) \\ &= \Delta x^\alpha \Delta x^\beta \eta_{\alpha\beta} \qquad \text{for every choice of } \Delta x^\alpha \end{aligned}$$

$$\Longrightarrow \|\eta_{\alpha\beta}\| \equiv \begin{Vmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{Vmatrix} \text{ in any Lorentz frame.} \tag{2.10}$$

(3) Calculate the scalar product of any two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ from

$$\begin{aligned}\boldsymbol{u} \cdot \boldsymbol{v} &= \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{g}(u^\alpha \boldsymbol{e}_\alpha, v^\beta \boldsymbol{e}_\beta) = u^\alpha v^\beta \boldsymbol{g}(\boldsymbol{e}_\alpha, \boldsymbol{e}_\beta); \\ \boldsymbol{u} \cdot \boldsymbol{v} &= u^\alpha v^\beta \eta_{\alpha\beta} = -u^0 v^0 + u^1 v^1 + u^2 v^2 + u^3 v^3. \end{aligned} \tag{2.11}$$

Scalar products computed from components of vectors

That one can classify directions and vectors in spacetime into *"timelike"* (negative squared length), *"spacelike"* (positive squared length), and *"null"* or *"lightlike"* (zero squared length) is made possible by the negative sign on the metric coefficient $\eta_{00}$.

Box 2.2 shows applications of the above ideas and notation to two elementary problems in special relativity theory.

## §2.5. DIFFERENTIAL FORMS

Vectors and the metric tensor are geometric objects that are already familiar from Chapter 1 and from elementary courses in special relativity. Not so familiar, yet equally important, is a third geometric object: the *"differential form"* or *"1-form."*

Consider the 4-momentum $\boldsymbol{p}$ of a particle, an electron, for example. To spell out one concept of momentum, start with the 4-velocity, $\boldsymbol{u} = d\mathscr{P}/d\tau$, of this electron ("spacetime displacement per unit of proper time along a straightline approximation of the world line"). This is a vector of unit length. Multiply by the mass $m$ of the particle to obtain the *momentum vector*

$$\boldsymbol{p} = m\boldsymbol{u}.$$

The 1-form illustrated by de Broglie waves

But physics gives also quite another idea of momentum. It associates a de Broglie wave with each particle. Moreover, this wave has the most direct possible physical significance. Diffract this wave from a crystal lattice. From the pattern of diffraction, one can determine not merely the length of the de Broglie waves, but also the pattern in space made by surfaces of equal, integral phase $\phi = 7$, $\phi = 8$, $\phi = 9, \ldots$. This

**Box 2.2 WORKED EXERCISES USING THE METRIC**

*Exercise:* Show that the squared length of a test particle's 4-velocity $\boldsymbol{u}$ is $-1$.
*Solution:* In any Lorentz frame, using the components (2.2), one calculates as follows

$$\begin{aligned}\boldsymbol{u}^2 = \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{u}) = u^\alpha u^\beta \eta_{\alpha\beta} &= -(u^0)^2 + (u^1)^2 + (u^2)^2 + (u^3)^2 \\ &= -\frac{1}{1 - \mathrm{v}^2} + \frac{\mathrm{v}^2}{1 - \mathrm{v}^2} = -1.\end{aligned}$$

*Exercise:* Show that the rest mass of a particle is related to its energy and momentum by the famous equation

$$(mc^2)^2 = E^2 - (\boldsymbol{p}c)^2$$

or, equivalently (geometrized units!),

$$m^2 = E^2 - \boldsymbol{p}^2.$$

*First Solution:* The 4-momentum is defined by $\boldsymbol{p} = m\boldsymbol{u}$, where $\boldsymbol{u}$ is the 4-velocity and $m$ is the rest mass. Consequently, its squared length is

$$\begin{aligned}\boldsymbol{p}^2 &= m^2\boldsymbol{u}^2 = -m^2 \\ &= -(mu^0)^2 + m^2\boldsymbol{u}^2 = -\underset{\substack{\uparrow \\ E^2}}{\frac{m^2}{1 - \mathrm{v}^2}} + \underset{\substack{\uparrow \\ \boldsymbol{p}^2}}{\frac{m^2\mathrm{v}^2}{1 - \mathrm{v}^2}}.\end{aligned}$$

*Second Solution:* In the frame of the observer, where $E$ and $\boldsymbol{p}$ are measured, the 4-momentum splits into time and space parts as

$$p^0 = E, \qquad p^1\boldsymbol{e}_1 + p^2\boldsymbol{e}_2 + p^3\boldsymbol{e}_3 = \boldsymbol{p};$$

hence, its squared length is

$$\boldsymbol{p}^2 = -E^2 + \boldsymbol{p}^2.$$

But in the particle's rest frame, $\boldsymbol{p}$ splits as

$$p^0 = m, \qquad p^1 = p^2 = p^3 = 0;$$

hence, its squared length is $\boldsymbol{p}^2 = -m^2$. But the squared length is a geometric object defined independently of any coordinate system; so it must be the same by whatever means one calculates it:

$$-\boldsymbol{p}^2 = m^2 = E^2 - \boldsymbol{p}^2.$$

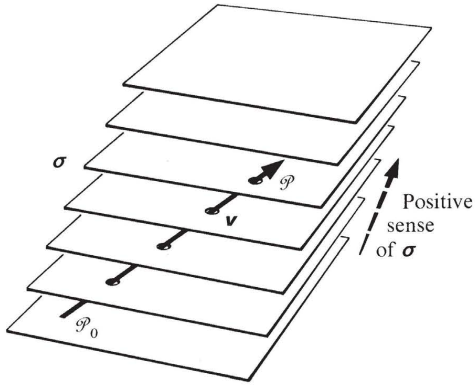



**Figure 2.4.**
The vector separation $\boldsymbol{v} = \mathcal{P} - \mathcal{P}_0$ between two neighboring events $\mathcal{P}_0$ and $\mathcal{P}$; a 1-form $\boldsymbol{\sigma}$; and the piercing of $\boldsymbol{\sigma}$ by $\boldsymbol{v}$ to give the number

$$\langle \boldsymbol{\sigma}, \boldsymbol{v} \rangle = (\text{number of surfaces pierced}) = 4.4$$

(4.4 "bongs of bell"). When $\boldsymbol{\sigma}$ is made of surfaces of constant phase, $\phi = 17$, $\phi = 18$, $\phi = 19, \ldots$ of the de Broglie wave for an electron, then $\langle \boldsymbol{\sigma}, \boldsymbol{v} \rangle$ is the phase difference between the events $\mathcal{P}_0$ and $\mathcal{P}$. Note that $\boldsymbol{\sigma}$ is not fully specified by its surfaces; an orientation is also necessary. Which direction from surface to surface is "positive"; i.e., in which direction does $\phi$ increase?

pattern of surfaces, given a name "$\tilde{\boldsymbol{k}}$," provides the simplest illustration one can easily find for a 1-form.

The pattern of surfaces in spacetime made by such a 1-form: what is it good for? Take two nearby points in spacetime, $\mathcal{P}$ and $\mathcal{P}_0$. Run an arrow $\boldsymbol{v} = \mathcal{P} - \mathcal{P}_0$ from $\mathcal{P}_0$ to $\mathcal{P}$. It will pierce a certain number of the de Broglie wave's surfaces of integral phase, with a bong of an imaginary bell at each piercing. The number of surfaces pierced (number of "bongs of bell") is denoted

Vector pierces 1-form

$$\langle \underset{\text{1-form pierced}\,\longrightarrow\uparrow}{\tilde{\boldsymbol{k}}}, \underset{\uparrow\!\!\longrightarrow\text{vector that pierces}}{\boldsymbol{v}} \rangle;$$

in this example it equals the phase difference between tail ($\mathcal{P}_0$) and tip ($\mathcal{P}$) of $\boldsymbol{v}$,

$$\langle \tilde{\boldsymbol{k}}, \boldsymbol{v} \rangle = \phi(\mathcal{P}) - \phi(\mathcal{P}_0).$$

See Figure 2.4.

Normally neither $\mathcal{P}_0$ nor $\mathcal{P}$ will lie at a point of integral phase. Therefore one can and will imagine, as uniformly interpolated between the surfaces of integral phase, an infinitude of surfaces with all the intermediate phase values. With their aid, the precise value of $\langle \tilde{\boldsymbol{k}}, \boldsymbol{v} \rangle = \phi(\mathcal{P}) - \phi(\mathcal{P}_0)$ can be determined.

To make the mathematics simple, regard $\tilde{\boldsymbol{k}}$ not as the global pattern of de Broglie-wave surfaces, but as a local pattern near a specific point in spacetime. Just as the vector $\boldsymbol{u} = d\mathcal{P}/d\tau$ represents the local behavior of a particle's world line (linear approximation to curved line in general), so the 1-form $\tilde{\boldsymbol{k}}$ represents the local form

The 1-form viewed as family of flat, equally spaced surfaces

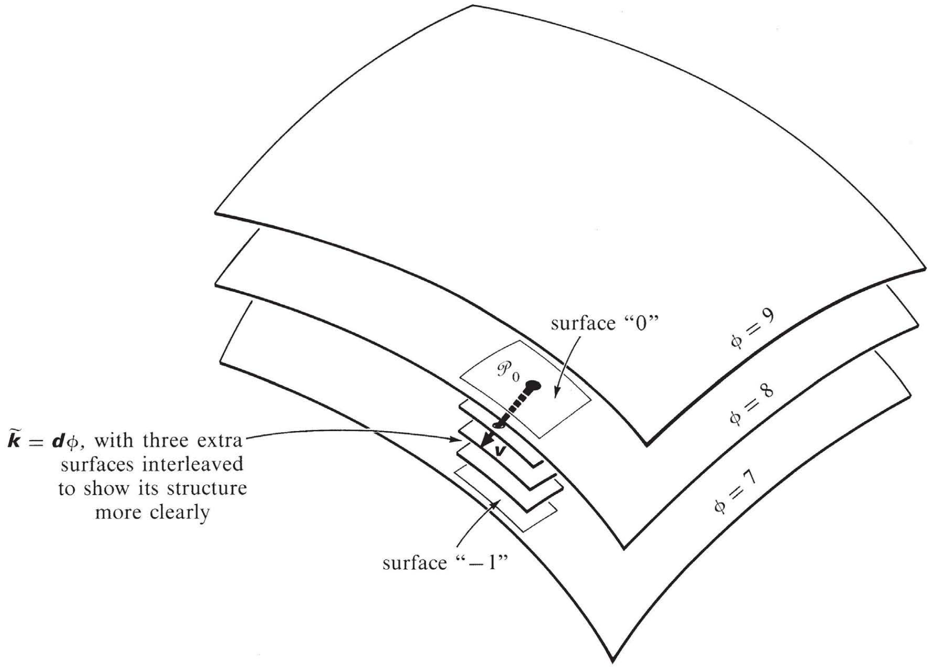



**Figure 2.5.**
This is a dual-purpose figure. (a) It illustrates the de Broglie wave 1-form $\tilde{\boldsymbol{k}}$ at an event $\mathscr{P}_0$ (family of equally spaced, flat surfaces, or "hyperplanes" approximating the surfaces of constant phase). (b) It illustrates the gradient $\boldsymbol{d}\phi$ of the function $\phi$ (concept defined in §2.6), which is the same oriented family of flat surfaces

$$\tilde{\boldsymbol{k}} = \boldsymbol{d}\phi.$$

At different events, $\tilde{\boldsymbol{k}} = \boldsymbol{d}\phi$ is different—different orientation of surfaces and different spacing. The change in $\phi$ between the tail and tip of the very short vector $\boldsymbol{v}$ is equal to the number of surfaces of $\boldsymbol{d}\phi$ pierced by $\boldsymbol{v}$, $\langle \boldsymbol{d}\phi, \boldsymbol{v}\rangle$; it equals $-0.5$ in this figure.

of the de Broglie wave's surfaces (linear approximation; surfaces flat and equally spaced; see Figure 2.5).

Regard the 1-form $\tilde{\boldsymbol{k}}$ as a machine into which vectors are inserted, and from which numbers emerge. Insertion of $\boldsymbol{v}$ produces as output $\langle \tilde{\boldsymbol{k}}, \boldsymbol{v}\rangle$. Since the surfaces of $\tilde{\boldsymbol{k}}$ are flat and equally spaced, the output is a linear function of the input:

$$\langle \tilde{\boldsymbol{k}}, a\boldsymbol{u} + b\boldsymbol{v}\rangle = a\langle \tilde{\boldsymbol{k}}, \boldsymbol{u}\rangle + b\langle \tilde{\boldsymbol{k}}, \boldsymbol{v}\rangle. \tag{2.12a}$$

The 1-form viewed as linear function of vectors

This, in fact, is the mathematical definition of a 1-form: *a 1-form is a linear, real-valued function of vectors;* i.e., a linear machine that takes in a vector and puts out a number. Given the machine $\tilde{\boldsymbol{k}}$, it is straightforward to draw the corresponding surfaces in spacetime. Pick a point $\mathscr{P}_0$ at which the machine is to reside. The surface of $\tilde{\boldsymbol{k}}$ that passes through $\mathscr{P}_0$ contains points $\mathscr{P}$ for which $\langle \tilde{\boldsymbol{k}}, \mathscr{P} - \mathscr{P}_0\rangle = 0$ (no bongs of bell). The other surfaces contain points with $\langle \tilde{\boldsymbol{k}}, \mathscr{P} - \mathscr{P}_0\rangle = \pm 1, \pm 2, \pm 3, \ldots .$

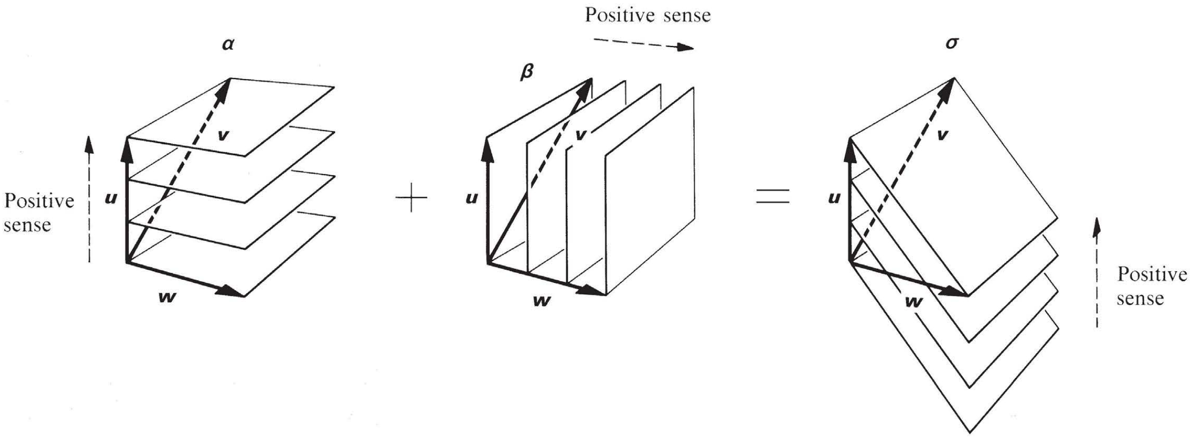



**Figure 2.6.**
The addition of two 1-forms, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, to produce the 1-form $\boldsymbol{\sigma}$. Required is a pictorial construction that starts from the surfaces of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, e.g., $\langle \boldsymbol{\alpha}, \mathscr{P} - \mathscr{P}_0 \rangle = \cdots -1, 0, 1, 2, \ldots$, and constructs those of $\boldsymbol{\sigma} = \boldsymbol{\alpha} + \boldsymbol{\beta}$. Such a construction, based on linearity (2.12b) of the addition process, is as follows. (1) Pick several vectors $\boldsymbol{u}, \boldsymbol{v}, \ldots$ that lie parallel to the surfaces of $\boldsymbol{\beta}$ (no piercing!), but pierce precisely 3 surfaces of $\boldsymbol{\alpha}$; each of these must then pierce precisely 3 surfaces of $\boldsymbol{\sigma}$:

$$\langle \boldsymbol{\sigma}, \boldsymbol{u} \rangle = \langle \boldsymbol{\alpha} + \boldsymbol{\beta}, \boldsymbol{u} \rangle = \langle \boldsymbol{\alpha}, \boldsymbol{u} \rangle = 3.$$

(2) Pick several other vectors $\boldsymbol{w}, \ldots$ that lie parallel to the surfaces of $\boldsymbol{\alpha}$ but pierce precisely 3 surfaces of $\boldsymbol{\beta}$; these will also pierce precisely 3 surfaces of $\boldsymbol{\sigma}$. (3) Construct that unique family of equally spaced surfaces in which $\boldsymbol{u}, \boldsymbol{v}, \ldots, \boldsymbol{w}, \ldots$ all have their tails on one surface and their tips on the third succeeding surface.

Sometimes 1-forms are denoted by boldface, sans-serif Latin letters with tildes over them, e.g., $\tilde{\boldsymbol{k}}$; but more often by boldface Greek letters, e.g., $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}$. The output of a 1-form $\boldsymbol{\sigma}$, when a vector $\boldsymbol{u}$ is inserted, is called *"the value of $\boldsymbol{\sigma}$ on $\boldsymbol{u}$"* or *"the contraction of $\boldsymbol{\sigma}$ with $\boldsymbol{u}$."*

Addition of 1-forms

Also, 1-forms, like any other kind of function, can be added. The 1-form $a\boldsymbol{\alpha} + b\boldsymbol{\beta}$ is that machine (family of surfaces) which puts out the following number when a vector $\boldsymbol{u}$ is put in:

$$\langle a\boldsymbol{\alpha} + b\boldsymbol{\beta}, \boldsymbol{u} \rangle = a\langle \boldsymbol{\alpha}, \boldsymbol{u} \rangle + b\langle \boldsymbol{\beta}, \boldsymbol{u} \rangle. \tag{2.12b}$$

Figure 2.6 depicts this addition in terms of surfaces.

One can verify that the set of all 1-forms at a given event is a "vector space" in the abstract, algebraic sense of the term.

Return to a particle and its de Broglie wave. Just as the arrow $\boldsymbol{p} = m d\mathscr{P}/d\tau$ represents the best *linear* approximation to the particle's actual world line near $\mathscr{P}_0$, so the flat surfaces of the 1-form $\tilde{\boldsymbol{k}}$ provide the best linear approximation to the curved surfaces of the particle's de Broglie wave, and $\tilde{\boldsymbol{k}}$ itself is the *linear function* that best approximates the de Broglie phase $\phi$ near $\mathscr{P}_0$:

$$\begin{aligned} \phi(\mathscr{P}) = \phi(\mathscr{P}_0) &+ \langle \tilde{\boldsymbol{k}}, \mathscr{P} - \mathscr{P}_0 \rangle \\ &+ \text{terms of higher order in } (\mathscr{P} - \mathscr{P}_0). \end{aligned} \tag{2.13}$$

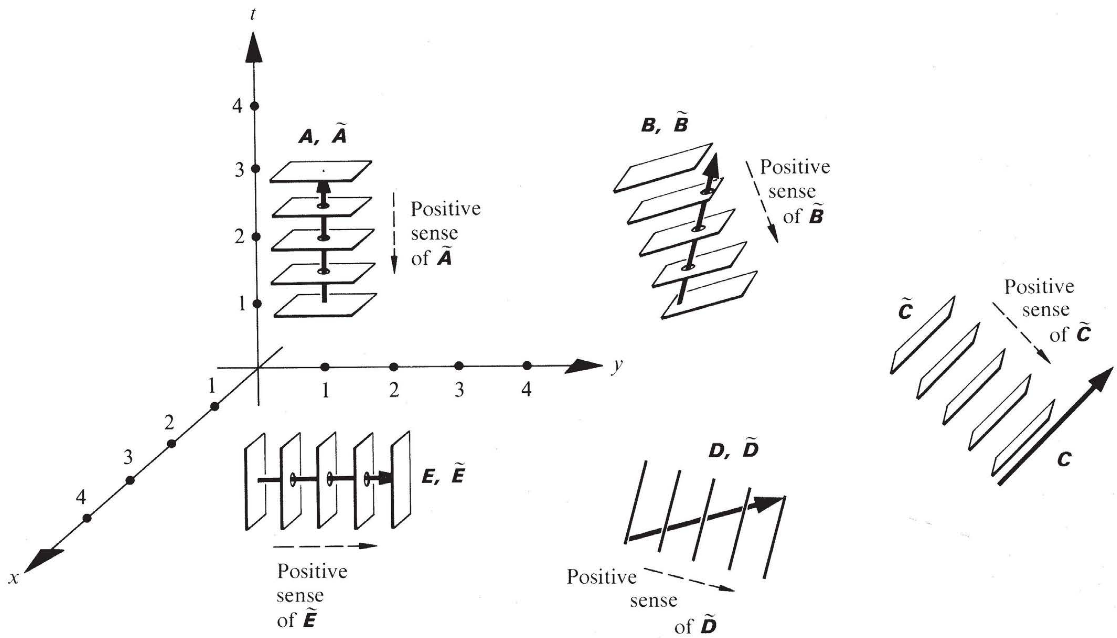



**Figure 2.7.**
Several vectors, $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, $\boldsymbol{E}$, and corresponding 1-forms $\tilde{\boldsymbol{A}}$, $\tilde{\boldsymbol{B}}$, $\tilde{\boldsymbol{C}}$, $\tilde{\boldsymbol{D}}$, $\tilde{\boldsymbol{E}}$. The process of drawing $\tilde{\boldsymbol{U}}$ corresponding to a given vector $\boldsymbol{U}$ is quite simple. (1) Orient the surfaces of $\tilde{\boldsymbol{U}}$ orthogonal to the vector $\boldsymbol{U}$. (Why? Because any vector $\boldsymbol{V}$ that is perpendicular to $\boldsymbol{U}$ must pierce no surfaces of $\tilde{\boldsymbol{U}}$ $(0 = \boldsymbol{U} \cdot \boldsymbol{V} = \langle \tilde{\boldsymbol{U}}, \boldsymbol{V} \rangle)$ and must therefore lie in a surface of $\tilde{\boldsymbol{U}}$.) (2) Space the surfaces of $\tilde{\boldsymbol{U}}$ so that the number of surfaces pierced by some arbitrary vector $\boldsymbol{Y}$ (e.g., $\boldsymbol{Y} = \boldsymbol{U}$) is equal to $\boldsymbol{Y} \cdot \boldsymbol{U}$.

Note that in the figure the surfaces of $\tilde{\boldsymbol{B}}$ are, indeed, orthogonal to $\boldsymbol{B}$; those of $\tilde{\boldsymbol{C}}$ are, indeed, orthogonal to $\boldsymbol{C}$, etc. If they do not look so, that is because the reader is attributing Euclidean geometry, not Lorentz geometry, to the spacetime diagram. He should recall, for example, that because $\boldsymbol{C}$ is a null vector, it is orthogonal to itself ($\boldsymbol{C} \cdot \boldsymbol{C} = 0$), so it must itself lie in a surface of the 1-form $\tilde{\boldsymbol{C}}$. Confused readers may review spacetime diagrams in a more elementary text, e.g., Taylor and Wheeler (1966).

Physical correspondence between 1-forms and vectors

Actually, the de Broglie 1-form $\tilde{\boldsymbol{k}}$ and the momentum vector $\boldsymbol{p}$ contain precisely the same information, both physically (via quantum theory) and mathematically. To see their relationship. relabel the surfaces of $\tilde{\boldsymbol{k}}$ by $\hbar \times$ phase, thereby obtaining the *"momentum 1-form"* $\tilde{\boldsymbol{p}}$. Pierce this 1-form with any vector $\boldsymbol{v}$, and find the result (exercise 2.1) that

$$\boldsymbol{p} \cdot \boldsymbol{v} = \langle \tilde{\boldsymbol{p}}, \boldsymbol{v} \rangle. \tag{2.14}$$

In words: the projection of $\boldsymbol{v}$ on the 4-momentum vector $\boldsymbol{p}$ equals the number of surfaces it pierces in the 4-momentum 1-form $\tilde{\boldsymbol{p}}$. *Examples:* Vectors $\boldsymbol{v}$ lying in a surface of $\tilde{\boldsymbol{p}}$ (no piercing) are perpendicular to $\boldsymbol{p}$ (no projection); $\boldsymbol{p}$ itself pierces $\boldsymbol{p}^2 = -m^2$ surfaces of $\tilde{\boldsymbol{p}}$.

Mathematical correspondence between 1-forms and vectors

Corresponding to any vector $\boldsymbol{p}$ there exists a unique 1-form (linear function of vectors) $\tilde{\boldsymbol{p}}$ defined by equation (2.14). And corresponding to any 1-form $\tilde{\boldsymbol{p}}$, there exists a unique vector $\boldsymbol{p}$ defined by its projections on all other vectors, by equation (2.14). Figure 2.7 shows several vectors and their corresponding 1-forms.

A single physical quantity can be described equally well by a vector $\boldsymbol{p}$ or by the corresponding 1-form $\tilde{\boldsymbol{p}}$. Sometimes the vector description is the simplest and most natural; sometimes the 1-form description is nicer. *Example:* Consider a 1-form representing the march of Lorentz coordinate time toward the future—surfaces $x^0 = \ldots, 7, 8, 9, \ldots$. The corresponding vector points toward the past [see Figure 2.7 or equation (2.14)]; its description of the forward march of time is not so nice!

One often omits the tilde from the 1-form $\tilde{\boldsymbol{p}}$ corresponding to a vector $\boldsymbol{p}$, and uses the same symbol $\boldsymbol{p}$ for both. Such practice is justified by the unique correspondence (both mathematical and physical) between $\tilde{\boldsymbol{p}}$ and $\boldsymbol{p}$.

---

**EXERCISE**

**Exercise 2.1.**

Show that equation (2.14) is in accord with the quantum-mechanical properties of a de Broglie wave,

$$\psi = e^{i\phi} = \exp\left[i(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)\right].$$

---

## §2.6. GRADIENTS AND DIRECTIONAL DERIVATIVES

Gradient of a function as a 1-form

There is no simpler 1-form than the *gradient,* "$\boldsymbol{d}f$," of a function $f$. Gradient a 1-form? How so? Hasn't one always known the gradient as a vector? Yes, indeed, but only because one was not familiar with the more appropriate 1-form concept. The more familiar gradient is the vector corresponding, via equation (2.14), to the 1-form gradient. The hyperplanes representing $\boldsymbol{d}f$ at a point $\mathscr{P}_0$ are just the level surfaces of $f$ itself, except for flattening and adjustment to equal spacing (Figure 2.5; identify $f$ here with $\phi$ there). More precisely, they are the level surfaces of the linear function that approximates $f$ in an infinitesimal neighborhood of $\mathscr{P}_0$.

Why the name "gradient"? Because $\boldsymbol{d}f$ describes the first order changes in $f$ in the neighborhood of $\mathscr{P}_0$:

$$f(\mathscr{P}) = f(\mathscr{P}_0) + \langle \boldsymbol{d}f, \mathscr{P} - \mathscr{P}_0\rangle + (\text{nonlinear terms}). \tag{2.15}$$

[Compare the fundamental idea of "derivative" of something as "best linear approximation to that something at a point"—an idea that works even for functions whose values and arguments are infinite dimensional vectors! See, e.g., Dieudonné (1960).]

Take any vector $\boldsymbol{v}$; construct the curve $\mathscr{P}(\lambda)$ defined by $\mathscr{P}(\lambda) - \mathscr{P}_0 = \lambda\boldsymbol{v}$; and differentiate the function $f$ along this curve:

$$\partial_{\boldsymbol{v}} f = (d/d\lambda)_{\lambda=0} f[\mathscr{P}(\lambda)] = (df/d\lambda)_{\mathscr{P}_0}. \tag{2.16a}$$

The "differential operator,"

$$\partial_{\boldsymbol{v}} = (d/d\lambda)_{\text{at } \lambda=0,\text{ along curve } \mathscr{P}(\lambda) - \mathscr{P}_0 = \lambda\boldsymbol{v}}, \tag{2.16b}$$

Directional derivative operator defined

which does this differentiating, is called the *"directional derivative operator along the vector* $\boldsymbol{v}$." The directional derivative $\partial_{\boldsymbol{v}} f$ and the gradient $\boldsymbol{d}f$ are intimately related, as one sees by applying $\partial_{\boldsymbol{v}}$ to equation (2.15) and evaluating the result at the point $\mathscr{P}_0$:

$$\partial_{\boldsymbol{v}} f = \langle \boldsymbol{d}f, d\mathscr{P}/d\lambda \rangle = \langle \boldsymbol{d}f, \boldsymbol{v} \rangle. \tag{2.17}$$

This result, expressed in words, is: $\boldsymbol{d}f$ is a linear machine for computing the rate of change of $f$ along any desired vector $\boldsymbol{v}$. Insert $\boldsymbol{v}$ into $\boldsymbol{d}f$; the output ("number of surfaces pierced; number of bongs of bell") is $\partial_{\boldsymbol{v}} f$—which, for sufficiently small $\boldsymbol{v}$, is simply the difference in $f$ between tip and tail of $\boldsymbol{v}$.

## §2.7. COORDINATE REPRESENTATION OF GEOMETRIC OBJECTS

In flat spacetime, special attention focuses on Lorentz frames. The coordinates $x^0(\mathscr{P})$, $x^1(\mathscr{P})$, $x^2(\mathscr{P})$, $x^3(\mathscr{P})$ of a Lorentz frame are functions; so their gradients can be calculated. Each of the resulting "basis 1-forms,"

Basis 1-forms

$$\boldsymbol{\omega}^\alpha = \boldsymbol{d}x^\alpha, \tag{2.18}$$

has as its hyperplanes the coordinate surfaces $x^\alpha = \text{const}$; see Figure 2.8. Consequently the basis vector $\boldsymbol{e}_\alpha$ pierces precisely one surface of the basis 1-form $\boldsymbol{\omega}^\alpha$,

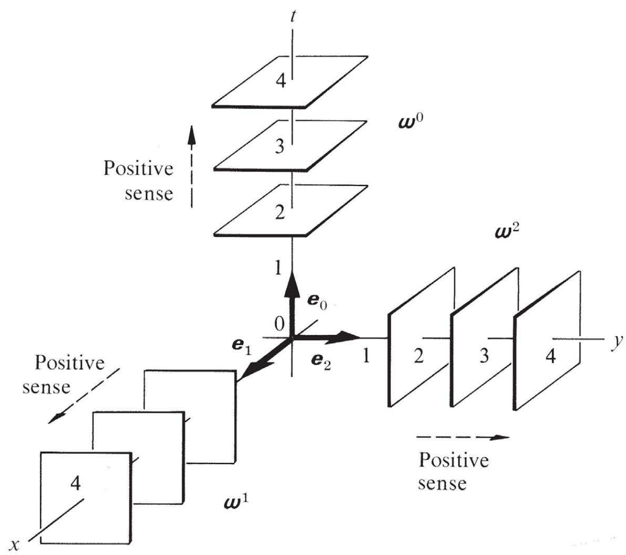



**Figure 2.8.**
The basis vectors and 1-forms of a particular Lorentz coordinate frame. The basis 1-forms are so laid out that

$$\langle \boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta \rangle = \delta^\alpha{}_\beta.$$

while the other three basis vectors lie parallel to the surfaces of $\boldsymbol{\omega}^\alpha$ and thus pierce none:

$$\langle \boldsymbol{\omega}^\alpha, \boldsymbol{e}_\beta \rangle = \delta^\alpha{}_\beta. \tag{2.19}$$

(One says that the set of basis 1-forms $\{\boldsymbol{\omega}^\alpha\}$ and the set of basis vectors $\{\boldsymbol{e}_\beta\}$ are the *"duals"* of each other if they have this property.)

Just as arbitrary vectors can be expanded in terms of the basis $\boldsymbol{e}_\alpha$, $\boldsymbol{v} = v^\alpha \boldsymbol{e}_\alpha$, so arbitrary 1-forms can be expanded in terms of $\boldsymbol{\omega}^\beta$:

Expansion of 1-form in terms of basis

$$\boldsymbol{\sigma} = \sigma_\beta \boldsymbol{\omega}^\beta. \tag{2.20}$$

The expansion coefficients $\sigma_\beta$ are called "the components of $\boldsymbol{\sigma}$ on the basis $\boldsymbol{\omega}^\beta$."

Calculation and manipulation of vector and 1-form components

These definitions produce an elegant computational formalism, thus: Calculate how many surfaces of $\boldsymbol{\sigma}$ are pierced by the basis vector $\boldsymbol{e}_\alpha$; equations (2.19) and (2.20) give the answer:

$$\langle \boldsymbol{\sigma}, \boldsymbol{e}_\alpha \rangle = \langle \sigma_\beta \boldsymbol{\omega}^\beta, \boldsymbol{e}_\alpha \rangle = \sigma_\beta \langle \boldsymbol{\omega}^\beta, \boldsymbol{e}_\alpha \rangle = \sigma_\beta \delta^\beta{}_\alpha;$$

i.e.,

$$\langle \boldsymbol{\sigma}, \boldsymbol{e}_\alpha \rangle = \sigma_\alpha. \tag{2.21a}$$

Similarly, calculate $\langle \boldsymbol{\omega}^\alpha, \boldsymbol{v} \rangle$ for any vector $\boldsymbol{v} = \boldsymbol{e}_\beta v^\beta$; the result is

$$\langle \boldsymbol{\omega}^\alpha, \boldsymbol{v} \rangle = v^\alpha. \tag{2.21b}$$

Multiply equation (2.21a) by $v^\alpha$ and sum, or multiply (2.21b) by $\sigma_\alpha$ and sum; the result in either case is

$$\langle \boldsymbol{\sigma}, \boldsymbol{v} \rangle = \sigma_\alpha v^\alpha. \tag{2.22}$$

This provides a way, using components, to calculate the coordinate-independent value of $\langle \boldsymbol{\sigma}, \boldsymbol{v} \rangle$.

Each Lorentz frame gives a coordinate-dependent representation of any geometric object or relation: $\boldsymbol{v}$ is represented by its components $v^\alpha$; $\boldsymbol{\sigma}$, by its components $\sigma_\alpha$; a point $\mathscr{P}$, by its coordinates $x^\alpha$; the relation $\langle \boldsymbol{\sigma}, \boldsymbol{v} \rangle = 17.3$ by $\sigma_\alpha v^\alpha = 17.3$.

To find the coordinate representation of the directional derivative operator $\partial_{\boldsymbol{v}}$, rewrite equation (2.16b) using elementary calculus

$$\partial_{\boldsymbol{v}} = \left(\frac{d}{d\lambda}\right)_{\mathscr{P}_0} = \underbrace{\left(\frac{dx^\alpha}{d\lambda}\right)_{\text{at } \mathscr{P}_0 \text{ along } \mathscr{P}(\lambda) - \mathscr{P}_0 = \lambda \boldsymbol{v}}}_{v^\alpha;\ \text{see equation (2.3)}} \left(\frac{\partial}{\partial x^\alpha}\right);$$

the result is

$$\partial_{\boldsymbol{v}} = v^\alpha\, \partial / \partial x^\alpha. \tag{2.23}$$

Directional derivative in terms of coordinates

In particular, the directional derivative along a basis vector $\boldsymbol{e}_\alpha$ (components $[\boldsymbol{e}_\alpha]^\beta = \langle \boldsymbol{\omega}^\beta, \boldsymbol{e}_\alpha \rangle = \delta^\beta{}_\alpha$) is

$$\partial_\alpha \equiv \partial_{\boldsymbol{e}_\alpha} = \partial / \partial x^\alpha. \tag{2.24}$$

This should also be obvious from Figure 2.8.

Components of gradient

The components of the gradient 1-form $\boldsymbol{d}f$, which are denoted $f_{,\alpha}$

$$\boldsymbol{d}f = f_{,\alpha}\boldsymbol{\omega}^{\alpha}, \tag{2.25a}$$

are calculated easily using the above formulas:

$$\begin{aligned} f_{,\alpha} &= \langle \boldsymbol{d}f, \boldsymbol{e}_{\alpha}\rangle && \text{[standard way to calculate components; equation (2.21a)]} \\ &= \partial_{\alpha} f && \text{[by relation (2.17) between directional derivative and gradient]} \\ &= \partial f/\partial x^{\alpha} && \text{[by equation (2.24)].} \end{aligned}$$

Thus, in agreement with the elementary calculus idea of gradient, the components of $\boldsymbol{d}f$ are just the partial derivatives along the coordinate axes:

$$f_{,\alpha} = \partial f/\partial x^{\alpha}; \qquad \text{i.e., } \boldsymbol{d}f = (\partial f/\partial x^{\alpha})\, \boldsymbol{d}x^{\alpha}. \tag{2.25b}$$

(Recall: $\boldsymbol{\omega}^{\alpha} = \boldsymbol{d}x^{\alpha}$.) The formula $\boldsymbol{d}f = (\partial f/\partial x^{\alpha})\, \boldsymbol{d}x^{\alpha}$ suggests, correctly, that $\boldsymbol{d}f$ is a rigorous version of the "differential" of elementary calculus; see Box 2.3.

Other important coordinate representations for geometric relations are explored in the following exercises.

---

## EXERCISES

Derive the following computationally useful formulas:

**Exercise 2.2. LOWERING INDEX TO GET THE 1-FORM CORRESPONDING TO A VECTOR**

The components $u_{\alpha}$ of the 1-form $\tilde{\boldsymbol{u}}$ that corresponds to a vector $\boldsymbol{u}$ can be obtained by "lowering an index" with the metric coefficients $\eta_{\alpha\beta}$:

$$u_{\alpha} = \eta_{\alpha\beta} u^{\beta}; \text{ i.e.,} \qquad u_0 = -u^0,\ u_k = u^k. \tag{2.26a}$$

**Exercise 2.3. RAISING INDEX TO RECOVER THE VECTOR**

One can return to the components of $\boldsymbol{u}$ by raising indices,

$$u^{\alpha} = \eta^{\alpha\beta} u_{\beta}; \tag{2.26b}$$

the matrix $\|\eta^{\alpha\beta}\|$ is defined as the inverse of $\|\eta_{\alpha\beta}\|$, and happens to equal $\|\eta_{\alpha\beta}\|$:

$$\eta^{\alpha\beta}\eta_{\beta\gamma} \equiv \delta^{\alpha}{}_{\gamma}; \qquad \eta^{\alpha\beta} = \eta_{\alpha\beta} \text{ for all } \alpha, \beta. \tag{2.27}$$

**Exercise 2.4. VARIED ROUTES TO THE SCALAR PRODUCT**

The scalar product of $\boldsymbol{u}$ with $\boldsymbol{v}$ can be calculated in any of the following ways:

$$\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = u^{\alpha} v^{\beta} \eta_{\alpha\beta} = u^{\alpha} v_{\alpha} = u_{\alpha} v_{\beta} \eta^{\alpha\beta}. \tag{2.28}$$

---

**Box 2.3 DIFFERENTIALS**

The "exterior derivative" or "gradient" $\boldsymbol{d}f$ of a function $f$ is a more rigorous version of the elementary concept of "differential."

In elementary textbooks, one is presented with the differential $df$ as representing "an infinitesimal change in the function $f(\mathscr{P})$" associated with some infinitesimal displacement of the point $\mathscr{P}$; but one will recall that the displacement of $\mathscr{P}$ is left arbitrary, albeit infinitesimal. Thus $df$ represents a change in $f$ in some unspecified direction.

But this is precisely what the exterior derivative $\boldsymbol{d}f$ represents. Choose a particular, infinitesimally long displacement $\boldsymbol{v}$ of the point $\mathscr{P}$. Let the displacement vector $\boldsymbol{v}$ pierce $\boldsymbol{d}f$ to give the number $\langle \boldsymbol{d}f, \boldsymbol{v}\rangle = \partial_{\boldsymbol{v}} f$. That number is the change of $f$ in going from the tail of $\boldsymbol{v}$ to its tip. Thus $\boldsymbol{d}f$, before it has been pierced to give a number, represents the change of $f$ in an unspecified direction. The act of piercing $\boldsymbol{d}f$ with $\boldsymbol{v}$ is the act of making explicit the direction in which the change is to be measured. The only failing of the textbook presentation, then, was its suggestion that $\boldsymbol{d}f$ was a scalar or a number; the explicit recognition of the need for specifying a direction $\boldsymbol{v}$ to reduce $\boldsymbol{d}f$ to a number $\langle \boldsymbol{d}f, \boldsymbol{v}\rangle$ shows that in fact $\boldsymbol{d}f$ is a 1-form, the gradient of $f$.

## §2.8. THE CENTRIFUGE AND THE PHOTON

Vectors, metric, 1-forms, functions, gradients, directional derivatives: all these geometric objects and more are used in flat spacetime to represent physical quantities; and all the laws of physics must be expressible in terms of such geometric objects.

Geometric objects in action: example of centrifuge and photon

As an example, consider a high-precision redshift experiment that uses the Mössbauer effect (Figure 2.9). The emitter and the absorber of photons are attached to

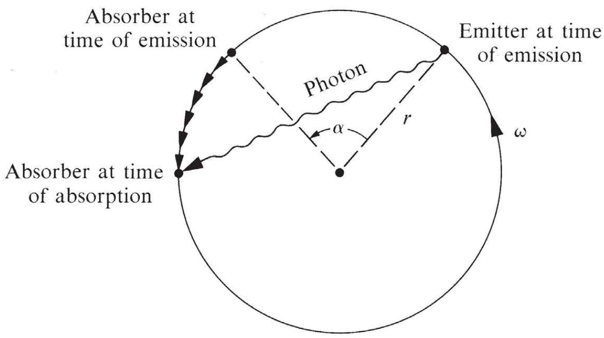



**Figure 2.9.**
The centrifuge and the photon.

the rim of a centrifuge at points separated by an angle $\alpha$, as measured in the inertial laboratory. The emitter and absorber are at radius $r$ as measured in the laboratory, and the centrifuge rotates with angular velocity $\omega$. PROBLEM: What is the redshift measured,

$$z = (\lambda_{\text{absorbed}} - \lambda_{\text{emitted}})/\lambda_{\text{emitted}},$$

in terms of $\omega$, $r$, and $\alpha$?

SOLUTION: Let $\boldsymbol{u}_e$ be the 4-velocity of the emitter at the event of emission of a given photon; let $\boldsymbol{u}_a$ be the 4-velocity of the absorber at the event of absorption; and let $\boldsymbol{p}$ be the 4-momentum of the photon. All three quantities are vectors defined without reference to coordinates. Equally coordinate-free are the photon energies $E_e$ and $E_a$ measured by emitter and absorber. No coordinates are needed to describe the fact that a specific emitter emitting a specific photon attributes to it the energy $E_e$; and no coordinates are required in the geometric formula

$$E_e = -\boldsymbol{p} \cdot \boldsymbol{u}_e \tag{2.29}$$

for $E_e$. [That this formula works can be readily verified by recalling that, in the emitter's frame, $u_e{}^0 = 1$ and $u_e{}^j = 0$; so

$$E_e = -p_\alpha u_e{}^\alpha = -p_0 = +p^0$$

in accordance with the identification "(time component of 4-momentum) = (energy."] Analogous to equation (2.29) is the purely geometric formula

$$E_a = -\boldsymbol{p} \cdot \boldsymbol{u}_a$$

for the absorbed energy.

The ratio of absorbed wavelength to emitted wavelength is the inverse of the energy ratio (since $E = h\nu = hc/\lambda$):

$$\frac{\lambda_a}{\lambda_e} = \frac{E_e}{E_a} = \frac{-\boldsymbol{p} \cdot \boldsymbol{u}_e}{-\boldsymbol{p} \cdot \boldsymbol{u}_a}.$$

This ratio is most readily calculated in the inertial laboratory frame

$$\frac{\lambda_a}{\lambda_e} = \frac{p^0 u_e{}^0 - p^j u_e{}^j}{p^0 u_a{}^0 - p^j u_a{}^j} \equiv \frac{p^0 u_e{}^0 - \boldsymbol{p} \cdot \boldsymbol{u}_e}{p^0 u_a{}^0 - \boldsymbol{p} \cdot \boldsymbol{u}_a}. \tag{2.30}$$

(Here and throughout we use boldface Latin letters for three-dimensional vectors in a given Lorentz frame; and we use the usual notation and formalism of three-dimensional, Euclidean vector analysis to manipulate them.) Because the magnitude of the ordinary velocity of the rim of the centrifuge, $v = \omega r$, is unchanging in time, $u_e{}^0$ and $u_a{}^0$ are equal, and the magnitudes—but not the directions—of $\boldsymbol{u}_e$ and $\boldsymbol{u}_a$ are equal:

$$u_e{}^0 = u_a{}^0 = (1 - v^2)^{-1/2}, \ |\boldsymbol{u}_e| = |\boldsymbol{u}_a| = v/(1 - v^2)^{1/2}.$$

From the geometry of Figure 2.9, one sees that $\boldsymbol{u}_e$ makes the same angle with $\boldsymbol{p}$ as does $\boldsymbol{u}_a$. Consequently, $\boldsymbol{p} \cdot \boldsymbol{u}_e = \boldsymbol{p} \cdot \boldsymbol{u}_a$, and $\lambda_{\text{absorbed}}/\lambda_{\text{emitted}} = 1$. *There is no redshift!*

Notice that this solution made no reference whatsoever to Lorentz transformations—they have not even been discussed yet in this book! The power of the geometric, coordinate-free viewpoint is evident!

## EXERCISES

One must have a variety of coordinate-free contacts between theory and experiment in order to use the geometric viewpoint. One such contact is the equation $E = -\boldsymbol{p} \cdot \boldsymbol{u}$ for the energy of a photon with 4-momentum $\boldsymbol{p}$, as measured by an observer with 4-velocity $\boldsymbol{u}$. Verify the following other points of contact.

### Exercise 2.5. ENERGY AND VELOCITY FROM 4-MOMENTUM

A particle of rest mass $m$ and 4-momentum $\boldsymbol{p}$ is examined by an observer with 4-velocity $\boldsymbol{u}$. Show that just as (a) the energy he measures is

$$E = -\boldsymbol{p} \cdot \boldsymbol{u}; \tag{2.31}$$

so (b) the rest mass he attributes to the particle is

$$m^2 = -\boldsymbol{p}^2; \tag{2.32}$$

(c) the momentum he measures has magnitude

$$|\boldsymbol{p}| = [(\boldsymbol{p} \cdot \boldsymbol{u})^2 + (\boldsymbol{p} \cdot \boldsymbol{p})]^{1/2}; \tag{2.33}$$

(d) the ordinary velocity $\boldsymbol{v}$ he measures has magnitude

$$|\boldsymbol{v}| = \frac{|\boldsymbol{p}|}{E}, \tag{2.34}$$

where $|\boldsymbol{p}|$ and $E$ are as given above; and (e) the 4-vector $\boldsymbol{v}$, whose components in the observer's Lorentz frame are

$$v^0 = 0, \qquad v^j = (dx^j/dt)_{\text{for particle}} = \text{ordinary velocity},$$

is given by

$$\boldsymbol{v} = \frac{\boldsymbol{p} + (\boldsymbol{p} \cdot \boldsymbol{u})\boldsymbol{u}}{-\boldsymbol{p} \cdot \boldsymbol{u}}. \tag{2.35}$$

### Exercise 2.6. TEMPERATURE GRADIENT

To each event $\mathscr{Q}$ inside the sun one attributes a temperature $T(\mathscr{Q})$, the temperature measured by a thermometer at rest in the hot gas there. Then $T(\mathscr{Q})$ is a function; no coordinates are required for its definition and discussion. A cosmic ray from outer space flies through the sun with 4-velocity $\boldsymbol{u}$. Show that, as measured by the cosmic ray's clock, the time derivative of temperature in its vicinity is

$$dT/d\tau = \partial_{\boldsymbol{u}} T = \langle \boldsymbol{d}T, \boldsymbol{u} \rangle. \tag{2.36}$$

In a local Lorentz frame inside the sun, this equation can be written

$$\frac{dT}{d\tau} = u^\alpha \frac{\partial T}{\partial x^\alpha} = \frac{1}{\sqrt{1 - \boldsymbol{v}^2}} \frac{\partial T}{\partial t} + \frac{v^j}{\sqrt{1 - \boldsymbol{v}^2}} \frac{\partial T}{\partial x^j}. \tag{2.37}$$

Why is this result reasonable?

## §2.9. LORENTZ TRANSFORMATIONS

To simplify computations, one often works with the components of vectors and 1-forms, rather than with coordinate-free language. Such component manipulations sometimes involve transformations from one Lorentz frame to another. The reader is already familiar with such Lorentz transformations; but the short review in Box 2.4 will refresh his memory and acquaint him with the notation used in this book.

Lorentz transformations: of coordinates

The key entities in the Lorentz transformation are the matrices $\|\Lambda^{\alpha'}{}_{\beta}\|$ and $\|\Lambda^{\beta}{}_{\alpha'}\|$; the first transforms coordinates from an unprimed frame to a primed frame, while the second goes from primed to unprimed

$$x^{\alpha'} = \Lambda^{\alpha'}{}_{\beta} x^{\beta}, \qquad x^{\beta} = \Lambda^{\beta}{}_{\alpha'} x^{\alpha'}. \tag{2.38}$$

Since they go in opposite directions, each of the two matrices must be the inverse of the other:

$$\Lambda^{\alpha'}{}_{\beta} \Lambda^{\beta}{}_{\gamma'} = \delta^{\alpha'}{}_{\gamma'}; \qquad \Lambda^{\beta}{}_{\alpha'} \Lambda^{\alpha'}{}_{\gamma} = \delta^{\beta}{}_{\gamma}. \tag{2.39}$$

From the coordinate-independent nature of 4-velocity, $\boldsymbol{u} = (dx^{\alpha}/d\tau)\boldsymbol{e}_{\alpha}$, one readily derives the expressions

Of basis vectors

$$\boldsymbol{e}_{\alpha'} = \boldsymbol{e}_{\beta} \Lambda^{\beta}{}_{\alpha'}, \qquad \boldsymbol{e}_{\beta} = \boldsymbol{e}_{\alpha'} \Lambda^{\alpha'}{}_{\beta} \tag{2.40}$$

for the basis vectors of one frame in terms of those of the other; and from other geometric equations, such as

$$\begin{aligned} \boldsymbol{v} &= \boldsymbol{e}_{\alpha} v^{\alpha} = \boldsymbol{e}_{\beta'} v^{\beta'}, \\ \langle \boldsymbol{\sigma}, \boldsymbol{v} \rangle &= \sigma_{\alpha} v^{\alpha} = \sigma_{\beta'} v^{\beta'}, \\ \boldsymbol{\sigma} &= \sigma_{\alpha} \boldsymbol{\omega}^{\alpha} = \sigma_{\beta'} \boldsymbol{\omega}^{\beta'}, \end{aligned}$$

one derives transformation laws

Of basis 1-forms

$$\boldsymbol{\omega}^{\alpha'} = \Lambda^{\alpha'}{}_{\beta} \boldsymbol{\omega}^{\beta}, \qquad \boldsymbol{\omega}^{\beta} = \Lambda^{\beta}{}_{\alpha'} \boldsymbol{\omega}^{\alpha'}; \tag{2.41}$$

Of components

$$v^{\alpha'} = \Lambda^{\alpha'}{}_{\beta} v^{\beta}, \qquad v^{\beta} = \Lambda^{\beta}{}_{\alpha'} v^{\alpha'}; \tag{2.42}$$

$$\sigma_{\alpha'} = \sigma_{\beta} \Lambda^{\beta}{}_{\alpha'}, \qquad \sigma_{\beta} = \sigma_{\alpha'} \Lambda^{\alpha'}{}_{\beta}. \tag{2.43}$$

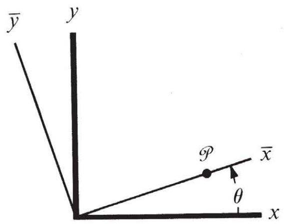
One need never memorize the index positions in these transformation laws. One need only line the indices up so that (1) free indices on each side of the equation are in the same position; and (2) summed indices appear once up and once down. Then all will be correct! (Note: the indices on $\Lambda$ always run "northwest to southeast.")

**Box 2.4 LORENTZ TRANSFORMATIONS**

### Rotation of Frame of Reference by Angle $\theta$ in $x$-$y$ Plane

$$\text{Slope } s = \tan\theta; \qquad \sin\theta = \frac{s}{(1+s^2)^{1/2}}; \qquad \cos\theta = \frac{1}{(1+s^2)^{1/2}}$$

$$\begin{aligned} t &= \bar t \\ x &= \bar x\cos\theta - \bar y\sin\theta \\ y &= \underbrace{\bar x\sin\theta} + \bar y\cos\theta \\ z &= \bar z \end{aligned}$$




$$\begin{aligned} \bar t &= t \\ \bar x &= x\cos\theta + y\sin\theta \\ \bar y &= -x\sin\theta + y\cos\theta \\ \bar z &= z \end{aligned}$$

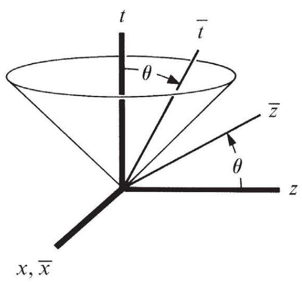
All signs follow from sign of this term. Positive by inspection of point $\mathscr{P}$.

### Combination of Two Such Rotations

$$s = \frac{s_1 + s_2}{1 - s_1 s_2} \qquad \text{or} \qquad \theta = \theta_1 + \theta_2$$

### Boost of Frame of Reference by Velocity Parameter $\alpha$ in $z$-$t$ Plane

$$\text{Velocity } \beta = \tanh\alpha; \qquad \sinh\alpha = \frac{\beta}{(1-\beta^2)^{1/2}}; \qquad \cosh\alpha = \frac{1}{(1-\beta^2)^{1/2}} = \text{“}\gamma\text{”}$$




$$\begin{aligned} \tan\theta &= \text{velocity } \beta \\ &= \tanh\alpha \end{aligned}$$

$$\begin{aligned} t &= \bar t\cosh\alpha + \bar z\sinh\alpha \\ x &= \bar x \\ y &= \bar y \\ z &= \underbrace{\bar t\sinh\alpha} + \bar z\cosh\alpha \end{aligned}$$

$$\begin{aligned} \bar t &= t\cosh\alpha - z\sinh\alpha \\ \bar x &= x \\ \bar y &= y \\ \bar z &= -t\sinh\alpha + z\cosh\alpha \end{aligned}$$

All signs follow from sign of this term. Positive because object at rest at $\bar z = 0$ in rocket frame moves in direction of increasing $z$ in lab frame.

