Relativity Matters
From Einstein’s EMC2 to
Laser Particle Acceleration and Quark-Gluon Plasma

With Exercises, Examples and Discussions

Johann Rafelski

July 31, 2016
Dedication

to: Walter Greiner
who motivated me to write about Special Relativity, and

In Memory

to: John S. Bell
who influenced my understanding of Special Relativity.
Preamble

This book presents Special Relativity (SR) in contemporary language. We take into consideration the way Einstein saw SR after 1915 as a part of the more general scientific context, with the newly formulated General Relativity (GR) influencing the way SR was now understood. This is complemented by the current perspective and connected to present day research topics. SR is presented such that nothing remains a paradox or just apparent, but rather is explained.

We first develop the basic principles of SR, and explore and discuss alternatives. Much of the first half of the book has the format of a discussion in which the teacher, and in particular his graduate student, will be challenged by a brilliant but web-self-taught student called ‘Simplicimus’.

These conversations are representative of both the foundational concepts in SR, and how students have challenged this author over the years. These conversations present the opportunity to explore what often remains unsaid when teaching SR. They often explain how one should think about SR. This is also in response to the realization that in many ‘Modern Physics’ texts there are a serious misunderstandings of the principles of SR. These find their way into web-based, and even some classroom teaching.

As the book progresses, the qualitative and historical discussion turns into textbook-style presentation, and at the end evolves into the concise and precise format of a physics research book. The final 100 pages reveal research topics and unresolved questions related to relativistic charged particle dynamics. I expect that the reader reaching the middle of this book has a good command of elementary algebra and the basic knowledge of calculus along with introductory knowledge of classical mechanics and, ultimately electrodynamics at the level of Maxwell’s equations.

A text of similar character, content, scope, has not been presented before. The search for clarity in the fundamental questions about SR, the developments after 1905, and the strong connection to current research topics are, in my view, the most important and original assets of this book. Each and every reader should keep in mind that I do not invent relativity, but report and interpret the development and the progress of the theoretical framework, with many conceptual developments reaching far beyond the initial ideas. Those who cherish Special Relativity of 1905 vintage should remember that in 1918-22 Einstein disavowed

\[1\] Simplicio appears in Galileo Galilei’s *Dialogue Concerning the Two Chief World Systems* (1632) comparing the Copernican with the Ptolemaic paradigm. The book is presented as a series of discussions among two philosophers and Simplicio, layman defender of the Aristotelian geocentric view on astronomy. Simplicius of Cilicia, c.490–c.560, was a Greco-Roman mathematician and philosopher who wrote extensively on the works of Aristotle. Simplicius Simplicissimus also appears in a Baroque style 1668 book written by Hans Jakob Christoffel von Grimmelshausen.
John S. Bell’s answer of March 12, 1985 about “How to teach Special Relativity”,
the title of the lengthy conference paper he sent along with the letter.

publication of his 1912 Special Relativity review, which had been delayed by the
outbreak of WWI. Looking at this manuscript after reading this book, the answer
to ‘why’ should be clear: by 1920 the scientific context had evolved. Today, of
course, it has evolved further.

\[2\text{The manuscript is published as a facsimile of the original handwritten document, with English translation}
\]
Background remarks

In the early 1980s when teaching at University of Frankfurt I wrote my first book on SR³. Published in Walter Greiner’s “Theoretical Physics” series, this volume was well received in three editions. Walter knew there were significant problems in many texts explaining SR; thus he encouraged and supported this project. Looking today at this 1980-83 effort, it is good but not complete. The current volume is very different, but has its basis in that first experience.

A few years later I asked John S. Bell, a friend and mentor, which English language book to use to teach relativity. I reproduce his letter and some key words are here: John said “...recommend ...my own paper ...Einstein approach is ...pedagogically dangerous...”. Between the lines John argues that the book I was seeking needed to be written. I of course agreed as my German language relativity book was well aligned with Bell’s thinking. In the past 30+ years I was on-off in respect to writing a new text, and I made sure to take John’s advice to follow the historical approach, clarifying why Einstein’s æther is different from the Lorentz and Larmor points of view.

Acknowledgments

During the past three decades, each time I taught a SR related class, I made progress in this new text. Several students at The University of Arizona contributed to the writing and vetting of this book. I thank in particular, (alphabetically) Jessica Bernier, Rebekah Cross, Stefan Evans, Kiel Howe, Taylor Kessinger, Will Parker, Daniel Rosser, and Per Schmidt. Kiel Howe helped me with the graphic material in the book, made valuable organizational comments, and challenged me in the way that led to the creation of the conversation Simplicius. I thank Victoria Grossack for her encouragement and editorial support over the past decade.

I benefited greatly from a long list of critical comments prepared in 2010 “all summer long” by Iwo Bialynicki-Birula, of the Centrum Fizyki Teoretycznej, Polish Academy of Sciences in Warsaw. Iwo’s crisp and critical mind was an invaluable help for me in realizing why the 2010 version of this book could be considerably improved. Iwo took deep interest in this first version of the manuscript, his criticism, comments, and questions influenced the precise final format of this volume. When he saw this edition, he exerted another important influence, making me switch to SI-units, the only units students know.

