Special and General Relativity

An introduction to spacetime and gravitation

Online at: https://doi.org/10.1088/2053-2571/aaf173

Special and General Relativity

An introduction to spacetime and gravitation

Rainer Dick

Department of Physics and Engineering, University of Saskatchewan, Saskatoon, Canada

Morgan & Claypool Publishers

ISBN 978-1-64327-380-8 (ebook)
ISBN 978-1-64327-377-8 (print)
ISBN 978-1-64327-378-5 (mobi)

DOI 10.1088/2053-2571/aaf173

Version: 20190201

IOP Concise Physics
ISSN 2053-2571 (online)
ISSN 2054-7307 (print)

A Morgan & Claypool publication as part of IOP Concise Physics
Published by Morgan & Claypool Publishers, 1210 Fifth Avenue, Suite 250, San Rafael, CA, 94901, USA

IOP Publishing, Temple Circus, Temple Way, Bristol BS1 6HG, UK

For Josephin, Fabian and Jonathan

Contents

Preface

The inception of relativity as a dynamical theory of spacetime was a major breakthrough of curiosity driven basic research, and it had important ramifications both for particle physics and for our understanding of the structure and evolution of the Universe on the largest scales. Training in special relativity is therefore an integral part of the education of every aspiring physicist, and training in general relativity is indispensable for every student who wishes to pursue graduate studies in theoretical physics, astrophysics, or cosmology.

Furthermore, only about half a century after the inception of relativity, the technological relevance of relativity emerged in the analysis of satellite signals and orbits. This was not noticed by the public yet, and both special and general relativity continued to be widely perceived as interesting theories with significant scientific impact, but little *practical* relevance. This perception started to change about three decades later, in the mid-1980s, when the global positioning system (GPS) became available for civil purposes and more people learned about the immediate relevance of relativity for the operation of GPS.

The wide ranging scientific implications and the modern technological relevance of relativity imply that learning relativity is one of the highlights in the training of physics students. In the North American physics curriculum this usually proceeds in two or three stages: a first introduction to relativistic mechanics is often provided as part of a second year course on the foundations of modern physics, while aspects of relativistic electrodynamics are discussed as part of a third or fourth year course on electromagnetism. This usually concludes the mandatory training in relativity for physics students and is limited to special relativity. A deeper introduction to special relativity along with an introduction to general relativity is then provided through a fourth year elective course or an introductory graduate course. The present book should be helpful at every stage of this traditional three-step approach to general relativity. However, it should also help to accelerate and streamline the process of learning relativity by unifying the introduction to special and general relativity in a single concise book, which can be used as a basis of a one-semester course taking students all the way from the foundations of special relativity to the derivation of the Schwarzschild metric and particle motion in general relativity. Furthermore, the immediate impact of relativity for time-keeping and signal transmission would make it prudent for engineering colleges to include introductory relativity courses as electives for students in electrical, computer, and aerospace engineering. It is therefore timely to make the theory of relativity more easily accessible to all students in science and engineering. This book intends to serve that purpose through its concise presentation and by making special relativity accessible to first year students, while the general relativity part should be understandable to third year students.

The book evolved from notes for the special relativity sections of my second year modern physics course, and from the lecture notes for my course on general relativity, gravitation, and cosmology. The latter course is offered biannually for third or fourth year undergraduate students and beginning graduate students.

The first three chapters do not require any preparation beyond a standard first year physics course. They can therefore be used as a reference for a first introduction to special relativity towards the end of the first year physics course, or as part of the second year modern physics course. Students who learn special relativity as part of their third or fourth year electromagnetism courses should also find these chapters useful.

Chapters 4–9 require that students have taken second year courses on vector calculus and mechanics, and second or third year courses on partial differential equations and electrodynamics. In North American universities, students in physics and engineering have usually acquired the necessary skills in vector calculus, intermediate mechanics, and partial differential equations after the first term of their third year, and they usually also take an electrodynamics course during their third year. The present book can therefore serve as a textbook for a course which is already offered in the second term of the third year, thus providing both students and instructors with a lot of flexibility whether they want to learn or teach general relativity in third year, fourth year, or at the introductory graduate level.