Matrix notation: $x^\mu = \Lambda^\mu{}_{\bar\nu} x^{\bar\nu}$, $\quad x^{\bar\nu} = \Lambda^{\bar\nu}{}_\mu x^\mu$

$$\|\Lambda^\mu{}_{\bar\nu}\| = \begin{Vmatrix} \cosh\alpha & 0 & 0 & \sinh\alpha \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh\alpha & 0 & 0 & \cosh\alpha \end{Vmatrix}, \quad \|\Lambda^{\bar\nu}{}_\mu\| = \begin{Vmatrix} \cosh\alpha & 0 & 0 & -\sinh\alpha \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh\alpha & 0 & 0 & \cosh\alpha \end{Vmatrix}$$

Box 2.4 (*continued*)

Energy-momentum 4-vector

$$E = \bar{E}\cosh\alpha + p^{\bar{z}}\sinh\alpha$$
$$p^x = p^{\bar{x}}$$
$$p^y = p^{\bar{y}}$$
$$p^z = \bar{E}\sinh\alpha + p^{\bar{z}}\cosh\alpha$$

Charge density-current 4-vector

$$\rho = \bar{\rho}\cosh\alpha + j^{\bar{z}}\sinh\alpha$$
$$j^x = j^{\bar{x}}$$
$$j^y = j^{\bar{y}}$$
$$j^z = \bar{\rho}\sinh\alpha + j^{\bar{z}}\cosh\alpha$$

Aberration, incoming photon:

$$\sin\theta = \frac{-p_\perp}{E} = \frac{(1-\beta^2)^{1/2}\sin\bar{\theta}}{1-\beta\cos\bar{\theta}} \qquad \sin\bar{\theta} = \frac{-\bar{p}_\perp}{\bar{E}} = \frac{(1-\beta^2)^{1/2}\sin\theta}{1+\beta\cos\theta}$$

$$\cos\theta = \frac{-p^z}{E} = \frac{\cos\bar{\theta}-\beta}{1-\beta\cos\bar{\theta}} \qquad \cos\bar{\theta} = \frac{-p^{\bar{z}}}{\bar{E}} = \frac{\cos\theta+\beta}{1+\beta\cos\theta}$$

$$\tan(\theta/2) = e^{\alpha}\tan(\bar{\theta}/2) \qquad \tan(\bar{\theta}/2) = e^{-\alpha}\tan(\theta/2)$$

## Combination of Two Boosts in Same Direction

$$\beta = \frac{\beta_1+\beta_2}{1+\beta_1\beta_2} \qquad \text{or} \qquad \alpha = \alpha_1 + \alpha_2.$$

## General Combinations of Boosts and Rotations

Spinor formalism of Chapter 41

## Poincaré Transformation

$$x^\mu = \Lambda^\mu{}_{\alpha'}x^{\alpha'} + a^\mu.$$

Condition on the Lorentz part of this transformation:

$$ds'^2 = \eta_{\alpha'\beta'}\,dx^{\alpha'}\,dx^{\beta'} = ds^2 = \eta_{\mu\nu}\Lambda^\mu{}_{\alpha'}\Lambda^\nu{}_{\beta'}\,dx^{\alpha'}\,dx^{\beta'}$$

or $\Lambda^T\eta\Lambda = \eta$ (matrix equation, with $T$ indicating "transposed," or rows and columns interchanged).

Effect of transformation on other quantities:

| | | |
|---|---|---|
| $u^\mu = \Lambda^\mu{}_{\alpha'}u^{\alpha'}$ | (4-velocity) | $u_{\alpha'} = u_\mu\Lambda^\mu{}_{\alpha'}$; |
| $p^\mu = \Lambda^\mu{}_{\alpha'}p^{\alpha'}$ | (4-momentum) | $p_{\alpha'} = p_\mu\Lambda^\mu{}_{\alpha'}$; |
| $F^{\mu\nu} = \Lambda^\mu{}_{\alpha'}\Lambda^\nu{}_{\beta'}F^{\alpha'\beta'}$ | (electromagnetic field) | $F_{\alpha'\beta'} = F_{\mu\nu}\Lambda^\mu{}_{\alpha'}\Lambda^\nu{}_{\beta'}$; |
| $\boldsymbol{e}_{\alpha'} = \boldsymbol{e}_\mu\Lambda^\mu{}_{\alpha'}$ | (basis vectors); | |
| $\boldsymbol{\omega}^{\alpha'} = \Lambda^{\alpha'}{}_\mu\boldsymbol{\omega}^\mu$ | (basis 1-forms); | |
| $\boldsymbol{u} = \boldsymbol{e}_{\alpha'}u^{\alpha'} = \boldsymbol{e}_\mu u^\mu = \boldsymbol{u}$ | (the 4-velocity *vector*). | |

**Exercise 2.7. BOOST IN AN ARBITRARY DIRECTION**

**EXERCISE**

An especially useful Lorentz transformation has the matrix components

$$\begin{aligned}
\Lambda^{0'}{}_0 &= \gamma \equiv \frac{1}{\sqrt{1-\beta^2}},\\
\Lambda^{0'}{}_j &= \Lambda^{j'}{}_0 = -\beta\gamma n^j,\\
\Lambda^{j'}{}_k &= \Lambda^{k'}{}_j = (\gamma - 1)n^j n^k + \delta^{jk},\\
\Lambda^{\mu}{}_{\nu'} &= (\text{same as } \Lambda^{\nu'}{}_{\mu} \text{ but with } \beta \text{ replaced by } -\beta),
\end{aligned} \tag{2.44}$$

where $\beta$, $n^1$, $n^2$, and $n^3$ are parameters, and $\boldsymbol{n}^2 \equiv (n^1)^2 + (n^2)^2 + (n^3)^2 = 1$. Show (a) that this does satisfy the condition $\Lambda^T \eta \Lambda = \eta$ required of a Lorentz transformation (see Box 2.4); (b) that the primed frame moves with ordinary velocity $\beta\boldsymbol{n}$ as seen in the unprimed frame; (c) that the unprimed frame moves with ordinary velocity $-\beta\boldsymbol{n}$ (i.e., $v^{1'} = -\beta n^1$, $v^{2'} = -\beta n^2$, $v^{3'} = -\beta n^3$) as seen in the primed frame; and (d) that for motion in the $z$ direction, the transformation matrices reduce to the familiar form

$$\|\Lambda^{\nu'}{}_{\mu}\| = \begin{Vmatrix} \gamma & 0 & 0 & -\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\beta\gamma & 0 & 0 & \gamma \end{Vmatrix}, \qquad \|\Lambda^{\mu}{}_{\nu'}\| = \begin{Vmatrix} \gamma & 0 & 0 & \beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \beta\gamma & 0 & 0 & \gamma \end{Vmatrix}. \tag{2.45}$$

## §2.10. COLLISIONS

Whatever the physical entity, whether it is an individual mass in motion, or a torrent of fluid, or a field of force, or the geometry of space itself, it is described in classical general relativity as a geometric object of its own characteristic kind. Each such object is built directly or by abstraction from identifiable points, and needs no coordinates for its representation. It has been seen how this coordinate-free description translates into, and how it can be translated out of, the language of coordinates and components, and how components in a local Lorentz frame transform under a Lorentz transformation. Turn now to two elementary applications of this mathematical machinery to a mass in motion. One has to do with short-range forces (collisions, this section); the other, with the long-range electromagnetic force (Lorentz force law, next chapter).

Scattering of particles

In a collision, all the change in momentum is concentrated in a time that is short compared to the time of observation. Moreover, the target is typically so small, and quantum mechanics so dominating, that a probabilistic description is the right one. A quantity

$$d\sigma = \left(\frac{d\sigma}{d\Omega}\right)_\theta d\Omega \tag{2.46}$$

gives the cross section ($\text{cm}^2$) for scattering into the element of solid angle $d\Omega$ at the deflection angle $\theta$; a more complicated expression gives the probability that the original particle will enter the aperture $d\Omega$ at a given polar angle $\theta$ and azimuth $\phi$ and with energy $E$ to $E + dE$, while simultaneously products of reaction also emerge into specified energy intervals and into specified angular apertures. It would be out of place here to enter into the calculation of such cross sections, though it is a fascinating branch of atomic physics. It is enough to note that the cross section is an area oriented perpendicular to the line of travel of the incident particle. Therefore it is unaffected by any boost of the observer in that direction, provided of course that energies and angles of emergence of the particles are transformed in accordance with the magnitude of that boost ("same events seen in an altered reference system").

Conservation of energy-momentum in a collision

Over and above any such detailed account of the encounter as follows from the local dynamic analysis, there stands the law of conservation of energy-momentum:

$$\sum_{\substack{\text{original}\\ \text{particles, } J}} \boldsymbol{p}_J = \sum_{\substack{\text{final}\\ \text{particles, } K}} \boldsymbol{p}_K \,. \tag{2.47}$$

Out of this relation, one wins without further analysis such simple results as the following. (1) A photon traveling as a plane wave through empty space cannot split (not true for a focused photon!). (2) When a high-energy electron strikes an electron at rest in an elastic encounter, and the two happen to come off sharing the energy equally, then the angle between their directions of travel is less than the Newtonian value of 90°, and the deficit gives a simple measure of the energy of the primary. (3) When an electron makes a head-on elastic encounter with a proton, the formula for the fraction of kinetic energy transferred has three rather different limiting forms, according to whether the energy of the electron is nonrelativistic, relativistic, or extreme-relativistic. (4) The threshold for the production of an $(e^+, e^-)$ pair by a photon in the field of force of a massive nucleus is $2m_e$. (5) The threshold for the production of an $(e^+, e^-)$ pair by a photon in an encounter with an electron at rest is $4m_e$ (or $4m_e - \epsilon$ when account is taken of the binding of the $e^+e^-e^-$ system in a very light "molecule"). All these results (topics for independent projects!) and more can be read out of the law of conservation of energy-momentum. For more on this topic, see Blaton (1950), Hagedorn (1964), and Chapter 4 and the last part of Chapter 5 of Sard (1970).

# CHAPTER 3
# THE ELECTROMAGNETIC FIELD

*The rotating armatures of every generator and every motor in this age of electricity are steadily proclaiming the truth of the relativity theory to all who have ears to hear.*

LEIGH PAGE (1941)

## §3.1. THE LORENTZ FORCE AND THE ELECTROMAGNETIC FIELD TENSOR

At the opposite extreme from an impulsive change of momentum in a collision (the last topic of Chapter 2) is the gradual change in the momentum of a charged particle under the action of electric and magnetic forces (the topic treated here).

Lorentz force as definer of fields and predicter of motions

Let electric and magnetic fields act on a system of charged particles. The accelerations of the particles reveal the electric and magnetic field strengths. In other words, the Lorentz force law, plus measurements on the components of acceleration of test particles, can be viewed as defining the components of the electric and magnetic fields. Once the field components are known from the accelerations of a few test particles, they can be used to predict the accelerations of other test particles (Box 3.1). Thus the Lorentz force law does double service (1) as definer of fields and (2) as predicter of motions.

*Here and elsewhere in science, as stressed not least by Henri Poincaré, that view is out of date which used to say, "Define your terms before you proceed." All the laws and theories of physics, including the Lorentz force law, have this deep and subtle character, that they both define the concepts they use (here* ***B*** *and* ***E****) and make statements about these concepts. Contrariwise, the absence of some body of theory, law, and principle deprives one of the means properly to define or even to use concepts. Any forward step in human knowledge is truly creative in this sense: that theory, concept, law, and method of measurement—forever inseparable—are born into the world in union.*

**Box 3.1 LORENTZ FORCE LAW AS BOTH DEFINER OF FIELDS AND PREDICTER OF MOTIONS**

How one goes about determining the components of the field from measurements of accelerations is not different in principle for electromagnetism and for gravitation. Compare the equations in the two cases:

$$\frac{d^2x^\alpha}{d\tau^2} = \frac{e}{m} F^\alpha{}_\beta u^\beta \text{ in a Lorentz frame,} \tag{1}$$

and

$$\frac{D^2\xi^\alpha}{d\tau^2} = -R^\alpha{}_{\beta\gamma\delta} u^\beta \xi^\gamma u^\delta \text{ in any coordinate system.} \tag{2}$$

To make explicit the simpler procedure for electromagnetism will indicate in broad outline how one similarly determines all the components of $R^\alpha{}_{\beta\gamma\delta}$ for gravity. Begin by asking how many test particles one needs to determine the three components of $\boldsymbol{B}$ and the three components of $\boldsymbol{E}$ in the neighborhood under study. For one particle, three components of acceleration are measurable; for a second particle, three more. Enough? No! The information from the one duplicates in part the information from the other. The proof? Whatever the state of motion of the first test particle, pick one's Lorentz frame to be moving the same way. Having zero velocity in this frame, the particle has a zero response to any magnetic field. The electric field alone acts on the particle. The three components of its acceleration give directly the three components $E_x$, $E_y$, $E_z$ of the electric field. The second test particle cannot be at rest if it is to do more than duplicate the information provided by the first test particle. Orient the $x$-axis of the frame of reference parallel to the direction of motion of this second particle, which will then respond to and measure the components $B_y$ and $B_z$ of the magnetic field. Not so $B_x$! The acceleration in the $x$-direction merely remeasures the already once measured $E_x$. To evaluate $B_x$, a third test particle is required, but it then gives duplicate information about the other field components. The alternative? Use all $N$ particles simultaneously and on the same democratic footing, both in the evaluation of the six $F_{\alpha\beta}$ and in the testing of the Lorentz force, by applying the method of least squares. Thus, write the discrepancy between predicted and observed acceleration of the $K$th particle in the form

$$\dot{u}_\alpha{}^K - \frac{e}{m} F_{\alpha\beta} u^{\beta,K} = \delta a_\alpha{}^K. \tag{3}$$

Take the squared magnitude of this discrepancy and sum over all the particles

$$S = \sum_k \eta^{\alpha\beta} \delta a_\alpha{}^K \delta a^K{}_\beta. \tag{4}$$

In this expression, everything is regarded as known except the six $F_{\alpha\beta}$. Minimize with respect to these six unknowns. In this way, arrive at six equations for the components of $\boldsymbol{B}$ and $\boldsymbol{E}$. These equations once solved, one goes back to (3) to test the Lorentz force law.

The $6 \times 6$ determinant of the coefficients in the equation for the $F_{\alpha\beta}$ automatically vanishes when there are only two test particles. The same line of reasoning permits one to determine the minimum number of test particles required to determine all the components of the Riemann curvature tensor.

The Lorentz force law, written in familiar three-dimensional notation, with $\boldsymbol{E}$ = electric field, $\boldsymbol{B}$ = magnetic field, $\boldsymbol{v}$ = ordinary velocity of particle, $\boldsymbol{p}$ = momentum of particle, $e$ = charge of particle, reads

The three-dimensional version of the Lorentz force law

$$(d\boldsymbol{p}/dt) = e(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}). \tag{3.1}$$

Useful though this version of the equation may be, it is far from the geometric spirit of Einstein. A fully geometric equation will involve the test particle's energy-momentum 4-vector, $\boldsymbol{p}$, not just the spatial part $\boldsymbol{p}$ as measured in a specific Lorentz frame; and it will ask for the rate of change of momentum not as measured by a specific Lorentz observer $(d/dt)$, but as measured by the only clock present *à priori* in the problem: the test particle's own clock $(d/d\tau)$. Thus, the lefthand side of a fully geometric equation will read

$$d\boldsymbol{p}/d\tau = \ .$$

The righthand side, the Lorentz 4-force, must also be a frame-independent object. It will be linear in the particle's 4-velocity $\boldsymbol{u}$, since the frame-dependent expression

$$\frac{d\boldsymbol{p}}{d\tau} = \frac{1}{\sqrt{1-\boldsymbol{v}^2}}\frac{d\boldsymbol{p}}{dt} = \frac{e}{\sqrt{1-\boldsymbol{v}^2}}(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}) = e(u^0\boldsymbol{E} + \boldsymbol{u} \times \boldsymbol{B}) \tag{3.2a}$$

is linear in the components of $\boldsymbol{u}$. Consequently, there must be a linear machine named ***Faraday***, or $\boldsymbol{F}$, or "electromagnetic field tensor," with a slot into which one inserts the 4-velocity of a test particle. The output of this machine, multiplied by the particle's charge, must be the electromagnetic 4-force that it feels:

Electromagnetic field tensor defined

Geometrical version of Lorentz force law

$$d\boldsymbol{p}/d\tau = e\boldsymbol{F}(\boldsymbol{u}). \tag{3.3}$$

By comparing this geometric equation with the original Lorentz force law (equation 3.2a), and with the companion energy-change law

$$\frac{dp^0}{d\tau} = \frac{1}{\sqrt{1-\boldsymbol{v}^2}}\frac{dE}{dt} = \frac{1}{\sqrt{1-\boldsymbol{v}^2}}e\boldsymbol{E}\cdot\boldsymbol{v} = e\boldsymbol{E}\cdot\boldsymbol{u}, \tag{3.2b}$$

one can read off the components of $\boldsymbol{F}$ in a specific Lorentz frame. The components of $d\boldsymbol{p}/d\tau$ are $dp^\alpha/d\tau$, and the components of $e\boldsymbol{F}(\boldsymbol{u})$ can be written (definition of $F^\alpha{}_\beta$!) $eF^\alpha{}_\beta u^\beta$. Consequently

$$dp^\alpha/d\tau = eF^\alpha{}_\beta u^\beta \tag{3.4}$$

must reduce to equations (3.2a,b). Indeed it does if one makes the identification

Components of electromagnetic field tensor

$$\|F^\alpha{}_\beta\| = \begin{array}{c} \\ \alpha=0 \\ \alpha=1 \\ \alpha=2 \\ \alpha=3 \end{array} \begin{array}{c} \begin{array}{cccc} \beta=0 & \beta=1 & \beta=2 & \beta=3 \end{array} \\ \left\| \begin{array}{cccc} 0 & E_x & E_y & E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{array} \right\| \end{array} . \tag{3.5}$$

More often seen in the literature are the "covariant components," obtained by lowering an index with the metric components:

$$F_{\alpha\beta} = \eta_{\alpha\gamma} F^{\gamma}{}_{\beta}; \tag{3.6}$$

$$\|F_{\alpha\beta}\| = \left\| \begin{matrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{matrix} \right\|. \tag{3.7}$$

This matrix equation demonstrates the unity of the electric and magnetic fields. Neither one by itself, $\boldsymbol{E}$ or $\boldsymbol{B}$, is a frame-independent, geometric entity. But merged together into a single entity, $\boldsymbol{F}$ = ***Faraday***, they acquire a meaning and significance that transcends coordinates and reference frames.

---

**EXERCISE**

**Exercise 3.1.**

Derive equations (3.5) and (3.7) for the components of ***Faraday*** by comparing (3.4) with (3.2a,b), and by using definition (3.6).

---

## §3.2. TENSORS IN ALL GENERALITY

Examples of tensors

A digression is in order. Now on the scene are several different tensors: the metric tensor $\boldsymbol{g}$ (§2.4), the Riemann curvature tensor ***Riemann*** (§1.6), the electromagnetic field tensor ***Faraday*** (§3.1). Each has been defined as a linear machine with input slots for vectors, and with an output that is either a real number, e.g., $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v})$, or a vector, e.g., ***Riemann*** $(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$ and ***Faraday*** $(\boldsymbol{u})$.

Should one make a distinction between tensors whose outputs are scalars, and tensors whose outputs are vectors? No! A tensor whose output is a vector can be reinterpreted trivially as one whose output is a scalar. Take, for example, ***Faraday*** = $\boldsymbol{F}$. Add a new slot for the insertion of an arbitrary 1-form $\boldsymbol{\sigma}$, and gears and wheels that guarantee the output

$$\boldsymbol{F}(\boldsymbol{\sigma}, \boldsymbol{u}) = \langle \boldsymbol{\sigma}, \boldsymbol{F}(\boldsymbol{u}) \rangle = \text{real number}. \tag{3.8}$$

Then permit the user to choose whether he inserts only a vector, and gets out the vector $\boldsymbol{F}(\ldots, \boldsymbol{u}) = \boldsymbol{F}(\boldsymbol{u})$, or whether he inserts a form and a vector, and gets out the number $\boldsymbol{F}(\boldsymbol{\sigma}, \boldsymbol{u})$. The same machine will do both jobs. Moreover, in terms of components in a given Lorentz frame, both jobs are achieved very simply:

$$\begin{aligned} &\boldsymbol{F}(\ldots, \boldsymbol{u}) \text{ is a vector with components } F^{\alpha}{}_{\beta} u^{\beta}; \\ &\boldsymbol{F}(\boldsymbol{\sigma}, \boldsymbol{u}) \text{ is the number } \langle \boldsymbol{\sigma}, \boldsymbol{F}(\ldots, \boldsymbol{u}) \rangle = \sigma_{\alpha} F^{\alpha}{}_{\beta} u^{\beta}. \end{aligned} \tag{3.9}$$

By analogy, one defines the most general tensor $\boldsymbol{H}$ and its rank $\binom{n}{m}$ as follows: $\boldsymbol{H}$ is a linear machine with $n$ input slots for $n$ 1-forms, and $m$ input slots for $m$ vectors; given the requested input, it puts out a real number denoted

Definition of tensor as linear machine that converts vectors and 1-forms into numbers

$$\boldsymbol{H}(\underbrace{\boldsymbol{\sigma}, \boldsymbol{\lambda}, \ldots, \boldsymbol{\beta}}_{n \text{ 1-forms}}, \underbrace{\boldsymbol{u}, \boldsymbol{v}, \ldots, \boldsymbol{w}}_{m \text{ vectors}}). \tag{3.10}$$

For most tensors the output changes when two input vectors are interchanged,

$$\boldsymbol{Riemann}(\boldsymbol{\sigma}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \neq \boldsymbol{Riemann}(\boldsymbol{\sigma}, \boldsymbol{v}, \boldsymbol{u}, \boldsymbol{w}), \tag{3.11}$$

or when two input 1-forms are interchanged.

Choose a specific tensor $\boldsymbol{S}$, of rank $\binom{2}{1}$ for explicitness. Into the slots of $\boldsymbol{S}$, insert the basis vectors and 1-forms of a specific Lorentz coordinate frame. The output is a "component of $\boldsymbol{S}$ in that frame":

$$S^{\alpha\beta}{}_{\gamma} \equiv \boldsymbol{S}(\boldsymbol{\omega}^{\alpha}, \boldsymbol{\omega}^{\beta}, \boldsymbol{e}_{\gamma}). \tag{3.12}$$

Components of a tensor

This defines components. Knowing the components in a specific frame, one can easily calculate the output produced from any input forms and vectors:

$$\begin{aligned}\boldsymbol{S}(\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{v}) &= \boldsymbol{S}(\sigma_{\alpha}\boldsymbol{\omega}^{\alpha}, \rho_{\beta}\boldsymbol{\omega}^{\beta}, v^{\gamma}\boldsymbol{e}_{\gamma}) = \sigma_{\alpha}\rho_{\beta}v^{\gamma}\boldsymbol{S}(\boldsymbol{\omega}^{\alpha}, \boldsymbol{\omega}^{\beta}, \boldsymbol{e}_{\gamma}) \\ &= S^{\alpha\beta}{}_{\gamma}\sigma_{\alpha}\rho_{\beta}v^{\gamma}.\end{aligned} \tag{3.13}$$

Tensor's machine action expressed in terms of components

And knowing the components of $\boldsymbol{S}$ in one Lorentz frame (unprimed), plus the Lorentz transformation matrices $||\Lambda^{\alpha'}{}_{\beta}||$ and $||\Lambda^{\beta}{}_{\alpha'}||$ which link that frame with another (primed), one can calculate the components in the new (primed) frame. As shown in exercise 3.2, one need only apply a matrix to each index of $\boldsymbol{S}$, lining up the matrix indices in the logical manner

$$S^{\mu'\nu'}{}_{\lambda'} = S^{\alpha\beta}{}_{\gamma}\Lambda^{\mu'}{}_{\alpha}\Lambda^{\nu'}{}_{\beta}\Lambda^{\gamma}{}_{\lambda'}. \tag{3.14}$$

Lorentz transformation of components

A slight change of the internal gears and wheels inside the tensor enables one of its 1-form slots to accept a vector. All that is necessary is a mechanism to convert an input vector $\boldsymbol{n}$ into its corresponding 1-form $\tilde{\boldsymbol{n}}$ and then to put that 1-form into the old machinery. Thus, denoting the modified tensor by the same symbol $\boldsymbol{S}$ as was used for the original tensor, one demands

$$\boldsymbol{S}(\boldsymbol{\sigma}, \boldsymbol{n}, \boldsymbol{v}) = \boldsymbol{S}(\boldsymbol{\sigma}, \tilde{\boldsymbol{n}}, \boldsymbol{v}); \tag{3.15}$$

Modifying a tensor to accept either a vector or a 1-form into each slot

or, in component notation

$$S^{\alpha}{}_{\beta\gamma}\sigma_{\alpha}n^{\beta}v^{\gamma} = S^{\alpha\beta}{}_{\gamma}\sigma_{\alpha}n_{\beta}v^{\gamma}. \tag{3.15$'$}$$

This is achieved if one raises and lowers the indices of $\boldsymbol{S}$ using the components of the metric:

$$S^{\alpha}{}_{\beta\gamma} = \eta_{\beta\mu}S^{\alpha\mu}{}_{\gamma}, \qquad S^{\alpha\mu}{}_{\gamma} = \eta^{\mu\beta}S^{\alpha}{}_{\beta\gamma}. \tag{3.16}$$

Raising and lowering indices

(See exercise 3.3 below.) By using the same symbol $\boldsymbol{S}$ for the original tensor and the modified tensor, one allows each slot to accept either a 1-form or a vector, so one loses sight of whether $\boldsymbol{S}$ is a $\binom{2}{1}$ tensor, or a $\binom{1}{2}$ tensor, or a $\binom{3}{0}$ tensor, or a $\binom{0}{3}$ tensor; one only distinguishes its total rank, 3. *Terminology:* an "upstairs index" is called "contravariant"; a "downstairs" index is called "covariant." Thus in $S^{\alpha}{}_{\beta\gamma}$, "$\alpha$" is a contravariant index, while "$\beta$" and "$\gamma$" are covariant indices.

Because tensors are nothing but functions, they can be added (if they have the same rank!) and multiplied by numbers in the usual way: the output of the rank-3 tensor $a\boldsymbol{S} + b\boldsymbol{Q}$, when vectors $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ are put in, is

$$(a\boldsymbol{S} + b\boldsymbol{Q})(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \equiv a\boldsymbol{S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) + b\boldsymbol{Q}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}). \tag{3.17}$$

Several other important operations on tensors are explored in the following exercises. They and the results of the exercises will be used freely in the material that follows.

---

## EXERCISES

### Exercise 3.2. TRANSFORMATION LAW FOR COMPONENTS OF A TENSOR

From the transformation laws for components of vectors and 1-forms, derive the transformation law (3.14).

### Exercise 3.3. RAISING AND LOWERING INDICES

Derive equations (3.16) from equation (3.15′) plus the law $n_\alpha = \eta_{\alpha\beta} n^\beta$ for getting the components of the 1-form $\tilde{\boldsymbol{n}}$ from the components of its corresponding vector $\boldsymbol{n}$.

### Exercise 3.4. TENSOR PRODUCT

Given any two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, one defines the second-rank tensor $\boldsymbol{u} \otimes \boldsymbol{v}$ ("tensor product of $\boldsymbol{u}$ with $\boldsymbol{v}$") to be a machine, with two input slots, whose output is the number

$$(\boldsymbol{u} \otimes \boldsymbol{v})(\boldsymbol{\sigma}, \boldsymbol{\lambda}) = \langle \boldsymbol{\sigma}, \boldsymbol{u} \rangle \langle \boldsymbol{\lambda}, \boldsymbol{v} \rangle \tag{3.18}$$

when 1-forms $\boldsymbol{\sigma}$ and $\boldsymbol{\lambda}$ are inserted. Show that the components of $\boldsymbol{T} = \boldsymbol{u} \otimes \boldsymbol{v}$ are the products of the components of $\boldsymbol{u}$ and $\boldsymbol{v}$:

$$T^{\alpha\beta} = u^\alpha v^\beta, \qquad T_\alpha{}^\beta = u_\alpha v^\beta, \qquad T_{\alpha\beta} = u_\alpha v_\beta. \tag{3.19}$$

Extend the definition to several vectors and forms,

$$(\boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{\beta} \otimes \boldsymbol{w})(\boldsymbol{\sigma}, \boldsymbol{\lambda}, \boldsymbol{n}, \boldsymbol{\zeta}) = \langle \boldsymbol{\sigma}, \boldsymbol{u} \rangle \langle \boldsymbol{\lambda}, \boldsymbol{v} \rangle \langle \boldsymbol{\beta}, \boldsymbol{n} \rangle \langle \boldsymbol{\zeta}, \boldsymbol{w} \rangle, \tag{3.20}$$

and show that the product rule for components still holds:

$$\begin{gathered} \boldsymbol{S} = \boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{\beta} \otimes \boldsymbol{w} \text{ has components} \\ S^{\mu\nu}{}_\lambda{}^\zeta = u^\mu v^\nu \beta_\lambda w^\zeta. \end{gathered} \tag{3.21}$$

### Exercise 3.5. BASIS TENSORS

Show that a tensor $\boldsymbol{M}$ with components $M^{\alpha\beta}{}_\gamma{}^\delta$ in a given Lorentz frame can be reconstructed from its components and from the basis 1-forms and vectors of that frame as follows:

$$\boldsymbol{M} = M^{\alpha\beta}{}_\gamma{}^\delta \boldsymbol{e}_\alpha \otimes \boldsymbol{e}_\beta \otimes \boldsymbol{\omega}^\gamma \otimes \boldsymbol{e}_\delta. \tag{3.22}$$

(For a special case of this, see Box 3.2.)

**Box 3.2 THE METRIC IN DIFFERENT LANGUAGES**

## A. Geometric Language

$\boldsymbol{g}$ is a linear, symmetric machine with two slots for insertion of vectors. When vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are inserted, the output of $\boldsymbol{g}$ is their scalar product:

$$\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u} \cdot \boldsymbol{v}.$$

## B. Component Language

$\eta_{\mu\nu}$ are the metric components. They are used to calculate the scalar product of two vectors from components in a specific Lorentz frame:

$$\boldsymbol{u} \cdot \boldsymbol{v} = \eta_{\mu\nu} u^{\mu} v^{\nu}.$$

## C. Coordinate-Based Geometric Language

The metric $\boldsymbol{g}$ can be written, in terms of basis 1-forms of a specific Lorentz frame, as

$$\boldsymbol{g} = \eta_{\mu\nu} \boldsymbol{\omega}^{\mu} \otimes \boldsymbol{\omega}^{\nu} = \eta_{\mu\nu}\, \boldsymbol{d}x^{\mu} \otimes \boldsymbol{d}x^{\nu}$$

[see equations (2.18) and (3.22)].

## D. Connection to the Elementary Concept of Line Element

Box 2.3 demonstrated the correspondence between the gradient $\boldsymbol{d}f$ of a function, and the elementary concept $df$ of a differential change of $f$ in some unspecified direction. There is a similar correspondence between the metric, written as $\eta_{\mu\nu}\, \boldsymbol{d}x^{\mu} \otimes \boldsymbol{d}x^{\nu}$, and the elementary concept of "line element," written as $ds^2 = \eta_{\mu\nu}\, dx^{\mu}\, dx^{\nu}$. This elementary line element, as expounded in many special relativity texts, represents the squared length of the displacement "$dx^{\mu}$" in an unspecified direction. The metric $\eta_{\mu\nu}\, \boldsymbol{d}x^{\mu} \otimes \boldsymbol{d}x^{\nu}$ does the same. Pick a specific infinitesimal displacement vector $\boldsymbol{\xi}$, and insert it into the slots of $\eta_{\mu\nu}\, \boldsymbol{d}x^{\mu} \otimes \boldsymbol{d}x^{\nu}$. The output will be $\boldsymbol{\xi}^2 = \eta_{\mu\nu}\xi^{\mu}\xi^{\nu}$, the squared length of the displacement. Before $\boldsymbol{\xi}$ is inserted, $\eta_{\mu\nu}\, \boldsymbol{d}x^{\mu} \otimes \boldsymbol{d}x^{\nu}$ has the potential to tell the squared length of any vector; the insertion of $\boldsymbol{\xi}$ converts potentiality into actuality: the numerical value of $\boldsymbol{\xi}^2$.

Because the metric $\eta_{\mu\nu}\, \boldsymbol{d}x^{\mu} \otimes \boldsymbol{d}x^{\nu}$ and the line element $ds^2 = \eta_{\mu\nu}\, dx^{\mu}\, dx^{\nu}$ perform this same function of representing the squared length of an unspecified infinitesimal displacement, there is no conceptual distinction between them. One sometimes uses the symbols $\boldsymbol{ds^2}$ to denote the metric; one sometimes gets pressed and writes it as $\boldsymbol{ds^2} = \eta_{\mu\nu}\, \boldsymbol{d}x^{\mu}\, \boldsymbol{d}x^{\nu}$, omitting the "$\otimes$"; and one sometimes even gets so pressed as to use nonbold characters, so that no notational distinction remains at all between metric and elementary line element:

$$\boldsymbol{g} = \boldsymbol{ds^2} = ds^2 = \eta_{\mu\nu}\, dx^{\mu}\, dx^{\nu}.$$

**Exercise 3.6.** ***Faraday*** **MACHINERY AT WORK**

An observer with 4-velocity $\boldsymbol{u}$ picks out three directions in spacetime that are orthogonal and purely spatial (no time part) as seen in his frame. Let $\boldsymbol{e}_{\hat{1}}$, $\boldsymbol{e}_{\hat{2}}$, $\boldsymbol{e}_{\hat{3}}$ be unit vectors in those directions and let them be oriented in a righthanded way ($\boldsymbol{e}_{\hat{1}} \cdot \boldsymbol{e}_{\hat{2}} \times \boldsymbol{e}_{\hat{3}} = +1$ in three-dimensional language). Why do the following relations hold?

$$\boldsymbol{e}_{\hat{j}} \cdot \boldsymbol{u} = 0, \qquad \boldsymbol{e}_{\hat{j}} \cdot \boldsymbol{e}_{\hat{k}} = \delta_{jk}.$$

What vectors are to be inserted in the two slots of the electromagnetic field tensor ***Faraday*** if one wants to get out the electric field along $\boldsymbol{e}_{\hat{j}}$ as measured by this observer? What vectors must be inserted to get the magnetic field he measures along $\boldsymbol{e}_{\hat{j}}$?

---

## §3.3. THREE-PLUS-ONE VIEW VERSUS GEOMETRIC VIEW

The power of the geometric view of physics

Great computational and conceptual power resides in Einstein's geometric view of physics. Ideas that seem complex when viewed in the everyday "space-plus-time" or "3 + 1" manner become elegant and simple when viewed as relations between geometric objects in four-dimensional spacetime. Derivations that are difficult in 3 + 1 language simplify in geometric language.

Example of electromagnetism

The electromagnetic field is a good example. In geometric language, it is described by a second-rank, antisymmetric tensor ("2-form") $\boldsymbol{F}$, which requires no coordinates for its definition. This tensor produces a 4-force on any charged particle given by

$$d\boldsymbol{p}/d\tau = e\boldsymbol{F}(\boldsymbol{u}).$$

It is all so simple!

By contrast, consider the "3 + 1" viewpoint. In a given Lorentz frame, there is an electric field $\boldsymbol{E}$ and a magnetic field $\boldsymbol{B}$. They push on a particle in accordance with

$$d\boldsymbol{p}/dt = e(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}).$$

Transformation law for electric and magnetic fields

But the values of $\boldsymbol{p}$, $\boldsymbol{E}$, $\boldsymbol{v}$, and $\boldsymbol{B}$ all change when one passes from the given Lorentz frame to a new one. For example, the electric and magnetic fields viewed from a rocket ship ("barred" frame) are related to those viewed in the laboratory ("unbarred" frame) by

$$\begin{aligned} \bar{\boldsymbol{E}}_{\parallel} &= \boldsymbol{E}_{\parallel}, \qquad \bar{\boldsymbol{E}}_{\perp} = \frac{1}{\sqrt{1-\beta^2}}(\boldsymbol{E}_{\perp} + \boldsymbol{\beta} \times \boldsymbol{B}_{\perp}), \\ \bar{\boldsymbol{B}}_{\parallel} &= \boldsymbol{B}_{\parallel}, \qquad \bar{\boldsymbol{B}}_{\perp} = \frac{1}{\sqrt{1-\beta^2}}(\boldsymbol{B}_{\perp} - \boldsymbol{\beta} \times \boldsymbol{E}_{\perp}). \end{aligned} \tag{3.23}$$

(Here "$\parallel$" means component along direction of rocket's motion; "$\perp$" means perpendicular component; and $\beta^j = dx^j{}_{\text{rocket}}/dt$ is the rocket's ordinary velocity.) The analogous transformation laws for the particle's momentum $\boldsymbol{p}$ and ordinary velocity

$\boldsymbol{v}$, and for the coordinate time $t$, all conspire—as if by magic, it seems, from the $3 + 1$ viewpoint—to maintain the validity of the Lorentz force law in all frames.

Not only is the geometric view far simpler than the $3 + 1$ view, it can even derive the $3 + 1$ equations with greater ease than can the $3 + 1$ view itself. Consider, for example, the transformation law (3.23) for the electric and magnetic fields. The geometric view derives it as follows: (1) Orient the axes of the two frames so their relative motion is in the $z$-direction. (2) Perform a simple Lorentz transformation (equation 2.45) on the components of the electromagnetic field tensor:

$$\begin{aligned} \bar{E}_{\|} &= \bar{E}_z = F_{\bar{3}\bar{0}} = \Lambda^{\alpha}{}_{\bar{3}}\Lambda^{\beta}{}_{\bar{0}}F_{\alpha\beta} = \gamma^2 F_{30} + \beta^2\gamma^2 F_{03} \\ &= (1 - \beta^2)\gamma^2 F_{30} = F_{30} = E_z = E_{\|}, \\ \bar{E}_x &= F_{\bar{1}\bar{0}} = \Lambda^{\alpha}{}_{\bar{1}}\Lambda^{\beta}{}_{\bar{0}}F_{\alpha\beta} = \gamma F_{10} + \beta\gamma F_{13} = \gamma(E_x - \beta B_y), \end{aligned} \tag{3.24}$$

etc.

By contrast, the $3 + 1$ view shows much more work. A standard approach is based on the Lorentz force law and energy-change law (3.2a,b), written in the slightly modified form

$$m\frac{d^2\bar{x}}{d\tau^2} = e\left(\bar{E}_x\frac{d\bar{t}}{d\tau} + 0\frac{d\bar{x}}{d\tau} + \bar{B}_z\frac{d\bar{y}}{d\tau} - \bar{B}_y\frac{d\bar{z}}{d\tau}\right), \tag{3.25}$$

. . . (three additional equations) . . . .

It proceeds as follows (details omitted because of their great length!):

(1) Substitute for the $d^2\bar{x}/d\tau^2$, etc., the expression for these quantities in terms of the $d^2x/d\tau^2, \ldots$ (Lorentz transformation).

(2) Substitute for the $d^2x/d\tau^2, \ldots$ the expression for these accelerations in terms of the laboratory $\boldsymbol{E}$ and $\boldsymbol{B}$ (Lorentz force law).

(3) In these expressions, wherever the components $dx/d\tau$ of the 4-velocity in the laboratory frame appear, substitute expressions in terms of the 4-velocities in the rocket frame (inverse Lorentz transformation).

(4) In (3.25) as thus transformed, demand equality of left and right sides for all values of the $d\bar{x}/d\tau$, etc. (validity for all test particles).

(5) In this way arrive at the expressions (3.23) for the $\bar{\boldsymbol{E}}$ and $\bar{\boldsymbol{B}}$ in terms of the $\boldsymbol{E}$ and $\boldsymbol{B}$.

The contrast in difficulty is obvious!

## §3.4. MAXWELL'S EQUATIONS

Turn now from the action of the field on a charge, and ask about the action of a charge on the field, or, more generally, ask about the dynamics of the electromagnetic

Magnetodynamics derived from magnetostatics

field, charge or no charge. Begin with the simplest of Maxwell's equations in a specific Lorentz frame, the one that says there are no free magnetic poles:

$$\nabla \cdot \boldsymbol{B} \equiv \operatorname{div} \boldsymbol{B} = \frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0. \tag{3.26}$$

This statement has to be true in all Lorentz frames. It is therefore true in the rocket frame:

$$\frac{\partial \bar{B}_x}{\partial \bar{x}} + \frac{\partial \bar{B}_y}{\partial \bar{y}} + \frac{\partial \bar{B}_z}{\partial \bar{z}} = 0. \tag{3.27}$$

For an infinitesimal Lorentz transformation in the $x$-direction (nonrelativistic velocity $\beta$), one has (see Box 2.4 and equations 3.23)

$$\bar{B}_x = B_x, \qquad \bar{B}_y = B_y + \beta E_z, \qquad \bar{B}_z = B_z - \beta E_y; \tag{3.28}$$

$$\frac{\partial}{\partial \bar{x}} = \frac{\partial}{\partial x} + \beta \frac{\partial}{\partial t}, \qquad \frac{\partial}{\partial \bar{y}} = \frac{\partial}{\partial y}, \qquad \frac{\partial}{\partial \bar{z}} = \frac{\partial}{\partial z}. \tag{3.29}$$

Substitute into the condition of zero divergence in the rocket frame. Recover the original condition of zero divergence in the laboratory frame, plus the following additional information (requirement for the vanishing of the coefficient of the arbitrary small velocity $\beta$):

$$\frac{\partial B_x}{\partial t} + \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = 0. \tag{3.30}$$

Had the velocity of transformation been directed in the $y$- or $z$-directions, a similar equation would have been obtained for $\partial B_y/\partial t$ or $\partial B_z/\partial t$. In the language of three-dimensional vectors, these three equations reduce to the one equation

$$\frac{\partial \boldsymbol{B}}{\partial t} + \nabla \times \boldsymbol{E} \equiv \frac{\partial \boldsymbol{B}}{\partial t} + \operatorname{curl} \boldsymbol{E} = 0. \tag{3.31}$$

Magnetodynamics and magnetostatics unified in one geometric law

How beautiful that (1) the principle of covariance (laws of physics are the same in every Lorentz reference system, which is equivalent to the geometric view of physics) plus (2) the principle that magnetic tubes of force never end, gives (3) Maxwell's dynamic law for the time-rate of change of the magnetic field! This suggests that the magnetostatic law $\nabla \cdot \boldsymbol{B} = 0$ and the magnetodynamic law $\partial \boldsymbol{B}/\partial t + \nabla \times \boldsymbol{E} = 0$ must be wrapped up together in a single frame-independent, geometric law. In terms of components of the field tensor $\boldsymbol{F}$, that geometric law must read

$$F_{\alpha\beta,\gamma} + F_{\beta\gamma,\alpha} + F_{\gamma\alpha,\beta} = 0, \tag{3.32}$$

since this reduces to $\nabla \cdot \boldsymbol{B} = 0$ when one takes $\alpha = 1, \beta = 2, \gamma = 3$; and it reduces to $\partial \boldsymbol{B}/\partial t + \nabla \times \boldsymbol{E} = 0$ when one sets any index, e.g., $\alpha$, equal to zero (see exercise 3.7 below). In frame-independent geometric language, this law is written (see §3.5, exercise 3.14, and Chapter 4 for notation)

$$\boldsymbol{dF} = 0, \text{ or, equivalently, } \boldsymbol{\nabla} \cdot {}^*\boldsymbol{F} = 0; \tag{3.33}$$

and it says, "Take the electromagnetic 2-form $\boldsymbol{F}$ (a geometric object defined even in absence of coordinates); from it construct a new geometric object $\boldsymbol{dF}$ (called the "exterior derivative of $\boldsymbol{F}$"); $\boldsymbol{dF}$ must vanish. The details of this coordinate-free process will be spelled out in exercise 3.15 and in §4.5 (track 2).

Two of Maxwell's equations remain: the electrostatic equation

$$\boldsymbol{\nabla} \cdot \boldsymbol{E} = 4\pi\rho, \tag{3.34}$$

and the electrodynamic equation

$$\partial \boldsymbol{E}/\partial t - \boldsymbol{\nabla} \times \boldsymbol{B} = -4\pi \boldsymbol{J}. \tag{3.35}$$

They, like the magnetostatic and magnetodynamic equations, are actually two different parts of a single geometric law. Written in terms of field components, that law says

Electrodynamics and electrostatics unified in one geometric law

$$F^{\alpha\beta}{}_{,\beta} = 4\pi J^\alpha, \tag{3.36}$$

where the components of the "4-current" $\boldsymbol{J}$ are

$$\begin{aligned} J^0 &= \rho = \text{charge density}, \\ (J^1, J^2, J^3) &= \text{components of current density}. \end{aligned} \tag{3.37}$$

Written in coordinate-free, geometric language, this electrodynamic law says

$$\boldsymbol{d}{}^*\boldsymbol{F} = 4\pi\, {}^*\boldsymbol{J} \text{ or, equivalently, } \boldsymbol{\nabla} \cdot \boldsymbol{F} = 4\pi \boldsymbol{J}. \tag{3.38}$$

(For full discussion, see exercise 3.15 and §4.5, which is on Track 2.)

---

**EXERCISE**

**Exercise 3.7. MAXWELL'S EQUATIONS**

Show, by explicit examination of components, that the geometric laws

$$F_{\alpha\beta,\gamma} + F_{\beta\gamma,\alpha} + F_{\gamma\alpha,\beta} = 0, \qquad F^{\alpha\beta}{}_{,\beta} = 4\pi J^\alpha,$$

do reduce to Maxwell's equations (3.26), (3.31), (3.34), (3.35), as claimed above.

---

## §3.5 WORKING WITH TENSORS

Another mathematical digression is needed. Given an arbitrary tensor field, $\boldsymbol{S}$, of arbitrary rank (choose rank = 3 for concreteness), one can construct new tensor fields by a variety of operations.

Ways to produce new tensors from old:

Gradient

One operation is the *gradient* $\boldsymbol{\nabla}$. (The symbol $\boldsymbol{d}$ is reserved for gradients of scalars, in which case $\boldsymbol{\nabla} f \equiv \boldsymbol{d}f$, and for "exterior derivatives of differential forms;" a Track-2 concept, on which see §4.5.) Like $\boldsymbol{S}$, $\boldsymbol{\nabla S}$ is a machine. It has four slots, whereas $\boldsymbol{S}$ has three. It describes how $\boldsymbol{S}$ changes from point to point. Specifically, if one desires to know how the number $\boldsymbol{S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$ for *fixed* $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ changes under a displacement $\boldsymbol{\xi}$, one inserts $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$, $\boldsymbol{\xi}$ into the four slots of $\boldsymbol{\nabla S}$:

$$\begin{aligned}\boldsymbol{\nabla S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{\xi}) &\equiv \partial_{\boldsymbol{\xi}} \boldsymbol{S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \text{ with } \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \text{ fixed} \\ &\simeq +[\text{value of } \boldsymbol{S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \text{ at tip of } \boldsymbol{\xi}] \\ &\quad -[\text{value of } \boldsymbol{S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \text{ at tail of } \boldsymbol{\xi}].\end{aligned} \tag{3.39}$$

In component notation in a Lorentz frame, this says

$$\begin{aligned}\boldsymbol{\nabla S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{\xi}) \equiv \partial_{\boldsymbol{\xi}}(S_{\alpha\beta\gamma} u^\alpha v^\beta w^\gamma) &= \left(\frac{\partial S_{\alpha\beta\gamma}}{\partial x^\delta}\xi^\delta\right) u^\alpha v^\beta w^\gamma \\ &= S_{\alpha\beta\gamma,\delta} u^\alpha v^\beta w^\gamma \xi^\delta.\end{aligned}$$

Thus, the Lorentz-frame components of $\boldsymbol{\nabla S}$ are nothing but the partial derivatives of the components of $\boldsymbol{S}$. Notice that the gradient raises the rank of a tensor by 1 (from 3 to 4 for $\boldsymbol{S}$).

Contraction

*Contraction* is another process that produces a new tensor from an old one. It seals off ("contracts") two of the old tensor's slots, thereby reducing the rank by two. Specifically, if $\boldsymbol{R}$ is a fourth-rank tensor and $\boldsymbol{M}$ is obtained by contracting the first and third slots of $\boldsymbol{R}$, then the output of $\boldsymbol{M}$ is given by (definition!)

$$\boldsymbol{M}(\boldsymbol{u}, \boldsymbol{v}) = \sum_{\alpha=0}^{3} \boldsymbol{R}(\boldsymbol{e}_\alpha, \boldsymbol{u}, \boldsymbol{\omega}^\alpha, \boldsymbol{v}). \tag{3.40}$$

Here $\boldsymbol{e}_\alpha$ and $\boldsymbol{\omega}^\alpha$ are the basis vectors and 1-forms of a specific but arbitrary Lorentz coordinate frame. It makes no difference which frame is chosen; the result will always be the same (exercise 3.8 below). In terms of components in any Lorentz frame, equation (3.40) says (exercise 3.8)

$$\boldsymbol{M}(\boldsymbol{u}, \boldsymbol{v}) = M_{\mu\nu} u^\mu v^\nu = R_{\alpha\mu}{}^{\alpha}{}_{\nu} u^\mu v^\nu,$$

so that

$$M_{\mu\nu} = R_{\alpha\mu}{}^{\alpha}{}_{\nu}. \tag{3.41}$$

Thus, in terms of components, contraction amounts to putting one index up and the other down, and then summing on them.

Divergence

*Divergence* is a third process for creating new tensors from old. It is accomplished by taking the gradient, then contracting the gradient's slot with one of the original slots:

$$(\text{divergence of } \boldsymbol{S} \text{ on first slot}) \equiv \boldsymbol{\nabla}\cdot\boldsymbol{S} \qquad \text{is a machine such that}$$

$$\boldsymbol{\nabla}\cdot\boldsymbol{S}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{\nabla S}(\boldsymbol{\omega}^\alpha, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{e}_\alpha) = S^\alpha{}_{\beta\gamma,\alpha} u^\beta v^\gamma; \tag{3.42}$$

i.e. $\boldsymbol{\nabla}\cdot\boldsymbol{S}$ has components $S^\alpha{}_{\beta\gamma,\alpha}$.

*Transpose* is a fourth, rather trivial process for creating new tensors. It merely interchanges two slots: Transpose

$$\boldsymbol{N} \text{ obtained by transposing second and third slots of } \boldsymbol{S} \Longrightarrow$$
$$\boldsymbol{N}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{S}(\boldsymbol{u}, \boldsymbol{w}, \boldsymbol{v}). \tag{3.43}$$

*Symmetrization* and *antisymmetrization* are fifth and sixth processes for producing new tensors from old. A tensor is completely symmetric if its output is unaffected by an interchange of two input vectors or 1-forms: Symmetrization and antisymmetrization

$$\boldsymbol{S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{S}(\boldsymbol{v}, \boldsymbol{u}, \boldsymbol{w}) = \boldsymbol{S}(\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{u}) = \cdots. \tag{3.44a}$$

It is completely antisymmetric if it reverses sign on each interchange of input

$$\boldsymbol{S}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = -\boldsymbol{S}(\boldsymbol{v}, \boldsymbol{u}, \boldsymbol{w}) = +\boldsymbol{S}(\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{u}) = \cdots. \tag{3.44b}$$

Any tensor can be symmetrized or antisymmetrized by constructing an appropriate linear combination of it and its transposes; see exercise 3.12.

*Wedge product* is a seventh process for producing new tensors from old. It is merely an antisymmetrized tensor product: given two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, their wedge product, the *"bivector"* $\boldsymbol{u} \wedge \boldsymbol{v}$, is defined by Wedge product Bivector

$$\boldsymbol{u} \wedge \boldsymbol{v} \equiv \boldsymbol{u} \otimes \boldsymbol{v} - \boldsymbol{v} \otimes \boldsymbol{u}; \tag{3.45a}$$

similarly, the *"2-form"* $\boldsymbol{\alpha} \wedge \boldsymbol{\beta}$ constructed from two 1-forms is 2-form

$$\boldsymbol{\alpha} \wedge \boldsymbol{\beta} \equiv \boldsymbol{\alpha} \otimes \boldsymbol{\beta} - \boldsymbol{\beta} \otimes \boldsymbol{\alpha}. \tag{3.45b}$$

From three vectors $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ one constructs the *"trivector"* Trivector

$$\begin{aligned} \boldsymbol{u} \wedge \boldsymbol{v} \wedge \boldsymbol{w} &\equiv (\boldsymbol{u} \wedge \boldsymbol{v}) \wedge \boldsymbol{w} \equiv \boldsymbol{u} \wedge (\boldsymbol{v} \wedge \boldsymbol{w}) \\ &= \boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{w} + \text{terms that guarantee complete antisymmetry} \\ &= \boldsymbol{u} \otimes \boldsymbol{v} \otimes \boldsymbol{w} + \boldsymbol{v} \otimes \boldsymbol{w} \otimes \boldsymbol{u} + \boldsymbol{w} \otimes \boldsymbol{u} \otimes \boldsymbol{v} \\ &\quad - \boldsymbol{v} \otimes \boldsymbol{u} \otimes \boldsymbol{w} - \boldsymbol{u} \otimes \boldsymbol{w} \otimes \boldsymbol{v} - \boldsymbol{w} \otimes \boldsymbol{v} \otimes \boldsymbol{u}. \end{aligned} \tag{3.45c}$$

From 1-forms $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ one similarly constructs the "3-forms" $\boldsymbol{\alpha} \wedge \boldsymbol{\beta} \wedge \boldsymbol{\gamma}$. The wedge product gives a simple way to test for coplanarity (linear dependence) of vectors: if $\boldsymbol{u}$ and $\boldsymbol{v}$ are collinear, so $\boldsymbol{u} = a\boldsymbol{v}$, then

$$\boldsymbol{u} \wedge \boldsymbol{v} = a\boldsymbol{v} \wedge \boldsymbol{v} = 0 \qquad \text{(by antisymmetry of "}\wedge\text{")}.$$

If $\boldsymbol{w}$ is coplanar with $\boldsymbol{u}$ and $\boldsymbol{v}$ so $\boldsymbol{w} = a\boldsymbol{u} + b\boldsymbol{v}$ ("collapsed box"), then

$$\boldsymbol{w} \wedge \boldsymbol{u} \wedge \boldsymbol{v} = a\boldsymbol{u} \wedge \boldsymbol{u} \wedge \boldsymbol{v} + b\boldsymbol{v} \wedge \boldsymbol{u} \wedge \boldsymbol{v} = 0.$$

The symbol "$\wedge$" is called a "hat" or "wedge" or "exterior product sign." Its properties are investigated in Chapter 4.

*Taking the dual* is an eighth process for constructing new tensors. It plays a fundamental role in Track 2 of this book, but since it is not needed for Track 1, its definition and properties are treated only in the exercises (3.14 and 3.15). Dual

Because the frame-independent geometric notation is somewhat ambiguous (which slots are being contracted? on which slot is the divergence taken? which slots are being transposed?), one often uses component notation to express coordinate-independent, geometric relations between geometric objects. For example,

$$J_{\beta\gamma} = S^{\alpha}{}_{\beta\gamma,\alpha}$$

means "$\boldsymbol{J}$ is a tensor obtained by taking the divergence on the first slot of the tensor $\boldsymbol{S}$". Also,

$$v^{\gamma} = (F_{\mu\nu}F^{\mu\nu})^{,\gamma} \equiv (F_{\mu\nu}F^{\mu\nu})_{,\beta}\eta^{\beta\gamma}$$

means "$\boldsymbol{v}$ is a vector obtained by (1) constructing the tensor product $\boldsymbol{F} \otimes \boldsymbol{F}$ of $\boldsymbol{F}$ with itself, (2) contracting $\boldsymbol{F} \otimes \boldsymbol{F}$ on its first and third slots, and also on its second and fourth, (3) taking the gradient of the resultant scalar function, (4) converting that gradient, which is a 1-form, into the corresponding vector."

