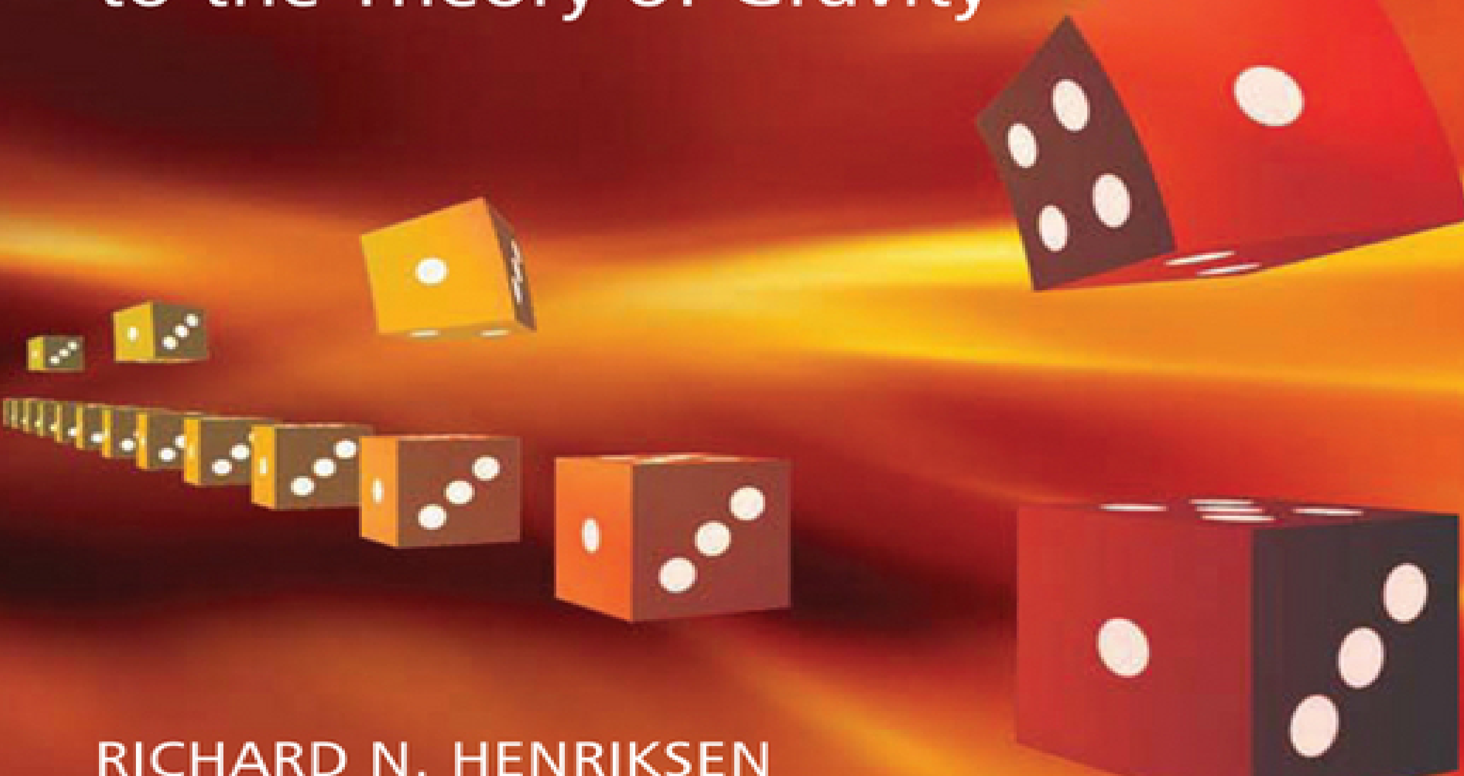
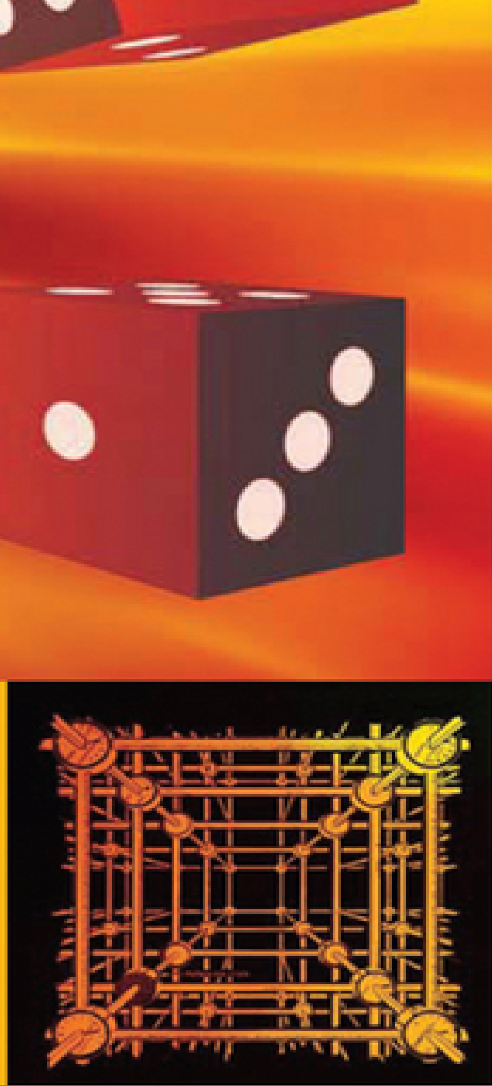
Practical Relativity