I thank all those involved for their kind help and interest. I am alone responsible for any errors, omissions in the contents presented here.

Johann Rafelski
Professor of Physics

### Table of frequently used abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMB</td>
<td>Cosmic microwave background</td>
</tr>
<tr>
<td>CMF</td>
<td>Center of momentum frame</td>
</tr>
<tr>
<td>CM-</td>
<td>Frame background = CM-frame</td>
</tr>
<tr>
<td>EM</td>
<td>Electromagnetic</td>
</tr>
<tr>
<td>GR</td>
<td>General Relativity</td>
</tr>
<tr>
<td>GT</td>
<td>Galilean transformation</td>
</tr>
<tr>
<td>LI</td>
<td>Lorentz invariant</td>
</tr>
<tr>
<td>LT</td>
<td>Lorentz transformation</td>
</tr>
<tr>
<td>MM</td>
<td>Michelson-Morley</td>
</tr>
<tr>
<td>SR</td>
<td>Special Relativity</td>
</tr>
</tbody>
</table>
## Contents

**Preamble** ................................................................. iii  
**Contents** ............................................................... xi  
**Guide to contents** ....................................................... xi  
**Many reasons to write this book** ........................................... xiii

### I  Space-Time, Light and the Æther

<table>
<thead>
<tr>
<th>TEST READER VERSION August 24, 2016</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong> Space-Time</td>
<td>3</td>
</tr>
<tr>
<td>1.1 Time, a new 4th coordinate</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Measuring space and time</td>
<td>6</td>
</tr>
<tr>
<td>1.3 Speed of light and the Æther</td>
<td>9</td>
</tr>
<tr>
<td><strong>2</strong> The Michelson-Morley Experiment</td>
<td>19</td>
</tr>
<tr>
<td>2.1 Earth’s motion and the Æther</td>
<td>19</td>
</tr>
<tr>
<td>2.2 Principle of Relativity</td>
<td>22</td>
</tr>
<tr>
<td>2.3 Cosmic microwave background frame of reference</td>
<td>25</td>
</tr>
<tr>
<td><strong>3</strong> Material Bodies in Special Relativity</td>
<td>27</td>
</tr>
<tr>
<td>3.1 Time dilation, body contraction</td>
<td>27</td>
</tr>
<tr>
<td>3.2 Reality of the Lorentz-FitzGerald body contraction</td>
<td>29</td>
</tr>
<tr>
<td>3.3 Path towards Lorentz coordinate transformations</td>
<td>38</td>
</tr>
<tr>
<td>3.4 Highlights: How did relativity ‘happen’?</td>
<td>40</td>
</tr>
</tbody>
</table>

### II  Time Dilation, and Lorentz Contraction

<table>
<thead>
<tr>
<th>TEST READER VERSION August 24, 2016</th>
<th>43</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>4</strong> Time Dilation</td>
<td>45</td>
</tr>
<tr>
<td>4.1 Proper time of a traveler</td>
<td>45</td>
</tr>
<tr>
<td>4.2 Relativistic light-clock</td>
<td>49</td>
</tr>
<tr>
<td><strong>5</strong> Lorentz-FitzGerald Body Contraction</td>
<td>59</td>
</tr>
<tr>
<td>5.1 Universality of time measurement</td>
<td>59</td>
</tr>
<tr>
<td>5.2 Parallel light-clock</td>
<td>59</td>
</tr>
<tr>
<td>5.3 Body contraction</td>
<td>61</td>
</tr>
</tbody>
</table>

### III  The Lorentz Transformation

<table>
<thead>
<tr>
<th>TEST READER VERSION August 24, 2016</th>
<th>73</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>6</strong> Relativistic Coordinate Transformation</td>
<td>75</td>
</tr>
<tr>
<td>6.1 Derivation of the form of the Lorentz transformation</td>
<td>75</td>
</tr>
<tr>
<td>6.2 Explicit form of the Lorentz transformation</td>
<td>80</td>
</tr>
</tbody>
</table>
## Contents