Index gymnastics

"Index gymnastics," the technique of extracting the content from geometric equations by working in component notation and rearranging indices as required, must be mastered if one wishes to do difficult calculations in relativity, special or general. Box 3.3 expounds some of the short cuts in index gymnastics, and exercises 3.8–3.18 offer practice.

---

## EXERCISES

**Exercise 3.8. CONTRACTION IS FRAME-INDEPENDENT**

Show that contraction, as defined in equation (3.40), does not depend on which Lorentz frame $\boldsymbol{e}_{\alpha}$ and $\boldsymbol{\omega}^{\alpha}$ are taken from. Also show that equation (3.40) implies

$$\boldsymbol{M}(\boldsymbol{u}, \boldsymbol{v}) = R_{\alpha\mu}{}^{\alpha}{}_{\nu}u^{\mu}v^{\nu}.$$

**Exercise 3.9. DIFFERENTIATION**

(a) Justify the formula

$$d(u_{\mu}v^{\nu})/d\tau = (du_{\mu}/d\tau)v^{\nu} + u_{\mu}(dv^{\nu}/d\tau).$$

by considering the special case $\mu = 0$, $\nu = 1$.

(b) Explain why

$$(T^{\alpha\beta}v_{\beta})_{,\mu} = T^{\alpha\beta}{}_{,\mu}v_{\beta} + T^{\alpha\beta}v_{\beta,\mu}.$$

**Exercise 3.10. MORE DIFFERENTIATION**

(a) Justify the formula,

$$d(u^{\mu}u_{\mu})/d\tau = 2u_{\mu}(du^{\mu}/d\tau),$$

by writing out the summation $u^{\mu}u_{\mu} \equiv \eta_{\mu\nu}u^{\mu}u^{\nu}$ explicitly.

(b) Let $\delta$ indicate a variation or small change, and justify the formula

$$\delta(F_{\alpha\beta}F^{\alpha\beta}) = 2F_{\alpha\beta}\delta F^{\alpha\beta}.$$

(c) Compute $(F_{\alpha\beta}F^{\alpha\beta})_{,\mu} = ?$

**Box 3.3 TECHNIQUES OF INDEX GYMNASTICS**

| *Equation* | *Name and Discussion* |
|---|---|
| $S^{\alpha}{}_{\beta\gamma} = S(\boldsymbol{\omega}^{\alpha}, \boldsymbol{e}_{\beta}, \boldsymbol{e}_{\gamma})$ | Computing components. |
| $S^{\alpha\beta}{}_{\gamma} = S(\boldsymbol{\omega}^{\alpha}, \boldsymbol{\omega}^{\beta}, \boldsymbol{e}_{\gamma})$ | Computing other components. |
| $\boldsymbol{S} = S^{\alpha}{}_{\beta\gamma}\boldsymbol{e}_{\alpha} \otimes \boldsymbol{\omega}^{\beta} \otimes \boldsymbol{\omega}^{\gamma}$ | Reconstructing the rank-$\binom{1}{2}$ version of $\boldsymbol{S}$. |
| $\boldsymbol{S} = S^{\alpha\beta\gamma}\boldsymbol{e}_{\alpha} \otimes \boldsymbol{e}_{\beta} \otimes \boldsymbol{e}_{\gamma}$ | Reconstructing the rank-$\binom{3}{0}$ version of $\boldsymbol{S}$. [Recall: one does not usually distinguish between the various versions; see equation (3.15) and associated discussion.] |
| $S^{\alpha\beta}{}_{\gamma} = \eta^{\beta\mu}S^{\alpha}{}_{\mu\gamma}$ | Raising an index. |
| $S^{\alpha}{}_{\mu\gamma} = \eta_{\mu\beta}S^{\alpha\beta}{}_{\gamma}$ | Lowering an index. |
| $M_{\mu} = S^{\alpha}{}_{\mu\alpha}$ | Contraction of $\boldsymbol{S}$ to form a new tensor $\boldsymbol{M}$. |
| $T^{\alpha\beta}{}_{\mu\nu} = S^{\alpha\beta}{}_{\mu}M_{\nu}$ | Tensor product of $\boldsymbol{S}$ with $\boldsymbol{M}$ to form a new tensor $\boldsymbol{T}$. |
| $\boldsymbol{A}^2 = A^{\alpha}A_{\alpha}$ | Squared length of vector $\boldsymbol{A}$ produced by forming tensor product $\boldsymbol{A} \otimes \boldsymbol{A}$ and then contracting, which is the same as forming the corresponding 1-form $\widetilde{\boldsymbol{A}}$ and then piercing: $\boldsymbol{A}^2 = \langle \widetilde{\boldsymbol{A}}, \boldsymbol{A} \rangle = A^{\alpha}A_{\alpha}$. |
| $\eta_{\alpha\beta}\eta^{\beta\gamma} = \delta_{\alpha}{}^{\gamma}$ | The matrix formed from the metric's "covariant components," $\lVert\eta_{\alpha\beta}\rVert$, is the inverse of that formed from its "contravariant components," $\lVert\eta^{\alpha\beta}\rVert$. Equivalently, raising one index of the metric $\eta_{\alpha\beta}$ produces the Kronecker delta. |
| $S^{\alpha}{}_{\beta\gamma} = N^{\alpha}{}_{\beta,\gamma}$ | Gradient of $\boldsymbol{N}$ to form a new tensor $\boldsymbol{S}$. |
| $R_{\beta} = N^{\alpha}{}_{\beta,\alpha}$ | Divergence of $\boldsymbol{N}$ to form a new tensor $\boldsymbol{R}$. |
| $N^{\alpha}{}_{\beta,\gamma} = (\eta_{\beta\mu}N^{\alpha\mu})_{,\gamma} = \eta_{\beta\mu}N^{\alpha\mu}{}_{,\gamma}$ | Taking gradients and raising or lowering indices are operations that commute. |
| $N^{\alpha}{}_{\beta}{}^{,\gamma} \equiv N^{\alpha}{}_{\beta,\mu}\eta^{\mu\gamma}$ | Contravariant index on a gradient is obtained by raising covariant index. |
| $(R_{\alpha}M_{\beta})_{,\gamma} = R_{\alpha,\gamma}M_{\beta} + R_{\alpha}M_{\beta,\gamma}$ | Gradient of a tensor product; says $\boldsymbol{\nabla}(\boldsymbol{R} \otimes \boldsymbol{M}) =$ Transpose $(\boldsymbol{\nabla R} \otimes \boldsymbol{M}) + \boldsymbol{R} \otimes \boldsymbol{\nabla M}$. |
| $G_{\alpha\beta} = F_{[\alpha\beta]} \equiv \frac{1}{2}(F_{\alpha\beta} - F_{\beta\alpha})$ | Antisymmetrizing a tensor $\boldsymbol{F}$ to produce a new tensor $\boldsymbol{G}$. |
| $H_{\alpha\beta} = F_{(\alpha\beta)} \equiv \frac{1}{2}(F_{\alpha\beta} + F_{\beta\alpha})$ | Symmetrizing a tensor $\boldsymbol{F}$ to produce a new tensor $\boldsymbol{H}$. |
| ${}^{*}J_{\alpha\beta\gamma} = J^{\mu}\varepsilon_{\mu\alpha\beta\gamma}$ | Forming the rank-3 tensor that is dual to a vector (exercise 3.14). |
| ${}^{*}F_{\alpha\beta} = \frac{1}{2}F^{\mu\nu}\varepsilon_{\mu\nu\alpha\beta}$ | Forming the antisymmetric rank-2 tensor that is dual to a given antisymmetric rank-2 tensor (exercise 3.14). |
| ${}^{*}B_{\alpha} = \frac{1}{6}B^{\lambda\mu\nu}\varepsilon_{\lambda\mu\nu\alpha}$ | Forming the 1-form that is dual to an antisymmetric rank-3 tensor (exercise 3.14). |

**Exercise 3.11. SYMMETRIES**

Let $A_{\mu\nu}$ be an antisymmetric tensor so that $A_{\mu\nu} = -A_{\nu\mu}$; and let $S^{\mu\nu}$ be a symmetric tensor so that $S^{\mu\nu} = S^{\nu\mu}$.

(a) Justify the equations $A_{\mu\nu}S^{\mu\nu} = 0$ in two ways: first, by writing out the sum explicitly (all sixteen terms) and showing how the terms in the sum cancel in pairs; second, by giving an argument to justify each equals sign in the following string:

$$A_{\mu\nu}S^{\mu\nu} = -A_{\nu\mu}S^{\mu\nu} = -A_{\nu\mu}S^{\nu\mu} = -A_{\alpha\beta}S^{\alpha\beta} = -A_{\mu\nu}S^{\mu\nu} = 0.$$

(b) Establish the following two identities for any arbitrary tensor $V_{\mu\nu}$:

$$V^{\mu\nu}A_{\mu\nu} = \frac{1}{2}(V^{\mu\nu} - V^{\nu\mu})A_{\mu\nu}, \qquad V^{\mu\nu}S_{\mu\nu} = \frac{1}{2}(V^{\mu\nu} + V^{\nu\mu})S_{\mu\nu}.$$

**Exercise 3.12. SYMMETRIZATION AND ANTISYMMETRIZATION**

To "symmetrize" a tensor, one averages it with all of its transposes. The components of the new, symmetrized tensor are distinguished by round brackets:

$$\begin{aligned} V_{(\mu\nu)} &\equiv \frac{1}{2}(V_{\mu\nu} + V_{\nu\mu}); \\ V_{(\mu\nu\lambda)} &\equiv \frac{1}{3!}(V_{\mu\nu\lambda} + V_{\nu\lambda\mu} + V_{\lambda\mu\nu} + V_{\nu\mu\lambda} + V_{\mu\lambda\nu} + V_{\lambda\nu\mu}). \end{aligned} \tag{3.46}$$

One "antisymmetrizes" a tensor (square brackets) similarly:

$$\begin{aligned} V_{[\mu\nu]} &\equiv \frac{1}{2}(V_{\mu\nu} - V_{\nu\mu}); \\ V_{[\mu\nu\lambda]} &\equiv \frac{1}{3!}(V_{\mu\nu\lambda} + V_{\nu\lambda\mu} + V_{\lambda\mu\nu} - V_{\nu\mu\lambda} - V_{\mu\lambda\nu} - V_{\lambda\nu\mu}). \end{aligned} \tag{3.47}$$

(a) Show that such symmetrized and antisymmetrized tensors are, indeed, symmetric and antisymmetric under interchange of the vectors inserted into their slots:

$$\begin{aligned} V_{(\alpha\beta\gamma)}u^\alpha v^\beta w^\gamma &= +V_{(\alpha\beta\gamma)}v^\alpha u^\beta w^\gamma = \cdots, \\ V_{[\alpha\beta\gamma]}u^\alpha v^\beta w^\gamma &= -V_{[\alpha\beta\gamma]}v^\alpha u^\beta w^\gamma = \cdots. \end{aligned}$$

(b) Show that a second-rank tensor can be reconstructed from its symmetric and antisymmetric parts,

$$V_{\mu\nu} = V_{(\mu\nu)} + V_{[\mu\nu]}, \tag{3.48}$$

but that a third-rank tensor cannot; $V_{(\alpha\beta\gamma)}$ and $V_{[\alpha\beta\gamma]}$ contain together "less information" than $V_{\alpha\beta\gamma}$. "Young diagrams" (see, e.g., Messiah [1961], appendix D) describe other symmetries, more subtle than these two, which contain the missing information.

(c) Show that the electromagnetic field tensor satisfies

$$F_{(\alpha\beta)} = 0, \qquad F_{\alpha\beta} = F_{[\alpha\beta]}. \tag{3.49a}$$

(d) Show that Maxwell's "magnetic" equations

$$F_{\alpha\beta,\gamma} + F_{\beta\gamma,\alpha} + F_{\gamma\alpha,\beta} = 0$$

can be rewritten in the form

$$F_{[\alpha\beta,\gamma]} = 0. \tag{3.49b}$$

**Exercise 3.13. LEVI-CIVITA TENSOR**

The "Levi-Civita tensor" $\boldsymbol{\varepsilon}$ in spacetime is a fourth-rank, completely antisymmetric tensor:

$$\boldsymbol{\varepsilon}(\boldsymbol{n}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \text{ changes sign when any two of the vectors are interchanged.} \tag{3.50a}$$

Choose an arbitrary but specific Lorentz frame, with $\boldsymbol{e}_0$ pointing toward the future and with $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ a righthanded set of spatial basis vectors. The covariant components of $\boldsymbol{\varepsilon}$ in this frame are

$$\varepsilon_{0123} = \boldsymbol{\varepsilon}(\boldsymbol{e}_0, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3) = +1. \tag{3.50b}$$

[Note: In an $n$-dimensional space, $\boldsymbol{\varepsilon}$ is the analogous completely antisymmetric rank-$n$ tensor. Its components are

$$\varepsilon_{12\ldots n} = \boldsymbol{\varepsilon}(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n) = +1, \tag{3.50c}$$

when computed on a "positively oriented," orthonormal basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$.]

(a) Use the antisymmetry to show that

$$\varepsilon_{\alpha\beta\gamma\delta} = 0 \text{ unless } \alpha, \beta, \gamma, \delta \text{ are all different,} \tag{3.50d}$$

$$\varepsilon_{\pi_0\pi_1\pi_2\pi_3} = \begin{cases} +1 \text{ for even permutations of } 0, 1, 2, 3, \text{ and} \\ -1 \text{ for odd permutations.} \end{cases} \tag{3.50e}$$

(b) Show that

$$\varepsilon^{\pi_0\pi_1\pi_2\pi_3} = -\varepsilon_{\pi_0\pi_1\pi_2\pi_3}. \tag{3.50f}$$

(c) By means of a Lorentz transformation show that $\varepsilon^{\bar{\alpha}\bar{\beta}\bar{\gamma}\bar{\delta}}$ and $\varepsilon_{\bar{\alpha}\bar{\beta}\bar{\gamma}\bar{\delta}}$ have these same values in any other Lorentz frame with $\boldsymbol{e}_{\bar{0}}$ pointing toward the future and with $\boldsymbol{e}_{\bar{1}}$, $\boldsymbol{e}_{\bar{2}}$, $\boldsymbol{e}_{\bar{3}}$ a righthanded set. Hint: show that

$$\varepsilon^{\alpha\beta\gamma\delta}\Lambda^{\bar{0}}{}_{\alpha}\Lambda^{\bar{1}}{}_{\beta}\Lambda^{\bar{2}}{}_{\gamma}\Lambda^{\bar{3}}{}_{\delta} = -\det|\Lambda^{\bar{\mu}}{}_{\nu}|; \tag{3.50g}$$

from $\Lambda^T\eta\Lambda = \eta$, show that $\det|\Lambda^{\bar{\mu}}{}_{\nu}| = \pm 1$; and verify that the determinant is $+1$ for transformations between frames with $\boldsymbol{e}_0$ and $\boldsymbol{e}_{\bar{0}}$ future-pointing, and with $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ and $\boldsymbol{e}_{\bar{1}}$, $\boldsymbol{e}_{\bar{2}}$, $\boldsymbol{e}_{\bar{3}}$ righthanded.

(d) What are the components of $\boldsymbol{\varepsilon}$ in a Lorentz frame with past-pointing $\boldsymbol{e}_{\bar{0}}$? with lefthanded $\boldsymbol{e}_{\bar{1}}$, $\boldsymbol{e}_{\bar{2}}$, $\boldsymbol{e}_{\bar{3}}$?

(e) From the Levi-Civita tensor, one can construct several "permutation tensors." In index notation:

$$\delta^{\alpha\beta\gamma}{}_{\mu\nu\lambda} \equiv -\varepsilon^{\alpha\beta\gamma\rho}\varepsilon_{\mu\nu\lambda\rho}; \tag{3.50h}$$

$$\delta^{\alpha\beta}{}_{\mu\nu} \equiv \frac{1}{2}\,\delta^{\alpha\beta\lambda}{}_{\mu\nu\lambda} = -\frac{1}{2}\varepsilon^{\alpha\beta\lambda\rho}\varepsilon_{\mu\nu\lambda\rho}; \tag{3.50i}$$

$$\delta^{\alpha}{}_{\mu} \equiv \frac{1}{3}\,\delta^{\alpha\beta}{}_{\mu\beta} = \frac{1}{6}\,\delta^{\alpha\beta\lambda}{}_{\mu\beta\lambda} = -\frac{1}{6}\,\varepsilon^{\alpha\beta\lambda\rho}\varepsilon_{\mu\beta\lambda\rho}. \tag{3.50j}$$

Show that:

$$\delta^{\alpha\beta\gamma}{}_{\mu\nu\lambda} = \begin{cases} +1 \text{ if } \alpha\beta\gamma \text{ is an even permutation of } \mu\nu\lambda, \\ -1 \text{ if } \alpha\beta\gamma \text{ is an odd permutation of } \mu\nu\lambda, \\ \;\;0 \text{ otherwise;} \end{cases} \tag{3.50k}$$

$$\delta^{\alpha\beta}{}_{\mu\nu} = \delta^{\alpha}{}_{\mu}\delta^{\beta}{}_{\nu} - \delta^{\alpha}{}_{\nu}\delta^{\beta}{}_{\mu}$$
$$= \begin{cases} +1 \text{ if } \alpha\beta \text{ is an even permutation of } \mu\nu, \\ -1 \text{ if } \alpha\beta \text{ is an odd permutation of } \mu\nu, \\ \ \ 0 \text{ otherwise;} \end{cases} \tag{3.50l}$$

$$\delta^{\alpha}{}_{\mu} = \begin{cases} +1 \text{ if } \alpha = \mu, \\ \ \ 0 \text{ otherwise.} \end{cases} \tag{3.50m}$$

**Exercise 3.14. DUALS**

From any vector $\boldsymbol{J}$, any second-rank antisymmetric tensor $\boldsymbol{F}(F_{\alpha\beta} = F_{[\alpha\beta]})$, and any third-rank antisymmetric tensor $\boldsymbol{B}(B_{\alpha\beta\gamma} = B_{[\alpha\beta\gamma]})$, one can construct new tensors defined by

$$^*J_{\alpha\beta\gamma} = J^{\mu}\varepsilon_{\mu\alpha\beta\gamma}, \qquad ^*F_{\alpha\beta} = \frac{1}{2}F^{\mu\nu}\varepsilon_{\mu\nu\alpha\beta}, \qquad ^*B_{\alpha} = \frac{1}{3!}B^{\lambda\mu\nu}\varepsilon_{\lambda\mu\nu\alpha}. \tag{3.51}$$

One calls $^*\boldsymbol{J}$ the *"dual"* of $\boldsymbol{J}$, $^*\boldsymbol{F}$ the dual of $\boldsymbol{F}$, and $^*\boldsymbol{B}$ the dual of $\boldsymbol{B}$. [A previous and entirely distinct use of the word "dual" (§2.7) called a *set* of basis one-forms $\{\boldsymbol{\omega}^{\alpha}\}$ dual to a *set* of basis vectors $\{\boldsymbol{e}_{\alpha}\}$ if $\langle\boldsymbol{\omega}^{\alpha}, \boldsymbol{e}_{\beta}\rangle = \delta^{\alpha}{}_{\beta}$. Fortunately there are no grounds for confusion between the two types of duality. One relates sets of vectors to sets of 1-forms. The other relates antisymmetric tensors of rank $p$ to antisymmetric tensors of rank $4 - p$.]

(a) Show that

$$^{**}\boldsymbol{J} = \boldsymbol{J}, \qquad ^{**}\boldsymbol{F} = -\boldsymbol{F}, \qquad ^{**}\boldsymbol{B} = \boldsymbol{B}. \tag{3.52}$$

so (aside from sign) one can recover any completely antisymmetric tensor $\boldsymbol{H}$ from its dual $^*\boldsymbol{H}$ by taking the dual once again, $^{**}\boldsymbol{H}$. This shows that $\boldsymbol{H}$ and $^*\boldsymbol{H}$ contain precisely the same information.

(b) Make explicit this fact of same-information-content by writing out the components $^*A^{\alpha\beta\gamma}$ in terms of $A^{\alpha}$, also $^*F^{\alpha\beta}$ in terms of $F^{\alpha\beta}$, also $^*B^{\alpha}$ in terms of $B^{\alpha\beta\gamma}$.

**Exercise 3.15. GEOMETRIC VERSIONS OF MAXWELL EQUATIONS**

Show that, if $\boldsymbol{F}$ is the electromagnetic field tensor, then $\boldsymbol{\nabla}\cdot{}^*\boldsymbol{F} = 0$ is a geometric frame-independent version of the Maxwell equations

$$F_{\alpha\beta,\gamma} + F_{\beta\gamma,\alpha} + F_{\gamma\alpha,\beta} = 0.$$

Similarly show that $\boldsymbol{\nabla}\cdot\boldsymbol{F} = 4\pi\boldsymbol{J}$ (divergence on second slot of $\boldsymbol{F}$) is a geometric version of $F^{\alpha\beta}{}_{,\beta} = 4\pi J^{\alpha}$.

**Exercise 3.16. CHARGE CONSERVATION**

From Maxwell's equations $F^{\alpha\beta}{}_{,\beta} = 4\pi J^{\alpha}$, derive the "equation of charge conservation"

$$J^{\alpha}{}_{,\alpha} = 0. \tag{3.53}$$

Show that this equation does, indeed, correspond to conservation of charge. It will be studied further in Chapter 5.

**Exercise 3.17. VECTOR POTENTIAL**

The vector potential $\boldsymbol{A}$ of electromagnetic theory generates the electromagnetic field tensor via the geometric equation

$$\boldsymbol{F} = -(\text{antisymmetric part of } \boldsymbol{\nabla}\boldsymbol{A}), \tag{3.54}$$

i.e.,

$$F_{\mu\nu} = A_{\nu,\mu} - A_{\mu,\nu}. \tag{3.54'}$$

(a) Show that the electric and magnetic fields in a specific Lorentz frame are given by

$$\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A}, \qquad \boldsymbol{E} = -\partial \boldsymbol{A}/\partial t - \boldsymbol{\nabla} A^0. \tag{3.55}$$

(b) Show that $\boldsymbol{F}$ will satisfy Maxwell's equations if and only if $\boldsymbol{A}$ satisfies

$$A^{\alpha,\mu}{}_{,\mu} - A^{\mu}{}_{,\mu}{}^{,\alpha} = -4\pi J^\alpha. \tag{3.56}$$

(c) Show that "gauge transformations"

$$\boldsymbol{A}_{\text{NEW}} = \boldsymbol{A}_{\text{OLD}} + \boldsymbol{d}\phi, \qquad \phi = \text{arbitrary function,} \tag{3.57}$$

leave $\boldsymbol{F}$ unaffected.

(d) Show that one can adjust the gauge so that

$$\boldsymbol{\nabla} \cdot \boldsymbol{A} = 0 \qquad \text{("Lorentz gauge"),} \tag{3.58a}$$

$$\Box \boldsymbol{A} = -4\pi \boldsymbol{J}. \tag{3.58b}$$

Here $\Box$ is the wave operator ("d'Alembertian"):

$$\Box \boldsymbol{A} = A^{\alpha,\mu}{}_{,\mu} \boldsymbol{e}_\alpha. \tag{3.59}$$

**Exercise 3.18. DIVERGENCE OF ELECTROMAGNETIC STRESS-ENERGY TENSOR**

From an electromagnetic field tensor $\boldsymbol{F}$, one constructs a second-rank, symmetric tensor $\boldsymbol{T}$ ("stress-energy tensor," to be studied in Chapter 5) as follows:

$$T^{\mu\nu} = \frac{1}{4\pi}\left(F^{\mu\alpha}F^{\nu}{}_{\alpha} - \frac{1}{4}\eta^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}\right). \tag{3.60}$$

As an exercise in index gymnastics:

(a) Show that $\boldsymbol{\nabla} \cdot \boldsymbol{T}$ has components

$$T^{\mu\nu}{}_{,\nu} = \frac{1}{4\pi}\left[F^{\mu\alpha}{}_{,\nu}F^{\nu}{}_{\alpha} + F^{\mu\alpha}F^{\nu}{}_{\alpha,\nu} - \frac{1}{2}F_{\alpha\beta}{}^{,\mu}F^{\alpha\beta}\right]. \tag{3.61}$$

(b) Manipulate this expression into the form

$$T_{\mu}{}^{\nu}{}_{,\nu} = \frac{1}{4\pi}\left[-F_{\mu\alpha}F^{\alpha\nu}{}_{,\nu} - \frac{1}{2}F^{\alpha\beta}(F_{\alpha\beta,\mu} + F_{\mu\alpha,\beta} + F_{\beta\mu,\alpha})\right]; \tag{3.62}$$

note that the first term of (3.62) arises directly from the second term of (3.61).

(c) Use Maxwell's equations to conclude that

$$T^{\mu\nu}{}_{,\nu} = -F^{\mu\alpha}J_\alpha. \tag{3.63}$$

2

CHAPTER 4

# ELECTROMAGNETISM AND DIFFERENTIAL FORMS

*The ether trembled at his agitations*
*In a manner so familiar that I only need to say,*
*In accordance with Clerk Maxwell's six equations*
*It tickled peoples' optics far away.*
*You can feel the way it's done,*
*You may trace them as they run—*
$d\gamma$ *by* $dy$ *less* $d\beta$ *by* $dz$ *is equal* $KdX/dt$ . . .

*While the curl of* (X, Y, Z) *is the*
*minus* d/dt *of the vector* (a, b, c).

From *The Revolution of the Corpuscle,*
written by A. A. Robb
(to the tune of *The Interfering Parrott*)
for a dinner of the research students
of the Cavendish Laboratory
in the days of the old mathematics.

**This chapter is all Track 2. It is needed as preparation for §§14.5 and 14.6 (computation of curvature using differential forms) and for Chapter 15 (Bianchi identities and boundary of a boundary), but is not needed for the rest of the book.**

## §4.1. EXTERIOR CALCULUS

Stacks of surfaces, individually or intersecting to make "honeycombs," "egg crates," and other such structures ("differential forms"), give unique insight into the geometry of electromagnetism and gravitation. However, such insight comes at some cost in time. Therefore, most readers should skip this chapter and later material that depends on it during a first reading of this book.

Analytically speaking, differential forms are completely antisymmetric tensors; pictorially speaking, they are intersecting stacks of surfaces. The mathematical formalism for manipulating differential forms with ease, called "exterior calculus," is summarized concisely in Box 4.1; its basic features are illustrated in the rest of this chapter by rewriting electromagnetic theory in its language. An effective way to tackle this chapter might be to (1) scan Box 4.1 to get the flavor of the formalism; (2) read the rest of the chapter in detail; (3) restudy Box 4.1 carefully; (4) get practice in manipulating the formalism by working the exercises.*

(*continued on page 99*)

*Exterior calculus is treated in greater detail than here by: É. Cartan (1945); de Rham (1955); Nickerson, Spencer, and Steenrod (1959); Hauser (1970); Israel (1970); especially Flanders (1963, relatively easy, with many applications); Spivak (1965, sophomore or junior level, but fully in tune with modern mathematics); H. Cartan (1970); and Choquet-Bruhat (1968a).

**Box 4.1 DIFFERENTIAL FORMS AND EXTERIOR CALCULUS IN BRIEF**

The fundamental definitions and formulas of exterior calculus are summarized here for ready reference. Each item consists of a general statement (at left of page) plus a leading application (at right of page). This formalism is applicable not only to spacetime, but also to more general geometrical systems (see heading of each section). No attempt is made here to demonstrate the internal consistency of the formalism, nor to derive it from any set of definitions and axioms. For a systematic treatment that does so, see, e.g., Spivak (1965), or Misner and Wheeler (1957).

## A. Algebra I (applicable to any vector space)

1. *Basis 1-forms.*
   a. Coordinate basis $\boldsymbol{\omega}^j = \boldsymbol{d}x^j$
      ($j$ tells which 1-form, not which component).
   b. General basis $\boldsymbol{\omega}^j = L^j{}_{k'}\,\boldsymbol{d}x^{k'}$.

*An application*
Simple basis 1-forms for analyzing Schwarzschild geometry around static spherically symmetric center of attraction:

$$\begin{aligned}\boldsymbol{\omega}^0 &= (1 - 2m/r)^{1/2}\,\boldsymbol{d}t;\\ \boldsymbol{\omega}^1 &= (1 - 2m/r)^{-1/2}\,\boldsymbol{d}r;\\ \boldsymbol{\omega}^2 &= r\,\boldsymbol{d}\theta;\\ \boldsymbol{\omega}^3 &= r\sin\theta\,\boldsymbol{d}\phi.\end{aligned}$$

2. *General* p*-form* (*or* p*-vector*) is a completely antisymmetric tensor of rank $\binom{0}{p}$ [or $\binom{p}{0}$]. It can be expanded in terms of wedge products (see §3.5 and exercise 4.12):

$$\begin{aligned}\boldsymbol{\alpha} &= \frac{1}{p!}\alpha_{i_1 i_2 \ldots i_p}\boldsymbol{\omega}^{i_1} \wedge \boldsymbol{\omega}^{i_2} \wedge \cdots \wedge \boldsymbol{\omega}^{i_p}\\ &\equiv \alpha_{|i_1 i_2 \ldots i_p|}\boldsymbol{\omega}^{i_1} \wedge \boldsymbol{\omega}^{i_2} \wedge \cdots \wedge \boldsymbol{\omega}^{i_p}.\end{aligned}$$

(Note: Vertical bars around the indices mean summation extends only over $i_1 < i_2 < \cdots < i_p$.)

*Two applications*
Energy-momentum 1-form is of type $\boldsymbol{\alpha} = \alpha_i \boldsymbol{\omega}^i$ or

$$\boldsymbol{p} = -E\,\boldsymbol{d}t + p_x\,\boldsymbol{d}x + p_y\,\boldsymbol{d}y + p_z\,\boldsymbol{d}z.$$

***Faraday*** is a 2-form of type $\boldsymbol{\beta} = \beta_{|\mu\nu|}\boldsymbol{\omega}^\mu \wedge \boldsymbol{\omega}^\nu$ or in flat spacetime

$$\begin{aligned}\boldsymbol{F} = &-E_x\,\boldsymbol{d}t \wedge \boldsymbol{d}x - E_y\,\boldsymbol{d}t \wedge \boldsymbol{d}y - E_z\,\boldsymbol{d}t \wedge \boldsymbol{d}z\\ &+ B_x\,\boldsymbol{d}y \wedge \boldsymbol{d}z + B_y\,\boldsymbol{d}z \wedge \boldsymbol{d}x + B_z\,\boldsymbol{d}x \wedge \boldsymbol{d}y\end{aligned}$$

**Box 4.1 (*continued*)**

3. *Wedge product.*
   *All* familiar rules of addition and multiplication hold, such as

$$(a\boldsymbol{\alpha} + b\boldsymbol{\beta}) \wedge \boldsymbol{\gamma} = a\boldsymbol{\alpha} \wedge \boldsymbol{\gamma} + b\boldsymbol{\beta} \wedge \boldsymbol{\gamma},$$
$$(\boldsymbol{\alpha} \wedge \boldsymbol{\beta}) \wedge \boldsymbol{\gamma} = \boldsymbol{\alpha} \wedge (\boldsymbol{\beta} \wedge \boldsymbol{\gamma}) \equiv \boldsymbol{\alpha} \wedge \boldsymbol{\beta} \wedge \boldsymbol{\gamma},$$

   *except* for a modified commutation law between a $p$-form $\boldsymbol{\alpha}$ and a $q$-form $\boldsymbol{\beta}$:

$$\underset{p}{\boldsymbol{\alpha}} \wedge \underset{q}{\boldsymbol{\beta}} = (-1)^{pq} \underset{q}{\boldsymbol{\beta}} \wedge \underset{p}{\boldsymbol{\alpha}}.$$

*Applications to 1-forms $\boldsymbol{\alpha}, \boldsymbol{\beta}$:*

$$\boldsymbol{\alpha} \wedge \boldsymbol{\beta} = -\boldsymbol{\beta} \wedge \boldsymbol{\alpha}, \qquad \boldsymbol{\alpha} \wedge \boldsymbol{\alpha} = 0;$$
$$\boldsymbol{\alpha} \wedge \boldsymbol{\beta} = (\alpha_j \boldsymbol{\omega}^j) \wedge (\beta_k \boldsymbol{\omega}^k) = \alpha_j \beta_k \boldsymbol{\omega}^j \wedge \boldsymbol{\omega}^k$$
$$= \frac{1}{2}(\alpha_j \beta_k - \beta_j \alpha_k) \boldsymbol{\omega}^j \wedge \boldsymbol{\omega}^k.$$

4. *Contraction of* p*-form on* p*-vector.*

$$\langle \underset{p}{\boldsymbol{\alpha}}, \underset{p}{\boldsymbol{A}} \rangle$$
$$= \alpha_{|i_1 \ldots i_p|} A^{|j_1 \ldots j_p|} \underbrace{\langle \boldsymbol{\omega}^{i_1} \wedge \cdots \wedge \boldsymbol{\omega}^{i_p}, \boldsymbol{e}_{j_1} \wedge \cdots \wedge \boldsymbol{e}_{j_p} \rangle}_{[\equiv \delta^{i_1 \ldots i_p}_{j_1 \ldots j_p} \text{ (see exercises 3.13 and 4.12)}]}$$
$$= \alpha_{|i_1 \ldots i_p|} A^{i_1 \ldots i_p}.$$

*Four applications*

a. Contraction of a particle's energy-momentum 1-form $\boldsymbol{p} = p_\alpha \boldsymbol{\omega}^\alpha$ with 4-velocity $\boldsymbol{u} = u^\alpha \boldsymbol{e}_\alpha$ of observer (a 1-vector):

$$-\langle \boldsymbol{p}, \boldsymbol{u} \rangle = -p_\alpha u^\alpha = \text{energy of particle}.$$

b. Contraction of ***Faraday*** 2-form $\boldsymbol{F}$ with bivector $\delta\mathscr{P} \wedge \Delta\mathscr{P}$ [where $\delta\mathscr{P} = (d\mathscr{P}/d\lambda_1)\Delta\lambda_1$ and $\Delta\mathscr{P} = (d\mathscr{P}/d\lambda_2)\Delta\lambda_2$ are two infinitesimal vectors in a 2-surface $\mathscr{P}(\lambda_1, \lambda_2)$, and the bivector represents the surface element they span] is the magnetic flux $\boldsymbol{\Phi} = \langle \boldsymbol{F}, \delta\mathscr{P} \wedge \Delta\mathscr{P} \rangle$ through that surface element.

c. More generally, a $p$-dimensional parallelepiped with vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_p$ for legs has an oriented volume described by the "*simple*" $p$-vector $\boldsymbol{a}_1 \wedge \boldsymbol{a}_2 \wedge \cdots \boldsymbol{a}_p$ (oriented because interchange of two legs changes its sign). An egg-crate type of structure with walls made from the hyperplanes of $p$ different 1-forms $\boldsymbol{\sigma}^1$,

$\boldsymbol{\sigma}^2, \ldots, \boldsymbol{\sigma}^p$ is described by the "*simple*" $p$-form $\boldsymbol{\sigma}^1 \wedge \boldsymbol{\sigma}^2 \wedge \cdots \wedge \boldsymbol{\sigma}^p$. The number of cells of $\boldsymbol{\sigma}^1 \wedge \boldsymbol{\sigma}^2 \wedge \cdots \wedge \boldsymbol{\sigma}^p$ sliced through by the infinitesimal $p$-volume $\boldsymbol{a}_1 \wedge \boldsymbol{a}_2 \wedge \cdots \wedge \boldsymbol{a}_p$ is

$$\langle \boldsymbol{\sigma}^1 \wedge \boldsymbol{\sigma}^2 \wedge \cdots \wedge \boldsymbol{\sigma}^p, \boldsymbol{a}_1 \wedge \boldsymbol{a}_2 \wedge \cdots \wedge \boldsymbol{a}_p \rangle.$$

d. The Jacobian determinant of a set of $p$ functions $f^k(x^1, \ldots, x^n)$ with respect to $p$ of their arguments is

$$\left\langle \boldsymbol{d}f^1 \wedge \boldsymbol{d}f^2 \wedge \cdots \wedge \boldsymbol{d}f^p, \frac{\partial \mathscr{P}}{\partial x^1} \wedge \frac{\partial \mathscr{P}}{\partial x^2} \wedge \cdots \wedge \frac{\partial \mathscr{P}}{\partial x^p} \right\rangle = \det \left\| \left( \frac{\partial f^k}{\partial x^j} \right) \right\| \equiv \frac{\partial(f^1, f^2, \ldots, f^p)}{\partial(x^1, x^2, \ldots, x^p)}.$$

5. *Simple forms.*
   a. A simple $p$-form is one that can be written as a wedge product of $p$ 1-forms:

$$\underset{p}{\boldsymbol{\sigma}} = \underbrace{\boldsymbol{\alpha} \wedge \boldsymbol{\beta} \wedge \cdots \wedge \boldsymbol{\gamma}}_{p \text{ factors}}.$$

   b. A simple $p$-form $\boldsymbol{\alpha} \wedge \boldsymbol{\beta} \wedge \cdots \wedge \boldsymbol{\gamma}$ is represented by the intersecting families of surfaces of $\boldsymbol{\alpha}, \boldsymbol{\beta}, \ldots, \boldsymbol{\gamma}$ (egg-crate structure) plus a sense of circulation (orientation).

*Applications:*

a. In four dimensions (e.g., spacetime) all 0-forms, 1-forms, 3-forms, and 4-forms are simple. A 2-form $\boldsymbol{F}$ is generally a sum of two simple forms, e.g., $\boldsymbol{F} = -e\, \boldsymbol{d}t \wedge \boldsymbol{d}x + h\, \boldsymbol{d}y \wedge \boldsymbol{d}z$; it is simple if and only if $\boldsymbol{F} \wedge \boldsymbol{F} = 0$.

b. A set of 1-forms $\boldsymbol{\alpha}, \boldsymbol{\beta}, \ldots, \boldsymbol{\gamma}$ is linearly dependent (one a linear combination of the others) if and only if

$$\boldsymbol{\alpha} \wedge \boldsymbol{\beta} \wedge \cdots \wedge \boldsymbol{\gamma} = 0 \qquad \text{(egg crate collapsed)}.$$

## B. Exterior Derivative (applicable to any "differentiable manifold," with or without metric)

1. $\boldsymbol{d}$ produces a $(p + 1)$-form $\boldsymbol{d\sigma}$ from a $p$-form $\boldsymbol{\sigma}$.
2. Effect of $\boldsymbol{d}$ is defined by induction using the

**Box 4.1** (*continued*)

(Chapter 2) definition of $\boldsymbol{d}f$, and $f$ a function (0-form), plus

$$\boldsymbol{d}(\underset{p}{\boldsymbol{\alpha}} \wedge \underset{q}{\boldsymbol{\beta}}) = \boldsymbol{d\alpha} \wedge \boldsymbol{\beta} + (-1)^p \boldsymbol{\alpha} \wedge \boldsymbol{d\beta},$$

$$\boldsymbol{d}^2 = \boldsymbol{dd} = 0.$$

*Two applications*

$$\boldsymbol{d}(\boldsymbol{\alpha} \wedge \boldsymbol{d\beta}) = \boldsymbol{d\alpha} \wedge \boldsymbol{d\beta}.$$

For the $p$-form $\boldsymbol{\phi}$, with

$$\boldsymbol{\phi} = \phi_{|i_1 \ldots i_p|}\, \boldsymbol{d}x^{i_1} \wedge \cdots \wedge \boldsymbol{d}x^{i_p},$$

one has (alternative and equivalent definition of $\boldsymbol{d\phi}$)

$$\boldsymbol{d\phi} = \boldsymbol{d}\phi_{|i_1 \ldots i_p|} \wedge \boldsymbol{d}x^{i_1} \wedge \cdots \wedge \boldsymbol{d}x^{i_p}.$$

## C. Integration (applicable to any "differentiable manifold," with or without metric)

1. *Pictorial interpretation.*
   Text and pictures of Chapter 4 interpret $\int \boldsymbol{\alpha}$ (integral of specified 1-form $\boldsymbol{\alpha}$ along specified curve from specified starting point to specified end point) as "number of $\boldsymbol{\alpha}$-surfaces pierced on that route"; similarly, they interpret $\int \boldsymbol{\phi}$ (integral of specified 2-form $\boldsymbol{\phi}$ over specified bit of surface on which there is an assigned sense of circulation or "orientation") as "number of cells of the honeycomb-like structure $\boldsymbol{\phi}$ cut through by that surface"; similarly for the egg-crate-like structures that represent 3-forms; etc.
2. *Computational rules for integration.*
   To evaluate $\int \boldsymbol{\alpha}$, the integral of a $p$-form

   $$\boldsymbol{\alpha} = \alpha_{|i_1 \ldots i_p|}(x^1, \ldots, x^n)\, \boldsymbol{d}x^{i_1} \wedge \cdots \wedge \boldsymbol{d}x^{i_p},$$

   over a $p$-dimensional surface, proceed in two steps.
   a. Substitute a parameterization of the surface,

   $$x^k(\lambda^1, \ldots, \lambda^p)$$

   into $\boldsymbol{\alpha}$, and collect terms in the form

   $$\boldsymbol{\alpha} = a(\lambda^j)\, \boldsymbol{d}\lambda^1 \wedge \cdots \wedge \boldsymbol{d}\lambda^p$$

   (this is $\boldsymbol{\alpha}$ viewed as a $p$-form in the $p$-dimensional surface);

b. Integrate

$$\int \boldsymbol{\alpha} = \int a(\lambda^j)\, d\lambda^1\, d\lambda^2 \ldots d\lambda^p$$

using elementary definition of integration.

*Example:* See equations (4.12) to (4.14).

3. *The differential geometry of integration.* Calculate $\int \boldsymbol{\alpha}$ for a $p$-form $\boldsymbol{\alpha}$ as follows.

a. Choose the $p$-dimensional surface $\mathcal{S}$ over which to integrate.

b. Represent $\mathcal{S}$ by a parametrization giving the generic point of the surface as a function of the parameters, $\mathcal{P}(\lambda^1, \lambda^2, \ldots \lambda^p)$. This fixes the orientation. The same function with $\lambda^1 \leftrightarrow \lambda^2$, $\mathcal{P}(\lambda^2, \lambda^1, \ldots, \lambda^p)$, describes a different (i.e., oppositely oriented) surface, $-\mathcal{S}$.

c. The infinitesimal parallelepiped

$$\left(\frac{\partial \mathcal{P}}{\partial \lambda^1}\,\Delta\lambda^1\right) \wedge \left(\frac{\partial \mathcal{P}}{\partial \lambda^2}\,\Delta\lambda^2\right) \wedge \cdots \wedge \left(\frac{\partial \mathcal{P}}{\partial \lambda^p}\,\Delta\lambda^p\right)$$

is tangent to the surface. The number of cells of $\boldsymbol{\alpha}$ it slices is

$$\left\langle \boldsymbol{\alpha}, \frac{\partial \mathcal{P}}{\partial \lambda^1} \wedge \cdots \wedge \frac{\partial \mathcal{P}}{\partial \lambda^p}\right\rangle \Delta\lambda^1 \ldots \Delta\lambda^p.$$

This number changes sign if two of the vectors $\partial\mathcal{P}/\partial\lambda^k$ are interchanged, as for an oppositely oriented surface.

d. The above provides an interpretation motivating the definition

$$\int \boldsymbol{\alpha} \equiv \int\!\!\int \cdots \int \left\langle \boldsymbol{\alpha}, \frac{\partial \mathcal{P}}{\partial \lambda^1} \wedge \frac{\partial \mathcal{P}}{\partial \lambda^2} \wedge \cdots \wedge \frac{\partial \mathcal{P}}{\partial \lambda^p}\right\rangle d\lambda^1\, d\lambda^2 \ldots d\lambda^p.$$

This definition is identified with the computational rule of the preceding section (C.2) in exercise 4.9.

*An application*

Integrate a gradient $\boldsymbol{d}f$ along a curve, $\mathcal{P}(\lambda)$ from $\mathcal{P}(0)$ to $\mathcal{P}(1)$:

$$\int \boldsymbol{d}f = \int_0^1 \langle \boldsymbol{d}f, d\mathcal{P}/d\lambda\rangle\, d\lambda = \int_0^1 (df/d\lambda)\, d\lambda$$
$$= f[\mathcal{P}(1)] - f[\mathcal{P}(0)].$$

e. Three different uses for symbol "$d$": *First,* lightface $d$ in explicit derivative expressions such as

**Box 4.1** ***(continued)***

$d/da$, or $df/da$, or $d\mathscr{P}/da$; neither numerator nor denominator alone has any meaning, but only the full string of symbols. *Second,* lightface $d$ inside an integral sign; e.g., $\int f\,da$. This is an instruction to perform integration, and has no meaning whatsoever without an integral sign; "$\int \ldots d \ldots$" lives as an indivisible unit. *Third,* sans-serif $\boldsymbol{d}$; e.g., $\boldsymbol{d}$ alone, or $\boldsymbol{d}f$, or $\boldsymbol{d\alpha}$. This is an exterior derivative, which converts a $p$-form into a $(p+1)$-form. Sometimes lightface $d$ is used for the same purpose. Hence, $d$ alone, or $df$, or $dx$, is always an exterior derivative unless coupled to an $\int$ sign (*second* use), or coupled to a / sign (*first* use).

4. *The generalized Stokes theorem* (see Box 4.6).
   a. Let $\partial\mathscr{V}$ be the closed $p$-dimensional boundary of a $(p+1)$-dimensional surface $\mathscr{V}$. Let $\boldsymbol{\sigma}$ be a $p$-form defined throughout $\mathscr{V}$.

   Then

$$\boxed{\int_{\mathscr{V}} \boldsymbol{d\sigma} = \int_{\partial\mathscr{V}} \boldsymbol{\sigma}}$$

   [integral of $p$-form $\boldsymbol{\sigma}$ over boundary $\partial\mathscr{V}$ equals integral of $(p+1)$-form $\boldsymbol{d\sigma}$ over interior $\mathscr{V}$].

   b. For the sign to come out right, orientations of $\mathscr{V}$ and $\partial\mathscr{V}$ must agree in this sense: choose coordinates $y^0, y^1, \ldots, y^p$ on a portion of $\mathscr{V}$, with $y^0$ specialized so $y^0 \leq 0$ in $\mathscr{V}$, and $y^0 = 0$ at the boundary $\partial\mathscr{V}$; then the orientation

$$\frac{\partial\mathscr{P}}{\partial y^0} \wedge \frac{\partial\mathscr{P}}{\partial y^1} \wedge \cdots \wedge \frac{\partial\mathscr{P}}{\partial y^p}$$

   for $\mathscr{V}$ demands the orientation

$$\frac{\partial\mathscr{P}}{\partial y^1} \wedge \cdots \wedge \frac{\partial\mathscr{P}}{\partial y^p}$$

   for $\partial\mathscr{V}$.

   c. Note: For a nonorientable surface, such as a Möbius strip, where a consistent and continuous choice of orientation is impossible, more intricate mathematics is required to give a definition of "$\partial$" for which the Stokes theorem holds.

*Applications:* Includes as special cases all integral theorems for surfaces of arbitrary dimension in spaces of arbitrary dimension, with or without metric, generalizing all versions of theorems of Stokes and Gauss. Examples:

a. $\mathscr{V}$ a curve, $\partial\mathscr{V}$ its endpoints, $\boldsymbol{\sigma} = f$ a 0-form (function):

$$\int_{\mathscr{V}} \boldsymbol{d}f = \int_0^1 (df/d\lambda)\, d\lambda = \int_{\partial\mathscr{V}} f = f(1) - f(0).$$

b. $\mathscr{V}$ a 2-surface in 3-space, $\partial\mathscr{V}$ its closed-curve boundary, $\boldsymbol{v}$ a 1-form; translated into Euclidean vector notation, the two integrals are

$$\int_{\mathscr{V}} \boldsymbol{dv} = \int_{\mathscr{V}} (\boldsymbol{\nabla} \times \boldsymbol{v}) \cdot \boldsymbol{dS}; \int_{\partial\mathscr{V}} \boldsymbol{v} = \int_{\partial\mathscr{V}} \boldsymbol{v} \cdot \boldsymbol{dl}.$$

c. Other applications in §§5.8, 20.2, 20.3, 20.5, and exercises 4.10, 4.11, 5.2, and below.

## D. Algebra II (applicable to any vector space with metric)

1. *Norm of a* p*-form.*

$\|\boldsymbol{\alpha}\|^2 \equiv \alpha_{|i_1 \ldots i_p|}\alpha^{i_1 \ldots i_p}.$

*Two applications:* Norm of a 1-form equals its squared length, $\|\boldsymbol{\alpha}\|^2 = \boldsymbol{\alpha} \cdot \boldsymbol{\alpha}$. Norm of electromagnetic 2-form or ***Faraday***: $\|\boldsymbol{F}\|^2 = \boldsymbol{B}^2 - \boldsymbol{E}^2$.

2. *Dual of a* p*-form.*

a. In an $n$-dimensional space, the dual of a $p$-form $\boldsymbol{\alpha}$ is the $(n-p)$-form $^*\boldsymbol{\alpha}$, with components

$$(^*\alpha)_{k_1 \ldots k_{n-p}} = \alpha^{|i_1 \ldots i_p|}\epsilon_{i_1 \ldots i_p k_1 \ldots k_{n-p}}.$$

b. Properties of duals:

$$^{**}\boldsymbol{\alpha} = (-1)^{p-1}\boldsymbol{\alpha} \text{ in spacetime;}$$
$$\boldsymbol{\alpha} \wedge {}^*\boldsymbol{\alpha} = \|\boldsymbol{\alpha}\|^2\boldsymbol{\varepsilon} \text{ in general.}$$

c. Note: the definition of $\boldsymbol{\varepsilon}$ (exercise 3.13) entails choosing an orientation of the space, i.e., deciding which orthonormal bases (1) are "right-handed" and thus (2) have $\boldsymbol{\varepsilon}(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n) = +1$.

*Applications*

a. For $f$ a 0-form, $^*f = f\boldsymbol{\varepsilon}$, and $\int f d(\text{volume}) = \int {}^*f$.

b. Dual of charge-current 1-form $\boldsymbol{J}$ is charge-current 3-form $^*\boldsymbol{J}$. The total charge $Q$ in a 3-dimensional hypersurface region $\mathcal{S}$ is

$$Q(\mathcal{S}) = \int_{\mathcal{S}} {}^*\boldsymbol{J}.$$

**Box 4.1** (*continued*)

Conservation of charge is stated locally by $\boldsymbol{d}{}^*\boldsymbol{J} = 0$. Stokes' Theorem goes from this differential conservation law to the integral conservation law,

$$0 = \int_{\mathcal{V}} \boldsymbol{d}{}^*\boldsymbol{J} = \int_{\partial\mathcal{V}} {}^*\boldsymbol{J}.$$

This law is of most interest when $\partial\mathcal{V} = \mathcal{S}_2 - \mathcal{S}_1$ consists of the future $\mathcal{S}_2$ and past $\mathcal{S}_1$ boundaries of a spacetime region, in which case it states $Q(\mathcal{S}_2) = Q(\mathcal{S}_1)$; see exercise 5.2.

c. Dual of electromagnetic field tensor $\boldsymbol{F} =$ ***Faraday*** is ${}^*\boldsymbol{F} =$ ***Maxwell***. From the $\boldsymbol{d}{}^*\boldsymbol{F} = 4\pi\,{}^*\boldsymbol{J}$ Maxwell equation, find $4\pi Q = 4\pi\int_{\mathcal{S}} {}^*\boldsymbol{J} = \int_{\mathcal{S}} \boldsymbol{d}{}^*\boldsymbol{F} = \int_{\partial\mathcal{S}} {}^*\boldsymbol{F}$.

3. *Simple forms revisited.*
   a. The dual of a simple form is simple.
   b. Egg crate of ${}^*\boldsymbol{\sigma}$ is perpendicular to egg crate of $\boldsymbol{\sigma} = \boldsymbol{\alpha} \wedge \boldsymbol{\beta} \wedge \cdots \wedge \boldsymbol{\mu}$ in this sense:
      (1) pick any vector $\boldsymbol{V}$ lying in intersection of surfaces of $\boldsymbol{\sigma}$

      $$(\langle \boldsymbol{\alpha}, \boldsymbol{V}\rangle = \langle \boldsymbol{\beta}, \boldsymbol{V}\rangle = \cdots = \langle \boldsymbol{\mu}, \boldsymbol{V}\rangle = 0);$$

      (2) pick any vector $\boldsymbol{W}$ lying in intersection of surfaces of ${}^*\boldsymbol{\sigma}$;
      (3) then $\boldsymbol{V}$ and $\boldsymbol{W}$ are necessarily perpendicular: $\boldsymbol{V} \cdot \boldsymbol{W} = 0$.

*Example:* $\boldsymbol{\sigma} = 3\,\boldsymbol{d}t$ is a simple 1-form in spacetime.
a. ${}^*\boldsymbol{\sigma} = -3\,\boldsymbol{d}x \wedge \boldsymbol{d}y \wedge \boldsymbol{d}z$ is a simple 3-form.
b. General vector in surfaces of $\boldsymbol{\sigma}$ is

$$\boldsymbol{V} = V^x\boldsymbol{e}_x + V^y\boldsymbol{e}_y + V^z\boldsymbol{e}_z.$$

c. General vector in intersection of surfaces of ${}^*\boldsymbol{\sigma}$ is

$$\boldsymbol{W} = W^t\boldsymbol{e}_t.$$

d. $\boldsymbol{W} \cdot \boldsymbol{V} = 0$.

2

## §4.2. ELECTROMAGNETIC 2-FORM AND LORENTZ FORCE

The electromagnetic field tensor, ***Faraday*** $= \boldsymbol{F}$, is an antisymmetric second-rank tensor (i.e., 2-form). Instead of expanding it in terms of the tensor products of basis 1-forms,

$$\boldsymbol{F} = F_{\alpha\beta}\, \boldsymbol{d}x^\alpha \otimes \boldsymbol{d}x^\beta,$$

the exterior calculus prefers to expand in terms of antisymmetrized tensor products (*"exterior products,"* exercise 4.1):

Electromagnetic 2-form expressed in terms of exterior products

$$\boldsymbol{F} = \frac{1}{2} F_{\alpha\beta}\, \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta, \tag{4.1}$$

$$\boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta \equiv \boldsymbol{d}x^\alpha \otimes \boldsymbol{d}x^\beta - \boldsymbol{d}x^\beta \otimes \boldsymbol{d}x^\alpha. \tag{4.2}$$

Any 2-form (antisymmetric, second-rank tensor) can be so expanded. The symbol "$\wedge$" is variously called a "wedge," a "hat," or an "exterior product sign"; and $\boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta$ are the "basis 2-forms" of a given Lorentz frame (see §3.5, exercise 3.12, and Box 4.1).