From First Principles to the Theory of Gravity



RICHARD N. HENRIKSEN

WILEY



Contents

Practical Relativity

From First Principles to the Theory of Gravity

RICHARD N. HENRIKSEN
Queen's University, Kingston, Ontario

WILEY
A John Wiley and Sons, Ltd., Publication

This edition first published 2011

© 2011 John Wiley & Sons Ltd

Registered office

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of fitness for a particular purpose. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for every situation. In view of ongoing research, equipment modifications, changes in governmental regulations, and the constant flow of information relating to the use of experimental reagents, equipment, and devices, the reader is urged to review and evaluate the information provided in the package insert or instructions for each chemical, piece of equipment, reagent, or device for, among other things, any changes in the instructions or indication of usage and for added warnings and precautions. The fact that an organization or Website is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or Website may provide or recommendations it may make. Further, readers should be aware that Internet Websites listed in this work may have changed or disappeared between when this work was written and when it is read. No warranty may be created or extended by any promotional statements for this work. Neither the publisher nor the author shall be liable for any damages arising herefrom.

Library of Congress Cataloguing-in-Publication Data

Henriksen, R. N.

Practical relativity : from first principles to the theory of gravity / Richard N. Henriksen.

p. cm.

Includes bibliographical references and index.

ISBN 978-0-470-74142-9 (cloth) – ISBN 978-0-470-74141-2 (pbk.) 1. Relativity (Physics) I. Title.

QC173.58.H36 2010

530.11 – dc22

2010023306 A catalogue record for this book is available from the British Library

ISBN: 978-0-470-74142-9 (H/B) 978-0-470-74141-2 (P/B)

Typeset in 10/12 Times by Laserwords Private Limited, Chennai, India

Printed in Singapore by Markono Print Media Pte Ltd

Cover images: Dice image from *Sights That Einstein Could Not Yet See – visualization of relativistic effects.* Ute Kraus, Hanns Ruder, Daniel Weiskopf, Corvin Zahn. *Latticework and Clocks* from Spacetime Physics, 2e by E.F Taylor and J. A. Wheeler © 1992 by Edwin Taylor and John Archibald Wheeler, used with permission.

To the memory of Gertrude and Norman, Emma and Halter;
and to Judith who is life’s force incarnate.

Preface

This book is entitled *Practical Relativity*. Many of you (I hope that there will be many) will wonder why, once confronted with the dense forest of equations. I can only say that in places I have used words to explain conceptual points. However I think that what makes it 'practical' rests primarily on two other things. I have started at the beginning of the subject and gone nearly to the end. Moreover, insofar as I have been able to navigate between tedium and necessity, I have included all of the steps that lead to important results. This is, I think, a characteristic of 'practicality'. Both of these themes should be appreciated by serious students. Problems are included that elucidate the ideas, and these should be appreciated by professors. A solutions manual, containing answers to the problems, is available at www.wiley.com/go/henriksen.

My approach has been to regard fundamental principles 'eye to eye', so that any possible alternatives to the traditional arguments may become evident. Most derivations are from first principles. I have not cluttered the book with every possible application of the theory, but the grand classics are present. I believe, however, and I hope that you will agree, that the presentation of the necessary techniques has been comprehensive.