Author biography

Rainer Dick

Rainer Dick studied physics at the Universities in Stuttgart, Karlsruhe and Hamburg, and received a PhD degree from the University of Hamburg in 1990. He worked at the University of Munich and the Institute for Advanced Study in Princeton before joining the University of Saskatchewan in 2000. Rainer's research interests span a wide range of topics from particle physics, cosmology and string theory to materials physics and quantum optics. Rainer has published over 100 papers in journals and conference proceedings, and a textbook on *Advanced Quantum Mechanics: Material and Photons*.

Chapter 1

Why relativity?

1.1 The Galilei invariance of Newtonian mechanics

Before we learn relativity, we all live our lives with an implicit *prejudice:* In agreement with our everyday experience, *we assume that every person always measures the same time* (up to conversions between time zones, but let's just assume that we all use Greenwich time). This prejudice can also be denoted as an *assumption of universal time*. If an astronaut on the space station and an astronomer on Earth observe the brightening and subsequent fading of a supernova, then we naively think that they would assign the same time interval between the occurrence and the fading of the supernova, and this point of view was in agreement with the science of mechanics for over two centuries.

To explain the connection with mechanics, we need to recognize that the *assumption of universal time* implies certain coordinate transformation equations between observers which move at constant velocity $\boldsymbol{v}$ relative to each other: Suppose two observers, Frank and Mary, each use their own coordinate system $\{x, y, z\}$ (Frank) and $\{x', y', z'\}$ (Mary), to describe the 'event' in figure 1.1. We also assume that Frank and Mary met at time $t = t' = 0$ and that Mary moves with velocity $\boldsymbol{v} = (v, 0, 0) = v\boldsymbol{e}_x$ relative to Frank, where $\boldsymbol{e}_x$ denotes the unit vector in the x-direction. The assumption $t' = t$ then implies that we should be able to read off from figure 1.1 that Frank's and Mary's coordinates for the event are related according to

$$x' = x - vt,\ y' = y,\ z' = z, \tag{1.1}$$

because after time $t = t'$ elapsed, Mary has moved a distance $vt = vt'$ away from Frank in the x-direction, such that her x'-coordinate to the event is smaller than Frank's x-coordinate by the amount vt.

This simple example illustrates already that the world is four-dimensional, even without relativity: We need four coordinates t, x, y, z to localize an event in space and time. If we explicitly remind ourselves that we use the assumption of universal

doi:10.1088/2053-2571/aaf173ch1

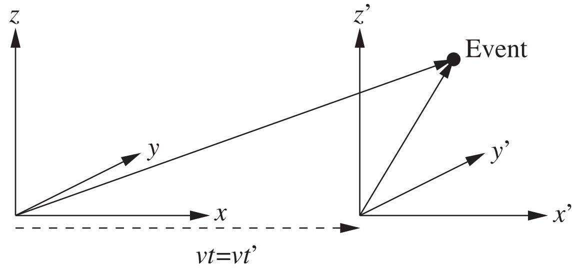

Figure 1.1. Two frames with relative velocity $\boldsymbol{v} = (v, 0, 0) = v\boldsymbol{e}_x$ and the assumption of universal time: $t = t'$.

time, the transformation equations (1.1) become the *Galilei transformation* for relative motion with velocity $\boldsymbol{v} = v\boldsymbol{e}_x$,

$$t' = t,\ x' = x - vt,\ y' = y,\ z' = z. \tag{1.2}$$

However, it is crucial to recognize that our inference of the equations (1.1) from figure 1.1 used the assumption of universal time, $t' = t$, in two ways: The assumption $t = t'$ implies the further assumption that both observers see the *same momentary picture* 1.1 when the times t and t' have elapsed on their clocks, and therefore we also infer a third assumption, viz that both observers assign the *same distance vector* with length $vt = vt'$ to their separation at times t and t'. However, all this is wrong! Once we are forced to surrender the assumption of universal time, we also have no more reason to believe that Frank's perception of the two frames at time t equals Mary's perception at time t', and the distance vectors $\boldsymbol{v}t$ and $\boldsymbol{v}t'$ will also not be the same any more. Therefore a basic mistake in figure 1.1 is the inclusion of the equation $vt = vt'$ and the corresponding identification of the distance vectors at times t and t'. We will discuss this in detail in chapter 3.