There is no simpler way to illustrate this 2-form representation of the electromagnetic field than to consider a magnetic field in the $x$-direction:

$$\begin{aligned} F_{yz} &= -F_{zy} = B_x, \\ \boldsymbol{F} &= B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z. \end{aligned} \tag{4.3}$$

The 1-form $\boldsymbol{d}y = \text{grad}\, y$ is the set of surfaces (actually hypersurfaces) $y = 18$ (all $t, x, z$), $y = 19$ (all $t, x, z$), $y = 20$ (all $t, x, z$), etc.; and surfaces uniformly interpolated between them. Similarly for the 1-form $\boldsymbol{d}z$. The intersection between these two sets of surfaces produces a honeycomb-like structure. That structure becomes a "2-form" when it is supplemented by instructions (see arrows in Figure 4.1) that give a "sense of circulation" to each tube of the honeycomb (order of factors in the "wedge product" of equation 4.2; $\boldsymbol{d}y \wedge \boldsymbol{d}z = -\boldsymbol{d}z \wedge \boldsymbol{d}y$). The 2-form $\boldsymbol{F}$ in the example differs from this "basis 2-form" $\boldsymbol{d}y \wedge \boldsymbol{d}z$ only in this respect, that where $\boldsymbol{d}y \wedge \boldsymbol{d}z$ had one tube, the field 2-form has $B_x$ tubes.

A 2-form as a honeycomb of tubes with a sense of circulation

When one considers a tubular structure that twists and turns on its way through spacetime, one must have more components to describe it. The 2-form for the general electromagnetic field can be written as

$$\begin{aligned} \boldsymbol{F} = {} & E_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t + E_y\, \boldsymbol{d}y \wedge \boldsymbol{d}t + E_z\, \boldsymbol{d}z \wedge \boldsymbol{d}t + B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z \\ & + B_y\, \boldsymbol{d}z \wedge \boldsymbol{d}x + B_z\, \boldsymbol{d}x \wedge \boldsymbol{d}y \end{aligned} \tag{4.4}$$

(6 components, 6 basis 2-forms).

A 1-form is a machine to produce a number out of a vector (bongs of a bell as the vector pierces successive surfaces). A 2-form is a machine to produce a number out of an oriented surface (surface with a sense of circulation indicated on it: Figure 4.1, lower right). The meaning is as clear here as it is in elementary magnetism:

A 2-form as a machine to produce a number out of an oriented surface

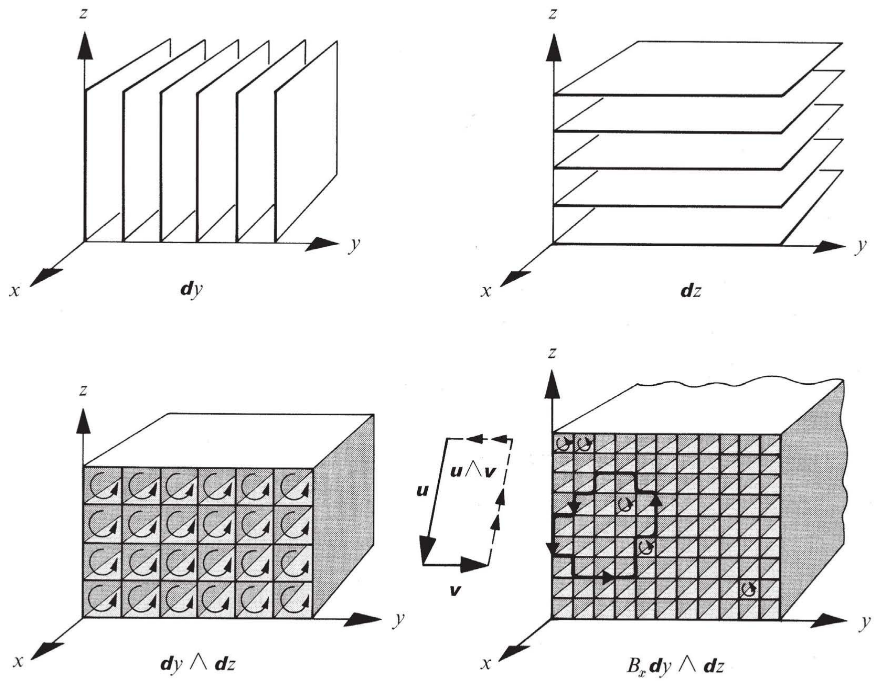



**Figure 4.1.**
Construction of the 2-form for the electromagnetic field $\boldsymbol{F} = B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z$ out of the 1-forms $\boldsymbol{d}y$ and $\boldsymbol{d}z$ by "wedge multiplication" (formation of honeycomb-like structure with sense of circulation indicated by arrows). A 2-form is a "machine to construct a number out of an oriented surface" (illustrated by sample surface enclosed by arrows at lower right; number of tubes intersected by this surface is

$$\int_{\text{(this surface)}} \boldsymbol{F} = 18;$$

Faraday's concept of "magnetic flux"). This idea of 2-form machinery can be connected to the "tensor-as-machine" idea of Chapter 3 as follows. The shape of the oriented surface over which one integrates $\boldsymbol{F}$ does not matter, for small surfaces. All that affects $\int \boldsymbol{F}$ is the area of the surface, and its orientation. Choose two vectors, $\boldsymbol{u}$ and $\boldsymbol{v}$, that lie in the surface. They form two legs of a parallelogram, whose orientation ($\boldsymbol{u}$ followed by $\boldsymbol{v}$) and area are embodied in the exterior product $\boldsymbol{u} \wedge \boldsymbol{v}$. Adjust the lengths of $\boldsymbol{u}$ and $\boldsymbol{v}$ so their parallelogram, $\boldsymbol{u} \wedge \boldsymbol{v}$, has the same area as the surface of integration. Then

$$\underbrace{\int_{\text{surface}} \boldsymbol{F} = \int_{\boldsymbol{u} \wedge \boldsymbol{v}} \boldsymbol{F}}_{\text{machinery idea of this chapter}} = \underbrace{\boldsymbol{F}(\boldsymbol{u}, \boldsymbol{v})}_{\text{machinery idea of Chapter 3}}.$$

*Exercise:* derive this result, for an infinitesimal surface $\boldsymbol{u} \wedge \boldsymbol{v}$ and for general $\boldsymbol{F}$, using the formalism of Box 4.1.

the number of Faraday tubes cut by that surface. The electromagnetic 2-form $\boldsymbol{F}$ or ***Faraday*** described by such a "tubular structure" (suitably abstracted; Box 4.2) has a reality and a location in space that is independent of all coordinate systems and all artificial distinctions between "electric" and "magnetic" fields. Moreover, those tubes provide the most direct geometric representation that anyone has ever been able to give for the machinery by which the electromagnetic field acts on a charged particle. Take a particle of charge $e$ and 4-velocity

$$\boldsymbol{u} = \frac{dx^\alpha}{d\tau}\boldsymbol{e}_\alpha. \tag{4.5}$$

Let this particle go through a region where the electromagnetic field is described by the 2-form

$$\boldsymbol{F} = B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z \tag{4.6}$$

Lorentz force as contraction of electromagnetic 2-form with particle's 4-velocity

of Figure 4.1. Then the force exerted on the particle (regarded as a 1-form) is the contraction of this 2-form with the 4-velocity (and the charge);

$$\dot{\boldsymbol{p}} = \boldsymbol{d}\boldsymbol{p}/d\tau = e\boldsymbol{F}(\boldsymbol{u}) \equiv e\langle \boldsymbol{F}, \boldsymbol{u}\rangle, \tag{4.7}$$

as one sees by direct evaluation, letting the two factors in the 2-form act in turn on the tangent vector $\boldsymbol{u}$:

$$\begin{aligned}
\dot{\boldsymbol{p}} &= eB_x\langle \boldsymbol{d}y \wedge \boldsymbol{d}z, \boldsymbol{u}\rangle \\
&= eB_x\{\boldsymbol{d}y\langle \boldsymbol{d}z, \boldsymbol{u}\rangle - \boldsymbol{d}z\langle \boldsymbol{d}y, \boldsymbol{u}\rangle\} \\
&= eB_x\{\boldsymbol{d}y\langle \boldsymbol{d}z, u^z\boldsymbol{e}_z\rangle - \boldsymbol{d}z\langle \boldsymbol{d}y, u^y\boldsymbol{e}_y\rangle\}
\end{aligned}$$

or

$$\dot{p}_\alpha\, \boldsymbol{d}x^\alpha = eB_x u^z\, \boldsymbol{d}y - eB_x u^y\, \boldsymbol{d}z. \tag{4.8}$$

Comparing coefficients of the separate basis 1-forms on the two sides of this equation, one sees reproduced all the detail of the Lorentz force exerted by the magnetic field $B_x$:

$$\begin{aligned}
\dot{p}_y &= \frac{dp_y}{d\tau} = eB_x\frac{dz}{d\tau}, \\
\dot{p}_z &= \frac{dp_z}{d\tau} = -eB_x\frac{dy}{d\tau}.
\end{aligned} \tag{4.9}$$

By simple extension of this line of reasoning to the general electromagnetic field, one concludes that *the time-rate of change of momentum (1-form) is equal to the charge multiplied by the contraction of the* ***Faraday*** *with the 4-velocity.* Figure 4.2 illustrates pictorially how the 2-form, $\boldsymbol{F}$, serves as a machine to produce the 1-form, $\dot{\boldsymbol{p}}$, out of the tangent vector, $e\boldsymbol{u}$.

*(continued on page 105)*

**Box 4.2 ABSTRACTING A 2-FORM FROM THE CONCEPT OF "HONEYCOMB-LIKE STRUCTURE," IN 3-SPACE AND IN SPACETIME**

Open up a cardboard carton containing a dozen bottles, and observe the honeycomb structure of intersecting north-south and east-west cardboard separators between the bottles. That honeycomb structure of "tubes" ("channels for bottles") is a fairly apt illustration of a 2-form in the context of everyday 3-space. It yields a number (number of tubes cut) for each choice of smooth element of 2-surface slicing through the three-dimensional structure. However, the intersecting cardboard separators are rather too specific. All that a true 2-form can ever give is the number of tubes sliced through, not the "shape" of the tubes. Slew the carton around on the floor by 45°. Then half the separators run NW-SE and the other half run NE-SW, but through a given bit of 2-surface fixed in 3-space the count of tubes is unchanged. Therefore, one should be careful to make the concept of tubes in the mind's eye abstract enough that one envisages direction of tubes (vertical in the example) and density of tubes, but not any specific location or orientation for the tube walls. Thus all the following representations give one and the same 2-form, $\boldsymbol{\sigma}$:

$$\boldsymbol{\sigma} = B\, \mathbf{d}x \wedge \mathbf{d}y;$$

$$\boldsymbol{\sigma} = B(2\, \mathbf{d}x) \wedge \left(\frac{1}{2}\, \mathbf{d}y\right)$$

(NS cardboards spaced twice as close as before; EW cardboards spaced twice as wide as before);

$$\boldsymbol{\sigma} = B\, \mathbf{d}\left(\frac{x-y}{\sqrt{2}}\right) \wedge \mathbf{d}\left(\frac{x+y}{\sqrt{2}}\right)$$

(cardboards rotated through 45°);

$$\boldsymbol{\sigma} = B \frac{\alpha\, \mathbf{d}x + \beta\, \mathbf{d}y}{(\alpha\delta - \beta\gamma)^{1/2}} \wedge \frac{\gamma\, \mathbf{d}x + \delta\, \mathbf{d}y}{(\alpha\delta - \beta\gamma)^{1/2}}$$

(both orientation and spacing of "cardboards" changing from point to point, with all four functions, $\alpha$, $\beta$, $\gamma$, and $\delta$, depending on position).

What has physical reality, and constitutes the real geometric object, is not any one of the 1-forms just encountered individually, but only the 2-form $\boldsymbol{\sigma}$ itself. This circumstance helps to explain why in the physical literature one sometimes refers to "tubes of force" and sometimes to "lines of force." The two terms for the same structure have this in common, that each yields a number when sliced by a bit of surface. The line-of-force picture has the advantage of not imposing on the mind any specific structure of "sheets of cardboard"; that is, any specific decomposition of the 2-form into the product of 1-forms. However, that very feature is also a disadvantage, for in a calculation one often finds it useful to have a well-defined representation of the 2-form as the wedge product of 1-forms. Moreover, the tube picture, abstract though it must be if it is to be truthful, also has this advantage, that the orientation of the elementary tubes (sense of circulation as indicated by arrows in Figures 4.1 and 4.5, for example) lends itself to ready visualization. Let the "walls" of the tubes therefore remain in all pictures drawn in this book as a reminder that 2-forms can be built out of 1-forms; but let it be understood here and hereafter how manyfold are the options for the individual 1-forms!

Turn now from three dimensions to four, and find that the concept of "honeycomb-like structure" must be made still more abstract. In three dimensions the arbitrariness of the decomposition of the 2-form into 1-forms showed in the slant and packing of the "cardboards," but had no effect on the verticality of the "channels for the bottles" ("direction of Faraday lines of force or tubes of force"); not so in four dimensions, or at least not in the generic case in four dimensions.

In special cases, the story is almost as simple in four dimensions as in three. An example of a special case is once again the 2-form $\boldsymbol{\sigma} = B\,\boldsymbol{d}x \wedge \boldsymbol{d}y$, with all the options for decomposition into 1-forms that have already been mentioned, but with every option giving the same "direction" for the tubes. If the word "direction" now rises in status from "tube walls unpierced by motion in the direction of increasing $z$" to "tube walls unpierced either by motion in the direction of increasing $z$, or by motion in the direction of increasing $t$, or by any linear combination of such motions," that is a natural enough consequence of adding the new dimension. Moreover, the same simplicity prevails for an electromagnetic plane wave. For example, let the wave be advancing in the $z$-direction, and let the electric polarization point in the $x$-direction; then for a monochromatic wave, one has

$$E_x = B_y = E_0 \cos \omega(z - t) = -F_{01} = F_{31},$$

and all components distinct from these equal zero. ***Faraday*** is

$$\begin{aligned} \boldsymbol{F} &= F_{01}\,\boldsymbol{d}t \wedge \boldsymbol{d}x + F_{31}\,\boldsymbol{d}z \wedge \boldsymbol{d}x \\ &= E_0 \cos \omega(z - t)\,\boldsymbol{d}(z - t) \wedge \boldsymbol{d}x, \end{aligned}$$

which is again representable as a single wedge product of two 1-forms.

Not so in general! The general 2-form in four dimensions consists of six distinct wedge products,

$$\boldsymbol{F} = F_{01}\,\boldsymbol{d}t \wedge \boldsymbol{d}x + F_{02}\,\boldsymbol{d}t \wedge \boldsymbol{d}y + \cdots + F_{23}\,\boldsymbol{d}y \wedge \boldsymbol{d}z.$$

It is too much to hope that this expression will reduce in the generic case to a single wedge product of two 1-forms ("*simple*" 2-form). It is not even true that it will. It is only remarkable that it can be reduced from six exterior products to two (details in exercise 4.1); thus,

$$\boldsymbol{F} = \boldsymbol{n}^1 \wedge \boldsymbol{\xi}^1 + \boldsymbol{n}^2 \wedge \boldsymbol{\xi}^2.$$

Each product $\boldsymbol{n}^i \wedge \boldsymbol{\xi}^i$ individually can be visualized as a honeycomb-like structure like those depicted in Figures 4.1, 4.2, 4.4, and 4.5. Each such structure individually can be pictured as built out of intersecting sheets (1-forms), but with such details as the tilt and packing of these 1-forms abstracted away. Each such structure individually gives a number when sliced by an element of surface. What counts for the 2-form $\boldsymbol{F}$, however, is neither the number of tubes of $\boldsymbol{n}^1 \wedge \boldsymbol{\xi}^1$ cut by the surface, nor the number of tubes of $\boldsymbol{n}^2 \wedge \boldsymbol{\xi}^2$ cut by the surface, but only the sum of the two. This sum is what is referred to in the text as the "number of tubes of $\boldsymbol{F}$" cut by the surface. The contribution of either wedge product individually is not well-defined, for a simple reason: the decomposition of a six-wedge-product object into two wedge products, miraculous though it seems, is actually far from unique (details in exercise 4.2).

In keeping with the need to have two products of 1-forms to represent the general 2-form note that the vanishing of $\boldsymbol{dF}$ ("no magnetic charges") does not automatically imply that $\boldsymbol{d}(\boldsymbol{n}^1 \wedge \boldsymbol{\xi}^1)$ or $\boldsymbol{d}(\boldsymbol{n}^2 \wedge \boldsymbol{\xi}^2)$ separately vanish. Note also that any spacelike slice through the general 2-form $\boldsymbol{F}$ (reduction from four dimensions to three) can always be represented in terms of a honeycomb-like structure ("simple" 2-form in three dimensions; Faraday's picture of magnetic tubes of force).

Despite the abstraction that has gone on in seeing in all generality what a 2-form is, there is no bar to continuing to use the term "honeycomb-like structure" in a broadened sense to describe this object; and that is the practice here and hereafter.

2

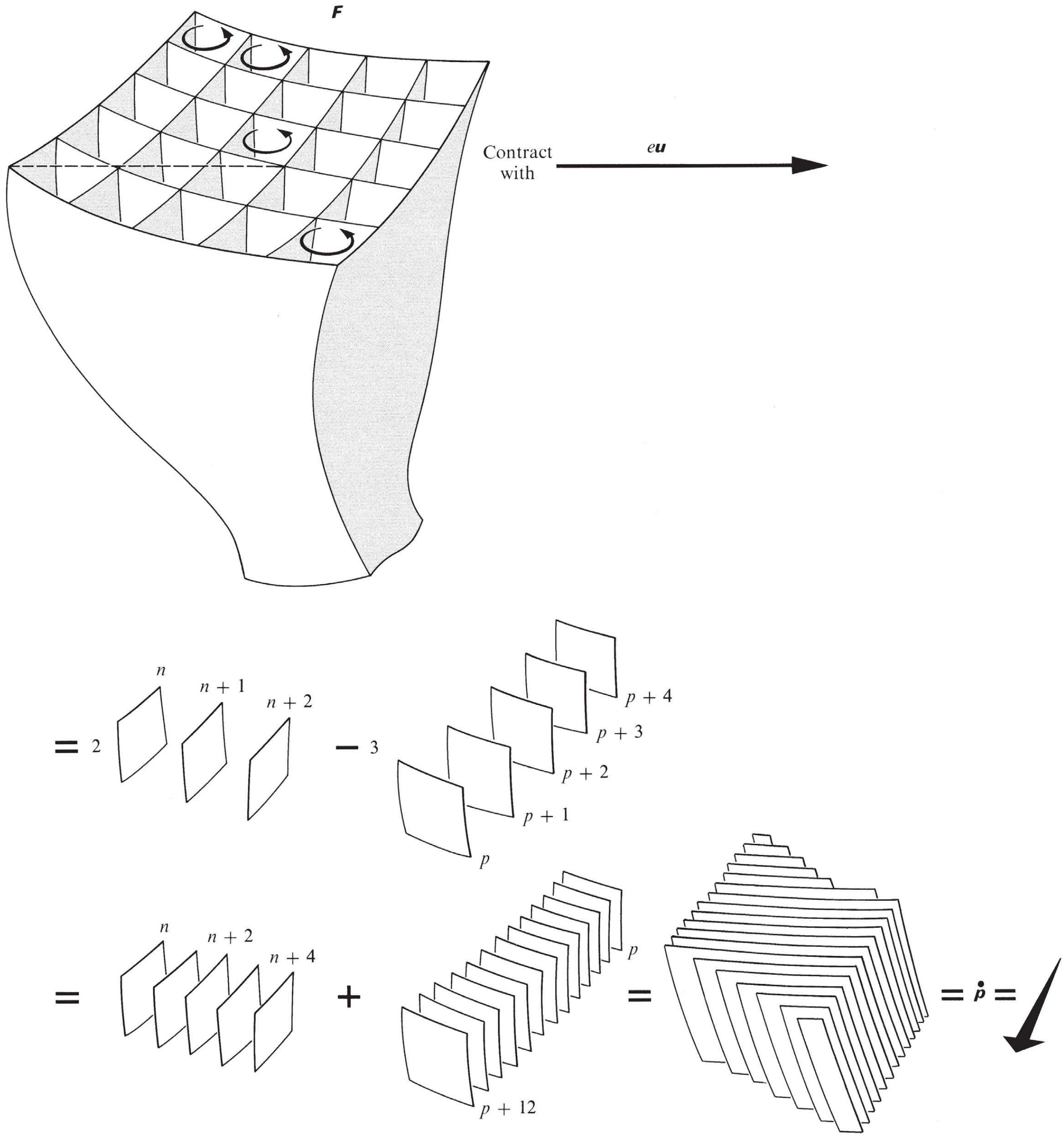



**Figure 4.2.**
The ***Faraday*** or 2-form $\boldsymbol{F}$ of the electromagnetic field is a machine to produce a 1-form (the time-rate of change of momentum $\dot{\boldsymbol{p}}$ of a charged particle) out of a tangent vector (product of charge $e$ of the particle and its 4-velocity $\boldsymbol{u}$). In spacetime the general 2-form is the "superposition" (see Box 4.2) of two structures like that illustrated at the top of this diagram, the tubes of the first being tilted and packed as indicated, the tubes of the second being tilted in another direction and having a different packing density.

## §4.3. FORMS ILLUMINATE ELECTROMAGNETISM, AND ELECTROMAGNETISM ILLUMINATES FORMS

All electromagnetism allows itself to be summarized in the language of 2-forms, honeycomb-like "structures" (again in the abstract sense of "structure" of Box 4.2) of tubes filling all spacetime, as well when spacetime is curved as when it is flat. In brief, there are two such structures, one ***Faraday*** = ***F***, the other ***Maxwell*** = ****F***, each dual ("perpendicular," the only place where metric need enter the discussion) to the other, each satisfying an elementary equation:

$$\boldsymbol{dF} = 0 \tag{4.10}$$

("no tubes of ***Faraday*** ever end") and

$$\boldsymbol{d}{}^*\boldsymbol{F} = 4\pi\, {}^*\boldsymbol{J} \tag{4.11}$$

("the number of tubes of ***Maxwell*** that end in an elementary volume is equal to the amount of electric charge in that volume"). To see in more detail how this machinery shows up in action, look in turn at: (1) the definition of a 2-form; (2) the appearance of a given electromagnetic field as ***Faraday*** and as ***Maxwell***; (3) the ***Maxwell*** structure for a point-charge at rest; (4) the same for a point-charge in motion; (5) the nature of the field of a charge that moves uniformly except during a brief instant of acceleration; (6) the ***Faraday*** structure for the field of an oscillating dipole; (7) the concept of exterior derivative; (8) Maxwell's equations in the language of forms; and (9) the solution of Maxwell's equations in flat spacetime, using a 1-form ***A*** from which the Liénard-Wiechert 2-form ***F*** can be calculated via ***F*** = ***dA***.

Preview of key points in electromagnetism

A 2-form, as illustrated in Figure 4.1, is a machine to construct a number ("net number of tubes cut") out of any "oriented 2-surface" (2-surface with "sense of circulation" marked on it):

A 2-form as machine for number of tubes cut

$$\begin{pmatrix}\text{number}\\ \text{of tubes}\\ \text{cut}\end{pmatrix} = \int_{\text{surface}} \boldsymbol{F}. \tag{4.12}$$

For example, let the 2-form be the one illustrated in Figure 4.1

Number of tubes cut calculated in one example

$$\boldsymbol{F} = B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z,$$

and let the surface of integration be the portion of the surface of the 2-sphere $x^2 + y^2 + z^2 = a^2$, $t =$ constant, bounded between $\theta = 70°$ and $\theta = 110°$ and between $\varphi = 0°$ and $\varphi = 90°$ ("Atlantic region of the tropics"). Write

$$\begin{aligned} y &= a \sin\theta \sin\varphi, \\ z &= a\cos\theta, \\ \boldsymbol{d}y &= a(\cos\theta\sin\varphi\, \boldsymbol{d}\theta + \sin\theta\cos\varphi\, \boldsymbol{d}\varphi), \\ \boldsymbol{d}z &= -a\sin\theta\, \boldsymbol{d}\theta, \\ \boldsymbol{d}y \wedge \boldsymbol{d}z &= a^2 \sin^2\theta \cos\varphi\, \boldsymbol{d}\theta \wedge \boldsymbol{d}\varphi. \end{aligned} \tag{4.13}$$

2

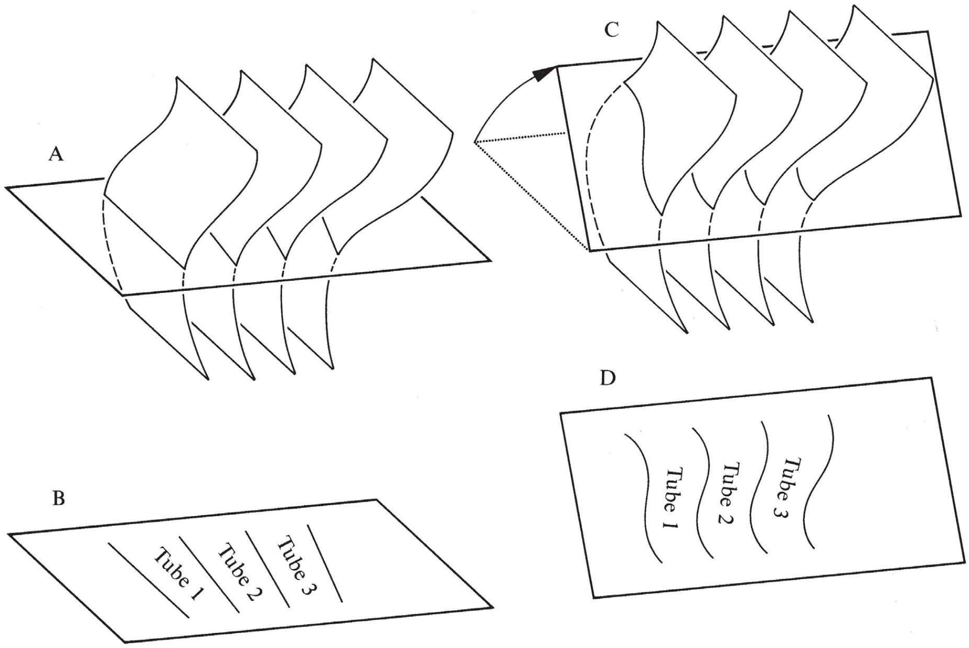



**Figure 4.3.**
Spacelike slices through ***Faraday***, the electromagnetic 2-form, a geometric object, a honeycomb of tubes that pervades all spacetime ("honeycomb" in the abstract sense spelled out more precisely in Box 4.2). The surfaces in the drawing do not look like a 2-form (honeycomb), because the second family of surfaces making up the honeycomb extends in the spatial direction that is suppressed from the drawing. Diagram A shows one spacelike slice through the 2-form (time increases upwards in the diagram). In diagram B, a projection of the 2-form on this spacelike hypersurface gives the Faraday tubes of magnetic force in this three-dimensional geometry (if the suppressed dimension were restored, the tubes would be tubes, not channels between lines). Diagram C shows another spacelike slice (hypersurface of simultaneity for an observer in a different Lorentz frame). Diagram D shows the very different pattern of magnetic tubes in this reference system. The demand that magnetic tubes of force shall not end ($\boldsymbol{\nabla} \cdot \boldsymbol{B} = 0$), repeated over and over for every spacelike slice through ***Faraday***, gives everywhere the result $\partial \boldsymbol{B}/\partial t = -\boldsymbol{\nabla} \times \boldsymbol{E}$. Thus (magnetostatics) + (covariance) $\longrightarrow$ (magnetodynamics). Similarly—see Chapters 17 and 21—(geometrostatics) + (covariance) $\longrightarrow$ (geometrodynamics).

The structure $\boldsymbol{d}\theta \wedge \boldsymbol{d}\theta$ looks like a "collapsed egg-crate" (Figure 1.4, upper right) and has zero content, a fact formally evident from the vanishing of $\boldsymbol{\alpha} \wedge \boldsymbol{\beta} = -\boldsymbol{\beta} \wedge \boldsymbol{\alpha}$ when $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are identical. The result of the integration, assuming constant $B_x$, is

$$\int_{\text{surface}} \boldsymbol{F} = a^2 B_x \int_{70^\circ}^{110^\circ} \sin^2\theta \, d\theta \int_{0^\circ}^{90^\circ} \cos\varphi \, d\varphi \tag{4.14}$$

It is not so easy to visualize a pure electric field by means of its 2-form $\boldsymbol{F}$ (Figure 4.4, left) as it is to visualize a pure magnetic field by means of its 2-form $\boldsymbol{F}$ (Figures 4.1, 4.2, 4.3). Is there not some way to treat the two fields on more nearly the same footing? Yes, construct the 2-form $^*\boldsymbol{F}$ (Figure 4.4, right) that is *dual* ("perpendicular"; Box 4.3; exercise 3.14) to $\boldsymbol{F}$.

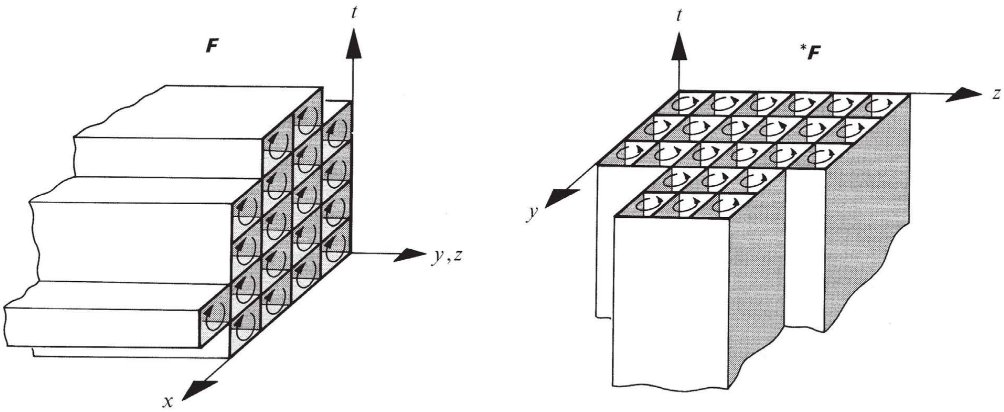



**Figure 4.4.**
The ***Faraday*** structure

$$\boldsymbol{F} = \frac{1}{2} F_{\mu\nu}\, \boldsymbol{d}x^{\mu} \wedge \boldsymbol{d}x^{\nu} = \frac{1}{2} F_{01}\, \boldsymbol{d}t \wedge \boldsymbol{d}x + \frac{1}{2} F_{10}\, \boldsymbol{d}x \wedge \boldsymbol{d}t = E_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t$$

associated with an electric field in the $x$-direction, and the dual ("perpendicular") ***Maxwell*** honeycomb-like 2-form

$$^{*}\boldsymbol{F} = \frac{1}{2}\, {}^{*} F_{\mu\nu}\, \boldsymbol{d}x^{\mu} \wedge \boldsymbol{d}x^{\nu} = {}^{*}F_{23}\, \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 = F^{01}\, \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 = F_{10}\, \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 = E_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z.$$

Represent in geometric form the field of a point-charge of strength $e$ at rest at the origin. Operate in flat space with spherical polar coordinates:

$$\begin{aligned} ds^2 &= -d\tau^2 = g_{\mu\nu}\, dx^{\mu}\, dx^{\nu} \\ &= -dt^2 + dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\varphi^2. \end{aligned} \tag{4.15}$$

The electric field in the $r$-direction being $E_r = e/r^2$, it follows that the 2-form $\boldsymbol{F}$ or ***Faraday*** is

$$\boldsymbol{F} = \frac{1}{2} F_{\mu\nu}\, \boldsymbol{d}x^{\mu} \wedge \boldsymbol{d}x^{\nu} = -E_r\, \boldsymbol{d}t \wedge \boldsymbol{d}r = -\frac{e}{r^2}\, \boldsymbol{d}t \wedge \boldsymbol{d}r. \tag{4.16}$$

Its dual, according to the prescription in exercise 3.14, is ***Maxwell***:

Pattern of tubes in dual structure ***Maxwell*** for point-charge at rest

$$\boldsymbol{Maxwell} = {}^{*}\boldsymbol{F} = e \sin\theta\, \boldsymbol{d}\theta \wedge \boldsymbol{d}\varphi, \tag{4.17}$$

as illustrated in Figure 4.5.

Take a tour in the positive sense around a region of the surface of the sphere illustrated in Figure 4.5. The number of tubes of $^{*}\boldsymbol{F}$ encompassed in the route will be precisely

$$\begin{pmatrix} \text{number} \\ \text{of tubes} \end{pmatrix} = e \begin{pmatrix} \text{solid} \\ \text{angle} \end{pmatrix}.$$

The whole number of tubes of $^{*}\boldsymbol{F}$ emergent over the entire sphere will be $4\pi e$, in conformity with Faraday's picture of tubes of force.

**Box 4.3 DUALITY OF 2-FORMS IN SPACETIME**

Given a general 2-form (containing six exterior or wedge products)

$$\boldsymbol{F} = E_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t + E_y\, \boldsymbol{d}y \wedge \boldsymbol{d}t + \cdots + B_z\, \boldsymbol{d}x \wedge \boldsymbol{d}y,$$

one gets to its dual ("perpendicular") by the prescription

$${}^*\boldsymbol{F} = -B_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t - \cdots + E_y\, \boldsymbol{d}z \wedge \boldsymbol{d}x + E_z\, \boldsymbol{d}x \wedge \boldsymbol{d}y.$$

### Duality Rotations

Note that the dual of the dual is the negative of the original 2-form; thus

$$^{**}\boldsymbol{F} = -E_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t - \cdots - B_z\, \boldsymbol{d}x \wedge \boldsymbol{d}y = -\boldsymbol{F}.$$

In this sense * has the same property as the imaginary number $i$: $** = ii = -1$. Thus one can write

$$e^{*\alpha} = \cos\alpha + {}^*\sin\alpha.$$

This operation, applied to $\boldsymbol{F}$, carries attention from the generic 2-form in its simplest representation (see exercise 4.1)

$$\boldsymbol{F} = E_x\, \boldsymbol{d}x \wedge \boldsymbol{d}t + B_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z$$

to another "duality rotated electromagnetic field"

$$e^{*\alpha}\boldsymbol{F} = (E_x \cos\alpha - B_x \sin\alpha)\, \boldsymbol{d}x \wedge \boldsymbol{d}t + (B_x \cos\alpha + E_x \sin\alpha)\, \boldsymbol{d}y \wedge \boldsymbol{d}z.$$

If the original field satisfied Maxwell's empty-space field equations, so does the new field. With suitable choice of the "complexion" $\alpha$, one can annul one of the two wedge products at any chosen point in spacetime and have for the other

$$(B_x{}^2 + E_x{}^2)^{1/2}\, \boldsymbol{d}y \wedge \boldsymbol{d}z.$$

Field of a point-charge in motion

How can one determine the structure of tubes associated with a charged particle moving at a uniform velocity? First express $^*\boldsymbol{F}$ in rectangular coordinates moving with the particle (barred coordinates in this comoving "rocket" frame of reference; unbarred coordinates will be used later for a laboratory frame of reference). The relevant steps can be listed:

(a)

$$^*\boldsymbol{F} = e \sin\bar\theta\, \boldsymbol{d}\bar\theta \wedge \boldsymbol{d}\bar\varphi = -e(\boldsymbol{d}\cos\bar\theta) \wedge \boldsymbol{d}\bar\varphi;$$

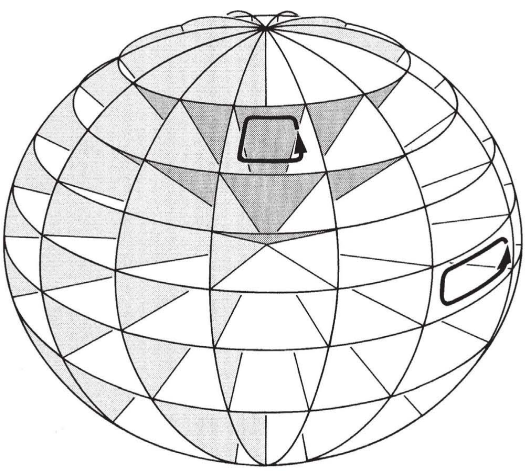


**Figure 4.5.**
The field of 2-forms ***Maxwell*** $= {}^*\boldsymbol{F} = e \sin\theta\, \boldsymbol{d}\theta \wedge \boldsymbol{d}\phi$ that describes the electromagnetic field of a charge $e$ at rest at the origin. This picture is actually the intersection of ${}^*\boldsymbol{F}$ with a 3-surface of constant time $t$; i.e., the time direction is suppressed from the picture.

(b)

$$\bar{\varphi} = \arctan\frac{\bar{y}}{\bar{x}}; \qquad \boldsymbol{d}\bar{\varphi} = \frac{\bar{x}\,\boldsymbol{d}\bar{y} - \bar{y}\,\boldsymbol{d}\bar{x}}{\bar{x}^2 + \bar{y}^2};$$

(c)

$$\cos\bar{\theta} = \frac{\bar{z}}{\bar{r}}; \qquad -\boldsymbol{d}(\cos\bar{\theta}) = \frac{-\boldsymbol{d}\bar{z}}{\bar{r}} + \frac{\bar{z}}{\bar{r}^3}(\bar{x}\,\boldsymbol{d}\bar{x} + \bar{y}\,\boldsymbol{d}\bar{y} + \bar{z}\,\boldsymbol{d}\bar{z});$$

(d) combine to find

$${}^*\boldsymbol{F} = (e/\bar{r}^3)(\bar{x}\,\boldsymbol{d}\bar{y} \wedge \boldsymbol{d}\bar{z} + \bar{y}\,\boldsymbol{d}\bar{z} \wedge \boldsymbol{d}\bar{x} + \bar{z}\,\boldsymbol{d}\bar{x} \wedge \boldsymbol{d}\bar{y}) \qquad (4.18)$$

(electromagnetic field of point charge in a comoving Cartesian system; spherically symmetric). Now transform to laboratory coordinates:

velocity parameter $\alpha$

velocity $\beta = \tanh\alpha$

$$\frac{1}{\sqrt{1-\beta^2}} = \cosh\alpha, \qquad \frac{\beta}{\sqrt{1-\beta^2}} = \sinh\alpha$$

(a)

$$\begin{cases} \bar{t} = t\cosh\alpha - x\sinh\alpha, \\ \bar{x} = -t\sinh\alpha + x\cosh\alpha, \\ \bar{y} = y \qquad \bar{z} = z; \end{cases}$$

(b)

$$\bar{r} = [(x\cosh\alpha - t\sinh\alpha)^2 + y^2 + z^2]^{1/2};$$

(c)

$$\begin{aligned} {}^*\boldsymbol{F} = (e/\bar{r}^3)[&(x\cosh\alpha - t\sinh\alpha)\,\boldsymbol{d}y \wedge \boldsymbol{d}z + y\,\boldsymbol{d}z \wedge \\ &(\cosh\alpha\,\boldsymbol{d}x - \sinh\alpha\,\boldsymbol{d}t) + z(\cosh\alpha\,\boldsymbol{d}x - \sinh\alpha\,\boldsymbol{d}t) \wedge \boldsymbol{d}y]; \qquad (4.19) \end{aligned}$$

(d) compare with the general dual 2-form,

$$\begin{aligned}{}^*\boldsymbol{F} = E_x\, \boldsymbol{d}y \wedge \boldsymbol{d}z + E_y\, \boldsymbol{d}z \wedge \boldsymbol{d}x + E_z\, \boldsymbol{d}x \wedge \boldsymbol{d}y \\ + B_x\, \boldsymbol{d}t \wedge \boldsymbol{d}x + B_y\, \boldsymbol{d}t \wedge \boldsymbol{d}y + B_z\, \boldsymbol{d}t \wedge \boldsymbol{d}z;\end{aligned}$$

and get the desired individual field components

(e)
$$\begin{cases} E_x = (e/\bar{r}^3)(x \cosh\alpha - t \sinh\alpha), & B_x = 0, \\ E_y = (e/\bar{r}^3)y \cosh\alpha, & B_y = -(e/\bar{r}^3)z \sinh\alpha, \\ E_z = (e/\bar{r}^3)z \cosh\alpha, & B_z = (e/\bar{r}^3)y \sinh\alpha. \end{cases} \tag{4.20}$$

One can verify that the invariants

$$\boldsymbol{B}^2 - \boldsymbol{E}^2 = \frac{1}{2} F_{\alpha\beta} F^{\alpha\beta}, \tag{4.21}$$

$$\boldsymbol{E} \cdot \boldsymbol{B} = \frac{1}{4} F_{\alpha\beta}{}^*F^{\alpha\beta} \tag{4.22}$$

have the same value in the laboratory frame as in the rocket frame, as required. Note that the honeycomb structure of the differential form is not changed when one goes from the rocket frame to the laboratory frame. What changes is only the mathematical formula that describes it.

## §4.4. RADIATION FIELDS

The ***Maxwell*** structure of tubes associated with a charge in uniform motion is more remarkable than it may seem at first sight, and not only because of the Lorentz contraction of the tubes in the direction of motion. The tubes arbitrarily far away move on in military step with the charge on which they center, despite the fact that there is no time for information "emitted" from the charge "right now" to get to the faraway tube "right now." The structure of the faraway tubes "right now" must therefore derive from the charge at an earlier moment on its uniform-motion, straight-line trajectory. This circumstance shows up nowhere more clearly than in what happens to the field in consequence of a sudden change, in a short time $\Delta\tau$, from one uniform velocity to another uniform velocity (Figure 4.6). The tubes have the standard patterns for the two states of motion, one pattern within a sphere of radius $r$, the other outside that sphere, where $r$ is equal to the lapse of time ("cm of light-travel time") since the acceleration took place. The necessity for the two patterns to fit together in the intervening zone, of thickness $\Delta r = \Delta\tau$, forces the field there to be multiplied up by a "stretching factor," proportional to $r$. This factor is responsible for the well-known fact that radiative forces fall off inversely only as the first power of the distance (Figure 4.6).

How an acceleration causes radiation

When the charge continuously changes its state of motion, the structure of the electromagnetic field, though based on the same simple principles as those illustrated in Figure 4.6, nevertheless looks more complex. The following is the ***Faraday*** 2-form

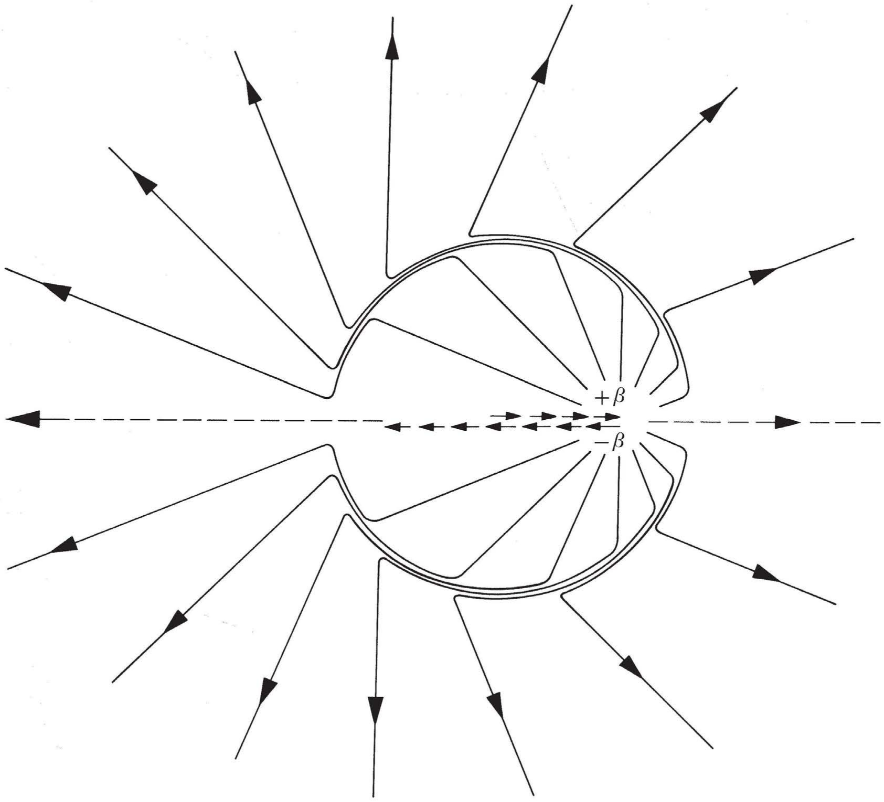



**Figure 4.6.**
*Mechanism of radiation.* J. J. Thomson's way to understand why the strength of an electromagnetic wave falls only as the inverse first power of distance $r$ and why the amplitude of the wave varies (for low velocities) as $\sin\theta$ (maximum in the plane perpendicular to the line of acceleration). The charge was moving to the left at uniform velocity. Far away from it, the lines of force continue to move as if this uniform velocity were going to continue forever (Coulomb field of point-charge in slow motion). However, closer up the field is that of a point-change moving to the right with uniform velocity ($1/r^2$ dependence of strength upon distance). The change from the one field pattern to another is confined to a shell of thickness $\Delta\tau$ located at a distance $r$ from the point of acceleration (amplification of field by "stretching factor" $r\sin\theta\,\Delta\beta/\Delta\tau$; see text). We thank C. Teitelboim for the construction of this diagram.

for the field of an electric dipole of magnitude $p_1$ oscillating up and down parallel to the $z$-axis:

Field of an oscillating dipole

$$\begin{aligned}\boldsymbol{F} = E_x\,\boldsymbol{d}x \wedge \boldsymbol{d}t + \cdots + B_x\,\boldsymbol{d}y \wedge \boldsymbol{d}z + \cdots = \text{real part of } \{p_1 e^{i\omega r - i\omega t} \\ \underbrace{[2\cos\theta\left(\frac{1}{r^3} - \frac{i\omega}{r^2}\right)}_{\text{gives } E_r}\boldsymbol{d}r \wedge \boldsymbol{d}t + \underbrace{\sin\theta\left(\frac{1}{r^3} - \frac{i\omega}{r^2} - \frac{\omega^2}{r}\right)}_{\text{gives } E_\theta} r\,\boldsymbol{d}\theta \wedge \boldsymbol{d}t \\ \underbrace{+ \sin\theta\left(\frac{-i\omega}{r^2} - \frac{\omega^2}{r}\right)}_{\text{gives } B_\phi}\boldsymbol{d}r \wedge r\,\boldsymbol{d}\theta]\}\end{aligned} \tag{4.23}$$

and the dual 2-form ***Maxwell*** $=$ ****F*** is

$$
\begin{aligned}
{}^{*}\boldsymbol{F} = -B_x\,\boldsymbol{d}x \wedge \boldsymbol{d}t - \cdots + E_x\,\boldsymbol{d}y \wedge \boldsymbol{d}z + \cdots = \text{real part of } \{p_1 e^{i\omega r - i\omega t}& \\
\underbrace{[\sin\theta\left(\frac{-i\omega}{r^2} - \frac{\omega^2}{r}\right)}_{\text{gives } B_\phi}\boldsymbol{d}t \wedge r\sin\theta\,\boldsymbol{d}\phi& \\
\underbrace{+\,2\cos\theta\left(\frac{1}{r^3} - \frac{i\omega}{r^2}\right)}_{\text{gives } E_r} r\,\boldsymbol{d}\theta \wedge r\sin\theta\,\boldsymbol{d}\phi& \\
\underbrace{+\,\sin\theta\left(\frac{1}{r^3} - \frac{i\omega}{r^2} - \frac{\omega^2}{r}\right)}_{\text{gives } E_\theta} r\sin\theta\,\boldsymbol{d}\phi \wedge \boldsymbol{d}r]\}.&
\end{aligned}
\tag{4.24}
$$

## §4.5. MAXWELL'S EQUATIONS

The general 2-form $\boldsymbol{F}$ is written as a superposition of wedge products with a factor $\frac{1}{2}$,

$$\boldsymbol{F} = \frac{1}{2} F_{\mu\nu}\,\boldsymbol{d}x^{\mu} \wedge \boldsymbol{d}x^{\nu}, \tag{4.25}$$

because the typical term appears twice, once as $F_{xy}\,\boldsymbol{d}x \wedge \boldsymbol{d}y$ and the second time as $F_{yx}\,\boldsymbol{d}y \wedge \boldsymbol{d}x$, with $F_{yx} = -F_{xy}$ and $\boldsymbol{d}y \wedge \boldsymbol{d}x = -\boldsymbol{d}x \wedge \boldsymbol{d}y$.

If differentiation ("taking the gradient"; the operator $\boldsymbol{d}$) produced out of a scalar a 1-form, it is also true that differentiation (again the operator $\boldsymbol{d}$, but now generally known under Cartan's name of "exterior differentiation") produces a 2-form out of the general 1-form; and applied to a 2-form produces a 3-form; and applied to a 3-form produces a 4-form, the form of the highest order that spacetime will accommodate. Write the general $f$-form as

$$\boldsymbol{\phi} = \frac{1}{f!}\,\phi_{\alpha_1\alpha_2\ldots\alpha_f}\,\boldsymbol{d}x^{\alpha_1} \wedge \boldsymbol{d}x^{\alpha_2} \wedge \cdots \wedge \boldsymbol{d}x^{\alpha_f} \tag{4.26}$$

Taking exterior derivative

where the coefficient $\phi_{\alpha_1\alpha_2\ldots\alpha_f}$, like the wedge product that follows it, is antisymmetric under interchange of any two indices. Then the exterior derivative of $\boldsymbol{\phi}$ is

$$\boldsymbol{d\phi} \equiv \frac{1}{f!}\,\frac{\partial\phi_{\alpha_1\alpha_2\ldots\alpha_f}}{\partial x^{\alpha_0}}\,\boldsymbol{d}x^{\alpha_0} \wedge \boldsymbol{d}x^{\alpha_1} \wedge \boldsymbol{d}x^{\alpha_2} \wedge \cdots \wedge \boldsymbol{d}x^{\alpha_f}. \tag{4.27}$$

Take the exterior derivative of ***Faraday*** according to this rule and find that it vanishes, not only for the special case of the dipole oscillator, but also for a general electromagnetic field. Thus, in the coordinates appropriate for a local Lorentz frame, one has

$$\begin{aligned}\boldsymbol{dF} &= \boldsymbol{d}(E_x\,\boldsymbol{dx}\wedge\boldsymbol{dt} + \cdots + B_x\,\boldsymbol{dy}\wedge\boldsymbol{dz} + \cdots)\\ &= \left(\frac{\partial E_x}{\partial t}\boldsymbol{dt} + \frac{\partial E_x}{\partial x}\boldsymbol{dx} + \frac{\partial E_x}{\partial y}\boldsymbol{dy} + \frac{\partial E_x}{\partial z}\boldsymbol{dz}\right)\wedge\boldsymbol{dx}\wedge\boldsymbol{dt}\\ &\quad + \cdots \text{ (5 more such sets of 4 terms each)}\ldots. \end{aligned} \tag{4.28}$$

Note that such a term as $\boldsymbol{dy}\wedge\boldsymbol{dy}\wedge\boldsymbol{dz}$ is automatically zero ("collapse of egg-crate cell when stamped on"). Collect the terms that do not vanish and find

$$\begin{aligned}\boldsymbol{dF} = &\left(\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z}\right)\boldsymbol{dx}\wedge\boldsymbol{dy}\wedge\boldsymbol{dz}\\ &+\left(\frac{\partial B_x}{\partial t} + \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right)\boldsymbol{dt}\wedge\boldsymbol{dy}\wedge\boldsymbol{dz}\\ &+\left(\frac{\partial B_y}{\partial t} + \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}\right)\boldsymbol{dt}\wedge\boldsymbol{dz}\wedge\boldsymbol{dx}\\ &+\left(\frac{\partial B_z}{\partial t} + \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}\right)\boldsymbol{dt}\wedge\boldsymbol{dx}\wedge\boldsymbol{dy}. \end{aligned} \tag{4.29}$$

Each term in this expression is familiar from Maxwell's equations

$$\operatorname{div}\boldsymbol{B} = \boldsymbol{\nabla}\cdot\boldsymbol{B} = 0$$

and

$$\operatorname{curl}\boldsymbol{E} = \boldsymbol{\nabla}\times\boldsymbol{E} = -\dot{\boldsymbol{B}}.$$

Each vanishes, and with their vanishing ***Faraday*** itself is seen to have zero exterior derivative:

$$\boldsymbol{dF} = 0. \tag{4.30}$$

In other words, "***Faraday*** is a closed 2-form"; "the tubes of $\boldsymbol{F}$ nowhere come to an end."

***Faraday*** structure: tubes nowhere end

A similar calculation gives for the exterior derivative of the dual 2-form ***Maxwell*** the result

$$\begin{aligned}\boldsymbol{d{}^*F} &= \boldsymbol{d}(-B_x\,\boldsymbol{dx}\wedge\boldsymbol{dt} - \cdots + E_x\,\boldsymbol{dy}\wedge\boldsymbol{dz} + \cdots)\\ &= \left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}\right)\boldsymbol{dx}\wedge\boldsymbol{dy}\wedge\boldsymbol{dz}\\ &\quad+\left(\frac{\partial E_x}{\partial t} - \frac{\partial B_z}{\partial y} + \frac{\partial B_y}{\partial z}\right)\boldsymbol{dt}\wedge\boldsymbol{dy}\wedge\boldsymbol{dz}\\ &\quad+\cdots\\ &= 4\pi(\rho\,\boldsymbol{dx}\wedge\boldsymbol{dy}\wedge\boldsymbol{dz}\\ &\qquad - J_x\,\boldsymbol{dt}\wedge\boldsymbol{dy}\wedge\boldsymbol{dz}\\ &\qquad - J_y\,\boldsymbol{dt}\wedge\boldsymbol{dz}\wedge\boldsymbol{dx}\\ &\qquad - J_z\,\boldsymbol{dt}\wedge\boldsymbol{dx}\wedge\boldsymbol{dy}) = 4\pi\,{}^*\boldsymbol{J};\end{aligned}$$

$$\boldsymbol{d{}^*F} = 4\pi\,{}^*\boldsymbol{J}. \tag{4.31}$$

***Maxwell*** structure: density of tube endings given by charge-current 3-form

In empty space this exterior derivative, too, vanishes; there ***Maxwell*** is a closed 2-form; the tubes of $\boldsymbol{{}^*F}$, like the tubes of $\boldsymbol{F}$, nowhere come to an end.

In a region where charge is present, the situation changes. Tubes of ***Maxwell*** take their origin in such a region. The density of endings is described by the 3-form $\boldsymbol{{}^*J} =$ ***charge***, a "collection of eggcrate cells" collected along bundles of world lines.

The two equations

$$\boldsymbol{dF} = 0$$

and

$$\boldsymbol{d{}^*F} = 4\pi\, \boldsymbol{{}^*J}$$

summarize the entire content of Maxwell's equations in geometric language. The forms $\boldsymbol{F} =$ ***Faraday***, and $\boldsymbol{{}^*F} =$ ***Maxwell***, can be described in any coordinates one pleases—or in a language (honeycomb and egg-crate structures) free of any reference whatsoever to coordinates. Remarkably, neither equation makes any reference whatsoever to *metric*. As Hermann Weyl was one of the most emphatic in stressing (see also Chapters 8 and 9), the concepts of form and exterior derivative are metric-free. Metric made an appearance only in one place, in the concept of duality ("perpendicularity") that carried attention from $\boldsymbol{F}$ to the dual structure $\boldsymbol{{}^*F}$.