I have not used the latest mathematical treatments of vectors and tensors, as found for example in the Cartan calculus. My approach has been to remain as close as possible to familiar concepts of vectors, tensors and reference systems in the hope of capitalizing on received wisdom. I believe that this is another practical aspect.

There are many books on this subject, and in the course of my writing I have enjoyed reading many of them. They are referred to throughout the book. I do believe that the present book is not quite the same as the others, mainly due to the attempt to distinguish the positivist approach from the theoretical. While one measures, the other imagines although in the end the loop must be closed. I have also attempted to cast light on dark corners. I have enjoyed exploring the corners, and I hope that this book will also help you to explore and enjoy them.

“Why, I’d like nothing better than to achieve some bold adventure worthy of our trip.”

Aristophanes, 450-385 BC

Acknowledgements

This book would not have been written without the opportunity to teach these subjects to many bright students, primarily at Queen's University. It could not have been written without the tradition of creative research that should be the pride of Queen's University. Colleagues set the standard, and I have been fortunate with these. The errors are of course mine. The inspiration is due to my wife Judith Irwin, who is a hard but loving taskmaster. Much credit is due to Wiley for overseeing the production process, as well as to Laserwords for laying out the pages elegantly.

RNH

Introduction

This book is written in six long chapters. The intention was to make each chapter a logical step on the way to relativistic electromagnetism and gravity, subjects that are the province of the last two chapters.

The first chapter starts from simple considerations of reference frames and vectors. The positivist attitude is emphasized. It is written entirely in the physical context of classical (Newtonian) space-time and mechanics. However, the notion of coordinate independence of the physical description leads inexorably to the apparatus of differential geometry. This is done deliberately, so that the reader will become accustomed to the formalism of relativity in an intuitive geometrical context.

The mystery of inertial frames is discussed at length in this chapter, with some connection to modern ideas. The discussion includes non-inertial frames, and the transformations between them. This leads to the introduction of time in the coordinate transformations and to a brief discussion of absolute time. Rotation matrices and angular velocity matrices are used to write Newton's second law in accelerated frames of reference. Contact is made with other, considerably less explicit, notation. Finally we emphasize that the necessity to define parallel transport of a vector already exists in Euclidian curvilinear coordinates. This is presented in a familiar (if awkward) notation, so that it is readily recognized later. This chapter assumes a familiarity with classical ideas at the level of advanced mechanics and neither is, nor was meant to be, gentle. It may be best to study it selectively.

The second chapter is devoted to the derivation of the Lorentz transformations in two distinct ways. The first method concentrates on the derivation of the Lorentz transformations as those transformations of space and time that leave the wave equation for light invariant. Considerable discussion is devoted to finding what the results would be, if other equations were taken as the source of the fundamental invariance. The essential step of allowing a time transformation, rather than insisting on absolute time, is shown to distinguish these transformations from earlier versions by Voigt and Poincaré.

After this derivation the question arises as to why such an invariance group should apply to all events in space-time. This question is answered by the operational or positivist derivation of the transformations first given by Einstein. We give a version that is based on light-clocks and the transformation of straight lines in a space-time diagram. We argue that such a linear transformation must be accompanied by a ‘units transformation’. These scale factors are the usual, time dilation and Lorentz-Fitzgerald contraction. Putting these two concepts together yields the Lorentz transformations. Because of the maximal and invariant nature of the speed of light that is assumed ‘a priori’ in this approach, it is really a theory of ideal measurement. So long as what one can measure is reality, the implications of the transformations are ‘real’.

The third chapter details many of the usual applications of the Lorentz transformations, together with some discussion that is perhaps rarer. Time dilation, the Döppler shift and the twin paradox start off the chapter. There are some astronomical applications. Time on a rotating disc is examined in the context of the Sagnac effect. The Lorentz-Fitzgerald contraction is discussed largely in terms of standard paradoxes, but once again the rotating disc is found to be instructive. Some gentle speculation is allowed here, since the topic has a history of errors. The velocity transformation is introduced and used to define the phenomenon of aberration of beams of relativistic particles. The limit is taken for photons and so the inherent transformation of angles appears, which is optical aberration.

Under the heading of geometrical optics we discuss such topics as Thomas precession and the appearance of moving objects. The derivations are not the most elegant possible! However, they do have the merit of revealing the essential unexpected phenomenon in the homogeneous Poincaré group. The astronomical phenomenon of 'light echos' is also introduced and then argued to be important using examples. A final topic in the chapter is dynamics with prescribed acceleration. This requires the transformation of particle acceleration between inertial observers. Hyperbolic motion is presented as an example of the horizon phenomenon.