However, for now let's stick with our naive assumption $t = t'$ and see how this seemed to be confirmed by Newtonian mechanics.

Suppose the 'event' in figure 1.1 is a moving particle of mass m at time t. If this particle moves under the influence of a force $\boldsymbol{F}(\boldsymbol{x}, t)$ (e.g. due to a local electric field which also changes with time), then Newtonian mechanics says that the particle satisfies the equation of motion

$$m\frac{d^2}{dt^2}\boldsymbol{x}(t) = \boldsymbol{F}(\boldsymbol{x}(t), t) \tag{1.3}$$

in the $\{t, \boldsymbol{x}\}$ frame. However, the Galilei transformation

$$t' = t,\ \boldsymbol{x}'(t') = \boldsymbol{x}(t) - \boldsymbol{v}t, \tag{1.4}$$

into the $\{t', \boldsymbol{x}'\}$ frame then implies that the particle also satisfies Newton's equation in that frame with a local force $\boldsymbol{F}'(\boldsymbol{x}'(t'), t')$,

$$m\frac{d^2}{dt'^2}\boldsymbol{x}'(t') = m\frac{d^2}{dt^2}\boldsymbol{x}(t) = \boldsymbol{F}(\boldsymbol{x}(t), t) = \boldsymbol{F}(\boldsymbol{x}'(t') + \boldsymbol{v}t', t') \\ = \boldsymbol{F}'(\boldsymbol{x}'(t'), t'), \tag{1.5}$$

with the transformation law for forces between the two reference frames,

$$\boldsymbol{F}'(\boldsymbol{x}', t') = \boldsymbol{F}(\boldsymbol{x}, t). \tag{1.6}$$

If we also include a rotation $\underline{R}$ of the coordinate axes and constant coordinate shifts T and $\boldsymbol{X}$ (which implies $\boldsymbol{v}' = \underline{R} \cdot \boldsymbol{v}$ and $\boldsymbol{X}' = \underline{R} \cdot \boldsymbol{X}$ for the relative velocity components and spatial shifts assigned by the two observers),

$$t' = t - T,\ \boldsymbol{x}'(t') = \underline{R} \cdot (\boldsymbol{x}(t) - \boldsymbol{v}t - \boldsymbol{X}),\ \boldsymbol{x}(t) = \underline{R}^T \cdot (\boldsymbol{x}'(t') + \boldsymbol{v}'t' + \boldsymbol{X}'), \tag{1.7}$$

then equations (1.5) and (1.6) become

$$m\frac{d^2}{dt'^2}\boldsymbol{x}'(t') = \boldsymbol{F}'(\boldsymbol{x}'(t'), t') \tag{1.8}$$

and[1]

$$\boldsymbol{F}'(\boldsymbol{x}', t') = \underline{R} \cdot \boldsymbol{F}(\boldsymbol{x}, t). \tag{1.9}$$

Equation (1.8) with the constant coordinate shifts T and $\boldsymbol{X}$ included is denoted as an *inhomogeneous Galilei transformation*, whereas the transformation (1.4) is denoted as a *Galilei boost*.

Equations (1.3) and (1.8) express the *form invariance* (or *invariance* for short) of Newton's equation under Galilei transformations. If Newtonian mechanics holds in a reference frame $\{t, \boldsymbol{x}\}$, then it also holds in every reference frame $\{t', \boldsymbol{x}'\}$ which relates to the frame $\{t, \boldsymbol{x}\}$ through a Galilei transformation (1.7).

How did physicists then learn that, in spite of the invariance of Newtonian mechanics under Galilei transformations, the assumption of universal time is wrong? We will not explain relativity theory in the remainder of this chapter (that requires a bit more preparation and has to wait until chapter 3 and the following chapters), but we will review the observations and developments which made relativity unavoidable in sections 1.2 and 1.3.