Duality: the only place in electromagnetism where metric must enter

## §4.6. EXTERIOR DERIVATIVE AND CLOSED FORMS

The words "honeycomb" and "egg crate" may have given some feeling for the geometry that goes with electrodynamics. Now to spell out these concepts more clearly and illustrate in geometric terms, with electrodynamics as subject matter, what it means to speak of "exterior differentiation." Marching around a boundary, yes; but how and why and with what consequences? It is helpful to return to functions and 1-forms, and see them and the 2-forms ***Faraday*** and ***Maxwell*** and the 3-form ***charge*** as part of an ordered progression (see Box 4.4). Two-forms are seen in this box to be of two kinds: (1) a special 2-form, known as a "closed" 2-form, which has the property that as many tubes enter a closed 2-surface as emerge from it (exterior derivative of 2-form zero; no 3-form derivable from it other than the trivial *zero* 3-form!); and (2) a general 2-form, which sends across a closed 2-surface a non-zero net number of tubes, and therefore permits one to define a nontrivial 3-form ("exterior derivative of the 2-form"), which has precisely as many egg-crate cells in any closed 2-surface as the net number of tubes of the 2-form emerging from that same closed 2-surface (generalization of Faraday's concept of tubes of force to the world of spacetime, curved as well as flat).

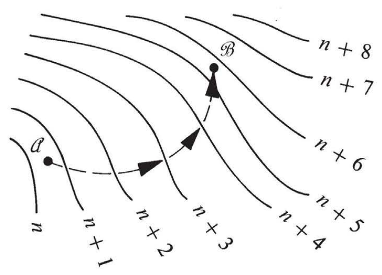
Closed 2-form contrasted with general 2-form

*(continued on page 120)*

**Box 4.4 THE PROGRESSION OF FORMS AND EXTERIOR DERIVATIVES**

### *0*-Form or Scalar, $f$

An example in the context of 3-space and Newtonian physics is temperature, $T(x, y, z)$, and in the context of spacetime, a scalar potential, $\phi(t, x, y, z)$.

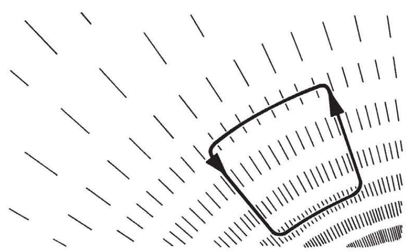



### From Scalar to *1*-Form

Take the gradient or "exterior derivative" of a scalar $f$ to obtain a special 1-form, $\boldsymbol{\gamma} = \boldsymbol{d}f$. Comments: (a) Any additive constant included in $f$ is erased in the process of differentiation; the quantity $n$ in the diagram at the left is unknown and irrelevant. (b) The 1-form $\boldsymbol{\gamma}$ is special in the sense that surfaces in one region "mesh" with surfaces in a neighboring region ("closed 1-form"). (c) Line integral $\int_{\mathscr{A}}^{\mathscr{B}} \boldsymbol{d}f$ is independent of path for any class of paths equivalent to one another under continuous deformation. (d) The 1-form is a machine to produce a number ("bongs of bell" as each successive integral surface is crossed) out of a displacement (approximation to concept of a tangent vector).

### General *1*-Form $\boldsymbol{\beta} = \beta_\alpha \, \boldsymbol{d}x^\alpha$

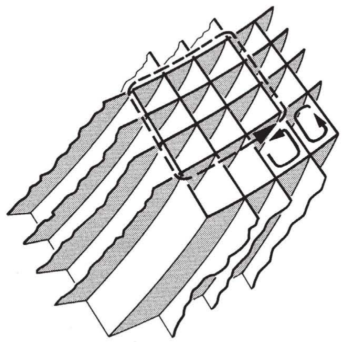
This is a pattern of surfaces, as illustrated in the diagram at the right; i.e., a machine to produce a number ("bongs of bell"; $\langle \boldsymbol{\beta}, \boldsymbol{u} \rangle$) out of a vector. A 1-form has a reality and position in space independent of all choice of coordinates. Surfaces do not ordinarily mesh. Integral $\int \boldsymbol{\beta}$ around indicated closed loop does not give zero ("more bongs than antibongs").



### From *1*-Form to *2*-Form $\boldsymbol{\xi} = \boldsymbol{d\beta} = \dfrac{\partial \beta_\alpha}{\partial x^\mu} \boldsymbol{d}x^\mu \wedge \boldsymbol{d}x^\alpha$

$\boldsymbol{\xi}$ is a pattern of honeycomb-like cells, with a direction of circulation marked on each, so stationed

**Box 4.4 (*continued*)**



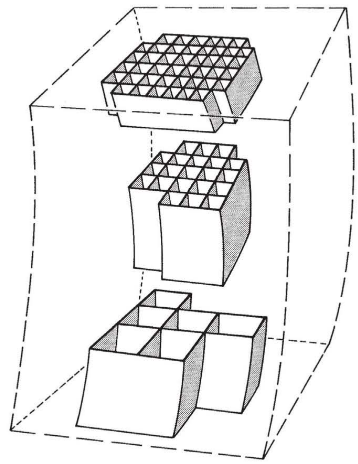
that the number of cells encompassed in the dotted closed path is identical to the net contribution (excess of bongs over antibongs) for the same path in the diagram of $\boldsymbol{\beta}$ above. The "exterior derivative" is *defined* so this shall be so; the generalized Stokes theorem codifies it. The word "exterior" comes from the fact that the path goes around the periphery of the region under analysis. Thus the 2-form is a machine to get a number (number of tubes, $\langle \boldsymbol{\xi}, \boldsymbol{u} \wedge \boldsymbol{v} \rangle$) out of a bit of surface ($\boldsymbol{u} \wedge \boldsymbol{v}$) that has a sense of circulation indicated upon it. The 2-form thus defined is special in this sense: a rubber sheet "supported around its edges" by the dotted curve or any other closed curve is crossed by the same number of tubes when: (a) it bulges up in the middle; (b) it is pushed down in the middle; (c) it experiences any other continuous deformation. The ***Faraday*** or 2-form $\boldsymbol{F}$ of electromagnetism, always expressible as $\boldsymbol{F} = \boldsymbol{d}\boldsymbol{A}$ ($\boldsymbol{A}$ = 4-potential, a 1-form), also has always this special property ("conservation of tubes").

### *0*-Form to *1*-Form to *2*-Form? No!

Go from scalar $f$ to 1-form $\boldsymbol{\gamma} = \boldsymbol{d}f$. The next step to a 2-form $\boldsymbol{\alpha}$ is vacuous. The net contribution of the line integral $\int \boldsymbol{\gamma}$ around the dotted closed path is automatically zero. To reproduce that zero result requires a zero 2-form. Thus $\boldsymbol{\alpha} = \boldsymbol{d}\boldsymbol{\gamma} = \boldsymbol{dd}f$ has to be the zero 2-form. This result is a special instance of the general result $\boldsymbol{dd} = 0$.

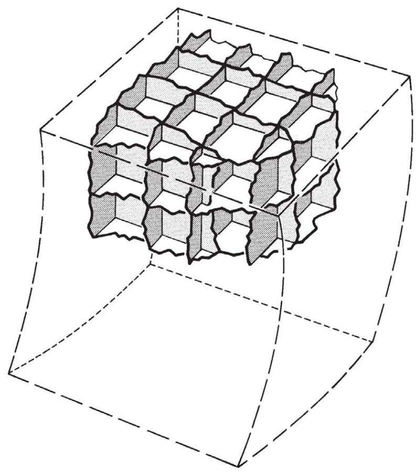
### General *2*-Form $\boldsymbol{\sigma} = \frac{1}{2}\sigma_{\alpha\beta}\, \boldsymbol{d}x^{\alpha} \wedge \boldsymbol{d}x^{\beta}$, with $\sigma_{\alpha\beta} = -\sigma_{\beta\alpha}$

Again, this is a honeycomb-like structure, and again a machine to get a number (number of tubes, $\langle \boldsymbol{\sigma}, \boldsymbol{u} \wedge \boldsymbol{v} \rangle$) out of a surface ($\boldsymbol{u} \wedge \boldsymbol{v}$) that has a sense of circulation indicated on it. It is general in the sense that the honeycomb structures in one region do not ordinarily mesh with those in a neighboring region. In consequence, a closed 2-surface, such as the box-like surface indicated by dotted lines at the right, is ordinarily crossed by a non-zero net number of tubes. The net number of tubes emerging from such a closed surface is, however, exactly zero when the 2-form is the exterior derivative of a 1-form.



**From *2*-Form to *3*-Form** $\boldsymbol{\mu} = \boldsymbol{d\sigma} = \dfrac{\partial \sigma_{|\alpha\beta|}}{\partial x^\gamma} \boldsymbol{d}x^\gamma \wedge \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta$,
**where** $\boldsymbol{d}x^\gamma \wedge \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta \equiv 3!\, \boldsymbol{d}x^{[\gamma} \otimes \boldsymbol{d}x^\alpha \otimes \boldsymbol{d}x^{\beta]}$

This egg-crate type of structure is a machine to get a number (number of cells $\langle \boldsymbol{\mu}, \boldsymbol{u} \wedge \boldsymbol{v} \wedge \boldsymbol{w} \rangle$) from a volume (volume $\boldsymbol{u} \wedge \boldsymbol{v} \wedge \boldsymbol{w}$ within which one counts the cells). A more complete diagram would provide each cell and the volume of integration itself with an indicator of orientation (analogous to the arrow of circulation shown for cells of the 2-form). The contribution of a given cell to the count of cells is $+1$ or $-1$, according as the orientation indicators have same sense or opposite sense. The number of egg-crate cells of $\boldsymbol{\mu} = \boldsymbol{d\sigma}$ in any given volume (such as the volume indicated by the dotted lines) is tailored to give precisely the same number as the net number of tubes of the 2-form $\boldsymbol{\sigma}$ (diagram above) that emerge from that volume (generalized Stokes theorem). For electromagnetism, the exterior derivative of ***Faraday*** or 2-form $\boldsymbol{F}$ gives a null 3-form, but the exterior derivative of ***Maxwell*** or 2-form $^*\boldsymbol{F}$ gives $4\pi$ times the 3-form $^*\boldsymbol{J}$ of charge:



$$\begin{aligned} {}^*\boldsymbol{J} = \rho\, \boldsymbol{d}x \wedge \boldsymbol{d}y \wedge \boldsymbol{d}z - J_x\, \boldsymbol{d}t \wedge \boldsymbol{d}y \wedge \boldsymbol{d}z \\ - J_y\, \boldsymbol{d}t \wedge \boldsymbol{d}z \wedge \boldsymbol{d}x - J_z\, \boldsymbol{d}t \wedge \boldsymbol{d}x \wedge \boldsymbol{d}y. \end{aligned}$$

**Box 4.4 (*continued*)**

### From *1*-Form to *2*-Form to *3*-Form? No!

Starting with a 1-form (electromagnetic 4-potential), construct its exterior derivative, the 2-form $\boldsymbol{F} = \boldsymbol{dA}$ (***Faraday***). The tubes in this honeycomb-like structure never end. So the number of tube endings in any elementary volume, and with it the 3-form $\boldsymbol{dF} = \boldsymbol{ddA}$, is automatically zero. This is another example of the general result that $\boldsymbol{dd} = 0$.

### From *2*-Form to *3*-Form to *4*-Form? No!

Starting with 2-form $^*\boldsymbol{F}$ (***Maxwell***), construct its exterior derivative, the 3-form $4\pi\,^*\boldsymbol{J}$. The cells in this egg-crate type of structure extend in a fourth dimension ("hypertube"). The number of these hypertubes that end in any elementary 4-volume, and with it the 4-form

$$\boldsymbol{d}(4\pi\,^*\boldsymbol{J}) = \boldsymbol{dd}^*\boldsymbol{F},$$

is automatically zero, still another example of the general result that $\boldsymbol{dd} = 0$. This result says that

$$\boldsymbol{d}^*\boldsymbol{J} = \left(\frac{\partial \rho}{\partial t} + \frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y} + \frac{\partial J_z}{\partial z}\right)\boldsymbol{d}t \wedge \boldsymbol{d}x \wedge \boldsymbol{d}y \wedge \boldsymbol{d}z = 0$$

("law of conservation of charge"). Note:

$$\boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta \wedge \boldsymbol{d}x^\gamma \wedge \boldsymbol{d}x^\delta \equiv 4!\, \boldsymbol{d}x^{[\alpha} \otimes \boldsymbol{d}x^\beta \otimes \boldsymbol{d}x^\gamma \otimes \boldsymbol{d}x^{\delta]}.$$

This implies $\boldsymbol{d}t \wedge \boldsymbol{d}x \wedge \boldsymbol{d}y \wedge \boldsymbol{d}z = \boldsymbol{\varepsilon}$.

### From *3*-Form to *4*-Form $\boldsymbol{\tau} = \boldsymbol{d\nu} = \dfrac{\partial \nu_{|\alpha\beta\gamma|}}{\partial x^\delta}\boldsymbol{d}x^\delta \wedge \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta \wedge \boldsymbol{d}x^\gamma$

This four-dimensional "super-egg-crate" type structure is a machine to get a number (number of cells, $\langle \boldsymbol{\tau}, \boldsymbol{n} \wedge \boldsymbol{u} \wedge \boldsymbol{v} \wedge \boldsymbol{w}\rangle$) from a 4-volume $\boldsymbol{n} \wedge \boldsymbol{u} \wedge \boldsymbol{v} \wedge \boldsymbol{w}$.

### From *4*-Form to *5*-Form? No!

Spacetime, being four-dimensional, cannot accommodate five-dimensional egg-crate structures. At least two of the $\boldsymbol{d}x^{\mu}$'s in

$$\boldsymbol{d}x^{\alpha} \wedge \boldsymbol{d}x^{\beta} \wedge \boldsymbol{d}x^{\gamma} \wedge \boldsymbol{d}x^{\delta} \wedge \boldsymbol{d}x^{\varepsilon}$$

must be the same; so, by antisymmetry of "$\wedge$," this "basis 5-form" must vanish.

### Results of Exterior Differentiation, Summarized

| | | | | | |
|---|---|---|---|---|---|
| 0-form | $f$ | | | | |
| 1-form | $\boldsymbol{d}f$ | $\boldsymbol{A}$ | | | |
| 2-form | $\boldsymbol{dd}f \equiv 0$ | $\boldsymbol{F} = \boldsymbol{dA}$ | $^*\boldsymbol{F}$ | | |
| 3-form | | $\boldsymbol{dF} = \boldsymbol{ddA} \equiv 0$ | $4\pi\, {}^*\boldsymbol{J} = \boldsymbol{d}\,{}^*\boldsymbol{F}$ | $\boldsymbol{\nu}$ | |
| 4-form | | | $\boldsymbol{d}(4\pi\, {}^*\boldsymbol{J}) = \boldsymbol{dd}\,{}^*\boldsymbol{F} \equiv 0$ | $\boldsymbol{\tau} = \boldsymbol{d\nu}$ | $\boldsymbol{\mu}$ |
| 5-form? | No! | | | $\boldsymbol{d\tau} \equiv 0$ | $\boldsymbol{d\mu} \equiv 0$ |

### New Forms from Old by Taking Dual (see exercise 3.14)

Dual of scalar $f$ is 4-form: $^*f = f\,\boldsymbol{d}x^0 \wedge \boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 = f\boldsymbol{\varepsilon}$.

Dual of 1-form $\boldsymbol{J}$ is 3-form: $$^*\boldsymbol{J} = J^0\,\boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 - J^1\,\boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3 \wedge \boldsymbol{d}x^0 + J^2\,\boldsymbol{d}x^3 \wedge \boldsymbol{d}x^0 \wedge \boldsymbol{d}x^1 - J^3\,\boldsymbol{d}x^0 \wedge \boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2.$$

Dual of 2-form $\boldsymbol{F}$ is 2-form: $^*\boldsymbol{F} = F^{|\alpha\beta|}\varepsilon_{\alpha\beta|\mu\nu|}\,\boldsymbol{d}x^{\mu} \wedge \boldsymbol{d}x^{\nu}$, where $F^{\alpha\beta} = \eta^{\alpha\lambda}\eta^{\beta\delta}F_{\lambda\delta}$.

Dual of 3-form $\boldsymbol{K}$ is 1-form: $^*\boldsymbol{K} = K^{012}\,\boldsymbol{d}x^3 - K^{123}\,\boldsymbol{d}x^0 + K^{230}\,\boldsymbol{d}x^1 - K^{301}\,\boldsymbol{d}x^2$, where $K^{\alpha\beta\gamma} = \eta^{\alpha\mu}\eta^{\beta\nu}\eta^{\gamma\lambda}K_{\mu\nu\lambda}$.

Dual of 4-form $\boldsymbol{L}$ is a scalar: $\boldsymbol{L} = L_{0123}\,\boldsymbol{d}x^0 \wedge \boldsymbol{d}x^1 \wedge \boldsymbol{d}x^2 \wedge \boldsymbol{d}x^3$; $^*\boldsymbol{L} = L^{0123} = -L_{0123}$.

Note 1: This concept of duality between *one form and another* is to be distinguished from the concept of duality between the *vector basis* $\boldsymbol{e}_{\alpha}$ and the *1-form basis* $\boldsymbol{\omega}^{\alpha}$ of a given frame. The two types of duality have nothing whatsoever to do with each other!

**Box 4.4 (*continued*)**

Note 2: In spacetime, the operation of taking the dual, applied twice, leads back to the original form for forms of odd order, and to the negative thereof for forms of even order. In Euclidean 3-space the operation reproduces the original form, regardless of its order.

### Duality Plus Exterior Differentiation

*Start with scalar* $\phi$. Its gradient $\boldsymbol{d}\phi$ is a 1-form. Take its dual, to get the 3-form $^*\boldsymbol{d}\phi$. Take its exterior derivative, to get the 4-form $\boldsymbol{d}{}^*\boldsymbol{d}\phi$. Take its dual, to get the scalar $\square\phi \equiv -{}^*\boldsymbol{d}{}^*\boldsymbol{d}\phi$. Verify by index manipulations that $\square$ as defined here is the wave operator; i.e., in any Lorentz frame, $\square\phi = \phi_{,\alpha}{}^{,\alpha} = -(\partial^2\phi/\partial t^2) + \nabla^2\phi$.

*Start with 1-form* $\boldsymbol{A}$. Get 2-form $\boldsymbol{F} = \boldsymbol{dA}$. Take its dual $^*\boldsymbol{F} = {}^*\boldsymbol{dA}$, also a 2-form. Take its exterior derivative, obtaining the 3-form $\boldsymbol{d}{}^*\boldsymbol{F}$ (has value $4\pi\, {}^*\boldsymbol{J}$ in electromagnetism). Take its dual, obtaining the 1-form $^*\boldsymbol{d}{}^*\boldsymbol{F} = {}^*\boldsymbol{d}{}^*\boldsymbol{dA} = 4\pi\boldsymbol{J}$ ("Wave equation for electromagnetic 4-potential"). Reduce in index notation to

$$F_{\mu\nu}{}^{,\nu} = A_{\nu,\mu}{}^{,\nu} - A_{\mu,\nu}{}^{,\nu} = 4\pi J_\mu.$$

[More in Flanders (1963) or Misner and Wheeler (1957); see also exercise 3.17.]

## §4.7. DISTANT ACTION FROM LOCAL LAW

Differential forms are a powerful tool in electromagnetic theory, but full power requires mastery of other tools as well. Action-at-a-distance techniques ("Green's functions," "propagators") are of special importance. Moreover, the passage from Maxwell field equations to electromagnetic action at a distance provides a preview of how Einstein's local equations will reproduce (approximately) Newton's $1/r^2$ law.

In flat spacetime and in a Lorentz coordinate system, express the coordinates of particle $A$ as a function of its proper time $\alpha$, thus:

$$a^\mu = a^\mu(\alpha), \qquad \frac{da^\mu}{d\alpha} = \dot{a}^\mu(\alpha), \qquad \frac{d^2a^\mu}{d\alpha^2} = \ddot{a}^\mu(\alpha). \tag{4.32}$$

Dirac found it helpful to express the distribution of charge and current for a particle of charge $e$ following such a motion as a superposition of charges that momentarily flash into existence and then flash out of existence. Any such flash has a localization in space and time that can be written as the product of four Dirac delta functions [see, for example, Schwartz (1950–1951), Lighthill (1958)]:

World line of charge regarded as succession of flash-on, flash-off charges

$$\delta^4(x^\mu - a^\mu) = \delta[x^0 - a^0(\alpha)]\,\delta[x^1 - a^1(\alpha)]\,\delta[x^2 - a^2(\alpha)]\,\delta[x^3 - a^3(\alpha)]. \quad (4.33)$$

Here any single Dirac function $\delta(x)$ ("symbolic function"; "distribution"; "limit of a Gauss error function" as width is made indefinitely narrow and peak indefinitely high, with integrated value always unity) both (1) vanishes for $x \neq 0$, and (2) has the integral $\int_{-\infty}^{+\infty} \delta(x)\,dx = 1$. Described in these terms, the density-current vector for the particle has the value ("superposition of flashes")

$$J^\mu = e \int \delta^4[x^\nu - a^\nu(\alpha)]\dot{a}^\mu(\alpha)\,d\alpha. \quad (4.34)$$

The density-current (4.34) drives the electromagnetic field, $\boldsymbol{F}$. Write $\boldsymbol{F} = \boldsymbol{dA}$ to satisfy automatically half of Maxwell's equations ($\boldsymbol{dF} = \boldsymbol{ddA} \equiv 0$):

$$F_{\mu\alpha} = \frac{\partial A_\alpha}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\alpha}. \quad (4.35)$$

In flat space, the remainder of Maxwell's equations ($\boldsymbol{d}{}^*\boldsymbol{F} = 4\pi\,{}^*\boldsymbol{J}$) become

$$\frac{\partial F_\mu{}^\nu}{\partial x^\nu} = 4\pi J_\mu$$

or

$$\frac{\partial}{\partial x^\mu}\frac{\partial A^\nu}{\partial x^\nu} - \eta^{\nu\alpha}\frac{\partial^2 A_\mu}{\partial x^\nu\,\partial x^\alpha} = 4\pi J_\mu. \quad (4.36)$$

Make use of the freedom that exists in the choice of 4-potentials $A^\nu$ to demand

$$\frac{\partial A^\nu}{\partial x^\nu} = 0 \quad (4.37)$$

(Lorentz gauge condition; see exercise 3.17). Thus get

$$\Box A_\mu = -4\pi J_\mu. \quad (4.38)$$

The electromagnetic wave equation

The density-current being the superposition of "flashes," the effect ($\boldsymbol{A}$) of this density-current can be expressed as the superposition of the effects $E$ of elementary flashes; thus

$$A^\mu(x) = \int E[x - a(\alpha)]\dot{a}^\mu(\alpha)\,d\alpha, \quad (4.39)$$

The solution of the wave equation

where the "elementary effect" $E$ ("kernel"; "Green's function") satisfies the equation

$$\Box E(x) = -4\pi\,\delta^4(x). \quad (4.40)$$

One solution is the "half-advanced-plus-half-retarded potential,"

$$E(x) = \delta(\eta_{\alpha\beta}x^\alpha x^\beta). \quad (4.41)$$

It vanishes everywhere except on the backward and forward light cones, where it has equal strength. Normally more useful is the retarded solution,

$$R(x) = \begin{cases} 2E(x) & \text{if } x^0 > 0, \\ 0 & \text{if } x^0 < 0, \end{cases} \tag{4.42}$$

which is obtained by doubling (4.41) in the region of the forward light cone and nullifying it in the region of the backward light cone. All electrodynamics (Coulomb forces, Ampère's law, electromagnetic induction, radiation) follows from the simple expression (4.39) for the vector potential [see, e.g., Wheeler and Feynman (1945) and (1949), also Rohrlich (1965)].

---

## EXERCISES

### Exercise 4.1. GENERIC LOCAL ELECTROMAGNETIC FIELD EXPRESSED IN SIMPLEST FORM

In the laboratory Lorentz frame, the electric field is $\boldsymbol{E}$, the magnetic field $\boldsymbol{B}$. Special cases are: (1) pure electric field ($\boldsymbol{B} = 0$); (2) pure magnetic field ($\boldsymbol{E} = 0$); and (3) "radiation field" or "null field" ($\boldsymbol{E}$ and $\boldsymbol{B}$ equal in magnitude and perpendicular in direction). All cases other than (1), (2), and (3) are "generic." In the generic case, calculate the Poynting density of flow of energy $\boldsymbol{E} \times \boldsymbol{B}/4\pi$ and the density of energy $(\boldsymbol{E}^2 + \boldsymbol{B}^2)/8\pi$. Define the direction of a unit vector $\boldsymbol{n}$ and the magnitude of a velocity parameter $\alpha$ by the ratio of energy flow to energy density:

$$\boldsymbol{n} \tanh 2\alpha = \frac{2\boldsymbol{E} \times \boldsymbol{B}}{\boldsymbol{E}^2 + \boldsymbol{B}^2}.$$

View the same electromagnetic field in a rocket frame moving in the direction of $\boldsymbol{n}$ with the velocity parameter $\alpha$ (not $2\alpha$; factor 2 comes in because energy flow and energy density are components, not of a vector, but of a tensor). By employing the formulas for a Lorentz transformation (equation 3.23), or otherwise, show that the energy flux vanishes in the rocket frame, with the consequence that $\bar{\boldsymbol{E}}$ and $\bar{\boldsymbol{B}}$ are parallel. No one can prevent the $\bar{z}$-axis from being put in the direction common to $\bar{\boldsymbol{E}}$ and $\bar{\boldsymbol{B}}$. Show that with this choice of direction, ***Faraday*** becomes

$$\boldsymbol{F} = \bar{E}_z\, \boldsymbol{d}\bar{z} \wedge \boldsymbol{d}\bar{t} + \bar{B}_z\, \boldsymbol{d}\bar{x} \wedge \boldsymbol{d}\bar{y}$$

(only two wedge products needed to represent the generic local field; "canonical representation"; valid in one frame, valid in any frame).

### Exercise 4.2. FREEDOM OF CHOICE OF 1-FORMS IN CANONICAL REPRESENTATION OF GENERIC LOCAL FIELD

Deal with a region so small that the variation of the field from place to place can be neglected. Write ***Faraday*** in canonical representation in the form

$$\boldsymbol{F} = \boldsymbol{d}p_I \wedge \boldsymbol{d}q^I + \boldsymbol{d}p_{II} \wedge \boldsymbol{d}q^{II},$$

where $p_A$ ($A = I$ or $II$) and $q^A$ are scalar functions of position in spacetime. Define a "canonical transformation" to new scalar functions of position $p_{\bar{A}}$ and $q^{\bar{A}}$ by way of the "equation of transformation"

$$p_A\, \boldsymbol{d}q^A = \boldsymbol{d}S + p_{\bar{A}}\, \boldsymbol{d}q^{\bar{A}},$$

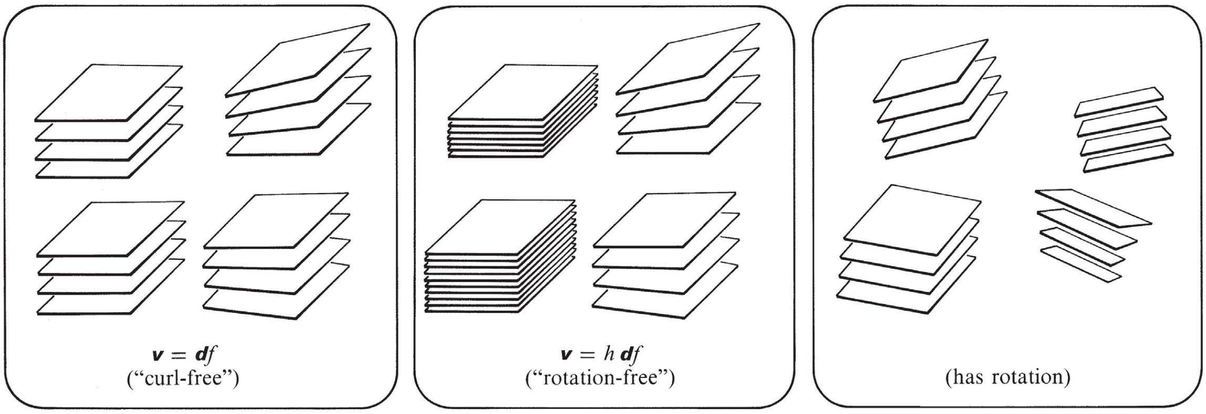



**Figure 4.7.**
Some simple types of 1-forms compared and contrasted.

where the "generating function" $S$ of the transformation is an arbitrary function of the $q^A$ and the $q^{\bar{A}}$:

$$\boldsymbol{d}S = (\partial S/\partial q^A)\,\boldsymbol{d}q^A + (\partial S/\partial q^{\bar{A}})\,\boldsymbol{d}q^{\bar{A}}.$$

(a) Derive expressions for the two $p_A$'s and the two $p_{\bar{A}}$'s in terms of $S$ by equating coefficients of $dq^I$, $dq^{II}$, $dq^{\bar{I}}$, $dq^{\bar{II}}$ individually on the two sides of the equation of transformation.

(b) Use these expressions for the $p_A$'s and $p_{\bar{A}}$'s to show that $\boldsymbol{F} = \boldsymbol{d}p_A \wedge \boldsymbol{d}q^A$ and $\bar{\boldsymbol{F}} = \boldsymbol{d}p_{\bar{A}} \wedge \boldsymbol{d}q^{\bar{A}}$, ostensibly different, are actually expressions for one and the same 2-form in terms of alternative sets of 1-forms.

### Exercise 4.3. A CLOSED OR CURL-FREE 1-FORM IS A GRADIENT

Given a 1-form $\boldsymbol{\sigma}$ such that $\boldsymbol{d\sigma} = 0$, show that $\boldsymbol{\sigma}$ can be expressed in the form $\boldsymbol{\sigma} = \boldsymbol{d}f$, where $f$ is some scalar. The 1-form $\boldsymbol{\sigma}$ is said to be "curl-free," a narrower category of 1-form than the "rotation-free" 1-form of the next exercise (expressible as $\boldsymbol{\sigma} = h\,\boldsymbol{d}f$), and it in turn is narrower (see Figure 4.7) than the category of "1-forms with rotation" (not expressible in the form $\boldsymbol{\sigma} = h\,\boldsymbol{d}f$). When the 1-form $\boldsymbol{\sigma}$ is expressed in terms of basis 1-forms $\boldsymbol{d}x^\alpha$, multiplied by corresponding components $\sigma_\alpha$, show that "curl-free" implies $\sigma_{[\alpha,\beta]} = 0$.

### Exercise 4.4. CANONICAL EXPRESSION FOR A ROTATION-FREE 1-FORM

In three dimensions a rigid body turning with angular velocity $\omega$ about the $z$-axis has components of velocity $v_y = \omega x$, and $v_x = -\omega y$. The quantity curl $\boldsymbol{v} = \boldsymbol{\nabla} \times \boldsymbol{v}$ has $z$-component equal to $2\omega$, and all other components equal to zero. Thus the scalar product of $\boldsymbol{v}$ and curl $\boldsymbol{v}$ vanishes:

$$v_{[i,j}v_{k]} = 0.$$

The same concept generalizes to four dimensions,

$$v_{[\alpha,\beta}v_{\gamma]} = 0,$$

and lends itself to expression in coordinate-free language, as the requirement that a certain 3-form must vanish:

$$\boldsymbol{dv} \wedge \boldsymbol{v} = 0.$$

Any 1-form $\boldsymbol{v}$ satisfying this condition is said to be "rotation-free." Show that a 1-form is rotation-free if and only if it can be written in the form

$$\boldsymbol{v} = h\,\boldsymbol{d}f,$$

where $h$ and $f$ are scalar functions of position (the "Frobenius theorem").

**Exercise 4.5. FORMS ENDOWED WITH POLAR SINGULARITIES**

List the principal results on how such forms are representable, such as

$$\boldsymbol{\Phi}_1 = \frac{\boldsymbol{d}S}{S} \wedge \boldsymbol{\psi}_1 + \boldsymbol{\theta}_1,$$

and the conditions under which each applies [for the meaning and answer to this exercise, see Lascoux (1968)].

**Exercise 4.6. THE FIELD OF THE OSCILLATING DIPOLE**

Verify that the expressions given for the electromagnetic field of an oscillating dipole in equations (4.23) and (4.24) satisfy $\boldsymbol{dF} = 0$ everywhere and $\boldsymbol{d}{}^*\boldsymbol{F} = 0$ everywhere except at the origin.

**Exercise 4.7. THE 2-FORM MACHINERY TRANSLATED INTO TENSOR MACHINERY**

This exercise is stated at the end of the legend caption of Figure 4.1.

**Exercise 4.8. PANCAKING THE COULOMB FIELD**

Figure 4.5 shows a spacelike slice, $t =$ const, through the ***Maxwell*** of a point-charge at rest. By the following pictorial steps, verify that the electric-field lines get compressed into the transverse direction when viewed from a moving Lorentz frame: (1) Draw a picture of an equatorial slice ($\theta = \pi/2$; $t, r, \phi$ variable) through ***Maxwell*** $= {}^*\boldsymbol{F}$. (2) Draw various spacelike slices, corresponding to constant time in various Lorentz frames, through the resultant geometric structure. (3) Interpret the intersection of ***Maxwell*** $= {}^*\boldsymbol{F}$ with each Lorentz slice in the manner of Figure 4.3.

**Exercise 4.9. COMPUTATION OF SURFACE INTEGRALS**

In Box 4.1 the definition

$$\int \boldsymbol{\alpha} = \int \dots \int \left\langle \boldsymbol{\alpha}, \frac{\partial \mathscr{P}}{\partial \lambda^1} \wedge \cdots \wedge \frac{\partial \mathscr{P}}{\partial \lambda^p} \right\rangle d\lambda^1 \dots d\lambda^p$$

is given for the integral of a $p$-form $\boldsymbol{\alpha}$ over a $p$-surface $\mathscr{P}(\lambda^1, \dots, \lambda^p)$ in $n$-dimensional space. From this show that the following computational rule (also given in Box 4.1) works: (1) substitute the equation for the surface,

$$x^k = x^k(\lambda^1, \dots, \lambda^p),$$

into $\boldsymbol{\alpha}$ and collect terms in the form

$$\boldsymbol{\alpha} = a(\lambda^1, \dots, \lambda^p)\, \boldsymbol{d}\lambda^1 \wedge \cdots \wedge \boldsymbol{d}\lambda^p;$$

(2) integrate

$$\int \boldsymbol{\alpha} = \int \dots \int a(\lambda^1, \dots, \lambda^p)\, d\lambda^1 \dots d\lambda^p$$

using the elementary definition of integration.

**Exercise 4.10. WHITAKER'S CALUMOID, OR, THE LIFE OF A LOOP**

Take a closed loop, bounding a 2-dimensional surface $\mathcal{S}$. It entraps a certain flux of ***Faraday*** $\Phi_F = \int_{\mathcal{S}} \boldsymbol{F}$ ("magnetic tubes") and a certain flux of ***Maxwell*** $\Phi_M = \int_{\mathcal{S}} {}^*\boldsymbol{F}$ ("electric tubes").

(a) Show that the fluxes $\Phi_F$ and $\Phi_M$ depend only on the choice of loop, and not on the choice of the surface $\mathcal{S}$ bounded by the loop, if and only if $\boldsymbol{dF} = \boldsymbol{d}{}^*\boldsymbol{F} = 0$ (no magnetic charge; no electric charge). *Hint:* use generalized Stokes theorem, Boxes 4.1 and 4.6.

(b) Move the loop in space and time so that it continues to entrap the same two fluxes. Move it forward a little more here, a little less there, so that it continues to do so. In this way trace out a 2-dimensional surface ("calumoid"; see E. T. Whitaker 1904) $\mathcal{P} = \mathcal{P}(a, b)$; $x^\mu = x^\mu(a, b)$. Show that the elementary bivector in this surface, $\boldsymbol{\Sigma} = \partial\mathcal{P}/\partial a \wedge \partial\mathcal{P}/\partial b$ satisfies $\langle \boldsymbol{F}, \boldsymbol{\Sigma} \rangle = 0$ and $\langle {}^*\boldsymbol{F}, \boldsymbol{\Sigma} \rangle = 0$.

(c) Show that these differential equations for $x^\mu(a, b)$ can possess a solution, with given initial condition $x^\mu = x^\mu(a, 0)$ for the initial location of the loop, if $\boldsymbol{dF} = 0$ and $\boldsymbol{d}{}^*\boldsymbol{F} = 0$ (no magnetic charge, no electric charge).

(d) Consider a static, uniform electric field $\boldsymbol{F} = -E_x\, \boldsymbol{d}t \wedge \boldsymbol{d}x$. Solve the equations, $\langle \boldsymbol{F}, \boldsymbol{\Sigma} \rangle = 0$ and $\langle {}^*\boldsymbol{F}, \boldsymbol{\Sigma} \rangle = 0$ to find the equation $\mathcal{P}(a, b)$ for the most general calumoid. [*Answer:* $y = y(a)$, $z = z(a)$, $x = x(b)$, $t = t(b)$.] Exhibit two special cases: (i) a calumoid that lies entirely in a hypersurface of constant time [loop moves at infinite velocity; analogous to super-light velocity of point of crossing for two blades of a pair of scissors]; (ii) a calumoid whose loop remains forever at rest in the $t$, $x$, $y$, $z$ Lorentz frame.

**Exercise 4.11. DIFFERENTIAL FORMS AND HAMILTONIAN MECHANICS**

Consider a dynamic system endowed with two degrees of freedom. For the definition of this system as a Hamiltonian system (special case: here the Hamiltonian is independent of time), one needs (1) a definition of canonical variables (see Box 4.5) and (2) a knowledge of the Hamiltonian $H$ as a function of the coordinates $q^1$, $q^2$ and the canonically conjugate momenta $p_1$, $p_2$. To derive the laws of mechanics, consider the five-dimensional space of $p_1$, $p_2$, $q^1$, $q^2$, and $t$, and a curve in this space leading from starting values of the five coordinates (subscript $A$) to final values (subscript $B$), and the value

$$I = \int_A^B p_1\, \boldsymbol{d}q^1 + p_2\, \boldsymbol{d}q^2 - H(p, q)\, \boldsymbol{d}t = \int_A^B \boldsymbol{\omega}$$

of the integral $I$ taken along this path. The difference of the integral for two "neighboring" paths enclosing a two-dimensional region $\mathcal{S}$, according to the theorem of Stokes (Boxes 4.1 and 4.6), is

$$\delta I = \oint \boldsymbol{\omega} = \int_{\mathcal{S}} \boldsymbol{d\omega}.$$

The principle of least action (principle of "extremal history") states that the representative point of the system must travel along a route in the five-dimensional manifold (route with tangent vector $d\mathcal{P}/dt$) such that the variation vanishes for this path; i.e.,

$$\boldsymbol{d\omega}(\ldots, d\mathcal{P}/dt) = 0$$

(2-form $\boldsymbol{d\omega}$ with a single vector argument supplied, and other slot left unfilled, gives the 1-form in 5-space that must vanish). This fixes only the direction of $d\mathcal{P}/dt$; its magnitude can be normalized by requiring $\langle \boldsymbol{d}t, d\mathcal{P}/dt \rangle = 1$.

(a) Evaluate $\boldsymbol{d\omega}$ from the expression $\boldsymbol{\omega} = p_j\, \boldsymbol{d}q^j - H\, \boldsymbol{d}t$.

(b) Set $d\mathcal{P}/dt = \dot{q}^j(\partial\mathcal{P}/\partial q^j) + \dot{p}_j(\partial\mathcal{P}/\partial p_j) + \dot{t}(\partial\mathcal{P}/\partial t)$, and expand $\boldsymbol{d\omega}(\ldots, d\mathcal{P}/dt) = 0$ in terms of the basis $\{\boldsymbol{d}p_j, \boldsymbol{d}q^k, \boldsymbol{d}t\}$.

**Box 4.5 METRIC STRUCTURE AND HAMILTONIAN OR "SYMPLECTIC STRUCTURE" COMPARED AND CONTRASTED**

| | *Metric structure* | *Symplectic structure* |
|---|---|---|
| 1. Physical application | Geometry of spacetime | Hamiltonian mechanics |
| 2. Canonical structure | $(\ldots\cdot\ldots) = \text{“}\boldsymbol{ds}^2\text{”} = -\boldsymbol{d}t \otimes \boldsymbol{d}t + \boldsymbol{d}x \otimes \boldsymbol{d}x + \boldsymbol{d}y \otimes \boldsymbol{d}y + \boldsymbol{d}z \otimes \boldsymbol{d}z$ | $\boldsymbol{\Theta} = \boldsymbol{d}p_1 \wedge \boldsymbol{d}q^1 + \boldsymbol{d}p_2 \wedge \boldsymbol{d}q^2$ |
| 3. Nature of "metric" | Symmetric | Antisymmetric |
| 4. Name for given coordinate system and any other set of four coordinates in which metric has same form | Lorentz coordinate system | System of "canonically" (or "dynamically") conjugate coordinates |
| 5. Field equation for this metric | $\boldsymbol{R}_{\mu\nu\alpha\beta} = 0$ (zero Riemann curvature; flat spacetime) | $\boldsymbol{d\Theta} = 0$ ("closed 2-form"; condition automatically satisfied by expression above). |
| 6. The four-dimensional manifold | Spacetime | Phase space |
| 7. Coordinate-free description of the structure of this manifold | ***Riemann*** $= 0$ | $\boldsymbol{d\Theta} = 0$ |
| 8. Canonical coordinates distinguished from other coordinates (allowable but less simple) | Make metric take above form (item 2) | Make metric take above form (item 2) |

(c) Show that this five-dimensional equation can be written in the 4-dimensional phase space of $\{q^j, p_k\}$ as

$$\boldsymbol{\Theta}(\ldots, d\mathscr{P}/dt) = \boldsymbol{d}H,$$

where $\boldsymbol{\Theta}$ is the 2-form defined in Box 4.5.

(d) Show that the components of $\boldsymbol{\Theta}(\ldots, d\mathscr{P}/dt) = \boldsymbol{d}H$ in the $\{q^j, p_k\}$ coordinate system are the familiar Hamilton equations. Note that this conclusion depends only on the form assumed for $\boldsymbol{\Theta}$, so that one also obtains the standard Hamilton equations in any other phase-space coordinates $\{\bar{q}^j, \bar{p}_k\}$ ("canonical variables") for which

$$\boldsymbol{\Theta} = \boldsymbol{d}\bar{p}_1 \wedge \boldsymbol{d}\bar{q}^1 + \boldsymbol{d}\bar{p}_2 \wedge \boldsymbol{d}\bar{q}^2.$$

**Exercise 4.12. SYMMETRY OPERATIONS AS TENSORS**

We define the meaning of square and round brackets enclosing a set of indices as follows:

$$V_{(\alpha_1\ldots\alpha_p)} \equiv \frac{1}{p!}\Sigma V_{\alpha_{\pi 1}\ldots\alpha_{\pi p}}, \qquad V_{[\alpha_1\ldots\alpha_p]} \equiv \frac{1}{p!}\Sigma(-1)^{\pi} V_{\alpha_{\pi 1}\ldots\alpha_{\pi p}}.$$

**Box 4.6 BIRTH OF STOKES' THEOREM**

Central to the mathematical formulation of electromagnetism are the theorems of Gauss (taken up in Chapter 5) and Stokes. Both today appear together as one unity when expressed in the language of forms. In earlier times the unity was not evident. Everitt (1970) recalls the history of Stokes' theorem: "The Smith's Prize paper set by [G. C.] Stokes [Lucasian Professor of Mathematics] and taken by Maxwell in [February] 1854 . . .

> 5. Given the centre and two points of an ellipse, and the length of the major axis, find its direction by a geometrical construction.
>
> 6. Integrate the differential equation
>
> $$(a^2 - x^2)\,dy^2 + 2xydydx + (a^2 - y^2)\,dx^2 = 0.$$
>
> Has it a singular solution?
>
> 7. In a double system of curves of double curvature, a tangent is always drawn at the variable point $P$; shew that, as $P$ moves away from an arbitrary fixed point $Q$, it must begin to move along a generating line of an elliptic cone having $Q$ for vertex in order that consecutive tangents may ultimately intersect, but that the conditions of the problem may be impossible.
>
> 8. If $X$, $Y$, $Z$ be functions of the rectangular co-ordinates $x$, $y$, $z$, $dS$ an element of any limited surface, $l$, $m$, $n$ the cosines of the inclinations of the normal at $dS$ to the axes, $ds$ an element of the bounding line, shew that
>
> $$\iint\left\{l\left(\frac{dZ}{dy}-\frac{dY}{dz}\right)+m\left(\frac{dX}{dz}-\frac{dZ}{dx}\right)+n\left(\frac{dY}{dx}-\frac{dX}{dy}\right)\right\}dS$$
> $$=\int\left(X\frac{dx}{ds}+Y\frac{dy}{ds}+Z\frac{dz}{ds}\right)ds,$$
>
> the differential coefficients of $X$, $Y$, $Z$ being partial, and the single integral being taken all round the perimeter of the surface

marks the first appearance in print of the formula connecting line and surface integrals now known as Stokes' theorem. This was of great importance to Maxwell's development of electromagnetic theory. The earliest explicit proof of the theorem appears to be that given in a letter from Thomson to Stokes dated July 2, 1850." [Quoted in Campbell and Garnett (1882), pp. 186–187.]

Here the sum is taken over all permutations $\pi$ of the numbers $1, 2, \ldots, p$, and $(-1)^\pi$ is $+1$ or $-1$ depending on whether the permutation is even or odd. The quantity $V$ may have other indices, not shown here, besides the set of $p$ indices $\alpha_1, \alpha_2, \ldots, \alpha_p$, but only this set of indices is affected by the operations described here. The numbers $\pi_1, \pi_2, \ldots, \pi_p$ are the numbers $1, 2, \ldots, p$ rearranged according to the permutation $\pi$. (Cases $p = 2, 3$ were treated in exercise 3.12.) We therefore have machinery to convert any rank-$p$ tensor with components $V_{\alpha_1 \ldots \alpha_p}$ into a new tensor with components

$$[\boldsymbol{Alt}(\boldsymbol{V})]_{\mu_1 \ldots \mu_p} = V_{[\mu_1 \ldots \mu_p]}.$$

Since this machinery $\boldsymbol{Alt}$ is linear, it can be viewed as a tensor which, given suitable arguments $\boldsymbol{u}, \boldsymbol{v}, \ldots, \boldsymbol{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \ldots, \boldsymbol{\gamma}$ produces a number

$$u^\mu v^\nu \ldots w^\lambda \alpha_{[\mu} \beta_\nu \ldots \gamma_{\lambda]}.$$

(a) Show that the components of this tensor are

$$(\boldsymbol{Alt})_{\beta_1 \ldots \beta_p}{}^{\alpha_1 \ldots \alpha_p} = (p!)^{-1}\, \delta^{\alpha_1 \ldots \alpha_p}_{\beta_1 \ldots \beta_p}$$

(Note: indices of $\delta$ are almost never raised or lowered, so this notation leads to no confusion.)

where

$$\delta^{\alpha_1 \ldots \alpha_p}_{\beta_1 \ldots \beta_p} = \begin{cases} +1 \text{ if } (\alpha_1, \ldots, \alpha_p) \text{ is an even permutation of } (\beta_1, \ldots, \beta_p), \\ -1 \text{ if } (\alpha_1, \ldots, \alpha_p) \text{ is an odd permutation of } (\beta_1, \ldots, \beta_p), \\ \;\;0 \text{ if (i) any two of the } \alpha\text{'s are the same}, \\ \;\;0 \text{ if (ii) any two of the } \beta\text{'s are the same}, \\ \;\;0 \text{ if (iii) the } \alpha\text{'s and } \beta\text{'s are different sets of integers.} \end{cases}$$

Note that the demonstration, and therefore these component values, are correct in any frame.

(b) Show for any "alternating" (i.e., "*completely antisymmetric*") tensor $A_{\alpha_1 \ldots \alpha_p} = A_{[\alpha_1 \ldots \alpha_p]}$ that

$$\begin{aligned} \frac{1}{p!} A_{\alpha_1 \ldots \alpha_p}\, \delta^{\alpha_1 \ldots \alpha_p \beta_1 \ldots \beta_q}_{\gamma_1 \ldots \gamma_p \gamma_{p+1} \ldots \gamma_{p+q}} &= \sum_{\alpha_1 < \alpha_2 < \ldots < \alpha_p} A_{\alpha_1 \ldots \alpha_p}\, \delta^{\alpha_1 \ldots \alpha_p \beta_1 \ldots \beta_q}_{\gamma_1 \ldots\ldots\ldots \gamma_{p+q}} \\ &\equiv A_{|\alpha_1 \ldots \alpha_p|}\, \delta^{\alpha_1 \ldots \alpha_p \beta_1 \ldots \beta_q}_{\gamma_1 \ldots\ldots\ldots \gamma_{p+q}}. \end{aligned}$$

The final line here introduces the convention that a summation over indices enclosed between vertical bars includes only terms with those indices in increasing order. Show, consequently or similarly, that

$$\delta^{\alpha_1 \ldots \alpha_p \beta_1 \ldots \beta_q}_{\gamma_1 \ldots\ldots\ldots \gamma_{p+q}}\, \delta^{\mu_1 \ldots \mu_q}_{|\beta_1 \ldots \beta_q|} = \delta^{\alpha_1 \ldots \alpha_p \mu_1 \ldots \mu_q}_{\gamma_1 \ldots\ldots\ldots \gamma_{p+q}}.$$

(c) *Define* the exterior ("wedge") product of any two alternating tensors by

$$(\boldsymbol{\alpha} \wedge \boldsymbol{\beta})_{\lambda_1 \ldots \lambda_{p+q}} = \delta^{\mu_1 \ldots \mu_p\, \nu_1 \ldots \nu_q}_{\lambda_1 \ldots \lambda_p \lambda_{p+1} \ldots \lambda_{p+q}}\, \alpha_{|\mu_1 \ldots \mu_p|}\, \beta_{|\nu_1 \ldots \nu_q|};$$

and similarly

$$(\boldsymbol{U} \wedge \boldsymbol{V})^{\lambda_1 \ldots \lambda_{p+q}} = \delta^{\lambda_1 \ldots \lambda_p \lambda_{p+1} \ldots \lambda_{p+q}}_{\mu_1 \ldots \mu_p \nu_1 \ldots \nu_q}\, U^{|\mu_1 \ldots \mu_p|}\, V^{|\nu_1 \ldots \nu_q|}.$$

Show that this implies equation (3.45b). Establish the associative law for this product rule by showing that

$$
\begin{aligned}
&[(\boldsymbol{\alpha} \wedge \boldsymbol{\beta}) \wedge \boldsymbol{\gamma}]_{\sigma_1\ldots\sigma_{p+q+r}} \\
&= \delta^{\lambda_1\ldots\lambda_p\mu_1\ldots\mu_q\nu_1\ldots\nu_r}_{\sigma_1\ldots\sigma_{p+q+r}}\alpha_{|\lambda_1\ldots\lambda_p|}\beta_{|\mu_1\ldots\mu_q|}\gamma_{|\nu_1\ldots\nu_r|} \\
&= [\boldsymbol{\alpha} \wedge (\boldsymbol{\beta} \wedge \boldsymbol{\gamma})]_{\sigma_1\ldots\sigma_{p+q+r}};
\end{aligned}
$$

and show that this reduces to the 3-form version of Equation (3.45c) when $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are all 1-forms.

(d) Derive the following formula for the components of the exterior product of $p$ vectors

$$
\begin{aligned}
(\boldsymbol{u}_1 \wedge \boldsymbol{u}_2 \wedge \cdots \wedge \boldsymbol{u}_p)^{\alpha_1\ldots\alpha_p} &= \delta^{\alpha_1\ldots\alpha_p}_{\mu\ldots\nu}(u_1)^{\mu}\ldots(u_p)^{\nu} \\
&= p!\,u_1{}^{[\alpha_1}u_2{}^{\alpha_2}\ldots u_p{}^{\alpha_p]} \\
&= \delta^{\alpha_1\alpha_2\ldots\alpha_p}_{1\;2\;\ldots\;p}\det[(u_\mu)^{\lambda}].
\end{aligned}
$$

1

# CHAPTER 5

# STRESS-ENERGY TENSOR AND CONSERVATION LAWS

## §5.1. TRACK-1 OVERVIEW

"Geometry tells matter how to move, and matter tells geometry how to curve." However, it will do no good to look into curvature (Part III) and Einstein's law for the production of curvature by mass-energy (Part IV) until a tool can be found to determine how much mass-energy there is in a unit volume. That tool is the stress-energy tensor. It is the focus of attention in this chapter.

The essential features of the stress-energy tensor are summarized in Box 5.1 for the benefit of readers who want to rush on into gravitation physics as quickly as possible. Such readers can proceed directly from Box 5.1 into Chapter 6—though by doing so, they close the door on several later portions of track two, which lean heavily on material treated in this chapter.

1

2

The rest of this chapter is Track 2.
It depends on no preceding Track-2 material.
It is needed as preparation for Chapter 20 (conservation laws for mass and angular momentum).
It will be extremely helpful in all applications of gravitation theory (Chapters 18–40).

## §5.2. THREE-DIMENSIONAL VOLUMES AND DEFINITION OF THE STRESS-ENERGY TENSOR

Spacetime contains a flowing "river" of 4-momentum. Each particle carries its 4-momentum vector with itself along its world line. Many particles, on many world lines, viewed in a smeared-out manner (continuum approximation), produce a continuum flow—a river of 4-momentum. Electromagnetic fields, neutrino fields, meson fields: they too contribute to the river.

How can the flow of the river be quantified? By means of a linear machine: the stress-energy tensor $\boldsymbol{T}$.

Choose a small, three-dimensional parallelepiped in spacetime with vectors $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ for edges (Figure 5.1). Ask how much 4-momentum crosses that volume in

Box 5.1 CHAPTER 5 SUMMARIZED

## A. STRESS-ENERGY TENSOR AS A MACHINE

At each event in spacetime, there exists a stress-energy tensor. It is a machine that contains a knowledge of the energy density, momentum density, and stress as measured by any and all observers at that event. Included are energy, momentum, and stress associated with all forms of matter and all nongravitational fields.

The stress-energy tensor is a linear, symmetric machine with two slots for the insertion of two vectors: $\boldsymbol{T}(\ldots,\ldots)$. Its output, for given input, can be summarized as follows.

(1) Insert the 4-velocity $\boldsymbol{u}$ of an observer into one of the slots; leave the other slot empty. The output is

$$\boldsymbol{T}(\boldsymbol{u},\ldots) = \boldsymbol{T}(\ldots,\boldsymbol{u}) = -\begin{pmatrix}\text{density of 4-momentum,}\\ \text{“}d\boldsymbol{p}/dV\text{,” i.e., 4-momentum}\\ \text{per unit of three-dimensional volume,}\\ \text{as measured in observer's}\\ \text{Lorentz frame at event where}\\ \boldsymbol{T}\text{ is chosen}\end{pmatrix};$$

i.e., $T^{\alpha}{}_{\beta}u^{\beta} = T_{\beta}{}^{\alpha}u^{\beta} = -(dp^{\alpha}/dV)$ for observer with 4-velocity $u^{\alpha}$.