### 6.3 The non-relativistic Galilean limit

85

### 6.4 The inverse Lorentz transformation

86

### 7 Properties of the Lorentz Transformation

88

#### 7.1 Relativistic addition of velocities

88

#### 7.2 Aberration of light

98

#### 7.3 Invariance of proper time

104

#### 7.4 Two Lorentz transformations in a sequence

107

#### 7.5 Rapidity

109

### IV Measurement

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Body Properties and Lorentz Transformations</td>
<td>119</td>
</tr>
<tr>
<td>8.1</td>
<td>Graphic representation of Lorentz transformation</td>
<td>119</td>
</tr>
<tr>
<td>8.2</td>
<td>Simultaneity and time dilation</td>
<td>120</td>
</tr>
<tr>
<td>9</td>
<td>Different Methods of Measuring Spatial Separation</td>
<td>123</td>
</tr>
<tr>
<td>9.1</td>
<td>Signal synchronized in the rest frame of the observer (S)</td>
<td>124</td>
</tr>
<tr>
<td>9.2</td>
<td>Signal synchronized in the rest frame of a body (S₀)</td>
<td>125</td>
</tr>
<tr>
<td>9.3</td>
<td>Light illumination emitted in the rest frame of the observer</td>
<td>127</td>
</tr>
<tr>
<td>9.4</td>
<td>Train in the tunnel: is the tunnel contracted?</td>
<td>131</td>
</tr>
<tr>
<td>10</td>
<td>The Bell Rockets</td>
<td>135</td>
</tr>
<tr>
<td>10.1</td>
<td>Rockets connected by a thread</td>
<td>135</td>
</tr>
<tr>
<td>10.2</td>
<td>The thread breaks</td>
<td>137</td>
</tr>
<tr>
<td>10.3</td>
<td>Lorentz-FitzGerald contraction measured</td>
<td>139</td>
</tr>
</tbody>
</table>

### V Time

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>The Light Cone</td>
<td>145</td>
</tr>
<tr>
<td>11.1</td>
<td>The future</td>
<td>147</td>
</tr>
<tr>
<td>11.2</td>
<td>The past</td>
<td>149</td>
</tr>
<tr>
<td>12</td>
<td>Causality, Twins</td>
<td>151</td>
</tr>
<tr>
<td>12.1</td>
<td>Timelike and spacelike event separation</td>
<td>151</td>
</tr>
<tr>
<td>12.2</td>
<td>Essay: Quantum entanglement and causality</td>
<td>159</td>
</tr>
<tr>
<td>12.3</td>
<td>Time dilation revisited</td>
<td>164</td>
</tr>
<tr>
<td>12.4</td>
<td>Spaceship travel in the Milky Way</td>
<td>167</td>
</tr>
<tr>
<td>13</td>
<td>Relativistic Doppler Shift</td>
<td>173</td>
</tr>
<tr>
<td>13.1</td>
<td>Introducing relativistic Doppler shift</td>
<td>173</td>
</tr>
<tr>
<td>13.2</td>
<td>Relativistic Doppler shift</td>
<td>177</td>
</tr>
<tr>
<td>14</td>
<td>Tests of Special Relativity</td>
<td>184</td>
</tr>
<tr>
<td>14.1</td>
<td>Tests of Lorentz transformation</td>
<td>184</td>
</tr>
<tr>
<td>14.2</td>
<td>Michelson-Morley experiment today</td>
<td>186</td>
</tr>
<tr>
<td>14.3</td>
<td>Modified coordinate transformations</td>
<td>188</td>
</tr>
<tr>
<td>14.4</td>
<td>Measurement of time</td>
<td>188</td>
</tr>
</tbody>
</table>
## VI Mass, Energy, Momentum

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>Mass and Energy</td>
<td>195</td>
</tr>
<tr>
<td>15.1</td>
<td>Proper body energy</td>
<td>195</td>
</tr>
<tr>
<td>15.2</td>
<td>Relativistic energy of a moving body</td>
<td>198</td>
</tr>
<tr>
<td>15.3</td>
<td>Mass of a body</td>
<td>200</td>
</tr>
<tr>
<td>16</td>
<td>Generalized Mass-Energy Equivalence</td>
<td>203</td>
</tr>
<tr>
<td>16.1</td>
<td>Where is energy coming from?</td>
<td>203</td>
</tr>
<tr>
<td>16.2</td>
<td>Mass equivalence for kinetic energy in a gas</td>
<td>203</td>
</tr>
<tr>
<td>16.3</td>
<td>Potential energy mass equivalence</td>
<td>206</td>
</tr>
<tr>
<td>16.4</td>
<td>Atomic mass defect</td>
<td>208</td>
</tr>
<tr>
<td>16.5</td>
<td>Rotational energy mass equivalence</td>
<td>209</td>
</tr>
<tr>
<td>16.6</td>
<td>Chemical energy mass defect</td>
<td>210</td>
</tr>
<tr>
<td>16.7</td>
<td>Nuclear mass defect</td>
<td>211</td>
</tr>
<tr>
<td>16.8</td>
<td>Origin of energy</td>
<td>213</td>
</tr>
<tr>
<td>17</td>
<td>Particle Momentum</td>
<td>214</td>
</tr>
<tr>
<td>17.1</td>
<td>Relation between energy and momentum</td>
<td>214</td>
</tr>
<tr>
<td>17.2</td>
<td>Particle rapidity</td>
<td>219</td>
</tr>
<tr>
<td>17.3</td>
<td>Relativistic rocket equation</td>
<td>225</td>
</tr>
</tbody>
</table>