In the fourth chapter we introduce Minkowski space-time and adapt all of the results of Chapter 1 regarding vectors and tensors to the four dimensions of space and time. At this stage we emphasize that it is not obligatory to conceive of space-time as a Minkowskian manifold, but that it is terribly convenient. We demonstrate this by rederiving the Lorentz transformations based on this principle in two ways. We assume first that metric moduli are invariant, and then that the metric itself should be invariant after synchronization. We also show that the four-vector treatment of velocity and acceleration allows previous results on their three-vector transformation properties to be readily obtained. These discussions serve principally to demonstrate the internal consistency of the Minkowskian metric space.

In the absence of real forces, we introduce the Lagrangian and Hamiltonian for a free particle and infer the momentum and energy. We show how one may impose constraints on the motion of a free particle to approximate relativistic forces. This is all done in generalized coordinates as well as Galilean coordinates. After deriving the action for a free particle, we observe that the Euler-Lagrange equations are equivalent to geodesic equations in the given metric. This leads to the equation of motion of a free particle that holds in any pure metric theory. Finally, the collisions of free particles are treated in terms of the conservation laws. The principles are extended to photons and applied to Compton and inverse-Compton scattering.

The fifth chapter is technically more challenging, but perhaps also more practical. The four-vector electromagnetic theory is presented in Galilean coordinates. Then in the traditional ‘three plus one’ split into space and time, Lagrangian and Hamiltonian methods for solving particle motion are introduced with examples. Many of the examples are important classics and some of them are solved in several different ways in order to elucidate the methods. The Lorentz equation of motion and the principle of relativity are used together to infer the transformation of electric and magnetic fields in an elementary way. One sees that these vectors are part of a larger object.

Next the three plus one split is abandoned, and Galilean four-vectors are used exclusively. This leads to the field tensor, electromagnetic field invariants and the tensor form of the field equations. The latter result requires, in part, a discussion of the action that holds for the matter coupled to the fields, when the vector potential is varied.

As a means of transiting to gravitational metrics, the Maxwell equations are generalized to metrics for which all components are in principle functions of space and time. Such dependence includes curvilinear coordinates and non-inertial coordinates, but the metric can also reflect a curved manifold. Finally, in this section, the Landau and Lifshitz approach based on a (locally) diagonalized metric is used to write the Maxwell equations in a recognizable form. This material is rather advanced and can be omitted without subsequent damage. It does, however, represent a useful exercise in the use of vectors, tensors and their duals. The final form of the equations can be used to discuss electromagnetism near rotating black holes and neutron stars.

Part II of the book deals with the implications of metrics that describe various gravitational fields. It is contained in one long Chapter 6. The chapter has a long prologue that is meant to introduce qualitatively the nature of the metric theory and its uses. The reader is free to pass on and let these speak for themselves.

The first major section explores the metric representation of a weak gravitational field. This is where the contact with Newtonian theory is established. The gravitational and cosmological frequency shifts are discussed in this section in order to form a complete set of such shifts, but their general nature is emphasized. Later the gravitational frequency shift is given a more general treatment. Simple tests are discussed briefly, with emphasis on the GPS system.

The next section deals with the general form of static and stationary metrics. The Lagrange equations are used to find the Christoffel symbols when the metric is spherically symmetric. The Hamilton-Jacobi and Eikonal equations are introduced for general metrics. These are used to discuss the energy of a particle and the proper frequency of a photon.

The third section presents the metrics for two of the best known and important strong gravitational fields, due to Schwarzschild and Kerr. Each metric has its own subsection. The nature of the Schwarzschild horizon is clarified by introducing freely-falling (inertial) observers, following LeMaître. We see that this field of inertial observers is completely determined by the metric in Schwarzschild form. The classic calculations of orbital precession and light deflection are given in detail in two independent approaches.

The Kerr metric is less manageable, but we find the meaning of its horizon and ergosphere. Frame dragging and energy extraction are discussed, as is the upper limit to the specific angular momentum.

In the last two sections we give first the conservation laws of matter as the true divergence of the energy (density)-momentum (flux) tensor. This is then used in the discussion of the matter sources of the gravitational metrics.