(2) Insert 4-velocity of observer into one slot; insert an arbitrary unit vector $\boldsymbol{n}$ into the other slot. The output is

$$\boldsymbol{T}(\boldsymbol{u},\boldsymbol{n}) = \boldsymbol{T}(\boldsymbol{n},\boldsymbol{u}) = -\begin{pmatrix}\text{component, “}\boldsymbol{n}\cdot d\boldsymbol{p}/dV\text{”, of}\\ \text{4-momentum density along the}\\ \boldsymbol{n}\text{ direction, as measured in}\\ \text{observer's Lorentz frame}\end{pmatrix};$$

i.e., $T_{\alpha\beta}u^{\alpha}n^{\beta} = T_{\alpha\beta}n^{\alpha}u^{\beta} = -n_{\mu}dp^{\mu}/dV$.

(3) Insert 4-velocity of observer into both slots. The output is the density of mass-energy that he measures in his Lorentz frame:

$$\boldsymbol{T}(\boldsymbol{u},\boldsymbol{u}) = \begin{pmatrix}\text{mass-energy per unit volume as measured}\\ \text{in frame with 4-velocity }\boldsymbol{u}\end{pmatrix}.$$

(4) Pick an observer and choose two spacelike basis vectors, $\boldsymbol{e}_j$ and $\boldsymbol{e}_k$, of his Lorentz frame. Insert $\boldsymbol{e}_j$ and $\boldsymbol{e}_k$ into the slots of $\boldsymbol{T}$. The output is the $j,k$ component of the stress as measured by that observer:

$$T_{jk} = \boldsymbol{T}(\boldsymbol{e}_j,\boldsymbol{e}_k) = T_{kj} = \boldsymbol{T}(\boldsymbol{e}_k,\boldsymbol{e}_j)$$
$$= \begin{pmatrix}j\text{-component of force acting}\\ \text{from side }x^k-\varepsilon\text{ to side }x^k+\varepsilon,\\ \text{across a unit surface area with}\\ \text{perpendicular direction }\boldsymbol{e}_k\end{pmatrix} = \begin{pmatrix}k\text{-component of force acting}\\ \text{from side }x^j-\varepsilon\text{ to side }x^j+\varepsilon,\\ \text{across a unit surface area with}\\ \text{perpendicular direction }\boldsymbol{e}_j\end{pmatrix}.$$

**Box 5.1** (*continued*)

## B. STRESS-ENERGY TENSOR FOR A PERFECT FLUID

One type of matter studied extensively later in this book is a *"perfect fluid."* A perfect fluid is a fluid or gas that (1) moves through spacetime with a 4-velocity $\boldsymbol{u}$ which may vary from event to event, and (2) exhibits a density of mass-energy $\rho$ and an isotropic pressure $p$ in the rest frame of each fluid element. Shear stresses, anisotropic pressures, and viscosity must be absent, or the fluid is not perfect. The stress-energy tensor for a perfect fluid at a given event can be constructed from the metric tensor, $\boldsymbol{g}$, the 4-velocity, $\boldsymbol{u}$, and the rest-frame density and pressure, $\rho$ and $p$:

$$\boldsymbol{T} = (\rho + p)\boldsymbol{u} \otimes \boldsymbol{u} + p\boldsymbol{g}, \qquad \text{or } T_{\alpha\beta} = (\rho + p)u_\alpha u_\beta + pg_{\alpha\beta}.$$

In the fluid's rest frame, the components of this stress-energy tensor have the expected form (insert into a slot of $\boldsymbol{T}$, as 4-velocity of observer, just the fluid's 4-velocity):

$$T^\alpha{}_\beta u^\beta = [(\rho + p)u^\alpha u_\beta + p\delta^\alpha{}_\beta]u^\beta = -(\rho + p)u^\alpha + pu^\alpha = -\rho u^\alpha;$$

i.e.,

$$T^0{}_\beta u^\beta = -\rho = -(\text{mass-energy density}) = -dp^0/dV,$$
$$T^j{}_\beta u^\beta = 0 = -(\text{momentum density}) = -dp^j/dV;$$

also

$$T_{jk} = \boldsymbol{T}(\boldsymbol{e}_j, \boldsymbol{e}_k) = p\delta_{jk} = \text{stress-tensor components}.$$

## C. CONSERVATION OF ENERGY-MOMENTUM

In electrodynamics the conservation of charge can be expressed by the differential equation

$$\partial(\text{charge density})/\partial t + \boldsymbol{\nabla} \cdot (\text{current density}) = 0;$$

i.e., $J^0{}_{,0} + \boldsymbol{\nabla} \cdot \boldsymbol{J} = 0$; i.e. $J^\alpha{}_{,\alpha} = 0$; i.e., $\boldsymbol{\nabla} \cdot \boldsymbol{J} = 0$. Similarly, conservation of energy-momentum can be expressed by the fundamental geometric law

$$\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0.$$

(Because $\boldsymbol{T}$ is symmetric, it does not matter on which slot the divergence is taken.) This law plays an important role in gravitation theory.

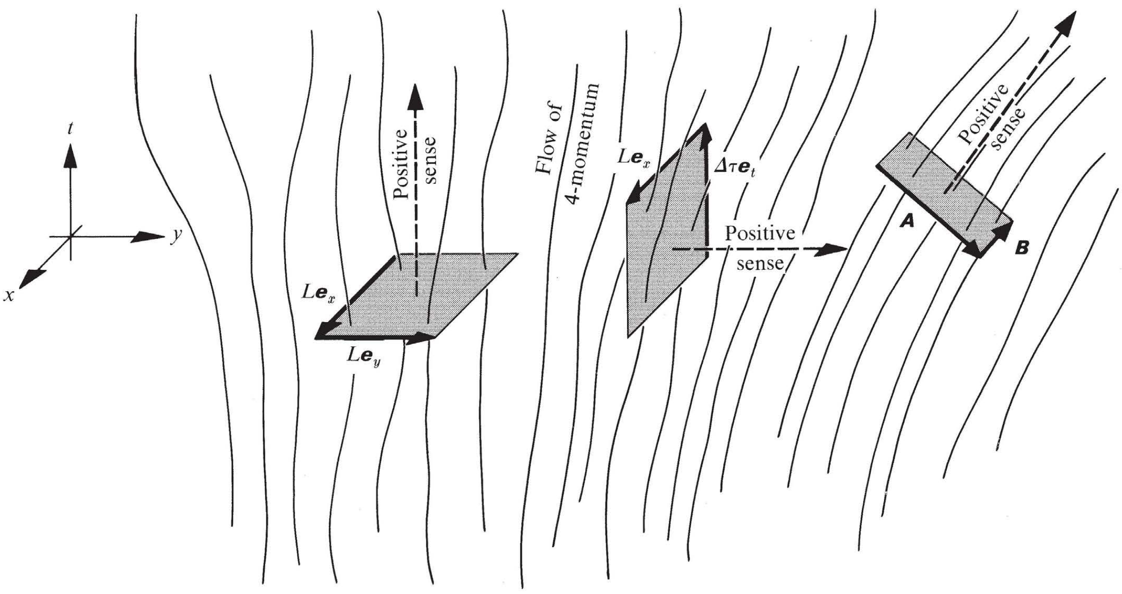



**Figure 5.1.**
The "river" of 4-momentum flowing through spacetime, and three different 3-volumes across which it flows. (One dimension is suppressed from the picture; so the 3-volumes look like 2-volumes.) The first 3-volume is the interior of a cubical soap box momentarily at rest in the depicted Lorentz frame. Its edges are $L\boldsymbol{e}_x$, $L\boldsymbol{e}_y$, $L\boldsymbol{e}_z$; and its volume 1-form, with "positive" sense toward future ("standard orientation"), is $\boldsymbol{\Sigma} = L^3\,\boldsymbol{d}t = -V\boldsymbol{u}(V = L^3 =$ volume as measured in rest frame; $\boldsymbol{u} = -\boldsymbol{d}t =$ 4-velocity of box). The second 3-volume is the "world sheet" swept out in time $\Delta\tau$ by the top of a second cubical box. The box top's edges are $L\boldsymbol{e}_x$ and $L\boldsymbol{e}_z$; and its volume 1-form, with "positive" sense away from the box's interior, in direction of increasing $y$, is $\boldsymbol{\Sigma} = L^2\,\Delta\tau\,\boldsymbol{d}y = \mathcal{A}\,\Delta\tau\boldsymbol{\sigma}$ ($\mathcal{A} = L^2 =$ area of box top; $\boldsymbol{\sigma} = \boldsymbol{d}y =$ unit 1-form containing world tube). The third 3-volume is an arbitrary one, with edges $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and volume 1-form $\Sigma_\mu = \epsilon_{\mu\alpha\beta\gamma}A^\alpha B^\beta C^\gamma$.

its positive sense (i.e., from its "negative side" toward its "positive side"). To calculate the answer: (1) Construct the *"volume 1-form"*

Mathematical representation of 3-volumes

$$\Sigma_\mu = +\epsilon_{\mu\alpha\beta\gamma}A^\alpha B^\beta C^\gamma; \tag{5.1}$$

the parallelepiped lies in one of the 1-form surfaces, and the positive sense across the parallelepiped is defined to be the positive sense of the 1-form $\boldsymbol{\Sigma}$. (2) Insert this volume 1-form into the second slot of the stress-energy tensor $\boldsymbol{T}$. The result is

Momentum crossing a 3-volume calculated, using stress-energy tensor

$$\boldsymbol{T}(\underbrace{\ldots}_{\text{empty slot}}, \boldsymbol{\Sigma}) = \boldsymbol{p} = \begin{pmatrix}\text{momentum crossing from} \\ \text{negative side toward positive side}\end{pmatrix}. \tag{5.2}$$

(3) To get the projection of the 4-momentum along a vector $\boldsymbol{w}$ or 1-form $\boldsymbol{\alpha}$, insert the volume 1-form $\boldsymbol{\Sigma}$ into the second slot and $\boldsymbol{w}$ or $\boldsymbol{\alpha}$ into the first:

$$\boldsymbol{T}(\boldsymbol{w}, \boldsymbol{\Sigma}) = \boldsymbol{w}\cdot\boldsymbol{p}, \qquad \boldsymbol{T}(\boldsymbol{\alpha}, \boldsymbol{\Sigma}) = \langle\boldsymbol{\alpha}, \boldsymbol{p}\rangle. \tag{5.3}$$

This defines the stress-energy tensor.

The key features of 3-volumes and the stress-energy tensor are encapsulated by the above three-step procedure. But encapsulation is not sufficient; deep understanding is also required. To gain it, one must study special cases, both of 3-volumes and of the operation of the stress-energy machinery.

### A Special Case

Interior of a soap box:

A soap box moves through spacetime. A man at an event $\mathscr{P}_0$ on the box's world line peers inside it, and examines all the soap, air, and electromagnetic fields it contains. He adds up all their 4-momenta to get a grand total $\boldsymbol{p}_{\text{box at } \mathscr{P}_0}$. How much is this grand total? One can calculate it by noting that the 4-momentum inside the box at $\mathscr{P}_0$ is precisely the 4-momentum crossing the box from past toward future there (Figure 5.1). Hence, the 4-momentum the man measures is

$$\boldsymbol{p}_{\text{box at } \mathscr{P}_0} = \boldsymbol{T}(\ldots, \boldsymbol{\Sigma}), \tag{5.4}$$

Its volume 1-form

where $\boldsymbol{\Sigma}$ is the box's volume 1-form at $\mathscr{P}_0$. But for such a soap box, $\boldsymbol{\Sigma}$ has a magnitude equal to the box's volume $V$ as measured by a man in its momentary rest frame, and the box itself lies in one of the hyperplanes of $\boldsymbol{\Sigma}$; equivalently,

$$\boldsymbol{\Sigma} = -V\boldsymbol{u}, \tag{5.5}$$

where $\boldsymbol{u}$ is the soap box's 4-velocity at $\mathscr{P}_0$ (minus sign because $\boldsymbol{u}$, regarded as a 1-form, has positive sense toward the past, $u_0 < 0$); see Box 5.2. Hence, the total 4-momentum inside the box is

Its 4-momentum content

$$\boldsymbol{p}_{\text{box at } \mathscr{P}_0} = \boldsymbol{T}(\ldots, -V\boldsymbol{u}) = -V\boldsymbol{T}(\ldots, \boldsymbol{u}), \tag{5.6}$$

or, in component notation,

$$(p^\alpha)_{\text{box at } \mathscr{P}_0} = -VT^{\alpha\beta}u_\beta. \tag{5.6$'$}$$

The energy in the box, as measured in its rest frame, is minus the projection of the 4-momentum on the box's 4-velocity:

$$E = -\boldsymbol{u} \cdot \boldsymbol{p}_{\text{box at } \mathscr{P}_0} = +VT^{\alpha\beta}u_\alpha u_\beta = V\boldsymbol{T}(\boldsymbol{u}, \boldsymbol{u});$$

so

Its energy density

$$\begin{pmatrix}\text{energy density as}\\ \text{measured in box's}\\ \text{rest frame}\end{pmatrix} = \frac{E}{V} = \boldsymbol{T}(\boldsymbol{u}, \boldsymbol{u}). \tag{5.7}$$

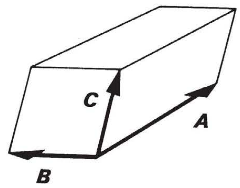
### Another Special Case

A man riding with the same soap box opens its top and pours out some soap. In a very small interval of time $\Delta\tau$, how much total 4-momentum flows out of the box?

## Box 5.2 THREE-DIMENSIONAL VOLUMES

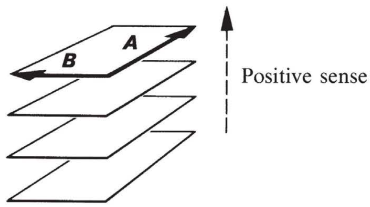
### A. General Parallelepiped

1. *Edges* of parallelepiped are three vectors $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$. One must order the edges; e.g., "$\boldsymbol{A}$ is followed by $\boldsymbol{B}$ is followed by $\boldsymbol{C}$."



(One dimension, that orthogonal to the parallelepiped, is suppressed here.)

2. *Volume trivector* is defined to be $\boldsymbol{A} \wedge \boldsymbol{B} \wedge \boldsymbol{C}$. It enters into the sophisticated theory of volumes (Chapter 4), but is not used much in the elementary theory.
3. *Volume 1-form* is defined by $\Sigma_\mu = \epsilon_{\mu\alpha\beta\gamma}A^\alpha B^\beta C^\gamma$. ($\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ must appear here in standard order as chosen in step 1.) Note that the vector "corresponding" to $\boldsymbol{\Sigma}$ and the volume trivector are related by $\boldsymbol{\Sigma} = -{}^*(\boldsymbol{A} \wedge \boldsymbol{B} \wedge \boldsymbol{C})$.
4. *Orientation* of the volume is defined to agree with the orientation of its 1-form $\boldsymbol{\Sigma}$. More specifically: the edges $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ lie in a hyperplane of $\boldsymbol{\Sigma}(\langle\boldsymbol{\Sigma},\boldsymbol{A}\rangle = \langle\boldsymbol{\Sigma},\boldsymbol{B}\rangle = \langle\boldsymbol{\Sigma},\boldsymbol{C}\rangle = 0$; no "bongs of bell"). Thus, *the volume itself is one of* $\boldsymbol{\Sigma}$*'s hyperplanes!* The positive sense moving away from the volume is defined to be the positive sense of $\boldsymbol{\Sigma}$. *Note:* reversing the order of $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ reverses the positive sense!



(One dimension, that along which $\boldsymbol{C}$ extends, is suppressed here.)

5. *The "standard orientation"* for a spacelike 3-volume has the positive sense of the 1-form $\boldsymbol{\Sigma}$ toward the future, corresponding to $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ forming a righthanded triad of vectors.

### B. 3-Volumes of Arbitrary Shape

Can be analyzed by being broken up into union of parallelepipeds.

### C. Interior of a Soap Box (Example)

1. *Analysis in soap box's rest frame.* Pick an event on the box's world line. The box's three edges there are three specific vectors $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$. In the box's rest frame they are purely spatial: $A^0 = B^0 = C^0 = 0$. Hence, the volume 1-form has components $\Sigma_j = 0$ and

**Box 5.2 (*continued*)**

$$\Sigma_0 = \epsilon_{0ijk} A^i B^j C^k = \det \begin{Vmatrix} A^1 & A^2 & A^3 \\ B^1 & B^2 & B^3 \\ C^1 & C^2 & C^3 \end{Vmatrix}$$

$= \boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C})$, in the standard notation of 3-dimensional vector analysis;
$= +V$ ($V =$ volume of box) if $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ are righthand ordered (positive sense of $\boldsymbol{\Sigma}$ toward future; standard orientation);
$= -V$ ($V =$ volume of box) if $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C})$ are lefthand ordered (positive sense of $\boldsymbol{\Sigma}$ toward past).

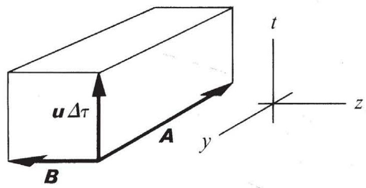
2. This result reexpressed in geometric language: Let $\boldsymbol{u}$ be the box's 4-velocity and $V$ be its volume, as measured in its rest frame. Then either

   $\boldsymbol{\Sigma} = -V\boldsymbol{u}$, in which case the "positive side" of the box's 3-surface is the future side, and its edges are ordered in a righthanded manner—the standard orientation;

   or else

   $\boldsymbol{\Sigma} = +V\boldsymbol{u}$, in which case the "positive side" is the past side, and the box's edges are ordered in a lefthanded manner.

## D. 3-Volume Swept Out in Time $\Delta\tau$ by Two-Dimensional Top of a Soap Box (Example)




1. *Analysis in box's rest frame:* Pick an event on box's world line. There the two edges of the box top are vectors $\boldsymbol{A}$ and $\boldsymbol{B}$. In the box's rest frame, orient the space axes so that $\boldsymbol{A}$ and $\boldsymbol{B}$ lie in the $y,z$-plane. During the lapse of a proper time $\Delta\tau$, the box top sweeps out a 3-volume whose third edge is $\boldsymbol{u}\,\Delta\tau$ ($\boldsymbol{u} =$ 4-velocity of box). In the box's rest-frame, with ordering "$\boldsymbol{A}$ followed by $\boldsymbol{B}$ followed by $\boldsymbol{u}\,\Delta\tau$," the volume 1-form has components

   $\Sigma_0 = \Sigma_2 = \Sigma_3 = 0$, and
   $\Sigma_1 = \epsilon_{1jk0} A^j B^k \,\Delta\tau u^0 = -\epsilon_{01jk} A^j B^k \,\Delta\tau$
   $= -\mathcal{A}\,\Delta\tau$ ($\mathcal{A} =$ area of box top) if $(\boldsymbol{e}_x, \boldsymbol{A}, \boldsymbol{B})$ are righthand ordered
   $= +\mathcal{A}\,\Delta\tau$ ($\mathcal{A} =$ area of box top) if $(\boldsymbol{e}_x, \boldsymbol{A}, \boldsymbol{B})$ are lefthand ordered.

   (Note: No standard orientation can be defined in this case, because $\boldsymbol{\Sigma}$ can be carried continuously into $-\boldsymbol{\Sigma}$ by purely spatial rotations.)

2. *This result reexpressed in geometric language:* Let $\mathcal{A}$ be the area of the box top as measured in its rest frame; and let $\boldsymbol{\sigma}$ be a unit 1-form, one of whose surfaces contains the box top and its 4-velocity (i.e., contains the box top's "world sheet"). Orient the positive sense of $\boldsymbol{\sigma}$ with the (arbitrarily chosen) positive sense of the box-top 3-volume. Then

$$\boldsymbol{\Sigma} = \mathcal{A}\,\Delta\tau\,\boldsymbol{\sigma}.$$

To answer this question, consider the three-dimensional volume swept out during $\Delta\tau$ by the box's opened two-dimensional top ("world sheet of top"). The 4-momentum asked for is the 4-momentum that crosses this world sheet in the positive sense (see Figure 5.1); hence, it is

The top of a soap box:

$$\boldsymbol{p}_{\text{flows out}} = \boldsymbol{T}(\ldots, \boldsymbol{\Sigma}), \tag{5.8}$$

where $\boldsymbol{\Sigma}$ is the world sheet's volume 1-form. Let $\mathcal{Q}$ be the area of the box top, and $\boldsymbol{\sigma}$ be the outward-oriented unit 1-form, whose surfaces contain the world sheet (i.e., contain the box top and its momentary 4-velocity vector). Then

Its volume 1-form

$$\boldsymbol{\Sigma} = \mathcal{Q}\, \Delta\tau\, \boldsymbol{\sigma} \tag{5.9}$$

(see Box 5.2); so the 4-momentum that flows out during $\Delta\tau$ is

Its 4-momentum that flows across

$$\boldsymbol{p}_{\text{flows out}} = \mathcal{Q}\, \Delta\tau\, \boldsymbol{T}(\ldots, \boldsymbol{\sigma}). \tag{5.10}$$

## §5.3. COMPONENTS OF STRESS-ENERGY TENSOR

Like all other tensors, the stress-energy tensor is a machine whose definition and significance transcend coordinate systems and reference frames. But any one observer, locked as he is into some one Lorentz frame, pays more attention to the components of $\boldsymbol{T}$ than to $\boldsymbol{T}$ itself. To each component he ascribes a specific physical significance. Of greatest interest, perhaps, is the "time-time" component. It is the total density of mass-energy as measured in the observer's Lorentz frame:

Physical interpretation of stress-energy tensor's components:

$$T_{00} = -T_0{}^0 = T^{00} = \boldsymbol{T}(\boldsymbol{e}_0, \boldsymbol{e}_0) = \text{density of mass-energy} \tag{5.11}$$

$T^{00}$: energy density

(cf. equation 5.7, with the observer's 4-velocity $\boldsymbol{u}$ replaced by the basis vector $\boldsymbol{e}_0 = \boldsymbol{u}$).

The "spacetime" components $T^{j0}$ can be interpreted by considering the interior of a soap box at rest in the observer's frame. If its volume is $V$, then its volume 1-form is $\boldsymbol{\Sigma} = -V\boldsymbol{u} = +V\,\boldsymbol{d}t$; and the $\mu$-component of 4-momentum inside it is

$$p^\mu = \langle \boldsymbol{d}x^\mu, \boldsymbol{p}\rangle = \boldsymbol{T}(\boldsymbol{d}x^\mu, \boldsymbol{\Sigma}) = V\boldsymbol{T}(\boldsymbol{d}x^\mu, \boldsymbol{d}t) = VT^{\mu 0}.$$

Thus, the 4-momentum per unit volume is

$$p^\mu/V = T^{\mu 0}, \tag{5.12a}$$

or, equivalently:

$$\begin{aligned} T^{00} = {} & \text{density of mass-energy} \\ & (\text{units: g/cm}^3\text{, or erg/cm}^3\text{, or cm}^{-2}); \end{aligned} \tag{5.13a}$$

$$\begin{aligned} T^{j0} = {} & \text{density of j-component of momentum} \\ & (\text{units: g (cm/sec) cm}^{-3}\text{, or cm}^{-2}). \end{aligned} \tag{5.13b}$$

$T^{j0}$: momentum density

The components $T^{\mu k}$ can be interpreted using a two-dimensional surface of area $\mathcal{Q}$, at rest in the observer's frame with positive normal pointing in the $k$-direction.

During the lapse of time $\Delta t$, this 2-surface sweeps out a 3-volume with volume 1-form $\boldsymbol{\Sigma} = \mathcal{Q}\, \Delta t\, \boldsymbol{d}x^k$ (see Box 5.2). The $\mu$-component of 4-momentum that crosses the 2-surface in time $\Delta t$ is

$$p^\mu = \boldsymbol{T}(\boldsymbol{d}x^\mu, \boldsymbol{\Sigma}) = \mathcal{Q}\, \Delta t\, \boldsymbol{T}(\boldsymbol{d}x^\mu, \boldsymbol{d}x^k) = \mathcal{Q}\, \Delta t\, T^{\mu k}.$$

Thus, the flux of 4-momentum (4-momentum crossing a unit surface oriented perpendicular to $\boldsymbol{e}_k$, in unit time) is

$$(p^\mu/\mathcal{Q}\, \Delta t)_{\text{crossing surface} \perp \text{to } \boldsymbol{e}_k} = T^{\mu k}, \tag{5.12b}$$

or, equivalently:

$T^{0k}$: energy flux

$T^{jk}$: stress

$$T^{0k} = k\text{-component of energy flux (units: erg/cm}^2\text{ sec, or cm}^{-2}); \tag{5.13c}$$

$$\begin{aligned} T^{jk} &= j, k \text{ component of “stress”} \\ &\equiv k\text{-component of flux of } j\text{-component of momentum} \\ &\equiv j\text{-component of force produced by fields and matter at } x^k - \epsilon \text{ acting} \\ &\quad \text{on fields and matter at } x^k + \epsilon \text{ across a unit surface, the perpendicular} \\ &\quad \text{to which is } \boldsymbol{e}_k \\ &\quad \text{(units: dynes/cm}^2\text{, or cm}^{-2}). \end{aligned} \tag{5.13d}$$

(Recall that “momentum transfer per second” is the same as “force.”)

The stress-energy tensor is necessarily symmetric, $T^{\alpha\beta} = T^{\beta\alpha}$; but the proof of this will be delayed until several illustrations have been examined.

## §5.4. STRESS-ENERGY TENSOR FOR A SWARM OF PARTICLES

Consider a swarm of particles. Choose some event $\mathcal{P}$ inside the swarm. Divide the particles near $\mathcal{P}$ into categories, $A = 1, 2, \ldots,$ in such a way that all particles in the same category have the same properties:

$$\begin{array}{ll} m_{(A)}, & \text{rest mass;} \\ \boldsymbol{u}_{(A)}, & \text{4-velocity;} \\ \boldsymbol{p}_{(A)} = m_{(A)}\boldsymbol{u}_{(A)}, & \text{4-momentum.} \end{array}$$

Number-flux vector for swarm of particles defined

Let $N_{(A)}$ be the number of category-$A$ particles per unit volume, as measured in the particles' own rest frame. Then the *“number-flux vector”* $\boldsymbol{S}_{(A)}$, defined by

$$\boldsymbol{S}_{(A)} \equiv N_{(A)}\boldsymbol{u}_{(A)}, \tag{5.14}$$

has components with simple physical meanings. In a frame where category-$A$ particles have ordinary velocity $\boldsymbol{v}_{(A)}$, the meanings are:

$$S^0_{(A)} = N_{(A)}u^0_{(A)} = \underbrace{N_{(A)}}_{\substack{\text{Number density in}\\ \text{particles' rest frame}}} \underbrace{[1 - v_{(A)}{}^2]^{-1/2}}_{\substack{\text{Lorentz contraction}\\ \text{factor for volume}}} = \text{number density}; \tag{5.15a}$$

$$\boldsymbol{S}_{(A)} = N_{(A)}\boldsymbol{u}_{(A)} = S^0_{(A)}\boldsymbol{v}_{(A)} = \text{flux of particles.} \tag{5.15b}$$

Consequently, the 4-momentum density has components

$$\begin{aligned} T^{\mu 0}_{(A)} &= p^{\mu}_{(A)}S^0_{(A)} = m_{(A)}u^{\mu}_{(A)}N_{(A)}u^0_{(A)} \\ &= m_{(A)}N_{(A)}u^{\mu}_{(A)}u^0_{(A)}; \end{aligned}$$

and the flux of $\mu$-component of momentum across a surface with perpendicular direction $\boldsymbol{e}_j$ is

$$\begin{aligned} T^{\mu j}_{(A)} &= p^{\mu}_{(A)}S^j_{(A)} = m_{(A)}u^{\mu}_{(A)}N_{(A)}u^j_{(A)} \\ &= m_{(A)}N_{(A)}u^{\mu}_{(A)}u^j_{(A)}. \end{aligned}$$

These equations are precisely the $\mu, 0$ and $\mu, j$ components of the geometric, frame-independent equation

Stress-energy tensor for swarm of particles

$$\boldsymbol{T}_{(A)} = m_{(A)}N_{(A)}\boldsymbol{u}_{(A)} \otimes \boldsymbol{u}_{(A)} = \boldsymbol{p}_{(A)} \otimes \boldsymbol{S}_{(A)}. \tag{5.16}$$

The total number-flux vector and stress-energy tensor for all particles in the swarm near $\mathscr{P}$ are obtained by summing over all categories:

$$\boldsymbol{S} = \sum_A N_{(A)}\boldsymbol{u}_{(A)}; \tag{5.17}$$

$$\boldsymbol{T} = \sum_A m_{(A)}N_{(A)}\boldsymbol{u}_{(A)} \otimes \boldsymbol{u}_{(A)} = \sum_A \boldsymbol{p}_{(A)} \otimes \boldsymbol{S}_{(A)}. \tag{5.18}$$

## §5.5. STRESS-ENERGY TENSOR FOR A PERFECT FLUID

Ideal gas defined

There is no simpler example of a fluid than a gas of noninteracting particles ("ideal gas") in which the velocities of the particles are distributed isotropically. In the Lorentz frame where isotropy obtains, symmetry argues equality of the diagonal space-space components of the stress-energy tensor,

$$T_{xx} = T_{yy} = T_{zz} = \sum_A \frac{m_{(A)}v_{x(A)}}{(1 - v_{(A)}{}^2)^{1/2}} \frac{N_{(A)}v_{x(A)}}{(1 - v_{(A)}{}^2)^{1/2}}, \tag{5.19}$$

and vanishing of all the off-diagonal components. Moreover, (5.19) represents a product: the number of particles per unit volume, multiplied by velocity in the $x$-direction (giving flux in the $x$-direction) and by momentum in the $x$-direction,

2



giving the standard kinetic-theory expression for the pressure, $p$. Therefore, the stress-energy tensor takes the form

$$T_{\alpha\beta} = \begin{Vmatrix} \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{Vmatrix} \tag{5.20}$$

in this special Lorentz frame—the "rest frame" of the gas. Here the quantity $\rho$ has nothing directly to do with the rest-masses of the constituent particles. It measures the density of rest-plus-kinetic energy of these particles.

Rewrite (5.20) in terms of the 4-velocity $u^\alpha = (1, 0, 0, 0)$ of the fluid in the gas's rest frame, and find

$$\begin{aligned} T_{\alpha\beta} &= \begin{Vmatrix} \rho & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{Vmatrix} + \begin{Vmatrix} 0 & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{Vmatrix} \\ &= \rho u_\alpha u_\beta + p(\eta_{\alpha\beta} + u_\alpha u_\beta), \end{aligned}$$

or, in frame-independent, geometric language

Stress-energy tensor for ideal gas or perfect fluid

$$T = p\boldsymbol{g} + (\rho + p)\boldsymbol{u} \otimes \boldsymbol{u}. \tag{5.21}$$

Perfect fluid defined

Expression (5.21) has general application. It is exact for the "ideal gas" just considered. It is also exact for any fluid that is "perfect" in the sense that it is free of such transport processes as heat conduction and viscosity, and therefore (in the rest frame) free of shear stress (diagonal stress tensor; diagonal components identical, because if they were not identical, a rotation of the frame of reference would reveal presence of shear stress). However, for a general perfect fluid, density $\rho$ of mass-energy as measured in the fluid's rest frame includes not only rest mass plus kinetic energy of particles, but also energy of compression, energy of nuclear binding, and all other sources of mass-energy [total density of mass-energy as it might be determined by an idealized experiment, such as that depicted in Figure 1.12, with the sample mass at the center of the sphere, and the test particle executing oscillations of small amplitude about that location, with $\omega^2 = (4\pi/3)\rho$].

## §5.6. ELECTROMAGNETIC STRESS-ENERGY

Faraday, with his picture of tensions along lines of force and pressures at right angles to them (Figure 5.2), won insight into new features of electromagnetism. In addition to the tension $\boldsymbol{E}^2/8\pi$ (or $\boldsymbol{B}^2/8\pi$) along lines of force, and an equal pressure at right angles, one has the Poynting flux $(\boldsymbol{E} \times \boldsymbol{B})/4\pi$ and the Maxwell expression for the

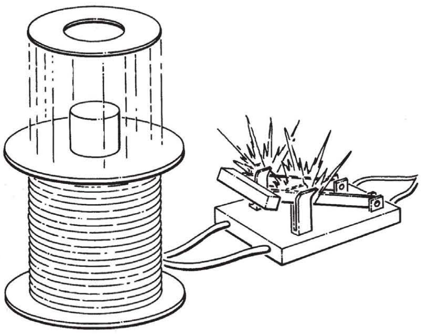


**Figure 5.2.**
Faraday stresses at work. When the electromagnet is connected to an alternating current, the aluminum ring flies into the air.

energy density, $(\boldsymbol{E}^2 + \boldsymbol{B}^2)/8\pi$. All these quantities find their places in the Maxwell stress-energy tensor, defined by

Stress-energy tensor for electromagnetic field

$$4\pi T^{\mu\nu} = F^{\mu\alpha}F^{\nu}{}_{\alpha} - \frac{1}{4}\eta^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}. \tag{5.22}$$

---

**EXERCISE**

**Exercise 5.1.**
Show that expression (5.22), evaluated in a Lorentz coordinate frame, gives

$$\begin{aligned} T^{00} &= (\boldsymbol{E}^2 + \boldsymbol{B}^2)/8\pi, \qquad T^{0j} = T^{j0} = (\boldsymbol{E} \times \boldsymbol{B})^j/4\pi, \\ T^{jk} &= \frac{1}{4\pi}\left[-(E^jE^k + B^jB^k) + \frac{1}{2}(\boldsymbol{E}^2 + \boldsymbol{B}^2)\,\delta^{jk}\right]. \end{aligned} \tag{5.23}$$

Show that the stress tensor does describe a tension $(\boldsymbol{E}^2 + \boldsymbol{B}^2)/8\pi$ along the field lines and a pressure $(\boldsymbol{E}^2 + \boldsymbol{B}^2)/8\pi$ perpendicular to the field lines, as stated in the text.

---

## §5.7. SYMMETRY OF THE STRESS-ENERGY TENSOR

All the stress-energy tensors explored above were symmetric. That they could not have been otherwise one sees as follows.

Proof that stress-energy tensor is symmetric

Calculate in a specific Lorentz frame. Consider first the momentum density (components $T^{j0}$) and the energy flux (components $T^{0j}$). They must be equal because energy = mass ("$E = Mc^2 = M$"):

$$\begin{aligned} T^{0j} &= (\text{energy flux}) \\ &= (\text{energy density}) \times (\text{mean velocity of energy flow})^j \\ &= (\text{mass density}) \times (\text{mean velocity of mass flow})^j \\ &= (\text{momentum density}) = T^{j0}. \end{aligned} \tag{5.24}$$

Only the stress tensor $T^{jk}$ remains. For it, one uses the same standard argument as in Newtonian theory. Consider a very small cube, of side $L$, mass-energy $T^{00}L^3$, and moment of inertia $\sim T^{00}L^5$. With the space coordinates centered at the cube, the expression for the $z$-component of torque exerted on the cube by its surroundings is

$$\begin{aligned}\tau^z = \; & \underbrace{(-T^{yx}L^2)}_{\begin{pmatrix}y\text{-component}\\ \text{of force on}\\ +x \text{ face}\end{pmatrix}} \underbrace{(L/2)}_{\begin{pmatrix}\text{lever}\\ \text{arm to}\\ +x \text{ face}\end{pmatrix}} + \underbrace{(T^{yx}L^2)}_{\begin{pmatrix}y\text{-component}\\ \text{of force on}\\ -x \text{ face}\end{pmatrix}} \underbrace{(-L/2)}_{\begin{pmatrix}\text{lever}\\ \text{arm to}\\ -x \text{ face}\end{pmatrix}} \\ & - \underbrace{(-T^{xy}L^2)}_{\begin{pmatrix}x\text{-component}\\ \text{of force on}\\ +y \text{ face}\end{pmatrix}} \underbrace{(L/2)}_{\begin{pmatrix}\text{lever}\\ \text{arm to}\\ +y \text{ face}\end{pmatrix}} - \underbrace{(T^{xy}L^2)}_{\begin{pmatrix}x\text{-component}\\ \text{of force on}\\ -y \text{ face}\end{pmatrix}} \underbrace{(-L/2)}_{\begin{pmatrix}\text{lever}\\ \text{arm to}\\ -y \text{ face}\end{pmatrix}} \\ = \; & (T^{xy} - T^{yx})L^3.\end{aligned}$$

Since the torque decreases only as $L^3$ with decreasing $L$, while the moment of inertia decreases as $L^5$, the torque will set an arbitrarily small cube into arbitrarily great angular acceleration—which is absurd. To avoid this, the stresses distribute themselves so the torque vanishes:

$$T^{yx} = T^{xy}.$$

Put differently, if the stresses were not so distributed, the resultant infinite angular accelerations would instantaneously redistribute them back to equilibrium. This condition of torque balance, repeated for all other pairs of directions, is equivalent to symmetry of the stresses:

$$T^{jk} = T^{kj}. \tag{5.25}$$

## §5.8. CONSERVATION OF 4-MOMENTUM: INTEGRAL FORMULATION

Energy-momentum conservation has been a cornerstone of physics for more than a century. Nowhere does its essence shine forth so clearly as in Einstein's geometric formulation of it (Figure 5.3,a). There one examines a four-dimensional region of spacetime $\mathcal{V}$ bounded by a closed, three-dimensional surface $\partial\mathcal{V}$. As particles and fields flow into $\mathcal{V}$ and later out, they carry 4-momentum. Inside $\mathcal{V}$ the particles collide, break up, radiate; radiation propagates, jiggles particles, produces pairs. But at each stage in this complex maze of physical processes, total energy-momentum is conserved. The energy-momentum lost by particles goes into fields; the energy-momentum lost by fields goes into particles. So finally, when the "river" of 4-momentum exits from $\mathcal{V}$, it carries out precisely the same energy-momentum as it carried in.

Integral conservation law for 4-momentum:
$\oint_{\partial\mathcal{V}} \boldsymbol{T}\cdot d^3\boldsymbol{\Sigma} = 0$

Restate this equality by asking for the total flux of 4-momentum *outward* across $\partial\mathcal{V}$. Count inflowing 4-momentum negatively. Then "inflow equals outflow" means "total outflow vanishes":

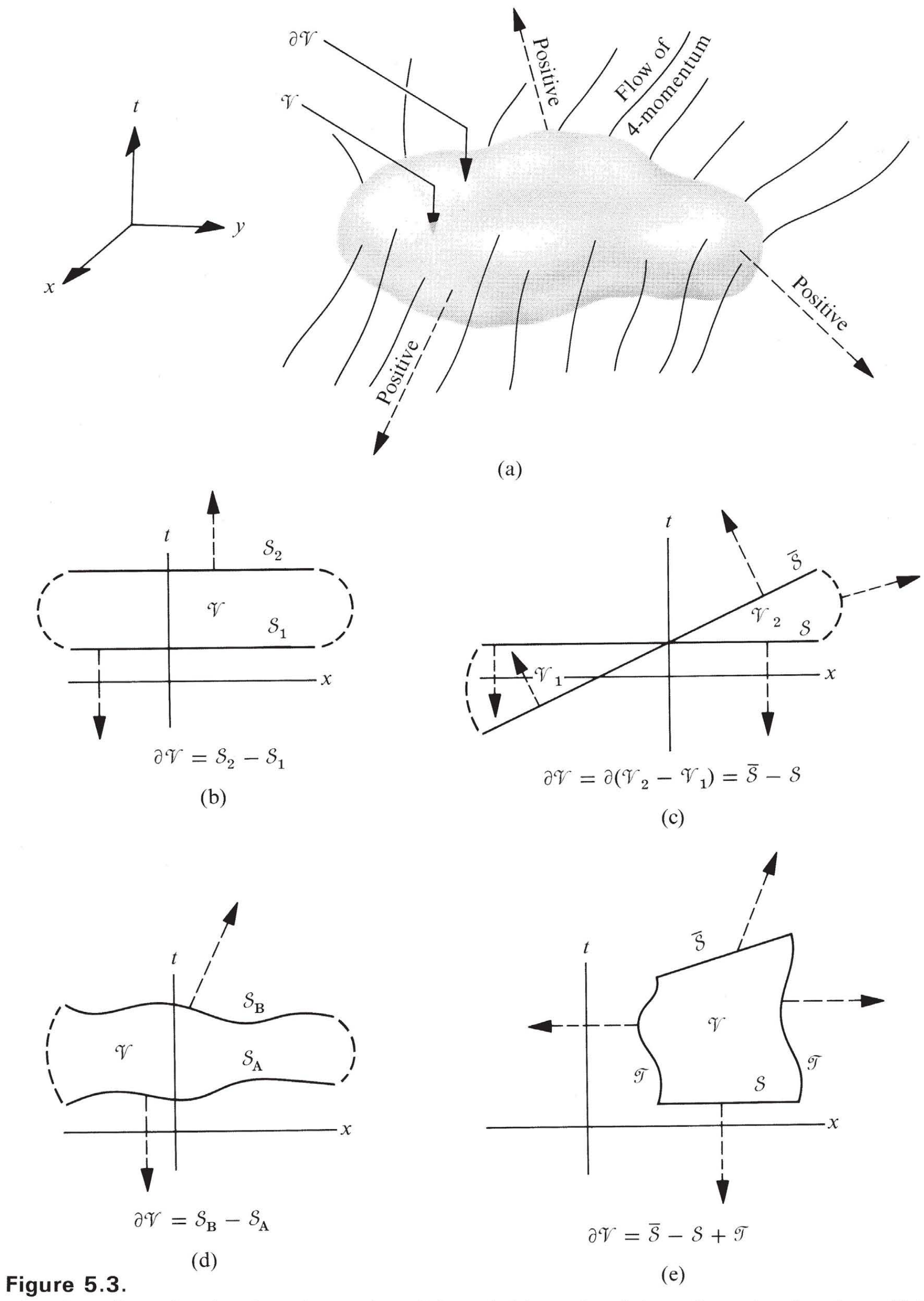



**Figure 5.3.**
(a) A four-dimensional region of spacetime $\mathcal{V}$ bounded by a closed three-dimensional surface $\partial\mathcal{V}$. The positive sense of $\partial\mathcal{V}$ is defined to be everywhere outward (away from $\mathcal{V}$). Conservation of energy-momentum demands that every bit of 4-momentum which flows into $\mathcal{V}$ through $\partial\mathcal{V}$ must somewhere flow back out; none can get lost inside; the interior contains no "sinks." Equivalently, the total flux of 4-momentum across $\partial\mathcal{V}$ in the positive (outward) sense must be zero:

$$\oint_{\partial\mathcal{V}} T^{\mu\alpha}\, d^3\Sigma_\alpha = 0.$$

Figures (b), (c), (d), and (e) depict examples to which the text applies this law of conservation of 4-momentum. All symbols $\mathcal{V}$ (or $\mathcal{S}$) in these figures mean spacetime volumes (or spacelike 3-volumes) with standard orientations. The dotted arrows indicate the positive sense of the closed surface $\partial\mathcal{V}$ used in the text's discussion of 4-momentum conservation. How $\partial\mathcal{V}$ is constructed from the surfaces $\mathcal{S}$ and $\mathcal{T}$ is indicated by formulas below the figures. For example, in case (b), $\partial\mathcal{V} = \mathcal{S}_2 - \mathcal{S}_1$ means that $\partial\mathcal{V}$ is made by joining together $\mathcal{S}_2$ with its standard orientation and $\mathcal{S}_1$ with reversed orientation.

*Total flux of 4-momentum outward across a closed three-dimensional surface must vanish.* (5.26)

To calculate the total outward flux in the most elementary of fashions, approximate the closed 3-surface $\partial\mathcal{V}$ by a large number of flat 3-volumes ("boiler plates") with positive direction oriented outward (away from $\mathcal{V}$). Then

$$\boldsymbol{p}_{\text{total out}} = \sum_{\text{boiler plates } A} \boldsymbol{T}(\ldots, \boldsymbol{\Sigma}_{(A)}) = 0, \tag{5.27}$$

where $\boldsymbol{\Sigma}_{(A)}$ is the volume 1-form of boiler plate $A$. Equivalently, in component notation

$$p^{\mu}{}_{\text{total out}} = \sum_{A} T^{\mu\alpha}\Sigma_{(A)\alpha}. \tag{5.27$'$}$$

To be slightly more sophisticated about the calculation, take the limit as the number of boiler plates goes to infinity and their sizes go to zero. The result is an integral (Box 5.3, at the end of this section),

$$p^{\mu}{}_{\text{total out}} = \oint_{\partial\mathcal{V}} T^{\mu\alpha}\, d^3\Sigma_{\alpha} = 0. \tag{5.28}$$

Think of this (like all component equations) as a convenient way to express a coordinate-independent statement:

$$\boldsymbol{p}_{\text{total out}} = \oint_{\partial\mathcal{V}} \boldsymbol{T} \cdot d^3\boldsymbol{\Sigma} = 0. \tag{5.29}$$

To be more sophisticated yet (not recommended on first reading of this book) and to simplify the computations in practical cases, interpret the integrands as exterior differential forms (Box 5.4, at the end of this section).

But however one calculates it, and however one interprets the integrands, the statement of the result is simple: the total flux of 4-momentum outward across a closed 3-surface must vanish.

Special cases of integral conservation law:

Several special cases of this "integral conservation law," shown in Figure 5.3, are instructive. There shown, in addition to the general case (a), are:

### Case (b)

The closed 3-surface $\partial\mathcal{V}$ is made up of two slices taken at constant time $t$ of a specific Lorentz frame, plus timelike surfaces at "infinity" that join the two slices together. The surfaces at infinity do not contribute to $\oint_{\partial\mathcal{V}} T^{\mu\alpha}\, d^3\Sigma_{\alpha}$ if the stress-energy tensor dies out rapidly enough there. The boundary $\partial\mathcal{V}$ of the standard-oriented 4-volume $\mathcal{V}$, by definition, has its positive sense away from $\mathcal{V}$. This demands nonstandard orientation of $\mathcal{S}_1$ (positive sense toward past), as is indicated by writing $\partial\mathcal{V} = \mathcal{S}_2 - \mathcal{S}_1$; and it produces a sign flip in the evaluation of the hypersurface integral

$$0 = \oint_{\partial\mathcal{V}} T^{\alpha\mu}\, d^3\Sigma_\mu = -\int_{\mathcal{S}_1} T^{\alpha 0}\, dx\, dy\, dz + \int_{\mathcal{S}_2} T^{\alpha 0}\, dx\, dy\, dz.$$

Because $T^{\alpha 0}$ is the density of 4-momentum, this equation says

Total 4-momentum conserved in time

$$\begin{aligned}\begin{pmatrix}\text{total 4-momentum in}\\ \text{all of space at time } t_1\end{pmatrix} &= \int_{\mathcal{S}_1} T^{\alpha 0}\, dx\, dy\, dz \\ &= \begin{pmatrix}\text{total 4-momentum in}\\ \text{all of space at time } t_2\end{pmatrix} = \int_{\mathcal{S}_2} T^{\alpha 0}\, dx\, dy\, dz.\end{aligned} \tag{5.30}$$

### Case (c)

Here one wants to compare hypersurface integrals over $\mathcal{S}$ and $\bar{\mathcal{S}}$, which are slices of constant time, $t =$ const and $\bar{t} =$ const in two different Lorentz frames. To form a closed surface, one adds time-like hypersurfaces at infinity and assumes they do not contribute to the integral. The orientations fit together smoothly and give a closed surface

$$\partial\mathcal{V} = \bar{\mathcal{S}} - \mathcal{S} + (\text{surfaces at infinity})$$

only if one takes $\mathcal{V} = \mathcal{V}_2 - \mathcal{V}_1$—i.e., only if one uses the nonstandard 4-volume orientation in $\mathcal{V}_1$. (See part A.1 of Box 5.3 for "standard" versus "non-standard" orientation.) The integral conservation law then gives

$$0 = \int_{\bar{\mathcal{S}}} \boldsymbol{T} \cdot d^3\boldsymbol{\Sigma} - \int_{\mathcal{S}} \boldsymbol{T} \cdot d^3\boldsymbol{\Sigma},$$

or, equivalently,

Total 4-momentum the same in all Lorentz frames

$$\begin{aligned}\int_{\bar{\mathcal{S}}} \boldsymbol{T} \cdot d^3\boldsymbol{\Sigma} &= (\text{total 4-momentum } \boldsymbol{p} \text{ on } \bar{\mathcal{S}}) \\ = \int_{\mathcal{S}} \boldsymbol{T} \cdot d^3\boldsymbol{\Sigma} &= (\text{total 4-momentum } \boldsymbol{p} \text{ on } \mathcal{S}).\end{aligned} \tag{5.31}$$

This says that observers in different Lorentz frames measure the same total 4-momentum $\boldsymbol{p}$. It does *not* mean that they measure the same components ($p^\alpha \neq p^{\bar{\alpha}}$); rather, it means they measure the same geometric vector

$$\boldsymbol{p}_{\text{on } \mathcal{S}} = p^\alpha \boldsymbol{e}_\alpha = \boldsymbol{p}_{\text{on } \bar{\mathcal{S}}} = p^{\bar{\alpha}} \boldsymbol{e}_{\bar{\alpha}},$$

a vector whose components are connected by the usual Lorentz transformation law

$$p^\alpha = \Lambda^\alpha{}_{\bar{\beta}}\, p^{\bar{\beta}}. \tag{5.32}$$

### Case (d)

Here the contribution to the integral comes entirely from two arbitrary spacelike hypersurfaces, $\mathcal{S}_A$ and $\mathcal{S}_B$, cutting all the way across spacetime. As in cases (a) and (b), the integral form of the conservation law says

Total 4-momentum independent of hypersurface where measured

$$\boldsymbol{p}_{\text{on } \mathcal{S}_A} = \boldsymbol{p}_{\text{on } \mathcal{S}_B}; \tag{5.33}$$

i.e., *the total 4-momentum on a spacelike slice through spacetime is independent of the specific slice chosen*—so long as the energy-momentum flux across the "hypersurface at infinity" connecting $\mathcal{S}_A$ and $\mathcal{S}_B$ is zero.

### Case (e)

This case concerns a box whose walls oscillate and accelerate as time passes. The three-dimensional boundary $\partial\mathcal{V}$ is made up of (1) the interior $\mathcal{S}$ of the box, at an initial moment of time $t =$ constant in the box's initial Lorentz frame, taken with nonstandard orientation; (2) the interior $\bar{\mathcal{S}}$ of the box, at $\bar{t} =$ constant in its final Lorentz frame, with standard orientation; (3) the 3-volume $\mathcal{T}$ swept out by the box's two-dimensional faces between the initial and final states, with positive sense oriented outward. The integral conservation law $\int_{\partial\mathcal{V}} \boldsymbol{T} \cdot d^3\boldsymbol{\Sigma} = 0$ says

Change with time of 4-momentum in a box equals flux of 4-momentum across its faces

$$\begin{aligned}\begin{pmatrix}\text{total 4-momentum}\\ \text{in box at } \bar{\mathcal{S}}\end{pmatrix} - \begin{pmatrix}\text{total 4-momentum}\\ \text{in box at } \mathcal{S}\end{pmatrix}& \\ = \begin{pmatrix}\text{total 4-momentum that enters box through}\\ \text{its faces between states } \mathcal{S} \text{ and } \bar{\mathcal{S}}\end{pmatrix}&.\end{aligned} \tag{5.34}$$

## §5.9. CONSERVATION OF 4-MOMENTUM: DIFFERENTIAL FORMULATION

Complementary to any "integral conservation law in flat spacetime" is a "differential conservation law" with identical information content. To pass back and forth between them, one can use Gauss's theorem.

Gauss's theorem in four dimensions, applied to the law of 4-momentum conservation, converts the surface integral of $T^{\mu\alpha}$ into a volume integral of $T^{\mu\alpha}{}_{,\alpha}$:

$$0 = \oint_{\partial\mathcal{V}} T^{\mu\alpha}\, d^3\Sigma_\alpha = \int_{\mathcal{V}} T^{\mu\alpha}{}_{,\alpha}\, dt\, dx\, dy\, dz. \tag{5.35}$$

(See Box 5.3 for elementary discussion; Box 5.4 for sophisticated discussion.) If the integral of $T^{\mu\alpha}{}_{,\alpha}$ is to vanish, as demanded, for any and every 4-volume $\mathcal{V}$, then $T^{\mu\alpha}{}_{,\alpha}$ must itself vanish everywhere in spacetime:

Differential conservation law for 4-momentum: $\boldsymbol{\nabla} \cdot \boldsymbol{T} = 0$

$$T^{\mu\alpha}{}_{,\alpha} = 0; \text{ i.e., } \boldsymbol{\nabla} \cdot \boldsymbol{T} = 0 \text{ everywhere.} \tag{5.36}$$

(*continued on page 152*)

**Box 5.3 VOLUME INTEGRALS, SURFACE INTEGRALS, AND GAUSS'S THEOREM IN COMPONENT NOTATION**

## A. Volume Integrals in Spacetime

1. By analogy with three-dimensional space, the volume of a "hyperparallelepiped" with vector edges $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$ is

$$\text{4-volume} \equiv \Omega \equiv \epsilon_{\alpha\beta\gamma\delta}A^\alpha B^\beta C^\gamma D^\delta = \det \begin{Vmatrix} A^0 & A^1 & A^2 & A^3 \\ B^0 & B^1 & B^2 & B^3 \\ C^0 & C^1 & C^2 & C^3 \\ D^0 & D^1 & D^2 & D^3 \end{Vmatrix}$$

$$= {}^*(\boldsymbol{A} \wedge \boldsymbol{B} \wedge \boldsymbol{C} \wedge \boldsymbol{D}).$$

Here, as for 3-volumes, orientation matters; interchange of any two edges reverses the sign of $\Omega$. The *standard orientation* for any 4-volume is the one which makes $\Omega$ positive; thus, $\boldsymbol{e}_0 \wedge \boldsymbol{e}_1 \wedge \boldsymbol{e}_2 \wedge \boldsymbol{e}_3$ has standard orientation if $\boldsymbol{e}_0$ points toward the future and $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ are a righthanded triad.