## VII Collisions, Decays

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>The Center of Momentum Frame (CM-Frame)</td>
<td>233</td>
</tr>
<tr>
<td>18.1</td>
<td>A preferred frame of reference</td>
<td>233</td>
</tr>
<tr>
<td>18.2</td>
<td>The Lorentz transformation to the CM-frame</td>
<td>235</td>
</tr>
<tr>
<td>18.3</td>
<td>Decay of a body in CM-frame</td>
<td>238</td>
</tr>
<tr>
<td>18.4</td>
<td>Decay energy balance in CM-frame</td>
<td>241</td>
</tr>
<tr>
<td>18.5</td>
<td>Decay of a body in flight</td>
<td>242</td>
</tr>
<tr>
<td>19</td>
<td>Particle Collisions</td>
<td>247</td>
</tr>
<tr>
<td>19.1</td>
<td>Two body reactions</td>
<td>247</td>
</tr>
<tr>
<td>19.2</td>
<td>Inelastic reaction threshold</td>
<td>248</td>
</tr>
<tr>
<td>19.3</td>
<td>Energy available in a collision</td>
<td>252</td>
</tr>
<tr>
<td>19.4</td>
<td>Inelastic collision and particle production</td>
<td>258</td>
</tr>
<tr>
<td>19.5</td>
<td>Elastic collisions</td>
<td>260</td>
</tr>
<tr>
<td>19.6</td>
<td>Compton Scattering</td>
<td>262</td>
</tr>
<tr>
<td>19.7</td>
<td>Elastic bounce from a moving wall</td>
<td>266</td>
</tr>
</tbody>
</table>

## VIII 4-Vectors and 4-Force

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>Four-Vectors in Minkowski Space</td>
<td>275</td>
</tr>
<tr>
<td>20.1</td>
<td>Lorentz invariants and covariant equations</td>
<td>275</td>
</tr>
<tr>
<td>20.2</td>
<td>The ‘position’ 4-vector</td>
<td>276</td>
</tr>
<tr>
<td>20.3</td>
<td>Metric in Minkowski space</td>
<td>277</td>
</tr>
<tr>
<td>20.4</td>
<td>Lorentz boosts as generalized rotation</td>
<td>280</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>20.5</td>
<td>Metric invariance</td>
<td>284</td>
</tr>
<tr>
<td>20.6</td>
<td>Finding new 4-vectors and invariants</td>
<td>287</td>
</tr>
<tr>
<td>21</td>
<td>Four-Velocity and Four-Momentum</td>
<td>290</td>
</tr>
<tr>
<td>21.1</td>
<td>Four-velocity $u^\mu$</td>
<td>290</td>
</tr>
<tr>
<td>21.2</td>
<td>Energy-momentum 4-vector</td>
<td>294</td>
</tr>
<tr>
<td>21.3</td>
<td>Properties of Mandelstam variables</td>
<td>295</td>
</tr>
<tr>
<td>22</td>
<td>Acceleration and Relativistic Mechanics</td>
<td>304</td>
</tr>
<tr>
<td>22.1</td>
<td>Small acceleration</td>
<td>304</td>
</tr>
<tr>
<td>22.2</td>
<td>Definition of four-acceleration</td>
<td>305</td>
</tr>
<tr>
<td>22.3</td>
<td>Relativistic form of Newton’s 2nd Law</td>
<td>309</td>
</tr>
<tr>
<td>22.4</td>
<td>The 4-force and work-energy theorem</td>
<td>313</td>
</tr>
<tr>
<td>IX</td>
<td>Motion of Charged Particles</td>
<td>317</td>
</tr>
<tr>
<td>23</td>
<td>The Lorentz Force</td>
<td>319</td>
</tr>
<tr>
<td>23.1</td>
<td>Motion in magnetic and electric fields</td>
<td>319</td>
</tr>
<tr>
<td>23.2</td>
<td>EM-potentials and homogeneous Maxwell equations</td>
<td>327</td>
</tr>
<tr>
<td>23.3</td>
<td>The Lorentz-force: from fields to potentials</td>
<td>332</td>
</tr>
<tr>
<td>23.4</td>
<td>Lorentz-force from variational principle</td>
<td>333</td>
</tr>
<tr>
<td>24</td>
<td>Electrons Riding a Plane Wave</td>
<td>345</td>
</tr>
<tr>
<td>24.1</td>
<td>Fields and potentials for a plane wave</td>
<td>345</td>
</tr>
<tr>
<td>24.2</td>
<td>Role of conservation laws</td>
<td>349</td>
</tr>
<tr>
<td>24.3</td>
<td>Surfing the plane wave</td>
<td>354</td>
</tr>
<tr>
<td>X</td>
<td>Covariant Force and Field</td>
<td>361</td>
</tr>
<tr>
<td>25</td>
<td>Covariant Formulation of Force</td>
<td>363</td>
</tr>
<tr>
<td>25.1</td>
<td>Lorentz-force in terms of 4-potential</td>
<td>363</td>
</tr>
<tr>
<td>25.2</td>
<td>Covariant variation principle</td>
<td>367</td>
</tr>
<tr>
<td>25.3</td>
<td>Covariant Hamiltonian for action principle</td>
<td>375</td>
</tr>
<tr>
<td>26</td>
<td>Covariant Fields and Invariants</td>
<td>378</td>
</tr>
<tr>
<td>26.1</td>
<td>EM-fields: relativistic form</td>
<td>378</td>
</tr>
<tr>
<td>26.2</td>
<td>LT of electromagnetic fields and field invariants</td>
<td>380</td>
</tr>
<tr>
<td>26.3</td>
<td>Constraints on field invariants</td>
<td>386</td>
</tr>
<tr>
<td>26.4</td>
<td>Covariant form of the Lorentz-force in terms of fields</td>
<td>389</td>
</tr>
<tr>
<td>26.5</td>
<td>The Poynting force</td>
<td>393</td>
</tr>
<tr>
<td>XI</td>
<td>Dynamics of Fields and Particles</td>
<td>399</td>
</tr>
<tr>
<td>27</td>
<td>Variational Principle for EM-fields</td>
<td>401</td>
</tr>
<tr>
<td>27.1</td>
<td>Maxwell equations with sources</td>
<td>401</td>
</tr>
<tr>
<td>27.2</td>
<td>Covariant gauge condition</td>
<td>408</td>
</tr>
<tr>
<td>27.3</td>
<td>Homogeneous Maxwell equations</td>
<td>415</td>
</tr>
<tr>
<td>27.4</td>
<td>Energy, momentum and mass of the EM-field</td>
<td>422</td>
</tr>
</tbody>
</table>
Guide to contents