Following Gauss and Riemann, the curvature is identified as the distinction between what is merely the Minkowski metric in generalized coordinates and the gravitational metrics. This leads to defining the Riemann, Ricci and Einstein tensors. The Riemann curvature tensor is shown to be equivalent to the non-commutation of the second-order true derivatives. The Einstein equations are given and the Bianchi identities are shown to be essential to the conservation laws of matter. Finally, a brief discussion of the modification necessary to include the cosmological constant (or vacuum energy density from another point of view) is given.

No detailed calculations using the Einstein equations are presented. These are left to other texts, although in principle the reader has the techniques with which to launch himself into the heart of this grand subject.

Part I

The World Without Gravity

1

Non-Relativity for Relativists

Dura lex, sed lex (The law is hard but it is the law)

1.1 Vectors and Reference Frames

In this section we discuss our fundamental concepts as drawn from experience. This ends in frustration since experience is approximate, most things are known relative to other things, and our concepts often seem to be defined in terms of themselves. Thus 'fundamental' argument resembles the circular snake devouring its tail (the *Ouroboros*). However we must make a beginning, and so we confront our first definition and its algebraic implications.

What is an inertial reference frame? I prefer to parse this question into two principal questions. By 'reference frame' we mean some well-defined system of assigning a measured time and a measured position to an 'event'. For the moment an 'event' is point-like, as for example the time at which a particle or the centre of mass of an extended body takes a particular spatial position. The reference frame also implies an 'observer' who records the measurements. The resulting numbers are the 'coordinates' of the event in this reference frame. By 'inertial' we mean a reference frame in which Newton's first law of motion[1] applies to sufficiently isolated bodies. This axiom requires not only that the coordinates of a body be determinable from moment to moment, but also that fixed spatial directions be defined.

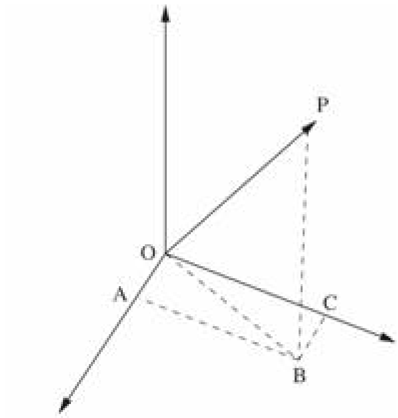
Neither one of these definitions is particularly exact or obvious and yet they are fundamental to our subject. Thus we continue their exploration in the next two sections.

1.1.1 Reference Frames

Although this is not strictly necessary, location is normally specified relative to a set of objects that have no relative motion between them. Some fixed point within this set of objects is chosen as the reference point or 'origin' from which all distances are measured. On small enough scales that we can reach continuously, the measurement is made by placing a standard length along a straight line between the points of interest. We call this standard length a 'ruler' or a 'unit' and we assume that we can determine a 'straight line'. On larger scales, various more subtle methods are required.

Our most familiar example is the Earth itself. On small scales we have no difficulty in establishing a rigid frame of reference by assuming Euclidean geometry. That is, we assume that the Earth is 'flat' so that trigonometry and an accurate ruler suffice to measure distance. When lasers are used we are assuming that even the near space above the surface of the Earth is Euclidean and that light follows the straight lines. On larger scales the Earth is found to be a sphere, so that its surface does not obey Euclidean geometry. Position has to be assigned by latitude and longitude, which requires the use of a combination of accurate clocks and astronomical observations in the measurements. Distance is computed between points using the rules of spherical trigonometry, rather than the Euclidean rule of Pythagoras[2] (see e.g. Figure 1.1).

Figure 1.1 *The three perpendicular axes emanating from O are reference directions. Each axis is rigid and the projections of OP on these axes furnish the Cartesian* *weights or components. The theorem of Pythagoras gives the distance OP in terms of these*



$$OB^2 = AB^2 + OA^2$$
$$OP^2 = OB^2 + BP^2$$
$$OP^2 = AB^2 + OA^2 + BP^2$$

The Earth is not exactly a rigid sphere, but a global reference frame precise enough to detect this fact became generally available only with the advent of the Global Positioning System (GPS) of satellites. This remarkable development, based on multiple one-way radar ranging, has allowed us to measure the ebb and flow of oceans and continents in a non-rigid, spheroidal global frame. However, it assumes principles that we have yet to examine, and that will be the subject of much of this book.