1.2 The need for special relativity

There were theoretical and experimental observations in the 1880s that something was amiss with Galilei transformations:

1. The laws of electromagnetism are not invariant under Galilei transformations, i.e. if Maxwell's equations hold in the $\{t, \boldsymbol{x}\}$ coordinate system, they

[1] We will learn in due course that a transformation law $F'(x') = F(x)$ of a quantity $F(x)$ under coordinate transformations $x \to x'$ is denoted as a *scalar* transformation. Equation (1.6) tells us that forces transform like scalars under Galilei boosts (1.4). However, equation (1.9) tells us that forces transform like *vectors* under the general Galilei transformation (1.7).

would not hold in the $\{t', \boldsymbol{x}'\}$ coordinate system if the coordinates are related by a Galilei transformation[2] (1.7).

This is in striking difference to the Galilei invariance of Newton's equation: For Newtonian mechanics the two frames are completely equivalent, but for electrodynamics they are inequivalent!

2. If Galilei transformations would be a correct description of the coordinate transformations, Frank and Mary would observe different speeds for light fronts:

   Suppose Frank triggers a light flash. In his frame this will generate a spherical light front moving with speed c in all directions.

   For Mary this light front would move with speed $c - v$ in x-direction, with speed $c + v$ in $(-x)$-direction, and with speed c orthogonal to the x-axis, if the transformation equations (1.2) between their coordinates were correct. *However, this prediction was shown to be false in the Michelson experiment*[3].

Both of these problems were solved by Einstein's inception of the special theory of relativity[4] (STR).

1. STR explained the unexpected and very puzzling outcome of the Michelson experiment. Michelson had observed that the Earth's motion around the Sun at a speed of about 30 km s^{-1} does not affect the speed of light observed in a terrestrial laboratory, irrespective of the direction in which the light wave travels. Although Einstein seemed to have been primarily motivated by the symmetry properties of electrodynamics, his work solved the Michelson problem by implying that different observers measure time and space intervals in such a way that they both find the same value c for the speed of a light wave in vacuum, irrespective of their mutual relative velocity $\boldsymbol{v}$. Specifically, this implies that space and time intervals in one coordinate system satisfy $c^2\Delta t^2 = \Delta\boldsymbol{x}^2$ if and only if they satisfy in another coordinate system the condition $c^2\Delta t'^2 = \Delta\boldsymbol{x}'^2$, if the other coordinate system moves with a constant velocity $\boldsymbol{v}$ relative to the first system. For relative velocity in x-direction, $\boldsymbol{v} = (v, 0, 0) \equiv v\boldsymbol{e}_x$, this yields transformation laws for the coordinate intervals and coordinates

$$\Delta x' = \gamma(\Delta x - \beta c\Delta t), \quad \Delta y' = \Delta y, \quad \Delta z' = \Delta z, \tag{1.10}$$

$$c\Delta t' = \gamma(c\Delta t - \beta\Delta x), \tag{1.11}$$

and

[2] Voigt W 1887 *Nachrichten von der Königlichen Gesellschaft der Wissenschaften und der Georg-Augusts-Universität zu Göttingen* **1887** (2) 41. This paper analyzes the symmetries of the electromagnetic wave equations and essentially contains the first incarnation of the Lorentz transformations (up to a scaling transformation).
[3] Michelson A A and Morley E W 1887 *Am. J. Sci.* 34 427; Michelson A A and Morley E W 1887 *Philos. Mag.* 24 463.
[4] Einstein A 1905 *Ann. Phys.* 17 891.

$$x' = \gamma[x - X - \beta c(t - T)], \quad y' = y - Y, \quad z' = z - Z, \tag{1.12}$$

$$ct' = \gamma[c(t - T) - \beta(x - X)], \tag{1.13}$$

where $\beta = v/c, \gamma = (1 - \beta^2)^{-1/2}$ and X, Y, Z, T are constant coordinate shifts.

Einstein's work provided an important re-interpretation and logical completion of previous work by Lorentz and Poincaré on the Michelson experiment, and the transformations (1.10), (1.11) and (1.12), (1.13) are known as Lorentz and Poincaré transformations, respectively.