i: General remarks

In colloquial speech the speed of light and the velocity of light are often used interchangeably. However, considering that ‘speed’ is the magnitude of a ‘velocity vector’, I refer in this book to the speed of light, with a few exceptions. For example: I must speak of velocity of light when we discuss physics questions such as the MM experiment in which we search for a direction-dependent result, or when discussing aberration of line of sight. However, the historical use of ‘the velocity of light’ in place of ‘the speed of light’ is deeply entrenched and may surface in this book inadvertently.

This book also aims to clarify that in general there is nothing “apparent” about any of the SR phenomena, and that there are no “paradoxes” other than in use of language aiming to dramatize results. If the words “apparent” and “paradox” appear in this text at all, it is mainly for the purpose of relating to the entrenched use introduced into the subject.

ii: What is included

This volume is limited in scope to SR and electromagnetic theory with specific focus on its relativistic formulation and relativistic charged particle motion. This book does not overlap with, and can be understood as an introduction to, my work on relativistic nuclear collisions and the formation of quark-gluon plasma.\footnote{J. Letessier and J. Rafelski, \textit{Hadrons and Quark-Gluon Plasma} Cambridge University Press (2002).}
iii: Organization of the book

Each of the 11 parts of this book includes a nearly page-long introduction to the contents and the reader is invited to look there for more detailed guidance to the book contents. Here follows a brief summary of these eleven overviews:

Part I offers a general overview of concepts and ideas leading to the formulation of the Special Theory of Relativity (SR). In this introduction the mathematical formulation is kept to a minimum, as these pages are a general motivation and preparation for approaching and thinking about the subject. Personalities key to the development of SR are introduced, including Galileo and Maxwell as well as the usual suspects: Einstein, Lorentz, Michelson, and Poincaré.

Part II presents the two key physics elements: time dilation and the Lorentz-FitzGerald body contraction. This is separated from the contents of the following part to make sure that the reader keeps in mind that we can study body properties without coordinate transformation know-how, as was historically the case. The Lorentz-FitzGerald body contraction is, and is presented as, a property of the body different from proper time ticking in the body. If there is a relation, it is that only in presence of both these independent effects will a ‘light clock’ tick the same way irrespective of its orientation in space.

Part III introduces the relativistic coordinate transformation, the Lorentz-transformation (LT) derived along the lines of argument put forward in the original 1905 work of Einstein. Several examples are presented in detail, such as the velocity addition theorem and the aberration of the line of sight.

Part IV is where the relativistic body properties we described in Part II and Lorentz coordinate transformations introduced in Part III are shown to be consistent with each other. Considerable effort is devoted to introducing John Bell’s rocket example of how space and body differ in relativity. We present in qualitative terms the design of a body-contraction measuring device, showing that one can build it and record the history of contraction effects just like one does with the time dilation effect measuring time using a clock.

Part V addresses important challenges that ‘time’ presents. We discuss causality, space travel and proper time. An essay on quantum entanglement and causality widens the scope of this discussion. The Doppler effect is discussed in depth in order to show how little it has to do with time dilation, and that it is reciprocal (unlike time dilation or the Lorentz-FitzGerald body contraction) once one remembers to account for the effect of the aberration of the line of sight.

Part VI presents the pivotal insight that mass and energy are connected to each other, with the rest mass describing the energy locked in material form. We show that the net energy of a body, after all internal binding and motion effects, determines the body mass. The relation of body energy, body momentum and body mass is introduced. The relativistic rocket equation is also introduced, showing by example how the energy-mass relationship works.
Part VII addresses the relativistic kinematics of a few particle system. The center of momentum frame of reference is introduced – analogous to the center of mass system. Kinematic examples involving particle collisions and decays show the key methods that one encounters. The example of relativistic back-scattering shows a multiplicative energy transfer from the wall to the back-scattered particle.