Thus the procedure to define a ‘rigid’ frame of spatial reference always involves assumptions about the nature of the world around us, and it is these that we must carefully examine subsequently. Moreover such a reference frame is always an idealization. Errors are involved in determining practical spatial coordinates on every scale, so that our knowledge of distance is always approximate. Moreover the degree of idealization increases with spatial extent of the reference frame, as it becomes progressively more difficult to maintain rigidity.

In parallel with spatial position, we have managed recently to establish a global measure of time that allows us to say whether or not events occurred simultaneously. This means that a single number can be assigned to a global point-like event (e.g. the onset of an earthquake in China or sunrise at Stonehenge on Midsummer's Day). The number is assigned by each of a network of synchronized atomic clocks distributed over the reference frame of the Earth. The sequence of such numbers defines 'coordinate time' for the Terrestrial Reference Frame. The difference between such numbers that encompass the beginning and end of an extended event (such as a lifetime) may be called a 'duration' for brevity. In practice, only durations of finite length are meaningful since no measurement can be made with infinite precision, but we normally assume that they can be arbitrarily small in principle. Figure 1.2 shows an ideal rigid reference frame with synchronized clocks at each spatial point.

The creation of a terrestrial coordinate time has been accomplished through the global synchronization of atomic clocks (within limits) rather than by astronomical measures such as day count and Sun angle. The latter does not establish a global reference time as any 'jet-lagged' traveller knows well! Once again this global clock synchronization involves principles and corrections that we have yet to discuss, but which will be one of our principal preoccupations.

Our direct experience of time tends, however, to be local rather than global. It is an event that includes oneself whose duration is measured by our clock, our schedule, our heart beat or our ageing process. Such local time is proper to us and is generally referred to as 'proper time'. The 'origin' of either coordinate or proper time is just as arbitrary as is the spatial origin, and may be chosen for convenience.

There are many reasons, however, why proper time does not run at the same rate as coordinate time. These reasons are physical as well as psychological. One physical reason is that our bodies age according to a thermodynamic time measured by increasing entropy, and the rate is different for different individuals. Another is the differing set of inertial frames that an individual occupies relative to the terrestrial reference frame. This unexpected dependence we shall explore in subsequent sections. The psychology of time is not within the competence of this author, but ‘apparent’ proper time is notoriously variable!

The complications involved in defining reference frames have been elegantly revealed by our exploration of the solar system. The planets do not form a rigid system of reference. A global reference frame on Mars moves relative to a global reference frame on the Earth, so that a rigid reference frame encompassing the two is not possible. One solution is to construct an imaginary rigid frame whose origin is at the centre of the Sun. The three independent directions required to encompass all space in the Cartesian fashion are not fixed in the Sun, which is not rigid either, but rather with respect to very distant objects in the Universe (such as quasars) that appear fixed to us. Coordinates determined along these directions are useful to determine the momentary position of the centres of mass of the planets. Ultimately, however, we are forced to have recourse to systems proper to each planet, such as latitude and longitude for the Earth, and these are neither fixed nor constantly oriented with respect to the Cartesian reference axes.

Time measurements in the solar system have also revealed difficulties with a panplanet coordinate time. For example, assuming nothing faster than our electromagnetic signals, Martian events happen later for an Earth observer than they do for a Martian observer such as a Martian Rover Vehicle (and vice versa for Earth events observed on Mars, such as the initiation of a command signal on the Earth). Electromagnetic signals propagate in a vacuum with the speed of ‘light’, which is almost universally labelled as c and which has the approximate numerical value 0.2998 metres per nanosecond (we know it to much greater accuracy). Thus although we can experience a Martian duration delayed by the travel time of our signals (and slightly distorted due to the motion of Mars relative to the Earth), we cannot share proper times. Moreover there can be no electromagnetic connection between the Earth and Mars during this travel time.

There is, then, since at present c is the fastest signal we know, a causality gap wherein nothing on Earth can affect Mars and vice versa. This a-causal gap varies roughly from 4 minutes to a little less than 12 minutes depending on the relative positions of Earth and Mars. We have met such an effect previously when astronauts were on the Moon, but the gap was only of the order of two seconds. Our intercontinental calls by way of satellites in synchronous orbit have an a-causal gap roughly equal to a third of a second, which is barely noticeable in conversation.

One might think that by using atomic clocks synchronized on Earth and Mars we could agree on simultaneous events after the fact at least, and so establish a pan-planet coordinate time, which would be ‘absolute’ in the solar system. However, we shall see that even the most perfect atomic clocks cannot remain synchronized in the presence of relative velocity between reference frames, provided that signals of only finite speed are available to us.