2. The "volume element" whose edges in a specific, standard-oriented Lorentz frame are

$$A^\alpha = (\Delta t, 0, 0, 0),\ B^\alpha = (0, \Delta x, 0, 0),\ C^\alpha = (0, 0, \Delta y, 0),\ D^\alpha = (0, 0, 0, \Delta z)$$

has a 4-volume, according to the above definition, given by

$$\Delta^4\Omega = \epsilon_{0123}\, \Delta t\, \Delta x\, \Delta y\, \Delta z = \Delta t\, \Delta x\, \Delta y\, \Delta z.$$

3. Thus, the volume integral of a tensor $\boldsymbol{S}$ over a four-dimensional region $\mathcal{V}$ of spacetime, defined as

$$\boldsymbol{M} \equiv \lim_{\begin{pmatrix}\text{number of}\\ \text{elementary}\\ \text{volumes}\\ \to\infty\end{pmatrix}} \sum_{\begin{pmatrix}\text{elementary}\\ \text{volumes } \mathcal{U}\\ \text{in } \mathcal{V}\end{pmatrix}} \boldsymbol{S}_{\text{at center of } \mathcal{U}}\,(\text{volume of } \mathcal{U}),$$

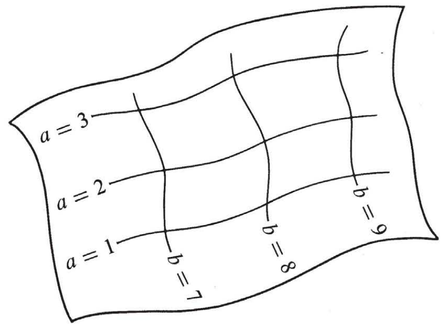
can be calculated in a Lorentz frame by

$$M^\alpha{}_{\beta\gamma} = \int_{\mathcal{V}} S^\alpha{}_{\beta\gamma}\, d^4\Omega = \int_{\mathcal{V}} S^\alpha{}_{\beta\gamma}\, dt\, dx\, dy\, dz.$$

**Box 5.3 (*continued*)**

## B. Integrals over *3*-Surfaces in Spacetime

1. Introduce arbitrary coordinates $a$, $b$, $c$ on the three-dimensional surface. The elementary volume bounded by coordinate surfaces




$$a_0 < a < a_0 + \Delta a, \qquad b_0 < b < b_0 + \Delta b,$$
$$c_0 < c < c_0 + \Delta c$$

has edges

$$A^\alpha = \frac{\partial x^\alpha}{\partial a}\Delta a,\ B^\beta = \frac{\partial x^\beta}{\partial b}\Delta b,\ C^\gamma = \frac{\partial x^\gamma}{\partial c}\Delta c;$$

so its volume 1-form is

$$\Delta^3\Sigma_\mu = \epsilon_{\mu\alpha\beta\gamma}\frac{\partial x^\alpha}{\partial a}\frac{\partial x^\beta}{\partial b}\frac{\partial x^\gamma}{\partial c}\Delta a\,\Delta b\,\Delta c.$$

2. The integral of a tensor $\boldsymbol{S}$ over the 3-surface $\mathcal{S}$ thus has components

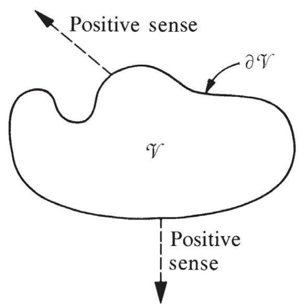
$$N^\alpha{}_\beta = \int_{\mathcal{S}} S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma = \int_{\mathcal{S}} S^\alpha{}_\beta{}^\gamma \epsilon_{\gamma\mu\nu\lambda}\frac{\partial x^\mu}{\partial a}\frac{\partial x^\nu}{\partial b}\frac{\partial x^\lambda}{\partial c}\, da\, db\, dc.$$

An equivalent formula involving a Jacobian is often used (see exercise 5.5):

$$N^\alpha{}_\beta = \int_{\mathcal{S}} S^\alpha{}_\beta{}^\gamma \frac{1}{3!}\epsilon_{\gamma\mu\nu\lambda}\frac{\partial(x^\mu, x^\nu, x^\lambda)}{\partial(a, b, c)}\, da\, db\, dc.$$

## C. Gauss's Theorem Stated

1. Consider a bounded four-dimensional region of spacetime $\mathcal{V}$ with closed boundary $\partial\mathcal{V}$. Orient the volume 1-forms on $\partial\mathcal{V}$ so that the "positive sense" is away from $\mathcal{V}$.
2. Choose a tensor field $\boldsymbol{S}$. Integrate its divergence over $\mathcal{V}$, and integrate it itself over $\partial\mathcal{V}$. The results must be the same (*Gauss's theorem*):




$$\int_{\mathcal{V}} S^\alpha{}_\beta{}^\gamma{}_{,\gamma}\, d^4\Omega = \oint_{\partial\mathcal{V}} S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma.$$

## D. Proof of Gauss's Theorem

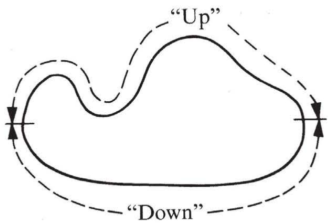
1. The indices $\alpha$ and $\beta$ of $S^{\alpha}{}_{\beta}{}^{\gamma}$ "go along for a free ride," so one can suppress them from the proof. Then the equation to be derived is

$$\int_{\mathcal{V}} S^{\gamma}{}_{,\gamma}\, dt\, dx\, dy\, dz = \oint_{\partial\mathcal{V}} S^{\gamma}\, d^3\Sigma_{\gamma}.$$

2. Since the integral of a derivative is just the original function, the volume integral of $S^0{}_{,0}$ is

$$\int_{\mathcal{V}} S^0{}_{,0}\, dt\, dx\, dy\, dz$$

$$= \int_{\text{"up"}} S^0\, dx\, dy\, dz - \int_{\text{"down"}} S^0\, dx\, dy\, dz.$$




3. The surface integral $\int_{\partial\mathcal{V}} S^0\, d^3\Sigma_0$ can be reduced to the same set of terms:
   a. Use $x, y, z$ as coordinates on $\partial\mathcal{V}$. On the "up" side, $d^3\Sigma_0$ must be positive to achieve a "positive" sense pointing away from $\mathcal{V}$, so (see part B above)

$$d^3\Sigma_0 = \epsilon_{0\alpha\beta\gamma} \frac{\partial x^{\alpha}}{\partial x} \frac{\partial x^{\beta}}{\partial y} \frac{\partial x^{\gamma}}{\partial z}\, dx\, dy\, dz = \epsilon_{0123}\, dx\, dy\, dz = dx\, dy\, dz.$$

   b. On the "down" side, $d^3\Sigma_0$ must be negative, so

$$d^3\Sigma_0 = -dx\, dy\, dz.$$

   c. Hence,

$$\int_{\partial\mathcal{V}} S^0\, d^3\Sigma_0 = \int_{\text{"up"}} S^0\, dx\, dy\, dz - \int_{\text{"down"}} S^0\, dx\, dy\, dz.$$

4. Equality is proved for the other components in the same manner. Adding components produces the result desired:

$$\int_{\mathcal{V}} S^{\gamma}{}_{,\gamma}\, d^4\Omega = \oint_{\partial\mathcal{V}} S^{\gamma}\, d^3\Sigma_{\gamma}.$$

2

FOR THE READER WHO HAS STUDIED CHAPTER 4

**Box 5.4 I. EVERY INTEGRAL IS THE INTEGRAL OF A FORM.**
**II. THE THEOREM OF GAUSS IN THE LANGUAGE OF FORMS.**

I. *Every integral encountered* in Chapter 5 can be interpreted as *the integral of an exterior differential form.* This circumstance shows up in fourfold and threefold integrals, for example, in the fact that

$$d^4\Omega = \boldsymbol{\varepsilon} = {}^*1 = \varepsilon_{0123}\, \mathbf{d}t \wedge \mathbf{d}x \wedge \mathbf{d}y \wedge \mathbf{d}z$$

and

$$d^3\Sigma_\mu = \varepsilon_{\mu|\alpha\beta\gamma|}\, \mathbf{d}x^\alpha \wedge \mathbf{d}x^\beta \wedge \mathbf{d}x^\gamma$$

are basis 4- and 3-forms. (Recall: the indices $\alpha\beta\gamma$ between vertical bars are to be summed only over $0 \leq \alpha < \beta < \gamma \leq 3$.) A more extensive glossary of notations is found in C below.

II. *Gauss's Theorem for a tensor integral* in flat space reads

$$\int_{\mathscr{V}} (\boldsymbol{\nabla} \cdot \boldsymbol{S})\, d^4\Omega = \oint_{\partial\mathscr{V}} \boldsymbol{S} \cdot d\boldsymbol{\Sigma}$$

for any tensor, such as $\boldsymbol{S} = S^\alpha{}_\beta{}^\gamma \boldsymbol{e}_\alpha \otimes \boldsymbol{\omega}^\beta \otimes \boldsymbol{e}_\gamma$ (see Box 5.3 for component form). It is an application of the generalized Stokes Theorem (Box 4.1), and *depends on the fact that the basis vectors $\boldsymbol{e}_\alpha$ and $\boldsymbol{\omega}^\beta$ of a global Lorentz frame are constants,* i.e., are independent of $x$. The definitions follow in A; the proof is in B.

A. *Tensor-valued integrals* can be defined in flat spaces because one uses constant basis vectors. Thus one defines

$$\int \boldsymbol{S} \cdot d^3\boldsymbol{\Sigma} = \boldsymbol{e}_\alpha \otimes \boldsymbol{\omega}^\beta \int S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma$$

for a tensor of the indicated rank. One justifies pulling basis vectors and forms outside the integral sign because they are constants, independent of location in spacetime. Each of the numbers $\int S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma$ (for $\alpha, \beta = 0, 1, 2, 3$) is then evaluated by substituting *any* properly oriented parametrization of the hypersurface into the 3-form $S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma$ as described in Box 4.1 (arbitrary curvilinear parametrization in the part of the calculation not involving the "free indices" $\alpha$ and $\beta$). In other words, $\boldsymbol{S} \cdot d^3\boldsymbol{\Sigma} = \boldsymbol{e}_\alpha \otimes \boldsymbol{\omega}^\beta \otimes S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma$ is considered a "tensor-valued 3-form." Under an integral sign, it is contracted with the hyperplane element tangent to the 3-surface $\mathscr{P}(\lambda^1, \lambda^2, \lambda^3)$ of integration to form the integral

$$\begin{aligned}
\int \boldsymbol{S} \cdot d^3\boldsymbol{\Sigma} &= \int \left\langle \boldsymbol{S} \cdot d^3\boldsymbol{\Sigma}, \frac{\partial\mathscr{P}}{\partial\lambda^1} \wedge \frac{\partial\mathscr{P}}{\partial\lambda^2} \wedge \frac{\partial\mathscr{P}}{\partial\lambda^3} \right\rangle d\lambda^1\, d\lambda^2\, d\lambda^3 \\
&= \boldsymbol{e}_\alpha \otimes \boldsymbol{\omega}^\beta \int S^\alpha{}_\beta{}^\gamma \varepsilon_{\gamma|\lambda\mu\nu|} \underbrace{\frac{\partial(x^\lambda, x^\mu, x^\nu)}{\partial(\lambda^1, \lambda^2, \lambda^3)}}_{\substack{\text{Jacobian}\\ \text{determinant}}} d\lambda^1\, d\lambda^2\, d\lambda^3.
\end{aligned}$$

Although constant basis vectors $\boldsymbol{e}_\alpha$, $\boldsymbol{\omega}^\beta$ derived from rectangular coordinates are essential here, *a completely general parametrization of the hypersurface may be used.*

B. *The proof of Gauss's Theorem* is a computation:

$$\oint_{\partial\mathcal{V}} \boldsymbol{S}\cdot d^3\boldsymbol{\Sigma} = \boldsymbol{e}_\alpha \otimes \boldsymbol{\omega}^\beta \oint_{\partial\mathcal{V}} S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma \qquad (\boldsymbol{e}_\alpha, \boldsymbol{\omega}^\beta \text{ are constant})$$

$$= \boldsymbol{e}_\alpha \otimes \boldsymbol{\omega}^\beta \int_{\mathcal{V}} \boldsymbol{d}(S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma) \qquad \text{(Stokes Theorem)}$$

$$= \boldsymbol{e}_\alpha \otimes \boldsymbol{\omega}^\beta \int_{\mathcal{V}} S^\alpha{}_\beta{}^\gamma{}_{,\gamma}\, {*1} \qquad \text{(see below)}$$

$$= \int_{\mathcal{V}} (\boldsymbol{\nabla}\cdot\boldsymbol{S})\, d^4\Omega. \qquad \text{(merely notation)}$$

The missing computational step above is

$$\begin{aligned}\boldsymbol{d}(S^\alpha{}_\beta{}^\gamma\, d^3\Sigma_\gamma) &= (\partial S^\alpha{}_\beta{}^\gamma/\partial x^\rho)\, \boldsymbol{d}x^\rho \wedge d^3\Sigma_\gamma \\ &= (\partial S^\alpha{}_\beta{}^\gamma/\partial x^\gamma)\, {*1}.\end{aligned}$$

Here the first step uses $\boldsymbol{d}(d^3\Sigma_\gamma) = 0$ (which follows from $\varepsilon_{\mu\alpha\beta\gamma} = \text{const}$ in flat spacetime). The second step uses

$$\boldsymbol{d}x^\rho \wedge d^3\Sigma_\gamma = \delta^\rho_\gamma\, {*1}.$$

[Write the lefthand side of this identity as $\varepsilon_{\gamma|\mu\nu\lambda|}\, \boldsymbol{d}x^\rho \wedge \boldsymbol{d}x^\mu \wedge \boldsymbol{d}x^\nu \wedge \boldsymbol{d}x^\lambda$. The only possible non-zero term in the sum over $\mu\nu\lambda$ is the one with $\mu < \nu < \lambda$ all different from $\rho$. The righthand side is the value of this term.]

C. *Glossary of notations.*

Charge density 3-form:

$$\begin{aligned}{*\boldsymbol{J}} &= J^\mu\, d^3\Sigma_\mu = \boldsymbol{J}\cdot d^3\boldsymbol{\Sigma} \\ &= \underbrace{J^\mu \varepsilon_{\mu\alpha\beta\gamma}}_{({*J})_{\alpha\beta\gamma}}\, \underbrace{\varepsilon_{\mu\alpha\beta\gamma}\, \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta \wedge \boldsymbol{d}x^\gamma/3!}_{d^3\Sigma_\mu}.\end{aligned}$$

Maxwell and Faraday 2-forms:

$${*\boldsymbol{F}} = \frac{1}{2} F^{\mu\nu}\, d^2S_{\mu\nu};$$

$$\boldsymbol{F} = \frac{1}{2} F_{\mu\nu}\, \boldsymbol{d}x^\mu \wedge \boldsymbol{d}x^\nu.$$

Basis 2-forms:

$$\boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta; \qquad \text{(one way to label)}$$

$$d^2S_{\mu\nu} = \varepsilon_{\mu\nu|\alpha\beta|}\, \boldsymbol{d}x^\alpha \wedge \boldsymbol{d}x^\beta. \qquad \text{(dual way to label)}$$

Energy-momentum density 3-form:

$$\boldsymbol{T}\cdot d^3\boldsymbol{\Sigma} \equiv \boldsymbol{e}_\mu T^{\mu\nu}\, d^3\Sigma_\nu \equiv \underbrace{{*\boldsymbol{T}};}_{\text{dual on last index, } ({*T})^\mu{}_{\alpha\beta\gamma} = T^{\mu\nu}\varepsilon_{\nu\alpha\beta\gamma}.}$$

Angular momentum density 3-form:

$$\boldsymbol{\mathcal{J}}\cdot d^3\boldsymbol{\Sigma} \equiv \frac{1}{2} \boldsymbol{e}_\mu \wedge \boldsymbol{e}_\nu \mathcal{J}^{\mu\nu\alpha}\, d^3\Sigma_\alpha \equiv {*\boldsymbol{\mathcal{J}}};$$

$$({*\mathcal{J}})^{\mu\nu}{}_{\alpha\beta\gamma} = \mathcal{J}^{\mu\nu\lambda}\varepsilon_{\lambda\alpha\beta\gamma}.$$

2



(In the frame-independent equation $\boldsymbol{\nabla}\cdot\boldsymbol{T}=0$, one need not worry about which slot of $\boldsymbol{T}$ to take the divergence on; the slots are symmetric, so either can be used.)

The equation $\boldsymbol{\nabla}\cdot\boldsymbol{T}=0$ is *the differential formulation of the law of 4-momentum conservation*. It is also called the *equation of motion for stress-energy*, because it places constraints on the dynamic evolution of the stress-energy tensor. To examine these constraints for simple systems is to realize the beauty and power of the equation $\boldsymbol{\nabla}\cdot\boldsymbol{T}=0$.

## §5.10. SAMPLE APPLICATIONS OF $\boldsymbol{\nabla}\cdot\boldsymbol{T}=0$

Newtonian fluid characterized by $|v^j| \ll 1$, $p \ll \rho$

The equation of motion $\boldsymbol{\nabla}\cdot\boldsymbol{T}=0$ makes contact with the classical (Newtonian) equations of hydrodynamics, when applied to a nearly Newtonian fluid. Such a fluid has low velocities relative to the Lorentz frame used, $|v^j| \ll 1$; and in its rest frame its pressure is small compared to its density of mass-energy, $p/\rho = p/\rho c^2 \ll 1$. For example, the air in a hurricane has

$$|v^j| \sim 100 \text{ km/hour} \sim 3{,}000 \text{ cm/sec} \sim 10^{-7}\, c = 10^{-7} \ll 1,$$

$$\frac{p}{\rho} \sim \frac{1 \text{ atmosphere}}{10^{-3}\text{ g/cm}^3} \sim \frac{10^6 \text{ dynes/cm}^2}{10^{-3}\text{ g/cm}^3} = 10^9\,\frac{\text{cm}^2}{\text{sec}^2} \sim 10^{-12}\, c^2 = 10^{-12} \ll 1.$$

Stress-energy tensor and equation of motion for a Newtonian fluid

The stress-energy tensor for such a fluid has components

$$T^{00} = (\rho + p)u^0u^0 - p \approx \rho, \tag{5.37a}$$

$$T^{0j} = T^{j0} = (\rho + p)u^0u^j \approx \rho v^j, \tag{5.37b}$$

$$T^{jk} = (\rho + p)u^ju^k + p\,\delta^{jk} \approx \rho v^jv^k + p\,\delta^{jk}; \tag{5.37c}$$

and the equation of motion $\boldsymbol{\nabla}\cdot\boldsymbol{T}=0$ has components

$$T^{00}{}_{,0} + T^{0j}{}_{,j} = \partial\rho/\partial t + \boldsymbol{\nabla}\cdot(\rho\boldsymbol{v}) = 0 \quad \text{("equation of continuity");} \tag{5.38a}$$

and

$$T^{j0}{}_{,0} + T^{jk}{}_{,k} = \partial(\rho v^j)/\partial t + \partial(\rho v^jv^k)/\partial x^k + \partial p/\partial x^j = 0,$$

or, equivalently (by combining with the equation of continuity),

$$\frac{\partial\boldsymbol{v}}{\partial t} + (\boldsymbol{v}\cdot\boldsymbol{\nabla})\boldsymbol{v} = -\frac{1}{\rho}\boldsymbol{\nabla}p \qquad \text{("Euler's equation").} \tag{5.38b}$$

Box 5.5 derives and discusses these results from the Newtonian viewpoint.

Application of $\boldsymbol{\nabla}\cdot\boldsymbol{T}=0$ to an electrically charged, vibrating rubber block

As a second application of $\boldsymbol{\nabla}\cdot\boldsymbol{T}=0$, consider a composite system: a block of rubber with electrically charged beads imbedded in it, interacting with an electromagnetic field. The block of rubber vibrates, and its accelerating beads radiate electromagnetic waves; at the same time, incoming electromagnetic waves push on the beads, altering the pattern of vibration of the block of rubber. The interactions shove 4-momentum back and forth between beaded block and electromagnetic field.

2

**Box 5.5 NEWTONIAN HYDRODYNAMICS REVIEWED**

Consider a classical, nonrelativistic, perfect fluid. Apply Newton's law $\boldsymbol{F} = m\boldsymbol{a}$ to a "fluid particle"; that is, to a small fixed mass of fluid followed in its progress through space:

$$\frac{d}{dt}(\text{momentum per unit mass}) = (\text{force per unit mass})$$

$$= \frac{(\text{force per unit volume})}{(\text{density})} = \frac{-(\text{gradient of pressure})}{(\text{density})}$$

or

$$\frac{d\boldsymbol{v}}{dt} = -\frac{1}{\rho}\nabla p. \tag{1}$$

Translate from time-rate of change following the fluid to time-rate of change as measured at a fixed location, finding

$$\begin{pmatrix}\text{rate of change}\\ \text{with time}\\ \text{following fluid}\end{pmatrix} = \begin{pmatrix}\text{rate of change}\\ \text{with time at}\\ \text{fixed location}\end{pmatrix} + \begin{pmatrix}\text{velocity}\\ \text{of fluid}\end{pmatrix} \cdot \begin{pmatrix}\text{rate of change}\\ \text{with position}\end{pmatrix}$$

or

$$\frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v} \cdot \nabla)\boldsymbol{v} = -\frac{1}{\rho}\nabla p \tag{2}$$

or

$$\frac{\partial v_i}{\partial t} + v_{i,k}v_k = -\frac{1}{\rho}p_{,i}.$$

(Latin indices run from 1 to 3; summation convention; upper and lower indices used indifferently for space dimensions in flat space!) This is *Euler's fundamental equation* for the hydrodynamics of a perfect fluid.

Two further equations are needed to complete the description of a perfect fluid. One states the absence of heat transfer by requiring that the specific entropy (entropy per unit mass) be constant for each fluid "particle":

$$\frac{ds}{dt} = 0, \qquad \text{or} \qquad \frac{\partial s}{\partial t} + (\boldsymbol{v} \cdot \nabla)s = 0. \tag{3}$$

The final equation expresses the conservation of mass:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) = 0, \tag{4}$$

or

$$\frac{\partial \rho}{\partial t} + (\rho v_k)_{,k} = 0;$$

2



**Box 5.5 (*continued*)**

it is analagous in every way to the equation that expresses conservation of charge in electrodynamics and that bears the same name, *"equation of continuity."*

The Newtonian stress-energy tensor, like its relativistic counterpart, is linked to conservation of momentum and mass. Therefore examine the time-rate of change of the density of fluid momentum, $\rho v_i$, contained in a unit volume; thus,

$$\partial(\rho v_i)/\partial t = -(\rho v_i v_k)_{,k} - p_{,i}. \tag{5}$$

Momentum flows into the little volume element on the left ("force equals time-rate of change of momentum") and out on the right; similarly at the other faces. Therefore the righthand side of (5) must represent the divergence of this momentum flux:

$$\partial(\rho v_i)/\partial t = -T_{ik,k}. \tag{6}$$

Consequently, we take for the momentum flux itself

$$T^{ik} = T_{ik} = \underbrace{\rho v_i v_k}_{\text{“convection”}} + \underbrace{\delta_{ik} p}_{\text{“push”}} . \tag{7}$$

For the momentum density, the Newtonian value is

$$T^{0i} = T^{i0} = \rho v_i. \tag{8}$$

With this notation, the equation for the time-rate of change of momentum becomes

$$\partial T^{i\mu}/\partial x^\mu = 0; \tag{9}$$

and with $T^{00} = \rho$, the equation of continuity reads

$$\partial T^{0\mu}/\partial x^\mu = 0. \tag{10}$$

In conclusion, these Newtonian considerations give a reasonable approximation to the relativistic stress-energy tensor:

$$\left\| \begin{array}{c:c} \rho & \rho v^j \\ \hdashline \rho v^i & \rho v^i v^j + \delta^{ij} p \end{array} \right\| \simeq \left\| \begin{array}{c:c} (p+\rho)u^0u^0 - p & (p+\rho)u^0u^j \\ \hdashline (p+\rho)u^0u^i & (p+\rho)u^iu^j + \delta^{ij}p \end{array} \right\| \tag{11}$$

The 4-momentum of neither block nor field is conserved; neither $\nabla \cdot \boldsymbol{T}_{\text{(block)}}$ nor $\nabla \cdot \boldsymbol{T}_{\text{(em field)}}$ vanishes. But total 4-momentum must be conserved, so

$$\nabla \cdot (\boldsymbol{T}_{\text{(block)}} + \boldsymbol{T}_{\text{(em field)}}) \text{ must vanish.} \tag{5.39}$$

For a general electromagnetic field interacting with any source, $\nabla \cdot \boldsymbol{T}_{\text{(em field)}}$ has the form

$$T^{\mu\nu}_{\text{(em field)},\nu} = -F^{\mu\alpha} J_{\alpha}. \tag{5.40}$$

(This was derived in exercise 3.18 by combining $T^{\mu\nu}{}_{,\nu} = 0$ with expression 5.22 for the electromagnetic stress-energy tensor, and with Maxwell's equations.) For our beaded block, $\boldsymbol{J}$ is the 4-current associated with the vibrating, charged beads, and $\boldsymbol{F}$ is the electromagnetic field tensor. The time component of equation (5.40) reads

$$\begin{aligned} T^{0\nu}_{\text{(em field)},\nu} &= -F^{0k} J_k = -\boldsymbol{E} \cdot \boldsymbol{J} \\ &= -\begin{pmatrix} \text{rate at which electric field } \boldsymbol{E} \text{ does work} \\ \text{on a unit volume of charged beads} \end{pmatrix}. \end{aligned} \tag{5.41}$$

For comparison, $T^{00}_{\text{(block)},0}$ is the rate at which the block's energy density changes with time, $-T^{0j}_{\text{(block)},j}$ is the contribution of the block's energy flux to this rate of change of energy density, and consequently their difference $T^{0\nu}_{\text{(block)},\nu}$ has the meaning

$$T^{0\nu}_{\text{(block)},\nu} = \begin{pmatrix} \text{rate at which mass-energy of block per} \\ \text{unit volume increases due to actions} \\ \text{other than internal mechanical forces} \\ \text{between one part of block and another} \end{pmatrix}. \tag{5.42}$$

Hence, the conservation law

$$(T^{0\nu}_{\text{(em field)}} + T^{0\nu}_{\text{(block)}})_{,\nu} = 0$$

says that the mass-energy of the block increases at precisely the same rate as the electric field does work on the beads. A similar result holds for momentum:

$$\begin{aligned} T^{k\nu}_{\text{(em field)},\nu} \boldsymbol{e}_k &= -F^{k\nu} J_\nu \boldsymbol{e}_k = -(J^0 \boldsymbol{E} + \boldsymbol{J} \times \boldsymbol{B}) \\ &= -\begin{pmatrix} \text{Lorentz force per unit volume} \\ \text{acting on beads} \end{pmatrix}, \end{aligned} \tag{5.43}$$

$$T^{k\nu}_{\text{(block)},\nu} \boldsymbol{e}_k = \begin{pmatrix} \text{rate at which momentum per unit volume} \\ \text{of block increases due to actions} \\ \text{other than its own stresses} \end{pmatrix}; \tag{5.44}$$

so the conservation law

$$(T^{k\nu}_{\text{(em field)}} + T^{k\nu}_{\text{(block)}})_{,\nu} = 0$$

says that the rate of change of the momentum of the block equals the force of the electromagnetic field on its beads.

2



## §5.11. ANGULAR MOMENTUM

Angular momentum defined and its integral conservation law derived

The symmetry, $T^{\mu\nu} = T^{\nu\mu}$, of the stress-energy tensor enables one to define a conserved angular momentum $J^{\alpha\beta}$, analogous to the linear momentum $p^\alpha$. The angular momentum is defined relative to a specific but arbitrary origin—an event $\mathscr{A}$ with coordinates, in a particular Lorentz frame,

$$x^\alpha(\mathscr{A}) = a^\alpha. \tag{5.45}$$

The angular momentum about $\mathscr{A}$ is defined using the tensor

$$\mathscr{J}^{\alpha\beta\gamma} = (x^\alpha - a^\alpha)T^{\beta\gamma} - (x^\beta - a^\beta)T^{\alpha\gamma}. \tag{5.46}$$

(Note that $x^\alpha - a^\alpha$ is the vector separation of the "field point" $x^\alpha$ from the "origin" $\mathscr{A}$; $T^{\alpha\gamma}$ is here evaluated at the "field point".) Because of the symmetry of $\boldsymbol{T}$, $\mathscr{J}^{\alpha\beta\gamma}$ has vanishing divergence:

$$\begin{aligned}\mathscr{J}^{\alpha\beta\gamma}{}_{,\gamma} &= \delta^\alpha{}_\gamma T^{\beta\gamma} + (x^\alpha - a^\alpha)\underbrace{T^{\beta\gamma}{}_{,\gamma}}_{0} - \delta^\beta{}_\gamma T^{\alpha\gamma} - (x^\beta - a^\beta)\underbrace{T^{\alpha\gamma}{}_{,\gamma}}_{0} \\ &= T^{\beta\alpha} - T^{\alpha\beta} = 0.\end{aligned} \tag{5.47}$$

Consequently, its integral over any closed 3-surface vanishes

$$\oint_{\partial\mathscr{V}} \mathscr{J}^{\alpha\beta\gamma}\, d^3\Sigma_\gamma = 0 \tag{5.48}$$

("*integral form of the law of conservation of angular momentum*").

The integral over a spacelike surface of constant time $t$ is

$$J^{\alpha\beta} = \int \mathscr{J}^{\alpha\beta 0}\, dx\, dy\, dz = \int [(x^\alpha - a^\alpha)T^{\beta 0} - (x^\beta - a^\beta)T^{\alpha 0}]\, dx\, dy\, dz. \tag{5.49}$$

Recalling that $T^{\beta 0}$ is momentum density, one sees that (5.49) has the same form as the equation "$\boldsymbol{J} = \boldsymbol{r} \times \boldsymbol{p}$" of Newtonian theory. Hence the name "total angular momentum" for $J^{\alpha\beta}$. Various aspects of this conserved angular momentum, including the tie to its Newtonian cousin, are explored in Box 5.6.

---

## EXERCISES

**Exercise 5.2. CHARGE CONSERVATION**

Exercise 3.16 revealed that the charge-current 4-vector $\boldsymbol{J}$ satisfies the differential conservation law $\boldsymbol{\nabla} \cdot \boldsymbol{J} = 0$. Write down the corresponding integral conservation law, and interpret it for the four closed surfaces of Fig. 5.3.

**Exercise 5.3. PARTICLE PRODUCTION**

Inside highly evolved, massive stars, the temperature is so high that electron-positron pairs are continually produced and destroyed. Let $\boldsymbol{S}$ be the number-flux vector for electrons and positrons, and denote its divergence by

$$\epsilon \equiv \boldsymbol{\nabla} \cdot \boldsymbol{S}. \tag{5.50}$$

**Box 5.6 ANGULAR MOMENTUM**

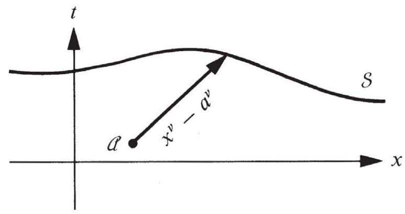
## A. Definition of Angular Momentum

(a) Pick an arbitrary spacelike hypersurface $\mathcal{S}$ and an arbitrary event $\mathcal{A}$ with coordinates $x^\alpha(\mathcal{A}) \equiv a^\alpha$. (Use globally inertial coordinates throughout.)

(b) Define *"total angular momentum on $\mathcal{S}$ about $\mathcal{A}$"* to be

$$J^{\mu\nu} \equiv \int_{\mathcal{S}} \mathcal{J}^{\mu\nu\alpha}\, d^3\Sigma_\alpha,$$

$$\mathcal{J}^{\mu\nu\alpha} \equiv (x^\mu - a^\mu)T^{\nu\alpha} - (x^\nu - a^\nu)T^{\mu\alpha}.$$




(c) If $\mathcal{S}$ is a hypersurface of constant time $t$, this becomes

$$J^{\mu\nu} = \int \mathcal{J}^{\mu\nu 0}\, dx\, dy\, dz.$$

## B. Conservation of Angular Momentum

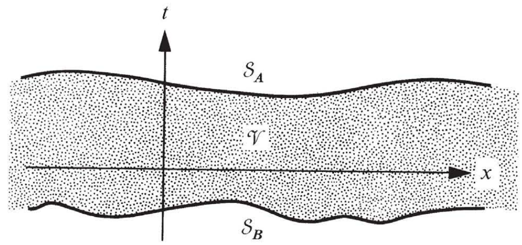
(a) $T^{\mu\nu}{}_{,\nu} = 0$ implies $\mathcal{J}^{\mu\nu\alpha}{}_{,\alpha} = 0$.

(b) This means that $J^{\mu\nu}$ *is independent of the hypersurface $\mathcal{S}$ on which it is calculated* (Gauss's theorem):

$$J^{\mu\nu}(\mathcal{S}_A) - J^{\mu\nu}(\mathcal{S}_B)$$

$$= \int_{\partial\mathcal{V}} \mathcal{J}^{\mu\nu\alpha}\, d^3\Sigma_\alpha$$

$$= \int_{\mathcal{V}} \mathcal{J}^{\mu\nu\alpha}{}_{,\alpha}\, d^4x = 0.$$




(Note: $\partial\mathcal{V} \equiv$ (boundary of $\mathcal{V}$) includes $\mathcal{S}_A$, $\mathcal{S}_B$, and timelike surfaces at spatial infinity; contribution of latter dropped—localized source.)

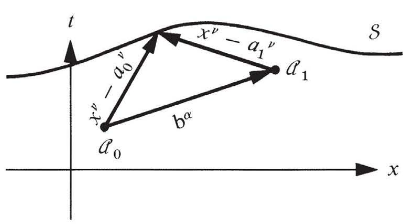
## C. Change of Point About Which Angular Momentum is Calculated

Let $b^\alpha$ be vector from $\mathcal{A}_0$ to $\mathcal{A}_1$: $b^\alpha = a_1{}^\alpha - a_0{}^\alpha$. Then

$$J^{\mu\nu}(\text{about } \mathcal{A}_1) - J^{\mu\nu}(\text{about } \mathcal{A}_0)$$

$$= -b^\mu \int_{\mathcal{S}} T^{\nu\alpha}\, d^3\Sigma_\alpha + b^\nu \int_{\mathcal{S}} T^{\mu\alpha}\, d^3\Sigma_\alpha$$

$$= -b^\mu P^\nu + b^\nu P^\mu,$$




where $P^\mu$ is total 4-momentum.

Box 5.6 (*continued*)

## D. Intrinsic Angular Momentum

(a) Work, for a moment, in the system's rest frame, where

$$P^0 = M, \qquad P^j = 0, \qquad x_{CM}{}^j = \frac{1}{M}\int x^j T^{00}\, d^3x = \text{location of center of mass.}$$

Intrinsic angular momentum is defined as angular momentum about any event $(a^0, x_{CM}{}^j)$ on center of mass's world line. Its components are denoted $S^{\mu\nu}$ and work out to be

$$S^{0j} = 0,\ S^{jk} = \epsilon^{jk\ell} S^\ell,$$

where

$$\begin{aligned} \boldsymbol{S} &\equiv \int (\boldsymbol{x} - \boldsymbol{x}_{CM}) \times (\textbf{momentum density})\, d^3x \\ &\equiv \text{“intrinsic angular momentum vector.”} \end{aligned}$$

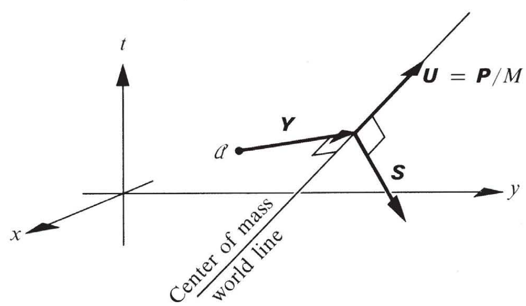
(b) Define "intrinsic angular momentum 4-vector" $S^\mu$ to be that 4-vector whose components in the rest frame are $(0, \boldsymbol{S})$; then the above equations say

$$\begin{aligned} S^{\mu\nu} &= U_\alpha S_\beta \epsilon^{\alpha\beta\mu\nu}, \\ U_\beta &\equiv P_\beta/M = \text{4-velocity of center of mass,} \\ U_\beta S^\beta &= 0. \end{aligned}$$

## E. Decomposition of Angular Momentum into Intrinsic and Orbital Parts

(a) Pick an arbitrary event $\mathscr{a}$, whose perpendicular displacement from center-of-mass world line is $-Y^\alpha$, so

$$U_\beta Y^\beta = 0.$$



(b) Then, by Part C, the angular momentum about $\mathscr{A}$ is

$$J^{\mu\nu} = \underbrace{U_\alpha S_\beta \epsilon^{\alpha\beta\mu\nu}}_{S^{\mu\nu}\text{ (intrinsic)}} + \underbrace{Y^\mu P^\nu - Y^\nu P^\mu}_{L^{\mu\nu}\text{ (orbital)}}.$$

(c) Knowing the angular momentum about $\mathscr{A}$, and the 4-momentum (and hence 4-velocity), one can calculate the vector from $\mathscr{A}$ to the center-of-mass world line,

$$Y^\mu = -J^{\mu\nu}P_\nu/M^2,$$

and the intrinsic angular momentum

$$S_\rho = \frac{1}{2}U^\sigma J^{\mu\nu}\epsilon_{\sigma\mu\nu\rho}.$$

Use Gauss's theorem to show that $\epsilon$ is the number of particles created (minus the number destroyed) in a unit four-dimensional volume of spacetime.

**Exercise 5.4. INERTIAL MASS PER UNIT VOLUME**

Consider a stressed medium in motion with ordinary velocity $|\boldsymbol{v}| \ll 1$ with respect to a specific Lorentz frame.

(a) Show by Lorentz transformations that the spatial components of the momentum density are

$$T^{0j} = \sum_k m^{jk}v^k, \tag{5.51}$$

where

$$m^{jk} = T^{\bar{0}\bar{0}}\,\delta^{jk} + T^{\bar{j}\bar{k}} \tag{5.52}$$

and $T^{\bar{\mu}\bar{\nu}}$ are the components of the stress-energy tensor in the rest frame of the medium. Throughout the solar system $T^{\bar{0}\bar{0}} \gg |T^{\bar{j}\bar{k}}|$ (see, e.g., discussion of hurricane in §5.10), so one is accustomed to write $T^{0j} = T^{\bar{0}\bar{0}}v^j$, i.e., "(momentum density) = (rest-mass density) $\times$ (velocity)". But inside a neutron star $T^{\bar{0}\bar{0}}$ may be of the same order of magnitude as $T^{\bar{j}\bar{k}}$, so one must replace "(momentum density) = (rest-mass density) $\times$ (velocity)" by equations (5.51) and (5.52), at low velocities.

(b) Derive equations (5.51) and (5.52) from Newtonian considerations plus the equivalence of mass and energy. (*Hint:* the total mass-energy carried past the observer by a volume $V$ of the medium includes both the rest mass $T^{\bar{0}\bar{0}}V$ and the work done by forces acting across the volume's faces as they "push" the volume through a distance.)

(c) As a result of relation (5.51), the force per unit volume required to produce an acceleration $dv^k/dt$ in a stressed medium, which is at rest with respect to the man who applies the force, is

$$F^j = dT^{0j}/dt = \sum_k m^{jk}\,dv^k/dt. \tag{5.53}$$

This equation suggests that one call $m^{jk}$ the "inertial mass per unit volume" of a stressed medium at rest. In general $m^{jk}$ is a symmetric 3-tensor. What does it become for the special case of a perfect fluid?

(d) Consider an isolated, stressed body at rest and in equilibrium ($T^{\alpha\beta}{}_{,0} = 0$) in the laboratory frame. Show that its total inertial mass, defined by

$$M^{ij} = \int\limits_{\substack{\text{stressed}\\ \text{body}}} m^{ij}\, dx\, dy\, dz, \tag{5.54}$$

is isotropic and equals the rest mass of the body

$$M^{ij} = \delta^{ij} \int T^{00}\, dx\, dy\, dz. \tag{5.55}$$

**Exercise 5.5. DETERMINANTS AND JACOBIANS**

(a) Write out explicitly the sum defining $d^2S_{01}$ in

$$d^2S_{\mu\nu} \equiv \epsilon_{\mu\nu\alpha\beta} \frac{\partial x^\alpha}{\partial a} \frac{\partial x^\beta}{\partial b}\, da\, db.$$

Thereby establish the formula

$$d^2S_{\mu\nu} = \epsilon_{\mu\nu|\alpha\beta|} \frac{\partial(x^\alpha, x^\beta)}{\partial(a, b)}\, da\, db = \frac{1}{2!} \epsilon_{\mu\nu\alpha\beta} \frac{\partial(x^\alpha, x^\beta)}{\partial(a, b)}\, da\, db.$$

(Expressions such as these should occur only under integral signs. In this exercise one may either supply an $\int \ldots$ wherever necessary, or else interpret the differentials in terms of the exterior calculus, $da\, db \longrightarrow \boldsymbol{da} \wedge \boldsymbol{db}$; see Box 5.4.) The notation used here for Jacobian determinants is

$$\frac{\partial(f, g)}{\partial(a, b)} = \begin{vmatrix} \dfrac{\partial f}{\partial a} & \dfrac{\partial f}{\partial b} \\[2ex] \dfrac{\partial g}{\partial a} & \dfrac{\partial g}{\partial b} \end{vmatrix}.$$

(b) By a similar inspection of a specific case, show that

$$d^3\Sigma_\mu \equiv \epsilon_{\mu\alpha\beta\gamma} \frac{\partial x^\alpha}{\partial a} \frac{\partial x^\beta}{\partial b} \frac{\partial x^\gamma}{\partial c}\, da\, db\, dc = \frac{1}{3!} \epsilon_{\mu\alpha\beta\gamma} \frac{\partial(x^\alpha, x^\beta, x^\gamma)}{\partial(a, b, c)}\, da\, db\, dc.$$

(c) Cite a precise definition of the value of a determinant as a sum of terms (with suitably alternating signs), with each term a product containing one factor from each row and simultaneously one factor from each column. Show that this definition can be stated (in the $4 \times 4$ case, with the $p \times p$ case an obvious extension) as

$$\det A \equiv \det \| A^\lambda{}_\rho \| = \epsilon_{\alpha\beta\gamma\delta} A^\alpha{}_0 A^\beta{}_1 A^\gamma{}_2 A^\delta{}_3.$$

(d) Show that

$$\det A = \frac{1}{4!} \delta^{\mu\nu\rho\sigma}_{\alpha\beta\gamma\delta} A^\alpha{}_\mu A^\beta{}_\nu A^\gamma{}_\rho A^\delta{}_\sigma$$

(for a definition of $\delta^{\mu\nu\rho\sigma}_{\alpha\beta\gamma\delta}$, see exercises 3.13 and 4.12).

(e) Use properties of the $\delta$-symbol to show that the matrix $A^{-1}$ inverse to $A$ has entries $(A^{-1})^{\mu}{}_{\alpha}$ given by

$$(A^{-1})^{\mu}{}_{\alpha}(\det A) = \frac{1}{3!}\,\delta^{\mu\nu\rho\sigma}_{\alpha\beta\gamma\delta}A^{\beta}{}_{\nu}A^{\gamma}{}_{\rho}A^{\delta}{}_{\sigma}.$$

(f) By an "index-mechanics" computation, from the formula for det $A$ in part (d) derive the following expression for the derivative of the logarithm of the determinant

$$\boldsymbol{d}\ln|\det A| = \operatorname{trace}(A^{-1}\,\boldsymbol{d}A).$$

Here $\boldsymbol{d}A$ is the matrix $\|\boldsymbol{d}A^{\alpha}{}_{\mu}\|$ whose entries are 1-forms.

### Exercise 5.6. CENTROIDS AND SIZES

Consider an isolated system with stress-energy tensor $T^{\mu\nu}$, total 4-momentum $P^{\alpha}$, magnitude of 4-momentum $M = (-\boldsymbol{P}\cdot\boldsymbol{P})^{1/2}$, intrinsic angular momentum tensor $S^{\alpha\beta}$, and intrinsic angular momentum vector $S^{\alpha}$. (See Box 5.6.) An observer with 4-velocity $u^{\alpha}$ defines the *centroid* of the system, at his Lorentz time $x^0 = t$ and in his own Lorentz frame, by

$$X^{j}_{\boldsymbol{u}}(t) = (1/P^0)\int_{x^0=t} x^j T^{00}\,d^3x \qquad \text{in Lorentz frame where } \boldsymbol{u} = \partial\mathscr{P}/\partial x^0. \qquad (5.56)$$

This centroid depends on (i) the particular system being studied, (ii) the 4-velocity $\boldsymbol{u}$ of the observer, and (iii) the time $t$ at which the system is observed.

(a) Show that the centroid moves with a uniform velocity

$$dX^{j}_{\boldsymbol{u}}/dt = P^j/P^0, \qquad (5.57)$$

corresponding to the 4-velocity

$$\boldsymbol{U} = \boldsymbol{P}/M. \qquad (5.57')$$

Note that this "4-velocity of centroid" is independent of the 4-velocity $\boldsymbol{u}$ used in defining the centroid.

(b) The centroid associated with the rest frame of the system (i.e., the centroid defined with $\boldsymbol{u} = \boldsymbol{U}$) is called the *center of mass;* see Box 5.6. Let $\boldsymbol{\xi}_{\boldsymbol{u}}$ be a vector reaching from any event on the center-of-mass world line to any event on the world line of the centroid associated with 4-velocity $\boldsymbol{u}$; thus the components of $\boldsymbol{\xi}_{\boldsymbol{u}}$ in any coordinate system are

$$\xi^{\alpha}_{\boldsymbol{u}} = X^{\alpha}_{\boldsymbol{u}} - X^{\alpha}_{\boldsymbol{U}}. \qquad (5.58)$$

Show that $\boldsymbol{\xi}_{\boldsymbol{u}}$ satisfies the equation

$$[(\xi^{\alpha}_{\boldsymbol{u}}P^{\beta} - P^{\alpha}\xi^{\beta}_{\boldsymbol{u}}) - S^{\alpha\beta}]u_{\beta} = 0. \qquad (5.59)$$

[*Hint:* perform the calculation in a Lorentz frame where $\boldsymbol{u} = \partial\mathscr{P}/\partial x^0$.]

(c) Show that, as seen in the rest-frame of the system at any given moment of time, the above equation reduces to the three-dimensional Euclidean equation

$$\boldsymbol{\xi}_{\boldsymbol{u}} = -(\boldsymbol{v}\times\boldsymbol{S})/M, \qquad (5.59')$$

where $\boldsymbol{v} = \boldsymbol{u}/u^0$ is the ordinary velocity of the frame associated with the centroid.

(d) Assume that the energy density measured by any observer anywhere in spacetime is non-negative ($\boldsymbol{u} \cdot \boldsymbol{T} \cdot \boldsymbol{u} \geq 0$ for all timelike $\boldsymbol{u}$). In the rest frame of the system, construct the smallest possible cylinder that is parallel to $\boldsymbol{S}$ and that contains the entire system ($T_{\alpha\beta} = 0$ everywhere outside the cylinder). Show that the radius $r_0$ of this cylinder is limited by

$$r_0 \geq |\boldsymbol{S}|/M. \tag{5.60}$$

Thus, a system with given intrinsic angular momentum $\boldsymbol{S}$ and given mass $M$ has a minimum possible size $r_{0\,\text{min}} = |\boldsymbol{S}|/M$ as measured in its rest frame.

CHAPTER 6

# ACCELERATED OBSERVERS

*The objective world simply* is; *it does not* happen. *Only to the gaze of my consciousness, crawling upward along the life line* [*world line*] *of my body, does a section of this world come to life as a fleeting image in space which continuously changes in time.*

HERMAN WEYL (1949, p. 116)

## §6.1. ACCELERATED OBSERVERS CAN BE ANALYZED USING SPECIAL RELATIVITY

It helps in analyzing gravitation to consider a situation where gravity is mocked up by acceleration. Focus attention on a region so far from any attracting matter, and so free of disturbance, that (to some proposed degree of precision) spacetime there can be considered to be flat and to have Lorentz geometry. Let the observer acquire the feeling that he is subject to gravity, either because of jet rockets strapped to his legs or because he is in a rocket-driven spaceship. How does physics look to him?

Accelerated motion and accelerated observers can be analyzed using special relativity

Dare one answer this question? At this early stage in the book, is one not too ignorant of gravitation physics to predict what physical effects will be measured by an observer who thinks he is in a gravitational field, although he is really in an accelerated spaceship? Quite the contrary; special relativity was developed precisely to predict the physics of accelerated objects—e.g., the radiation from an accelerated charge. Even the fantastic accelerations

$$a_{\text{nuclear}} \sim v^2/r \sim 10^{31}\ \text{cm/sec}^2 \sim 10^{28}\ \text{“earth gravities”}$$

suffered by a neutron bound in a nucleus, and the even greater accelerations met in high-energy particle-scattering events, are routinely and accurately treated within

1

**Box 6.1 GENERAL RELATIVITY IS BUILT ON SPECIAL RELATIVITY**

A tourist in a powered interplanetary rocket feels "gravity." Can a physicist by local effects convince him that this "gravity" is bogus? Never, says Einstein's principle of the local equivalence of gravity and accelerations. But then the physicist will make no errors if he deludes himself into treating true gravity as a local illusion caused by acceleration. Under this delusion, he barges ahead and solves gravitational problems by using special relativity: if he is clever enough to divide every problem into a network of local questions, each solvable under such a delusion, then he can work out all influences of any gravitational field. Only three basic principles are invoked: special-relativity physics, the equivalence principle, and the local nature of physics. They are simple and clear. To apply them, however, imposes a double task: (1) take spacetime apart into locally flat pieces (where the principles are valid), and (2) put these pieces together again into a comprehensible picture. To undertake this dissection and reconstitution, to see curved dynamic spacetime inescapably take form, and to see the consequences for physics: that is general relativity.

the framework of special relativity. The theoretician who confidently applies special relativity to antiproton annihilations and strange-particle resonances is not about to be frightened off by the mere illusions of a rocket passenger who gullibly believed the travel brochures advertising "earth gravity all the way." When spacetime is flat, move however one will, special relativity can handle the job. (It can handle bigger jobs too; see Box 6.1.) The essential features of *how* special relativity handles the job are summarized in Box 6.2 for the benefit of the Track-1 reader, who can skip the rest of the chapter, and also for the benefit of the Track-2 reader, who will find it useful background for the rest of the chapter.

**Box 6.2 ACCELERATED OBSERVERS IN BRIEF**

An accelerated observer can carry clocks and measuring rods with him, and can use them to set up a reference frame (coordinate system) in his neighborhood.

His clocks, if carefully chosen so their structures are affected negligibly by acceleration (e.g., atomic clocks), will tick at the same rate as unaccelerated clocks moving momentarily along with him:

$$\Delta\tau \equiv \begin{pmatrix} \text{time interval ticked off} \\ \text{by observer's clocks as he} \\ \text{moves a vector displacement} \\ \boldsymbol{\xi} \text{ along his world line} \end{pmatrix} = [-\boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{\xi})]^{1/2}.$$

And his rods, if chosen to be sufficiently rigid, will measure the same lengths as momentarily comoving, unaccelerated rods do. (For further discussion, see §16.4, and Boxes 16.2 to 16.4.)

Let the observer's coordinate system be a Cartesian latticework of rods and clocks, with the origin of the lattice always on his world line. He must keep his latticework small:

$$l \equiv \begin{pmatrix}\text{spatial dimensions}\\ \text{of lattice}\end{pmatrix} \ll \begin{pmatrix}\text{the acceleration measured}\\ \text{by accelerometers he carries}\end{pmatrix}^{-1} \equiv \frac{1}{g}.$$

At distances $l$ away from his world line, strange things of dimensionless magnitude $gl$ happen to his lattice—e.g., the acceleration measured by accelerometers differs from $g$ by a fractional amount $\sim gl$ (exercise 6.7); also, clocks initially synchronized with the clock on his world line get out of step (tick at different rates) by a fractional amount $\sim gl$ (exercise 6.6). (Note that an acceleration of one "earth gravity" corresponds to

$$g^{-1} \sim 10^{-3}\ \text{sec}^2/\text{cm} \sim 10^{18}\ \text{cm} \sim 1\ \text{light-year},$$

so the restriction $l \ll 1/g$ is normally not severe.)

To deduce the results of experiments and observations performed by an accelerated observer, one can analyze them in coordinate-independent, geometric terms, and then project the results onto the basis vectors of his accelerated frame. Alternatively, one can analyze the experiments and observations in a Lorentz frame, and then transform to the accelerated frame.

As deduced in this manner, the results of experiments performed locally (at $l \ll 1/g$) by an accelerated observer differ from the results of the same experiments performed in a Lorentz frame in only three ways:

(1) There are complicated fractional differences of order $gl \ll 1$ mentioned above, that can be made negligible by making the accelerated frame small enough.

(2) There are Coriolis forces of precisely the same type as are encountered in Newtonian theory (exercise 6.8). These the observer can get rid of by carefully preventing his latticework from rotating—e.g., by tying it to gyroscopes that he accelerates with himself by means of forces applied to their centers of mass (no torque!). Such a nonrotating latticework has "Fermi-Walker transported" basis vectors (§6.5),

$$\frac{d\boldsymbol{e}_{\alpha'}}{d\tau} = \boldsymbol{u}(\boldsymbol{a} \cdot \boldsymbol{e}_{\alpha'}) - \boldsymbol{a}(\boldsymbol{u} \cdot \boldsymbol{e}_{\alpha'}), \tag{1}$$

where $\boldsymbol{u}$ = 4-velocity, and $\boldsymbol{a} = d\boldsymbol{u}/d\tau$ = 4-acceleration.

(3) There are inertial forces of precisely the same type as are encountered in Newtonian theory (exercise 6.8). These are due to the observer's acceleration, and he cannot get rid of them except by stopping his accelerating.

## §6.2. HYPERBOLIC MOTION

1

2

The rest of this chapter is Track 2.

It depends on no preceding Track-2 material.

It is needed as preparation for

(1) the mathematical analysis of gyroscopes in curved spacetime (exercise 19.2, §40.7), and

(2) the mathematical theory of the proper reference frame of an accelerated observer (§13.6).

It will be helpful in many applications of gravitation theory (Chapters 18–40).