Part VIII moves by the way of introduction of 4-vectors to the introduction of the relativistic form of the Newton force equation and thus the generalization of body acceleration. As initial examples we show the modification of electron dynamics for simple cases of charged particle motion.

Part IX introduces the motion of charged particles subject to Lorentz-force, beginning with example solutions. The description of relativistic motion is presented in this chapter in non-covariant way. Among topics is the full treatment of a relativistic charged particle surfing the plane wave, a problem motivated by laser acceleration of charged particles.

Part X develops the covariant theory of particle motion and Lorentz-force. We study in depth the different formats of relativistic variational principle, including the covariant Hamiltonian. We obtain the covariant form of electromagnetic fields and use this to present the covariant field based form of the Lorentz-force. We develop an approximate solution of the Lorentz-force in covariant form, identifying the origin of the Poynting-vector oriented force.

Part XI develops the covariant format of Maxwell equations and presents a first exploration of the dynamics of fields. The energy-momentum content of the electromagnetic field is explored. The Born-Infeld nonlinear form of electromagnetism is described. The concept of strong acceleration is introduced and radiation friction described. Variants of radiation friction force are introduced.

Many reasons to write this book

More than one hundred and fifty years ago, what we now call Special Relativity (SR) became a part of our scientific context. To be specific, in a letter dated 5th of January 1865, James Clerk Maxwell wrote to his cousin about his latest scientific work\(^5\): *I have also a paper afloat containing an electromagnetic theory of light, which, till I am convinced to the contrary, I hold to be great guns.* Maxwell completed a theory that unified what were, at the time, three fundamental phenomena in physics: electricity, magnetism, and light. The unified theory of Maxwell opened the door to relativity, because it is inconsistent with Galilean transformations. In order to describe charges in motion as well as charges at rest one needs the Lorentz coordinate transformations of SR. This challenge together

with theories of the carrier of light waves, the (presumed – before Einstein – ma-
material) æther\[^6\] challenged the greatest minds of physics for the nearly four decades
that followed Maxwell’s theory; that is, till Einstein’s paper of mid-1905.

Einstein finds the solution to the riddle by formulating SR. His paper is accord-
ingly titled[^7] “On the electrodynamics of moving bodies”. Today, the classroom
experience with SR begins with Einstein’s derivation of what became known as
the Lorentz coordinate transformation. A few months later Einstein presented
the $E = mc^2$ argument[^8] “Does the inertia of a body depend upon its energy
content” – and for many students of science this formula is also where the teach-
ing of and about SR ends. However, students often find the subject of sufficient
importance to seek greater understanding on their own. Much of this book is
written for them. It makes an effort to explain SR using several communication
methods in order to limit the misunderstanding of concepts, and takes the subject
to the present day level of insight and experimental verification.

This is, as already reported, my second effort to write on relativity. The first
text spurred my interest during the decades that followed to collect SR research
results and to look at other texts about relativity. Thus I have come to realize that
in regard to principles and their applications, SR is not always explained correctly.
This is particularly the case when we strive to comprehend what happens to
extended material bodies. It is in this context that the Lorentz-Fitzgerald body
contraction emerges as a pivotal concept. But what is contracted? When posing
this question I am told most often that space is contracted. This is a blatantly
wrong answer. That is why in this book I aim to say ‘body contraction’ as often
as it makes good sense in the context.

It is hard to understand how and where in popular and/or introductory lit-
erature the view, ‘space is contracted’, made inroads. Perhaps some confusion
with the Lorentz transformation (LT) crept in. LT describes the reference frame
transformation and naturally the Lorentz-Fitzgerald body contraction must be
consistent with a reference frame transformation. However, a Lorentz-Fitzgerald
contraction cannot be a contraction of space for the simple reason that SR does
not address the properties of the space-time in which we live. A later theory,
Einstein’s General Theory of Relativity of 1915 (GR), looks at this question in
order to create a relativistic form of Newton’s law of gravity. I have yet to find
a popular book that puts clearly on paper: SR has no impact on space-time. To
make sure this situation is remedied I added a few dozen pages to this volume.

The question ‘...is it a body or a space contraction?’ has an interesting twist.
Seeing this idea mix up has prompted me to check what students and peers think

[^6]: Ancient Greek: Pure air breathed by Gods
of moving bodies,”) Annalen der Physik 17 891 (1905); received by publishers on June 30, 1905.
[^8]: A. Einstein, “Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?” (trans-
lated: “Does the inertia of a body depend upon its energy content?”), Annalen der Physik 187,
639 (1905); received by publishers on September 27, 1905.
about the principles of SR. To my surprise I was quizzed in return by John S. Bell, a friend known for his clear and unencumbered thinking, and who single-handedly changed the way we think about quantum mechanics. Up to this reciprocal SR pop quiz event I did not realize that John also had vested interest in the teaching of SR. I discovered that he liked to ask a very specific question regarding the Lorentz-Fitzgerald body contraction. My first, from-the-hip response was not, on second thought, 100% adequate. To my amazement, John reflected that only T.D. Lee had done better on his pop quiz – to be compared to T.D. Lee by John S. Bell was for me a big, positive jolt.