The sort of reference frame that we can construct at the centre of the Sun is composed of an inferred origin plus geometric straight lines and it has no proper ‘observer’. Time and space in this frame are constructed from events measured by atomic clocks and observers located elsewhere, after correction to the solar origin. These corrections are an example of a general mapping from local coordinates to ‘generalized’ coordinates at the centre of the Sun. Such mappings will be discussed in greater detail below. Although useful as fictitious standards and widely used in the theory of gravity, these virtual reference frames are distinct from a tangible reference frame that is inhabited by 'observers' capable of measuring the coordinates of events directly.

The conclusion to this discussion so far may be summarized algebraically by stating that a reference frame allows an observer to assign coordinates to point-like events according to

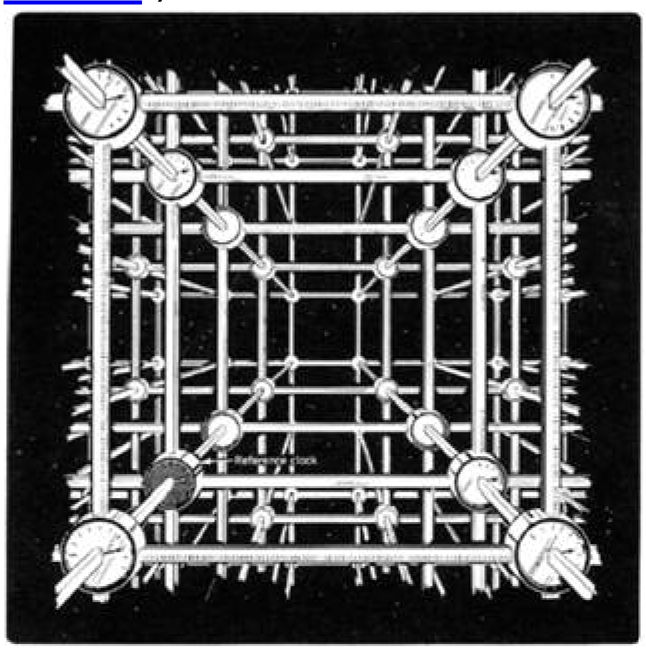
$$\{x^a\} \equiv \begin{pmatrix} t \\ q^1 \\ q^2 \\ q^3 \end{pmatrix}. \quad (1.1)$$

The notation on the left indicates a set of four quantities as a takes on the successive values {0, 1, 2, 3}, equal to the set of quantities in the column four-vector on the right (in order beginning at the top). Thus $x^0 = t$, $x^1 = q^1$ and so on. If there is any danger of confusing the raised indices with powers in a given context, we will enclose them in brackets. For brevity we write the column vector usually as x^a.

The quantity t is simply the coordinate time for the reference system and the set $\{q^i\}$ where $i = 1, 2, 3$ give the spatial position. Generally curly brackets are meant to indicate a set, but more usually they are simply understood. These may be the familiar Cartesian set $\{x, y, z\}$ (see Figure 1.3) or they might be spherical polar coordinates $\{r, \theta, \phi\}$ (θ is co-latitude, ϕ is longitude and r is the distance from the origin); or in fact any other set of three numbers that defines a spatial position. As such they are 'generalized coordinates'.

We shall use this convention whereby letters early in the alphabet (before h) shall take on four values 0, 1, 2, 3 as above for x^a, while those later in the alphabet will run from 1 to 3, as above for q^i. All four quantities in x^a may be taken as pure numbers, each giving the value of the corresponding quantity in terms of standard units when length or time is involved, or giving the radian measure for angles.

Figure 1.2 *After a rigid spatial frame of reference is established locally by measurement and synchronization, it might appear as shown in this cartoon. Each ruler indicates a unit of distance and any point on the grid is located with three numbers giving the three independent spatial steps relative to the reference point. The fourth number is the coordinate time, which is the same over the grid. The reference point is shown as having the reference clock with which all of the other clocks are synchronized. Extended to infinity, the grid is the instantaneous world of the reference observer O and friends. It is their inertial frame of reference. Source: Reproduced with permission from Taylor & Wheeler, Spacetime Physics* (*1966*) *W.H. Freeman & Company* (*See Plate 1*.)



By space we mean primarily the relative position of events, and especially the distance between them.[3] We can locate a particular point or position by a three vector, called