Einstein's reasoning implied that the vacuum speed of light c is absolute, but time is not absolute! The relativity of time for different observers, and the corresponding relativity of the notions of space and time, provided the motivation for the name *relativity theory* for the new theory.

2. STR explained why the wave equations derived from Maxwell's equations are invariant under Lorentz transformations instead of the classically expected Galilei transformations. It was then demonstrated by Minkowski a few years after the invention of STR that the Lorentz transformations are in fact symmetries of the full Maxwell's equations[5].

1.3 The need for general relativity

Explaining the universality of the vacuum speed of light and the symmetries of Maxwell's equations were outstanding successes, and yet Einstein had to face two new challenges as a consequence of STR:

1. While STR explained the symmetry properties of Maxwell's equations, the very successful theory of Newtonian gravity was now incompatible with the basic symmetry principles of the theory.
2. The problem of 'proper' and 'improper' coordinate frames, which had been around in mechanics ever since the publication of Newton's *Principia*, now appeared in a different guise: STR had to assume that there exists a fixed four-dimensional flat spacetime and 'inertial frames', i.e. coordinate systems which move uniformly through this spacetime. The basic equations of STR and electrodynamics were assumed to hold in these distinguished coordinate systems, and the challenge was to identify these special coordinate systems.

For the latter problem physicists ever since Newton referred to the rest frame of the 'fixed stars' as an example of an inertial frame, and any other frame that was related to this frame through Galilei transformations (mechanics) or Lorentz transformations (STR) was also considered to be an inertial frame.

In modern terms the rest frame of the 'fixed stars' would rather be denoted as the *CMB frame* (cosmic microwave background frame) or the *cosmic rest frame*.

[5] Minkowski H 1908 *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* **1908** (1) 53.

Nevertheless, before the inception of the general theory of relativity (GTR) we encounter two (supposedly equivalent, but actually inequivalent) definitions of inertial frames:

1. The cosmic rest frame and any non-rotating frame which moves with constant velocity $\boldsymbol{v}$ relative to the cosmic rest frame are inertial frames.
2. A coordinate system in which every trajectory $\boldsymbol{x}(t)$ of a free particle satisfies

$$\frac{d^2\boldsymbol{x}}{dt^2} = 0, \tag{1.14}$$

is an inertial frame. In different terms: the motion of a free particle is inert in an inertial frame (hence the designation 'inertial frame').

The supposed (or naively presumed) equivalence of these definitions assumes that free particles satisfy equation (1.14) in the cosmic rest frame. However, the Universe expands, and therefore the cosmic rest frame turns out *not* to be an inertial frame, because free particles do *not* satisfy an equation like (1.14) in this frame[6].

The second definition by itself, without reference to a specific example of an inertial frame, is actually meaningless unless we can find a definition of 'free particle' which does not require the notion of inertial frames to start with (if we want to use (2) as a definition of inertial frames, we better not try to define free particles with a statement like 'free particles are particles which have constant velocity, i.e. $d^2\boldsymbol{x}/dt^2 = 0$, in an inertial frame').

General relativity resolved this whole conundrum with the definition and presumed importance of inertial frames in the following way:

1. Inertial frames are actually not required for the proper formulation of the laws of nature. With a proper understanding of differential geometry, the laws of nature assume the same form in any coordinate system.
2. As a consequence of **1**, we can give a definition of free particles which works in every coordinate system and does not rely on the notion of inertial frames.
3. Indeed, inertial frames can be defined as coordinate systems where the equation of motion of free particles takes the simple form (1.14), but these inertial frames turn out to have limited extension in spacelike directions.

In a nutshell, general relativity provided a way to define free particles without reference to inertial frames, but at the same time also deprived inertial frames of their prominent role and identified limitations to the construction of inertial frames.

[6] In more precise terms, there is no global coordinate system for the Universe where all free particles satisfy an equation like (1.14) everywhere at all times. However, if we study only phenomena on time scales which are very small compared to the lifetime $t_0 \simeq 13.8$ billion years $\simeq 4.35 \times 10^{17}$ s of the Universe (i.e. if we do anything but large scale astronomy or cosmology), then a corresponding spacelike slice through the cosmic rest frame with small duration $\Delta t \ll t_0$ provides a very good approximation to an inertial frame as long as we stay away from neutron stars or black holes.