Today, I continue the social experiment of John S. Bell, both directly and indirectly. My question stated above: ‘Is the Lorentz contraction that of space or of a body?’ is much simpler. My question must be simpler since I normally ask students while John was quizzing eminent scientists; he was asking ‘will the thread between two rockets tear?’ The answer is in this book, explained, see section 10.

The fact that we were questioning peers and the results that we observed, suggests a complicated situation. I know today that we often do not do a great job teaching SR in the classroom. SR is a very subtle topic needing a dedicated class, and yet we sometimes barely teach SR at all. During my decades of teaching in the US, I have met many entering graduate physics students whose SR background was a one or two week component in a ‘Modern Physics’ class working with a textbook that makes claims which would be failing my students.

Another disconcerting point that this book tackles head-on are the evolving views of Einstein about the æther. Many GR books aiming at a more advanced, graduate student group begin with a terse and accurate introduction to SR. Such presentations are consistent with the point of view expressed as of 1919/20 by Einstein about the æther; they have to be since in GR the metric tensor represents mathematically Einstein’s conceptual view of the æther. However, ‘Modern Physics’ textbooks do not explain that this is also the right way to think about SR. Therefore, there is need for an introduction to SR, which addresses both the conceptual questions tangential to relativistically invariant non-material Einstein’s ‘æther’ and the modern physical situations. I hope this book fulfills this need.

Let me be explicit what Einstein had to say about æther. In spring 1920 Albert Einstein in his renowned Lorentz Lecture retreated from his earlier criticism of the æther, and concluded: To summarize we can say that according to

---

9Tsung-Dao Lee, born 1926, won the Nobel Prize in Physics with C. N. Yang for their work on the violation of parity law in weak interaction, since 1953 Professor (today Emeritus) at Columbia University.

the general theory of relativity, space is endowed with physical qualities; in this sense the æther exists. According to the general theory of relativity, space without æther is unthinkable: without æther light could not only not propagate, but also there could be no measuring rods and clocks, resulting in nonexistence of space-time distance as a physical concept. On the other hand, this æther cannot be thought to possess properties characteristic of ponderable matter, such as having parts trackable in time. Motion cannot be inherent to the æther.[11]

As I was working on this text a further concern surfaced, adding to my motivation to complete this book: in the internet era, unvetted information often permeates the web, just as it often dominates the contents of popular-scientific books. I performed many web searches preparing classes and/or writing this book. Unlike ‘real’ books, the web-based information shows, on average, a lesser level of concern for correctness and precision. My searches produced a view of SR in page-long snippets, written by different people, with different notations, and at times offering seemingly contradictory explanations of the same phenomena. I believe this is what students also find and how they see the subject when they attempt to explore on their own. Therefore I tried to offer here a coherent presentation, connected and embedded into the rich context of information offered by the world wide web.

Many will have heard how the cutting edge of new physics discovery is now divided into ‘energy’ and ‘intensity’ frontiers, according to the question whether the particle beam at an accelerator where the discoveries are to be made excels in its energy content or in number of particles that can be used. It is quite possible that there is another, a third frontier of foundational physics discovery, the ‘acceleration frontier’. I will address questions related to acceleration and show in the final pages of this book the shortcomings of the theoretical understanding. The identification of acceleration as an important element of SR is the specific strength of the more advanced parts of this book.

[11] Only the original text conveys the exact meaning intended by the author. Thus I reproduce this pivotal paragraph in the original German: Zusammenfassend können wir sagen: Nach der allgemeinen Relativitätstheorie ist der Raum mit physikalischen Qualitäten ausgestattet; es existiert also in diesem Sinne ein Äther. Gemäß der allgemeinen Relativitätstheorie ist ein Raum ohne Äther undenkbar; denn in einem solchen gäbe es nicht nur keine Lichtfortpflanzung, sondern auch keine Existenzmöglichkeit von Maßstäben und Uhren, also auch keine räumlich-zeitlichen Entfernungen im Sinne der Physik. Dieser Äther darf aber nicht mit der für ponderable Medien charakteristischen Eigenschaft ausgestattet gedacht werden, aus durch die Zeit verfolgbaren Teilen zu bestehen; der Bewegungsbegriff darf auf ihn nicht angewendet werden.
Part I

Space-Time, Light and the Æther
Introductory remarks to Part I

Part I of this book is a general introduction explaining the background and context of the theory of Special Relativity at a level appropriate for a wide range of science-oriented readers. Characterization of key principles governing relativity in an approachable manner, and placement of these in a greater context makes this a distinct part of the book.