Study a rocket passenger who feels "gravity" because he is being accelerated in flat spacetime. Begin by describing his motion relative to an inertial reference frame. His 4-velocity satisfies the condition $\boldsymbol{u}^2 = -1$. To say that it is fixed in magnitude is to say that the 4-acceleration,

$$\boldsymbol{a} = d\boldsymbol{u}/d\tau, \tag{6.1}$$

is orthogonal to the 4-velocity:

$$0 = (d/d\tau)(-1/2) = (d/d\tau)\left(\frac{1}{2}\boldsymbol{u}\cdot\boldsymbol{u}\right) = \boldsymbol{a}\cdot\boldsymbol{u}. \tag{6.2}$$

This equation implies that $a^0 = 0$ in the rest frame of the passenger (that Lorentz frame, where, at the instant in question, $\boldsymbol{u} = \boldsymbol{e}_0$); in this frame the space components of $a^\mu$ reduce to the ordinary definition of acceleration, $a^i = d^2x^i/dt^2$. From the components $a^\mu = (0; a^i)$ in the rest frame, then, one sees that the magnitude of the acceleration in the rest frame can be computed as the simple invariant

$$\boldsymbol{a}^2 = a^\mu a_\mu = (d^2\mathbf{x}/dt^2)^2_{\text{as measured in rest frame}}.$$

Consider, for simplicity, an observer who feels always a constant acceleration $g$. Take the acceleration to be in the $x^1$ direction of some inertial frame, and take $x^2 = x^3 = 0$. The equations for the motion of the observer in that inertial frame become

$$\frac{dt}{d\tau} = u^0, \qquad \frac{dx}{d\tau} = u^1; \qquad \frac{du^0}{d\tau} = a^0, \qquad \frac{du^1}{d\tau} = a^1. \tag{6.3}$$

Write out the three algebraic equations

$$u^\mu u_\mu = -1,$$
$$u^\mu a_\mu = -u^0a^0 + u^1a^1 = 0,$$

and

$$a^\mu a_\mu = g^2.$$

Solve for the acceleration, finding

$$a^0 = \frac{du^0}{d\tau} = gu^1, \qquad a^1 = \frac{du^1}{d\tau} = gu^0. \tag{6.4}$$

These linear differential equations can be solved immediately. The solution, with a suitable choice of the origin, reads

$$t = g^{-1}\sinh g\tau, \qquad x = g^{-1}\cosh g\tau. \tag{6.5}$$

Uniformly accelerated observer moves on hyperbola in spacetime diagram

Note that $x^2 - t^2 = g^{-2}$. The world line is a hyperbola in a spacetime diagram ("hyperbolic motion"; Figure 6.1). Several interesting aspects of this motion are

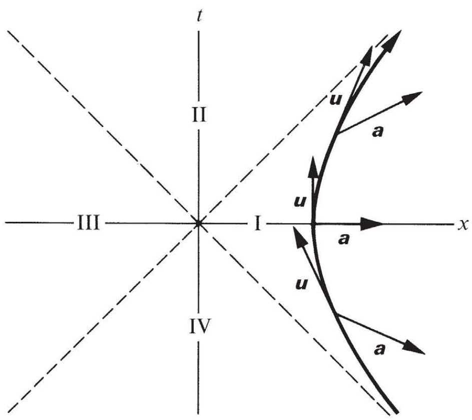



**Figure 6.1.**
Hyperbolic motion. World line of an object that (or an observer who) experiences always a fixed acceleration $g$ with respect to an inertial frame that is instantaneously comoving (different inertial frames at different instants!). The 4-acceleration $\boldsymbol{a}$ is everywhere orthogonal (Lorentz geometry!) to the 4-velocity $\boldsymbol{u}$.

treated in the exercises. Let the magnitude of the constant acceleration $g$ be the acceleration of gravity, $g = 980\ \mathrm{cm/sec^2}$ experienced on earth: $g \simeq (10^3\ \mathrm{cm/sec^2})/(3 \times 10^{10}\ \mathrm{cm/sec})^2 = (3 \times 10^7\ \mathrm{sec} \cdot 3 \times 10^{10}\ \mathrm{cm/sec})^{-1} = (1\ \text{light-year})^{-1}$. Thus the observer will attain relativistic velocities after maintaining this acceleration for something like one year of his own proper time. He can outrun a photon if he has a head start on it of one light-year or more.

---

## EXERCISES

**Exercise 6.1. A TRIP TO THE GALACTIC NUCLEUS**

Compute the proper time required for the occupants of a rocket ship to travel the ~30,000 light-years from the Earth to the center of the Galaxy. Assume that they maintain an acceleration of one "earth gravity" ($10^3\ \mathrm{cm/sec^2}$) for half the trip, and then decelerate at one earth gravity for the remaining half.

**Exercise 6.2. ROCKET PAYLOAD**

What fraction of the initial mass of the rocket can be payload for the journey considered in exercise 6.1? Assume an ideal rocket that converts rest mass into radiation and ejects all the radiation out the back of the rocket with 100 per cent efficiency and perfect collimation.

**Exercise 6.3. TWIN PARADOX**

(a) Show that, of all timelike world lines connecting two events $\mathscr{A}$ and $\mathscr{B}$, the one with the *longest* lapse of proper time is the unaccelerated one. (*Hint:* perform the calculation in the inertial frame of the unaccelerated world line.)

(b) One twin chooses to move from $\mathscr{A}$ to $\mathscr{B}$ along the unaccelerated world line. Show that the other twin, by an appropriate choice of accelerations, can get from $\mathscr{A}$ to $\mathscr{B}$ in arbitrarily small proper time.

(c) If the second twin prefers to ride in comfort, with the acceleration he feels never exceeding one earth gravity, $g$, what is the shortest proper time-lapse he can achieve between $\mathscr{A}$ and $\mathscr{B}$? Express the answer in terms of $g$ and the proper time-lapse $\Delta\tau$ measured by the unaccelerated twin.

(d) Evaluate the answer numerically for several interesting trips.

2



**Exercise 6.4. RADAR SPEED INDICATOR**

A radar set measures velocity by emitting a signal at a standard frequency and comparing it with the frequency of the signal reflected back by another object. This redshift measurement is then converted, using the standard special-relativistic formula, into the corresponding velocity, and the radar reads out this velocity. How useful is this radar set as a velocity-measuring instrument for a uniformly accelerated observer?

(a) Consider this problem first for the special case where the object and the radar set are at rest with respect to each other at the instant the radar pulse is reflected. Compute the redshift $1 + z = \omega_e/\omega_0$ that the radar set measures in this case, and the resulting (incorrect) velocity it infers. Simplify by making use of the symmetries of the situation.

(b) Now consider the situation where the object has a non-zero velocity in the momentary rest frame of the observer at the instant it reflects the radar pulse. Compute the ratio of the actual relative velocity to the velocity read out by the radar set.

**Exercise 6.5. RADAR DISTANCE INDICATOR**

Use radar as a distance-measuring device. The radar set measures its proper time $\tau$ between the instant at which it emits a pulse and the later instant when it receives the reflected pulse. It then performs the simple computation $L_0 = \tau/2$ and supplies as output the "distance" $L_0$. How accurate is the output reading of the radar set for measuring the actual distance $L$ to the object, when used by a uniformly accelerated observer? ($L$ is defined as the distance in the momentary rest frame of the observer at the instant the pulse is reflected, which is at the observer's proper time halfway between emitting and receiving the pulse.) Give a correct formula relating $L_0 \equiv \tau/2$ to the actual distance $L$. Show that the reading $L_0$ becomes infinite as $L$ approaches $g^{-1}$, where $g$ is the observer's acceleration, as measured by an accelerometer he carries.

---

## §6.3. CONSTRAINTS ON SIZE OF AN ACCELERATED FRAME

Difficulties in constructing "the coordinate system of an accelerated observer":

It is very easy to put together the words "the coordinate system of an accelerated observer," but it is much harder to find a concept these words might refer to. The most useful first remark one can make about these words is that, if taken seriously, they are self-contradictory. The definite article "the" in this phrase suggests that one is thinking of some unique coordinate system naturally associated with some specified accelerated observer, such as one whose world line is given in equation (6.5). If the coordinate system is indeed natural, one would expect that the coordinates of any event could be determined by a sufficiently ingenious observer by sending and receiving light signals. But from Figure 6.1 it is clear that the events composing one quarter of all spacetime (Zone III) can neither send light signals to, nor receive light signals from, the specified observer. Another half of spacetime suffers lesser disabilities in this respect: Zone II cannot send to the observer, Zone IV cannot receive from him. It is hard to see how the observer could define in any natural way a coordinate system covering events with which he has no causal relationship, which he cannot see, and from which he cannot be seen!

Breakdown in communication between observer and events at distance $l >$ (acceleration)$^{-1}$

Difficulties also occur when one considers an observer who begins at rest in one frame, is accelerated for a time, and maintains thereafter a constant velocity, at rest in some other inertial coordinate system. Do his motions define in any natural way

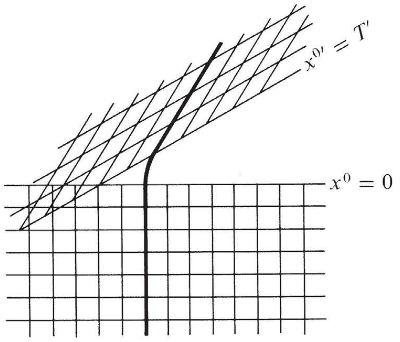



**Figure 6.2.**
World line of an observer who has undergone a brief period of acceleration. In each phase of motion at constant velocity, an inertial coordinate system can be set up. However, there is no way to reconcile these discordant coordinates in the region of overlap (beginning at distance $g^{-1}$ to the left of the region of acceleration).

a coordinate system? Then this coordinate system (1) should be the inertial frame $x^\mu$ in which he was at rest for times $x^0$ less than 0, and (2) should be the other inertial frame $x^{\mu'}$ for times $x^{0'} > T'$ in which he was at rest in that other frame. Evidently some further thinking would be required to decide how to define the coordinates in the regions not determined by these two conditions (Figure 6.2). More serious, however, is the fact that these two conditions are inconsistent for a region of spacetime that satisfies simultaneously $x^0 < 0$ and $x^{0'} > T'$. In both examples of accelerated motion (Figures 6.1 and 6.2), the serious difficulties about defining a coordinate system begin only at a finite distance $g^{-1}$ from the world line of the accelerated observer. The problem evidently has no solution for distances from the world line greater than $g^{-1}$. It does possess a natural solution in the immediate vicinity of the observer. This solution goes under the name of "Fermi-Walker transported orthonormal tetrad." The essential idea lends itself to simple illustration for hyperbolic motion, as follows.

Natural coordinates inconsistent at distance $l >$ (acceleration)$^{-1}$

## §6.4. THE TETRAD CARRIED BY A UNIFORMLY ACCELERATED OBSERVER

An infinitesimal version of a coordinate system is supplied by a "tetrad," or "moving frame" (Cartan's "repère mobile"), or set of basis vectors $\boldsymbol{e}_{0'}$, $\boldsymbol{e}_{1'}$, $\boldsymbol{e}_{2'}$, $\boldsymbol{e}_{3'}$ (subscript tells which vector, not which component of one vector!) Let the time axis be the time axis of a comoving inertial frame in which the observer is momentarily at rest. Thus the zeroth basis vector is identical with his 4-velocity: $\boldsymbol{e}_{0'} = \boldsymbol{u}$. The space axes $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$ are not affected by Lorentz transformations in the 1-direction. Therefore take $\boldsymbol{e}_{2'}$ and $\boldsymbol{e}_{3'}$ to be the unit basis vectors of the all-encompassing Lorentz frame relative to which the hyperbolic motion of the observer has already been described in equations (6.5): $\boldsymbol{e}_{2'} = \boldsymbol{e}_2$; $\boldsymbol{e}_{3'} = \boldsymbol{e}_3$. The remaining basis vector, $\boldsymbol{e}_{1'}$, orthogonal to the other three, is parallel to the acceleration vector, $\boldsymbol{e}_{1'} = g^{-1}\boldsymbol{a}$ [see equation (6.4)]. There is a more satisfactory way to characterize this moving frame: the time axis $\boldsymbol{e}_{0'}$ is the observer's 4-velocity, so he is always at rest in this frame; and the

Orthonormal tetrad of basis vectors carried by uniformly accelerated observer

2



other three vectors $\boldsymbol{e}_{1'}$ are chosen in such a way as to be (1) orthogonal and (2) nonrotating. These basis vectors are:

$$\begin{aligned}(e_{0'})^\mu &= (\cosh g\tau;\ \sinh g\tau,\ 0,\ 0);\\ (e_{1'})^\mu &= (\sinh g\tau;\ \cosh g\tau,\ 0,\ 0);\\ (e_{2'})^\mu &= (0;\ 0,\ 1,\ 0);\\ (e_{3'})^\mu &= (0;\ 0,\ 0,\ 1).\end{aligned} \tag{6.6}$$

There is a simple prescription to obtain these four basis vectors. Take the four basis vectors $\boldsymbol{e}_0$, $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ of the original global Lorentz reference frame, and apply to them a simple boost in the 1-direction, of such a magnitude that $\boldsymbol{e}_{0'}$ comes into coincidence with the 4-velocity of the observer. The fact that these vectors are all orthogonal to each other and of unit magnitude is formally stated by the equation

$$\boldsymbol{e}_{\mu'} \cdot \boldsymbol{e}_{\nu'} = \eta_{\mu'\nu'}. \tag{6.7}$$

## §6.5. THE TETRAD FERMI-WALKER TRANSPORTED BY AN OBSERVER WITH ARBITRARY ACCELERATION

Orthonormal tetrad of arbitrarily accelerated observer: should be "nonrotating"

Turn now from an observer, or an object, in hyperbolic motion to one whose acceleration, always finite, varies arbitrarily with time. Here also we impose three criteria on the moving, infinitesimal reference frame, or tetrad: (1) the basis vectors $\boldsymbol{e}_{\mu'}$ of the tetrad must remain orthonormal [equation (6.7)]; (2) the basis vectors must form a rest frame for the observer at each instant ($\boldsymbol{e}_{0'} = \boldsymbol{u}$); and (3) the tetrad should be "nonrotating."

This last criterion requires discussion. The basis vectors of the tetrad at any proper time $\tau$ must be related to the basis vectors $\boldsymbol{e}_0$, $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ of some given inertial frame by a Lorentz transformation $\boldsymbol{e}_{\mu'}(\tau) = \Lambda^\nu{}_{\mu'}(\tau)\boldsymbol{e}_\nu$. Therefore the basis vectors at two successive instants must also be related to each other by a Lorentz transformation. But a Lorentz transformation can be thought of as a "rotation" in spacetime. The 4-velocity $\boldsymbol{u}$, always of unit magnitude, changes in direction. The very concept of acceleration therefore implies "rotation" of velocity 4-vector. How then is the requirement of "no rotation" to be interpreted? Demand that the tetrad $\boldsymbol{e}_{\mu'}(\tau)$ change from instant to instant by precisely that amount implied by the rate of change of $\boldsymbol{u} = \boldsymbol{e}_{0'}$, and by no additional arbitrary rotation. In other words, (1) accept the inevitable pseudorotation in the timelike plane defined by the velocity 4-vector and the acceleration, but (2) rule out any ordinary rotation of the three space vectors.

"Nonrotating" means rotation only in timelike plane of 4-velocity and 4-acceleration

Mathematics of rotation in 3-space

Nonrelativistic physics describes the rotation of a vector (components $v_i$) by an instantaneous angular velocity vector (components $\omega_i$). This angular velocity appears in the formula for the rate of change of $\boldsymbol{v}$,

$$(dv_i/dt) = (\boldsymbol{\omega} \times \boldsymbol{v})_i = \varepsilon_{ijk}\omega_j v_k. \tag{6.8}$$

For the extension to four-dimensional spacetime, it is helpful to think of the rotation

2

as occurring in the plane perpendicular to the angular velocity vector $\boldsymbol{\omega}$. Thus rewrite (6.8) as

$$dv_i/dt = -\Omega_{ik}v_k, \tag{6.9}$$

where

$$\Omega_{jk} = -\Omega_{kj} = \omega_i \varepsilon_{ijk} \tag{6.10}$$

has non-zero components only in the plane of the rotation. In other words, to speak of "a rotation in the (1, 2)-plane" is more useful than to speak of a rotation about the 3-axis. The concept of "plane of rotation" carries over to four dimensions. There a rotation in the (1, 2)-plane will leave constant not only the $v_3$ but also the $v_0$ component of the velocity. The four-dimensional definition of a rotation is

Mathematics of rotation in spacetime

$$\frac{dv^\mu}{d\tau} = -\Omega^{\mu\nu}v_\nu, \qquad \text{with} \qquad \Omega^{\mu\nu} = -\Omega^{\nu\mu}. \tag{6.11}$$

To test the appropriateness of this definition of a generalized rotation or infinitesimal Lorentz transformation, verify that it leaves invariant the length of the 4-vector:

$$d(v_\mu v^\mu)/d\tau = 2v_\mu(dv^\mu/d\tau) = -2\Omega^{\mu\nu}v_\mu v_\nu = 0. \tag{6.12}$$

The last expression vanishes because $\Omega^{\mu\nu}$ is antisymmetric, whereas $v_\mu v_\nu$ is symmetric. Note also that the antisymmetric tensor $\Omega^{\mu\nu}$ ("rotation matrix"; "infinitesimal Lorentz transformation") has $4 \times 3/2 = 6$ independent components. This number agrees with the number of components in a finite Lorentz transformation (three parameters for rotations, plus three parameters for the components of a boost). The "infinitesimal Lorentz transformation" here must (1) generate the appropriate Lorentz transformation in the timelike plane spanned by the 4-velocity and the 4-acceleration, and (2) exclude a rotation in any other plane, in particular, in any spacelike plane. The unique answer to these requirements is

$$\Omega^{\mu\nu} = a^\mu u^\nu - a^\nu u^\mu; \qquad \text{i.e., } \boldsymbol{\Omega} = \boldsymbol{a} \wedge \boldsymbol{u}. \tag{6.13}$$

Apply this rotation to a spacelike vector $\boldsymbol{w}$ orthogonal to $\boldsymbol{u}$ and $\boldsymbol{a}$, ($\boldsymbol{u} \cdot \boldsymbol{w} = 0$ and $\boldsymbol{a} \cdot \boldsymbol{w} = 0$). Immediately compute $\Omega^{\mu\nu}w_\nu = 0$. Thus verify the absence of any space rotation. Now check the over-all normalization of $\Omega^{\mu\nu}$ in equation (6.13). Apply the infinitesimal Lorentz transformation to the velocity 4-vector $\boldsymbol{u}$ of the observer. Thus insert $v^\mu = u^\mu$ in (6.11). It then reads

$$du^\mu/d\tau \equiv a^\mu = u^\mu(a^\nu u_\nu) - a^\mu(u^\nu u_\nu) = a^\mu.$$

This result is an identity, since $\boldsymbol{u} \cdot \boldsymbol{u} = -1$ and $\boldsymbol{u} \cdot \boldsymbol{a} = 0$.

A vector $\boldsymbol{v}$ that undergoes the indicated infinitesimal Lorentz transformation,

Fermi-Walker law of transport for "nonrotating" tetrad of basis vectors carried by an accelerated observer

$$dv^\mu/d\tau = (u^\mu a^\nu - u^\nu a^\mu)v_\nu, \tag{6.14}$$

is said to experience "Fermi-Walker transport" along the world line of the observer.

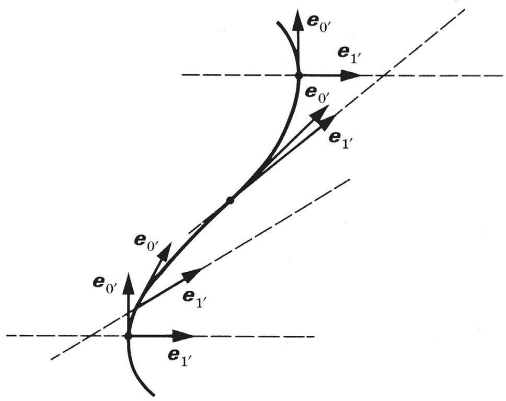



**Figure 6.3.**
Construction of spacelike hyperplanes (dashed) orthogonal to the world line (heavy line) of an accelerated particle at selected moments along that world line. Note crossing of hyperplanes at distance $g^{-1}(\tau)$ (time-dependent acceleration!) from the world line.

*The natural moving frame associated with an accelerated observer consists of four orthonormal vectors, each of which is Fermi-Walker transported along the world line and one of which is* $\boldsymbol{e}_{0'} = \boldsymbol{u}$ (the 4-velocity of the observer). Fermi-Walker transport of the space basis vectors $\boldsymbol{e}_{j'}$ can be achieved in practice by attaching them to gyroscopes (see Box 6.2 and exercise 6.9).

## §6.6. THE LOCAL COORDINATE SYSTEM OF AN ACCELERATED OBSERVER

Tetrad used to construct ''local coordinate system of accelerated observer''

Extend this moving frame or "infinitesimal coordinate system" to a "local coordinate system" covering a finite domain. Such local coordinates can escape none of the problems encountered in "hyperbolic motion" (Figure 6.1) and "briefly accelerated motion" (Figure 6.2). Therefore the local coordinate system has to be restricted to a region within a distance $g^{-1}$ of the observer, where these problems do not arise. Figure 6.3 illustrates the construction of the local coordinates $\xi^{\mu'}$. At any given proper time $\tau$ the observer sits at a specific event $\mathscr{P}(\tau)$ along his world line. Let the displacement vector, from the origin of the original inertial frame to his position $\mathscr{P}(\tau)$, be $\boldsymbol{z}(\tau)$. At $\mathscr{P}(\tau)$ the observer has three spacelike basis vectors $\boldsymbol{e}_{1'}(\tau)$, $\boldsymbol{e}_{2'}(\tau)$, $\boldsymbol{e}_{3'}(\tau)$. The point $\mathscr{P}(\tau)$ plus those basis vectors define a spacelike hyperplane. The typical point of this hyperplane can be represented in the form

$$\begin{aligned}\boldsymbol{x} &= \xi^{1'}\boldsymbol{e}_{1'}(\tau) + \xi^{2'}\boldsymbol{e}_{2'}(\tau) + \xi^{3'}\boldsymbol{e}_{3'}(\tau) + \boldsymbol{z}(\tau)\\ &= \text{(separation vector from origin of original inertial frame).}\end{aligned} \tag{6.15}$$

Here the three numbers $\xi^{k'}$ play the role of Euclidean coordinates in the hyperplane. This hyperplane advances as proper time unrolls. Eventually the hyperplane cuts through the event $\mathscr{P}_0$ to which it is desired to assign coordinates. *Assign to this event as coordinates the numbers* $\xi^{0'} = \tau$, $\xi^{k'}$ *given by (6.15). Call these four numbers* *"coordinates relative to the accelerated observer."* In detail, the prescription for the determination of these four coordinates consists of the four equations

$$x^\mu = \xi^{k'}[e_{k'}(\tau)]^\mu + z^\mu(\tau), \tag{6.16}$$

in which the $x^\mu$ are considered as known, and the coordinates $\tau$, $\xi^{k'}$ are considered unknowns.

At a certain distance from the accelerated world line, successive spacelike hyperplanes, instead of advancing with increasing $\tau$, will be retrogressing. At this distance, and at greater distances, the concept of "coordinates relative to the accelerated observer" becomes ambiguous and has to be abandoned. To evaluate this distance, note that any sufficiently short section of the world line can be approximated by a hyperbola ("hyperbolic motion with acceleration $g$"), where the time-dependent acceleration $g(\tau)$ is given by the equation $g^2 = a^\mu a_\mu$.

Local coordinate system for uniformly accelerated observer

Apply the above general prescription to hyperbolic motion, arriving at the equations

$$\begin{aligned}
x^0 &= (g^{-1} + \xi^{1'})\sinh(g\xi^{0'}),\\
x^1 &= (g^{-1} + \xi^{1'})\cosh(g\xi^{0'}),\\
x^2 &= \xi^{2'},\\
x^3 &= \xi^{3'}.
\end{aligned} \tag{6.17}$$

The surfaces of constant $\xi^{0'}$ are the hyperplanes with $x^0/x^1 = \tanh g\xi^{0'}$ sketched in Figure 6.4. Substitute expressions (6.17) into the Minkowski formula for the line element to find

$$\begin{aligned}
ds^2 &= \eta_{\mu\nu}\, dx^\mu\, dx^\nu\\
&= -(1 + g\xi^{1'})^2(d\xi^{0'})^2 + (d\xi^{1'})^2 + (d\xi^{2'})^2 + (d\xi^{3'})^2.
\end{aligned} \tag{6.18}$$

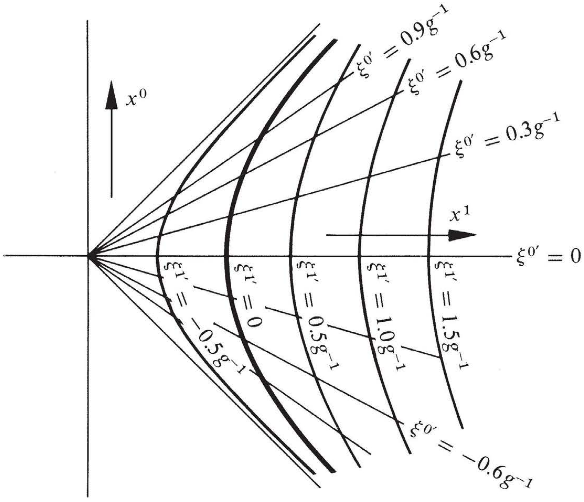



**Figure 6.4.**
Local coordinate system associated with an observer in hyperbolic motion (heavy black world line). The local coordinate system fails for $\xi^{1'}$ less than $-g^{-1}$.

The coefficients of $d\xi^{\mu'}\, d\xi^{\nu'}$ in this expansion are not the standard Lorentz metric components. The reason is clear. The $\xi^{\mu'}$ do not form an inertial coordinate system. However, at the position of the observer, $\xi^{1'} = 0$, the coefficients reduce to the standard form. Therefore these "local coordinates" approximate a Lorentz coordinate system in the immediate neighborhood of the observer.

---

## EXERCISES

**Exercise 6.6. CLOCK RATES VERSUS COORDINATE TIME IN ACCELERATED COORDINATES**

Let a clock be attached to each grid point, $(\xi^{1'}, \xi^{2'}, \xi^{3'})$ = constant, of the local coordinate system of an accelerated observer. Assume for simplicity that the observer is in hyperbolic motion. Use equation (6.18) to show that proper time as measured by a lattice clock differs from coordinate time at its lattice point:

$$d\tau/d\xi^{0'} = 1 + g\xi^{1'}.$$

(Of course, very near the observer, at $\xi^{1'} \ll g^{-1}$, the discrepancy is negligible.)

**Exercise 6.7. ACCELERATION OF LATTICE POINTS IN ACCELERATED COORDINATES**

Let an accelerometer be attached to each grid point of the local coordinates of an observer in hyperbolic motion. Calculate the magnitude of the acceleration measured by the accelerometer at $(\xi^{1'}, \xi^{2'}, \xi^{3'})$.

**Exercise 6.8. OBSERVER WITH ROTATING TETRAD**

An observer moving along an arbitrarily accelerated world line chooses *not* to Fermi-Walker transport his orthonormal tetrad. Instead, he allows it to rotate. The antisymmetric rotation tensor $\boldsymbol{\Omega}$ that enters into his transport law

$$d\boldsymbol{e}_{\alpha'}/d\tau = -\boldsymbol{\Omega}\cdot\boldsymbol{e}_{\alpha'} \tag{6.19}$$

splits into a Fermi-Walker part plus a spatial rotation part:

$$\Omega^{\mu\nu} = \underbrace{a^{\mu}u^{\nu} - a^{\nu}u^{\mu}}_{\Omega^{\mu\nu}_{(FW)}} + \underbrace{u_{\alpha}\omega_{\beta}\epsilon^{\alpha\beta\mu\nu}}_{\Omega^{\mu\nu}_{(SR)}} \tag{6.20}$$

$$\boldsymbol{\omega} = \text{a vector orthogonal to 4-velocity } \boldsymbol{u}.$$

(a) The observer chooses his time basis vector to be $\boldsymbol{e}_{0'} = \boldsymbol{u}$. Show that this choice is permitted by his transport law (6.19), (6.20).

(b) Show that $\Omega^{\mu\nu}_{(SR)}$ produces a rotation in the plane perpendicular to $\boldsymbol{u}$ and $\boldsymbol{\omega}$—i.e., that

$$\boldsymbol{\Omega}_{(SR)}\cdot\boldsymbol{u} = 0, \qquad \boldsymbol{\Omega}_{(SR)}\cdot\boldsymbol{\omega} = 0. \tag{6.21}$$

(c) Suppose the accelerated observer Fermi-Walker transports a second orthonormal tetrad $\boldsymbol{e}_{\alpha''}$. Show that the space vectors of his first tetrad rotate relative to those of his second tetrad with angular velocity vector equal to $\boldsymbol{\omega}$. *Hint:* At a moment when the tetrads coincide, show that (in three-dimensional notation, referring to the 3-space orthogonal to the observer's world line):

$$d(\boldsymbol{e}_{j'} - \boldsymbol{e}_{j''})/d\tau = \boldsymbol{\omega}\times\boldsymbol{e}_{j'}. \tag{6.22}$$

(d) The observer uses the same prescription [equation (6.16)] to set up local coordinates based on his rotating tetrad as for his Fermi-Walker tetrad. Pick an event $\mathscr{Q}$ on the observer's world line, set $\tau = 0$ there, and choose the original inertial frame of prescription (6.16) so (1) it comoves with the accelerated observer at $\mathscr{Q}$, (2) its origin is at $\mathscr{Q}$, and (3) its axes coincide with the accelerated axes at $\mathscr{Q}$. Show that these conditions translate into

$$z^{\mu}(0) = 0, \qquad \boldsymbol{e}_{\alpha'}(0) = \boldsymbol{e}_{\alpha}. \tag{6.23}$$

(e) Show that near $\mathscr{Q}$, equations (6.16) for the rotating, accelerated coordinates reduce to:

$$\begin{aligned} x^0 &= \xi^{0'} + a_{k'}\xi^{k'}\xi^{0'} + O([\xi^{\alpha'}]^3); \\ x^j &= \xi^{j'} + \frac{1}{2}a^j\xi^{0'2} + \epsilon^{jk\ell}\omega^k\xi^{\ell'}\xi^{0'} + O([\xi^{\alpha'}]^3). \end{aligned} \tag{6.24}$$

(f) A freely moving particle passes through the event $\mathscr{Q}$ with ordinary velocity $\boldsymbol{v}$ as measured in the inertial frame. By transforming its straight world line $x^j = v^j x^0$ to the accelerated, rotating coordinates, show that its coordinate velocity and acceleration there are:

$$\begin{aligned} (d\xi^{j'}/d\xi^{0'})_{\text{at }\mathscr{Q}} &= v^j; \\ (d^2\xi^{j'}/d\xi^{0'2})_{\text{at }\mathscr{Q}} &= \underbrace{-a^j}_{\substack{\text{inertial}\\ \text{acceleration}}} - \underbrace{2\epsilon^{jk\ell}\omega^k v^\ell}_{\substack{\text{Coriolis}\\ \text{acceleration}}} + \underbrace{2v^j a^k v^k}_{\substack{\text{relativisitc}\\ \text{correction to}\\ \text{inertial acceleration}}}. \end{aligned} \tag{6.25}$$

**Exercise 6.9. THOMAS PRECESSION**

Consider a spinning body (gyroscope, electron, . . .) that accelerates because forces act at its center of mass. Such forces produce no torque; so they leave the body's intrinsic angular-momentum vector $\boldsymbol{S}$ unchanged, except for the unique rotation in the $\boldsymbol{u} \wedge \boldsymbol{a}$ plane required to keep $\boldsymbol{S}$ orthogonal to the 4-velocity $\boldsymbol{u}$. Mathematically speaking, the body Fermi-Walker transports its angular momentum (no rotation in planes other than $\boldsymbol{u} \wedge \boldsymbol{a}$):

$$d\boldsymbol{S}/d\tau = (\boldsymbol{u} \wedge \boldsymbol{a}) \cdot \boldsymbol{S}. \tag{6.26}$$

This transport law applies to a spinning electron that moves in a circular orbit of radius $r$ around an atomic nucleus. As seen in the laboratory frame, the electron moves in the $x$, $y$-plane with constant angular velocity, $\omega$. At time $t = 0$, the electron is at $x = r$, $y = 0$; and its spin (as treated classically) has components

$$S^0 = 0, \qquad S^x = \frac{1}{\sqrt{2}}\hbar, \qquad S^y = 0, \qquad S^z = \frac{1}{2}\hbar.$$

*Calculate* the subsequent behavior of the spin as a function of laboratory time, $S^{\mu}(t)$. Answer:

$$\begin{aligned} S^x &= \frac{1}{\sqrt{2}}\hbar\,(\cos\omega t\cos\omega\gamma t + \gamma\sin\omega t\sin\omega\gamma t); \\ S^y &= \frac{1}{\sqrt{2}}\hbar\,(\sin\omega t\cos\omega\gamma t - \gamma\cos\omega t\sin\omega\gamma t); \\ S^z &= \frac{1}{2}\hbar; \qquad S^0 = -\frac{1}{\sqrt{2}}\hbar\,v\gamma\sin\omega\gamma t; \\ v &= \omega r; \qquad \gamma = (1 - v^2)^{-1/2}. \end{aligned} \tag{6.27}$$

Rewrite the time-dependent spatial part of this as

$$S^x + iS^y = \frac{\hbar}{\sqrt{2}}[e^{-i(\gamma-1)\omega t} + i(1-\gamma)\sin(\omega\gamma t)e^{i\omega t}]. \tag{6.28}$$

The first term rotates steadily in a retrograde direction with angular velocity

$$\begin{aligned}\omega_{\text{Thomas}} &= (\gamma - 1)\omega \\ &\approx \frac{1}{2}v^2\omega \ll \omega \text{ if } v \ll 1.\end{aligned} \tag{6.29}$$

It is called the Thomas precession. The second term rotates in a righthanded manner for part of an orbit ($0 < \omega\gamma t < \pi$) and in a lefthanded manner for the rest ($\pi < \omega\gamma t < 2\pi$). Averaged in time, it does nothing. Moreover, in an atom it is very small ($\gamma - 1 \ll 1$). It must be present, superimposed on the Thomas precession, in order to keep

$$\mathbf{S}\cdot\mathbf{u} = \boldsymbol{S}\cdot\boldsymbol{u} - S^0u^0 = 0, \tag{6.30}$$

and

$$\mathbf{S}^2 = \boldsymbol{S}^2 - (S^0)^2 = 3\hbar^2/4 = \text{constant}. \tag{6.31}$$

It comes into play with righthanded rotation when $\boldsymbol{S}\cdot\boldsymbol{u}$ is negative; it goes out of play when $\boldsymbol{S}\cdot\boldsymbol{u} = 0$; and it returns with lefthanded rotation when $\boldsymbol{S}\cdot\boldsymbol{u}$ turns positive.

The Thomas precession can be understood, alternatively, as a spatial rotation that results from the combination of successive boosts in slightly different directions. [See, e.g., exercise 103 of Taylor and Wheeler (1966).] For an alternative derivation of the Thomas precession (6.29) from "spinor formalism," see §41.4.

CHAPTER 7

# INCOMPATIBILITY OF GRAVITY AND SPECIAL RELATIVITY

## §7.1. ATTEMPTS TO INCORPORATE GRAVITY INTO SPECIAL RELATIVITY

This chapter is entirely Track 2.
It depends on no preceding Track-2 material.
It is not needed as preparation for any later chapter, but will be helpful in Chapter 18 (weak gravitational fields), and in Chapters 38 and 39 (experimental tests and other theories of gravity).

The discussion of special relativity so far has consistently assumed an absence of gravitational fields. Why must gravity be ignored in special relativity? This chapter describes the difficulties that gravitational fields cause in the foundations of special relativity. After meeting these difficulties, one can appreciate fully the curved-spacetime methods that Einstein introduced to overcome them.

Start, then, with what one already knows about gravity, Newton's formulation of its laws:

$$d^2x^i/dt^2 = -\partial\Phi/\partial x^i, \tag{7.1}$$

$$\nabla^2\Phi = 4\pi G\rho. \tag{7.2}$$

Newton's gravitational laws must be modified into four-dimensional, geometric form

These equations cannot be incorporated as they stand into special relativity. The equation of motion (7.1) for a particle is in three-dimensional rather than four-dimensional form; it requires modification into a four-dimensional vector equation for $d^2x^\mu/d\tau^2$. Likewise, the field equation (7.2) is not Lorentz-invariant, since the appearance of a three-dimensional Laplacian operator instead of a four-dimensional d'Alembertian operator means that the potential $\Phi$ responds instantaneously to changes in the density $\rho$ at arbitrarily large distances away. In brief, Newtonian gravitational fields propagate with infinite velocity.

All straightforward modifications are unsatisfactory

One's first reaction to these problems might be to think that they are relatively straightforward to correct. Exercises at the end of this section study some relatively straightforward generalizations of these equations, in which the gravitational potential $\Phi$ is taken to be first a scalar, then a vector, and finally a symmetric tensor field. Each of these theories has significant shortcomings, and all fail to agree with observations. The best of them is the tensor theory (exercise 7.3, Box 7.1), which, however,

2



Best modification (tensor theory in flat spacetime) is internally inconsistent; when repaired, it becomes general relativity.

is internally inconsistent and admits no exact solutions. This difficulty has been attacked in recent times by Gupta (1954, 1957, 1962), Kraichnan (1955), Thirring (1961), Feynman (1963), Weinberg (1965), Deser (1970). They show how the flat-space tensor theory may be modified within the spirit of present-day relativistic field theory to overcome these inconsistencies. By this field-theory route (part 5 of Box 17.2), they arrive uniquely at standard 1915 general relativity. Only at this end point does one finally recognize, from the mathematical form of the equations, that what ostensibly started out as a flat-space theory of gravity is really Einstein's theory, with gravitation being a manifestation of the curvature of spacetime. This book follows Einstein's line of reasoning because it keeps the physics to the fore.

---

## EXERCISES

### EXERCISES ON FLAT-SPACETIME THEORIES OF GRAVITY

The following three exercises provide a solid challenge. Happily, all three require similar techniques, and a solution to the most difficult one (exercise 7.3) is presented in Box 7.1. Therefore, it is reasonable to proceed as follows. (a) Work either exercise 7.1 (scalar theory of gravity) or 7.2 (vector theory of gravity), skimming exercise 7.3 and Box 7.1 (tensor theory of gravity) for outline and method, not for detail, whenever difficulties arise. (b) Become familiar with the results of the other exercise (7.2 or 7.1) by discussing it with someone who has worked it in detail. (c) Read in detail the solution to exercise 7.3 as presented in Box 7.1, and compare with the computed results for the other two theories. (d) Develop computational power by checking some detailed computations from Box 7.1.

**Exercise 7.1. SCALAR GRAVITATIONAL FIELD, $\Phi$**

A. Consider the variational principle $\delta I = 0$, where

$$I = -m \int e^{\Phi} \left( -\eta_{\alpha\beta} \frac{dz^{\alpha}}{d\lambda} \frac{dz^{\beta}}{d\lambda} \right)^{1/2} d\lambda, \tag{7.3}$$

Here $m$ = (rest mass) and $z^{\alpha}(\lambda)$ = (parametrized world line) for a test particle in the scalar gravitational field $\Phi$. By varying the particle's world line, derive differential equations governing the particle's motion. Write them using the particle's proper time as the path parameter,

$$d\tau = \left( -\eta_{\alpha\beta} \frac{dz^{\alpha}}{d\lambda} \frac{dz^{\beta}}{d\lambda} \right)^{1/2} d\lambda,$$

so that $u^{\alpha} = dz^{\alpha}/d\tau$ satisfies $u^{\alpha}u^{\beta}\eta_{\alpha\beta} = -1$.

B. Obtain the field equation for $\Phi(\boldsymbol{x})$ implied by the variational principle $\delta I = 0$, where $I = \int \mathcal{L}\, d^4x$ and

$$\mathcal{L} = -\frac{1}{8\pi G} \eta^{\alpha\beta} \frac{\partial \Phi}{\partial x^{\alpha}} \frac{\partial \Phi}{\partial x^{\beta}} - \int m e^{\Phi}\, \delta^4[\boldsymbol{x} - \boldsymbol{z}(\tau)]\, d\tau. \tag{7.4}$$

Show that the second term here gives the same integral as that studied in part A (equation 7.3).

*Discussion:* The field equations obtained describe how a single particle of mass $m$ generates the scalar field. If many particles are present, one includes in $\mathcal{L}$ a term $-\int m e^{\Phi}\, \delta^4[\boldsymbol{x} - \boldsymbol{z}(\tau)]\, d\tau$ for each particle.

C. Solve the field equation of part B, assuming a single source particle at rest. Also assume that $e^{\Phi} = 1$ is an adequate approximation in the neighborhood of the particle. Then check this assumption from your solution; i.e., what value does it assign to $e^{\Phi}$ at the surface of the earth? (Units with $c = 1$ are used throughout; one may also set $G = 1$, if one wishes.)

D. Now treat the static, spherically symmetric field $\Phi$ from part C as the field of the sun acting as a given external field in the variational principle of part A, and study the motion of a planet determined by this variational principle. Constants of motion are available from the spherical symmetry and time-independence of the integrand. Use spherical coordinates and assume motion in a plane. Derive a formula for the perihelion precession of a planet.

E. Pass to the limit of a zero rest-mass particle in the equations of motion of part A. Do this by using a parameter $\lambda$ different from proper time, so chosen that $k^\mu = dx^\mu/d\lambda$ is the energy-momentum vector, and by taking the limit $m \longrightarrow 0$ with $k^0 = \gamma m = E$ remaining finite (so $u^0 = \gamma \longrightarrow \infty$). Use these equations to show that the quantities $q^\mu = k^\mu e^\Phi$ are constants of motion, and from this deduce that there is no bending of light by the sun in this scalar theory.

**Exercise 7.2. VECTOR GRAVITATIONAL FIELD, $\Phi_\mu$**

A. Verify that the variational principle $\delta I = 0$ gives Maxwell's equations by varying $A_\mu$, and the Lorentz force law by varying $z^\mu(\tau)$, when

$$I = \frac{-1}{16\pi}\int F_{\mu\nu}F^{\mu\nu}\,d^4x + \frac{1}{2}m\int \frac{dz^\mu}{d\tau}\frac{dz_\mu}{d\tau}\,d\tau + e\int \frac{dz^\mu}{d\tau}A_\mu(z)\,d\tau. \tag{7.5}$$

Here $F_{\mu\nu}$ is an abbreviation for $A_{\nu,\mu} - A_{\mu,\nu}$. Hint: to vary $A_\mu(\boldsymbol{x})$, rewrite the last term as a spacetime integral by introducing a delta function $\delta^4[\boldsymbol{x} - \boldsymbol{z}(\tau)]$ as in exercise 7.1, parts A and B.

B. Define, by analogy to the above, a vector gravitational field $\Phi_\mu$ with $G_{\mu\nu} \equiv \Phi_{\nu,\mu} - \Phi_{\mu,\nu}$ using a variational principle with

$$I = +\frac{1}{16\pi G}\int G_{\mu\nu}G^{\mu\nu}\,d^4x + \frac{1}{2}m\int \frac{dz^\mu}{d\tau}\frac{dz_\mu}{d\tau}\,d\tau + m\int \Phi_\mu \frac{dz^\mu}{d\tau}\,d\tau. \tag{7.6}$$

(Note: if many particles are present, one must augment $I$ by terms $\frac{1}{2}m\int(dz^\mu/d\tau)(dz_\mu/d\tau)\,d\tau + m\int\Phi_\mu(dz^\mu/d\tau)\,d\tau$ for each particle.) Find the "Coulomb" law in this theory, and verify that the coefficients of the terms in the variational principle have been chosen reasonably.

C. Compute the perihelion precession in this theory.

D. Compute the bending of light in this theory (i.e., scattering of a highly relativistic particle $u^0 = \gamma \longrightarrow \infty$), as it passes by the sun, because of the sun's $\Phi_\mu$ field.

E. Obtain a formula for the total field energy corresponding to the Lagrangian implicit in part B. Use the standard method of Hamiltonian mechanics, with

$$I_{\text{field}} = \frac{1}{16\pi G}\int G_{\mu\nu}G^{\mu\nu}\,d^4x \equiv \int \mathcal{L}\,d^4x;$$

$\mathcal{L}$ is the Lagrangian density and $L \equiv \int\mathcal{L}\,d^3x$ is the Lagrangian. The corresponding Hamiltonian density ($\equiv$ energy density) is

$$\mathcal{H} = \sum_\mu \Phi_{\mu,0}\frac{\partial\mathcal{L}}{\partial\Phi_{\mu,0}} - \mathcal{L}.$$

Show that vector gravitational waves carry negative energy.

**Exercise 7.3. SYMMETRIC TENSOR GRAVITATIONAL FIELD, $h_{\mu\nu} = h_{\nu\mu}$**

Here the action principle is, as for the vector field, $\delta I = 0$, with $I = I_{\text{field}} + I_{\text{particle}} + I_{\text{interaction}}$. $I_{\text{particle}}$ is the same as for the vector field:

$$I_{\text{particle}} = \frac{1}{2}m\int \frac{dz^\mu}{d\tau}\frac{dz_\mu}{d\tau}\,d\tau. \tag{7.7}$$

2



However, $I_{\text{field}}$ and $I_{\text{interaction}}$ are different:

$$I_{\text{field}} = \int \mathcal{L}_f \, d^4x, \tag{7.8a}$$

$$\mathcal{L}_f = \frac{-1}{32\pi G}\left(\frac{1}{2}\, h_{\nu\beta,\alpha}\bar{h}^{\nu\beta,\alpha} - \bar{h}_{\mu\alpha}{}^{,\alpha}\bar{h}^{\mu\beta}{}_{,\beta}\right) \quad \left[\begin{array}{l}\text{Note that}\\ \text{one } h \text{ here}\\ \text{is not an } \bar{h}\end{array}\right], \tag{7.8b}$$

with

$$\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h^{\sigma}{}_{\sigma}; \tag{7.8c}$$

$$I_{\text{interaction}} = \frac{1}{2}\int h_{\mu\nu}T^{\mu\nu}\, d^4x. \tag{7.9}$$

Here $T^{\mu\nu}$ is the stress-energy tensor for all nongravitational fields and matter present. For a system of point particles (used throughout this exercise),

$$T^{\mu\nu}(\boldsymbol{x}) = \int m \frac{dz^{\mu}}{d\tau}\frac{dz^{\nu}}{d\tau}\,\delta^4[\boldsymbol{x} - \boldsymbol{z}(\tau)]\, d\tau. \tag{7.10}$$

A. Obtain the equations of motion of a particle by varying $z^{\mu}(\tau)$ in $\delta(I_{\text{particle}} + I_{\text{interaction}}) = 0$. Express your result in terms of the "gravitational force field"

$$\Gamma_{\nu\alpha\beta} = \frac{1}{2}(h_{\nu\alpha,\beta} + h_{\nu\beta,\alpha} - h_{\alpha\beta,\nu}) \tag{7.11}$$

derived from the tensor gravitational potentials $h_{\mu\nu} = h_{\nu\mu}$.

B. Obtain the field equations from $\delta(I_{\text{field}} + I_{\text{interaction}}) = 0$; express them in terms of

$$-H^{\mu\alpha\nu\beta} \equiv \bar{h}^{\mu\nu}\eta^{\alpha\beta} + \bar{h}^{\alpha\beta}\eta^{\mu\nu} - \bar{h}^{\alpha\nu}\eta^{\mu\beta} - \bar{h}^{\mu\beta}\eta^{\alpha\nu}. \tag{7.12}$$

Discuss gauge invariance, and the condition $\bar{h}^{\mu\alpha}{}_{,\alpha} = 0$.

C. Find the tensor gravitational potentials $h_{\mu\nu}$ due to the sun (treated as a point mass).

D. Compute the perihelion precession.

E. Compute the bending of light.

F. Consider a gravitational wave

$$\bar{h}^{\mu\nu} = A^{\mu\nu}\exp(ik_{\alpha}x^{\alpha}). \tag{7.13}$$

What conditions are imposed by the field equations? By the gauge condition

$$\bar{h}^{\mu\alpha}{}_{,\alpha} = 0? \tag{7.14}$$

Show that, by further gauge transformations

$$h_{\mu\nu} \longrightarrow h_{\mu\nu} + \xi_{\mu,\nu} + \xi_{\nu,\mu} \tag{7.15}$$

that preserve the $\bar{h}^{\mu\alpha}{}_{,\alpha} = 0$ restrictions, further conditions

$$u_{\alpha}\bar{h}^{\alpha\mu} = 0, \qquad \bar{h}^{\alpha}{}_{\alpha} = 0 \tag{7.16}$$

can be imposed, where $u^{\alpha}$ is a fixed, timelike vector. It is sufficient to consider the case, obtained by a suitable choice of reference frame, where $u^{\alpha} = (1; 0, 0, 0)$ and $k^{\alpha} = (\omega; 0, 0, \omega)$.

G. From the Hamiltonian density

$$\mathcal{H} \equiv \dot{h}_{\mu\nu}(\partial\mathcal{L}/\partial\dot{h}_{\mu\nu}) - \mathcal{L} \tag{7.17}$$

for the field, show that the energy density of the waves considered in part F is positive.

H. Compute $T^{\mu\nu}{}_{,\nu}$ for the stress-energy tensor of particles $T^{\mu\nu}$ that appears in the action integral $I$. Does $T^{\mu\nu}{}_{,\nu}$ vanish (e.g., for the earth in orbit around the sun)? Why? Show that the coupled equations for fields and particles obtained from $\delta I = 0$ have no solutions.

*(continued on page 187)*

2

### Box 7.1 AN ATTEMPT TO DESCRIBE GRAVITY BY A SYMMETRIC TENSOR FIELD IN FLAT SPACETIME [Solution to exercise 7.3]

Attempts to describe gravity within the framework of special relativity would naturally begin by considering the gravitational field to be a scalar (exercise 7.1) as it is in Newtonian theory, or a vector (exercise 7.2) by analogy to electromagnetism. Only after these are found to be deficient (e.g., no bending of light in either theory; negative-energy waves in the vector theory) would one face the computational complexities of a symmetric tensor gravitational potential, $h_{\mu\nu} = h_{\nu\mu}$, which has more indices.

The foundations of the most satisfactory of all tensor theories of gravity in flat spacetime are laid out at the beginning of exercise 7.3. The choice of the Lagrangian made there (equations 7.8) is dictated by the demand that $h_{\mu\nu}$ be a "Lorentz covariant, massless, spin-two field." The meaning of this demand, and the techniques of special relativity required to translate it into a set of field equations, are customarily found in books on elementary particle physics or quantum field theory; see, e.g., Wentzel (1949), Feynman (1963), or Gasiorowicz (1966). Fierz and Pauli (1939) were the first to write down this Lagrangian and investigate the resulting theory. The conclusions of the theory are spelled out here in the form of a solution to exercise 7.3.

### A. Equation of Motion for a Test Particle (exercise 7.3A)

Carry out the integration in equation (7.9), using the particle stress-energy tensor of equation (7.10), to find

$$I_{p+i} \equiv I_{\text{particle}} + I_{\text{interaction}} = \frac{1}{2} m \int (\eta_{\mu\nu} + h_{\mu\nu}) \dot{z}^\mu \dot{z}^\nu \, d\tau, \tag{1}$$

where

$$\dot{z}^\mu \equiv dz^\mu/d\tau.$$

Then compute $\delta I_{p+i}$, and find that the coefficient of the arbitrary variation in path $\delta z^\mu$ vanishes if and only if

$$(d/d\tau)[(\eta_{\mu\nu} + h_{\mu\nu})\dot{z}^\nu] - \frac{1}{2} h_{\alpha\beta,\mu} \dot{z}^\alpha \dot{z}^\beta = 0.$$

Rewrite this equation of motion in the form

$$(\eta_{\mu\nu} + h_{\mu\nu})\ddot{z}^\nu + \Gamma_{\mu\alpha\beta} \dot{z}^\alpha \dot{z}^\beta = 0, \tag{2}$$

where $\Gamma_{\mu\alpha\beta}$ is defined in equation (7.11).

### $B_1$. Field Equations (exercise 7.3B)

Use $I_{\text{field}}$ and $I_{\text{interaction}}$ in the forms given in equations (7.8) and (7.9); but for the quickest and least messy derivation, do *not* use the standard Euler-Lagrange equations. Instead, compute directly the first-order change $\delta \mathcal{L}_f$ produced by a small

**Box 7.1 (*continued*)**

variation $\delta h_{\alpha\beta}$ of the field. For the second term of $\mathcal{L}_f$, it is clear (by relabeling dummy indices as needed) that varying each factor gives the same result, so the two terms from the product rule combine:

$$\delta(\bar{h}_{\mu\alpha}{}^{,\alpha}\,\bar{h}^{\mu\beta}{}_{,\beta}) = 2\bar{h}^{\mu\beta}{}_{,\beta}\,\delta\bar{h}_{\mu\alpha}{}^{,\alpha}.$$

A similar result holds for the first term of $\mathcal{L}_f$, in view of the identity $a_{\mu\nu}\bar{b}^{\mu\nu} = \bar{a}_{\mu\nu}b^{\mu\nu}$, which holds for the "bar" operation of equations (7.8); each side here is just $a_{\mu\nu}b^{\mu\nu} - \frac{1}{2}a^{\mu}{}_{\mu}b^{\nu}{}_{\nu}$. Consequently,

$$-(32\pi G)\,\delta\mathcal{L}_f = \bar{h}^{\nu\beta,\alpha}\,\delta h_{\nu\beta,\alpha} - 2\bar{h}^{\mu\beta}{}_{,\beta}\,\delta\bar{h}_{\mu\alpha}{}^{,\alpha}. \tag{3}$$

Next use this expression in $\delta I_{\text{field}}$; and, by an integration by parts, remove the derivatives from $\delta h_{\mu\nu}$, giving

$$\delta I_{\text{field}} = (32\pi G)^{-1}\int [\bar{h}^{\nu\beta,\alpha}{}_{,\alpha}\,\delta h_{\nu\beta} - 2\bar{h}^{\mu\beta}{}_{,\beta}{}^{,\alpha}\,\delta\bar{h}_{\mu\alpha}]\,d^4x.$$

To find the coefficient of $\delta h_{\mu\nu}$ in this expression, write (from equation 7.8c)

$$\delta\bar{h}_{\alpha\beta} = (\delta^{\mu}{}_{\alpha}\,\delta_{\nu\beta} - \frac{1}{2}\eta_{\alpha\beta}\eta^{\mu\nu})\,\delta h_{\mu\nu};$$

and then rearrange and relabel dummy (summation) indices to obtain

$$\delta I_{\text{field}} = (32\pi G)^{-1}\int [\bar{h}^{\mu\beta,\alpha}{}_{,\alpha}\,\delta h_{\nu\beta} - 2\bar{h}^{\mu\beta}{}_{,\beta}{}^{,\alpha}\,\delta\bar{h}_{\mu\alpha}]\,d^4x.$$

By combining this with $\delta I_{\text{interaction}} = \frac{1}{2}T^{\mu\nu}\,\delta h_{\mu\nu}\,d^4x$, and by using the symmetry $\delta h_{\mu\nu} = \delta h_{\nu\mu}$, obtain

$$-\bar{h}^{\mu\nu}{}_{,\alpha}{}^{,\alpha} - \eta^{\mu\nu}\bar{h}^{\alpha\beta}{}_{,\alpha\beta} + \bar{h}^{\mu\alpha}{}_{,\alpha}{}^{,\nu} + \bar{h}^{\nu\alpha}{}_{,\alpha}{}^{,\mu} = 16\pi G T^{\mu\nu}. \tag{4}$$

The definition made in equation (7.12) allows this to be rewritten as

$$H^{\mu\alpha\nu\beta}{}_{,\alpha\beta} = 16\pi G T^{\mu\nu}. \tag{4$'$}$$

### $B_2$. Gauge Invariance (exercise 7.3B, continued)

The symmetries,

$$H^{\mu\alpha\nu\beta} = H^{[\mu\alpha][\nu\beta]} = H^{\nu\beta\mu\alpha},$$

of $H^{\mu\alpha\nu\beta}$ imply an identity

$$H^{\mu\alpha\nu\beta}{}_{,\alpha\beta\nu} = H^{\mu\alpha[\nu\beta]}{}_{,\alpha(\beta\nu)} \equiv 0$$

analogous to $F^{\mu\nu}{}_{,\nu\mu} \equiv 0$ in electromagnetism.

Thus $T^{\mu\nu}{}_{,\nu} = 0$ is required of the sources, just as is $J^{\mu}{}_{,\mu} = 0$ in electromagnetism (exercise 3.16). These identities make the field equations (4′) too weak to fix $h_{\mu\nu}$