The equivalence of inertial and gravitational mass helped Einstein to solve the problem to reconcile gravity and relativity[7]. To understand this equivalence, recall that Newton[8] had solved the problem to find a dynamical explanation of Kepler's laws of planetary motion (Kepler 1609, 1619) by relating forces to acceleration, and by assuming his famous force law for the gravitational forces between masses m and M at locations $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$:

$$m\frac{d^2}{dt^2}\boldsymbol{r}_1 = -M\frac{d^2}{dt^2}\boldsymbol{r}_2 = G\frac{mM}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|^2}\frac{\boldsymbol{r}_2 - \boldsymbol{r}_1}{|\boldsymbol{r}_2 - \boldsymbol{r}_1|}. \tag{1.15}$$

This contains a peculiarity which appears as a mere coincidence in classical mechanics: The acceleration $d^2\boldsymbol{r}_1/dt^2$ of the mass m is independent of m, since the 'inertial mass' m in $\boldsymbol{F} = m\boldsymbol{a}$ cancels the 'gravitational mass' m in $GmM/|\boldsymbol{r}_1 - \boldsymbol{r}_2|^2$.

Einstein realized in a famous 'Gedankenexperiment' ('thought experiment') that this implies a striking similarity between gravitational forces and inertial forces (pseudo-forces). Assume an observer standing in an elevator at rest in a homogeneous gravitational field, with the gravitational acceleration $\boldsymbol{g}$ pointing downwards. If the observer releases two different objects with different masses, both objects will hit the floor of the elevator at the same time, since they experience the same acceleration $\boldsymbol{g}$. Now assume that the elevator is far away from gravitational fields, but accelerated upwards with acceleration $\boldsymbol{a} = -\boldsymbol{g}$. Again the observer releases the two objects, which now *appear* to be accelerated downwards with acceleration $\boldsymbol{g}$, and again would hit the elevator floor simultaneously.

The outcome in both experiments is identical, and the observer cannot decide whether the force acting on the two objects was due to an external homogeneous gravitational field, or whether the force was an inertial force due to the observer's accelerated motion.

Einstein concluded that *locally we cannot separate gravitational forces from inertial forces*. But if gravitational forces are so similar to pseudo-forces, *maybe they should not be considered as genuine forces at all! Maybe what we perceive as motion of particles under the influence of gravitational forces is actually only the motion of free particles through the curved geometry of spacetime!*

The stage was set for the invention of general relativity: Gravity is a manifestation of the geometry of spacetime.

Note that the equivalence of inertial and gravitational mass plays a key role in Einstein's Gedankenexperiment: If inertial and gravitational mass were different, then two different objects would experience *different* accelerations in a homogeneous gravitational field, but they would still appear to undergo the *same* acceleration $-\boldsymbol{a}$ relative to an accelerated coordinate frame. The observer could then unambiguously distinguish between gravitational forces and inertial forces.

[7] For a historical account of the development of mechanics, see Mach E 1902 *The Science of Mechanics: a Critical and Historical Account of Its Development* 2nd edn (Chicago: Open Court Publishing Company) (Engl. transl. T J McCormack).

[8] Newton assembled all the pieces of the puzzle in several works from 1666 to 1687. This is also discussed in Mach E 1902 loc. cit.

Einstein's identification of motion in 'gravitational fields' as free motion through curved spacetime explains the coincidence of inertial and gravitational mass: The mass of a particle should have no impact on its force free motion, and therefore it should have no effect on the free motion which we perceive as motion in a gravitational field. And we will also see that it naturally solved the problems which Einstein was facing as a consequence of STR: The need to extend STR to a geometric theory of a curved spacetime automatically incorporates gravity, and it deprives inertial frames of their prominent role.

We will see this explicitly in chapters 7–9. However, first we need to acquire a better understanding of geometry and of special relativity.