Foundational ideas such as 1+3-dimensional Minkowski space, world line, proper time, causality, the process of measurement of space and time, and several centuries of effort to determine the speed of light form a core of Part I. These topics are rounded off with an introduction to the principle of relativity, Lorentz-FitzGerald body contraction, and time dilation. The æther saga is described. Discussion of the Michelson-Morley experiment is complemented with latest research results showing the continued interest to test SR.

Much of this discussion follows the historic path. This allows to see why Maxwell did not solve a problem he created, and why Einstein chose the title “On the electrodynamics of moving bodies” for his revolutionary paper where he presents SR. This title in essence explains the question posed leading to the invention of SR theory. For 40 years between the eras of Maxwell and Einstein the greatest minds were unable to make decisive progress. Arguably this was so since the contemporary way of thinking required a material ‘luminiferous’ æther.

Einstein’s elaborate views about the æther are recalled in this Part I. Most scholars have heard that in his 1905 publication Albert Einstein denies the existence of an æther. About 15 years later Einstein however refines his point of view as we described previously, see page xv in Preamble for full paragraphs of text and reference. With general relativity, as Einstein explains, space becomes part of the physical reality. Einstein places restrictions on æther properties; it is important also to know that in a private communication with Lorentz, Einstein explains that in his 1905 paper he should have restricted his statement to nonexistence of ‘æther velocity’, see Ref. 15 on page 15. This new type of æther is a relativistically invariant, non-material æther.

There is no effort spared in Part I and for that matter a few more times in this book, to explain that more than a century after its creation, SR suffers today from complex misunderstandings which infect many introductory books and often the introductory physics classes. Hearing this, well you may think, who can tell if this author is himself not wrong? No effort is spared to make sure that the reader can answer this question for herself, and in order to do this convincingly she will need to progress in her study beyond Part I.
1 Space-Time

1.1 Time, a new 4th coordinate

We deal daily with time as a coordinate just like any other; this is how time enters relativity. We use three coordinates to share information about a spatial location: “I am standing on the second story of the building located on the corner of 4th street and 5th avenue.” This is shown as a point in figure [1-1]. The bottom pane presents, using the spatial 3-dimensional $x$, $y$, and $z$ coordinates, a specific point in space.

Suppose that besides describing a location, I want to meet you for lunch. In addition, I say, “Let’s meet at noon.” Now we have, aside from $x$, $y$, $z$, also the fourth time $t$ coordinate, which we illustrate in figure [1-1] in the form of spaced horizontal panels drawn for three different times: origin $t = 0$ at the bottom, the half-way point at 2 min, and the meeting event after a 4 min walk.

The time coordinate is required for a complete description of an event, in this case the lunch meeting. The series of values of these four coordinates describing our positions in space-time is called “a world line.” Each point, including the event where we meet in the 3+1-dimensional space-time is called “an event.”
At the age of 15, Einstein moved to Switzerland (1901 Swiss citizen). After studying physics at ETH in Zürich, Einstein worked from 1902 to 1909 at the patent office in Bern, obtaining his Ph.D. in 1905. After brief tenures as a professor at the Universities in Prague and Zurich, he moved to Berlin. From 1914 till 1933 he was a member of the Prussian Academy of Sciences and Director of the Kaiser Wilhelm Institute for Physics in Berlin. In 1922, for his pivotal contribution to Theoretical Physics, Einstein was awarded the Nobel Prize for physics. In 1933 he left Nazi Germany and settled in the USA at the Institute for Advanced Study in Princeton where he worked until his death.

Much of his work while at the patent office in Bern related to questions about the transmission of electric signals and electromechanical synchronization of time: problems that show up in the thought experiments which spurred Einstein to the formulation of special relativity in 1905. By 1916 he formulated the general theory of relativity based on the equivalence principle of inertial and gravitating mass. Eddington’s announcement of the observation (1919) of the predicted bending of starlight by the gravitation field of the sun galvanized Einstein’s world-wide fame.

Einstein is also known for many other contributions to physics: Brownian motion, light quantum hypothesis, lattice vibration theory of specific heat, and quantum statistics. Einstein was deeply involved in the creation of quantum mechanics, yet he could not accept the probabilistic interpretations which followed. For decades Einstein worked toward a unified field theory; this legacy continues today. Through his work, which revolutionized foundations of physics, Einstein is acclaimed as the most influential physicist of the 20th century.
Time as the 4th coordinate of an event

In the special theory of relativity (SR) we join the coordinates of time and space form the ‘Minkowski Space’\(^1\). “The new notions about space and time, which I would like to develop for you, are founded in experimental physics. Therein lies the argument’s strength. The outcome is a new paradigm. From now on, space and time considered separately sink into obscurity, henceforth a union of the two is the new and fully adequate reality.”\(^2\)

Following Minkowski we rely on the four-dimensional space-time framework, also referred to as 3+1 and more often as 1(time)+3(space)-dimensional space-time or simply (4-dimensional) space. The relation between time and space coordinates will be better appreciated once we consider how different observers measure the coordinates of events.

Proper time

Before the development of special relativity (SR), ‘time’ was assumed to be universal, and to advance at the same rate for everyone, whether in London, in New York, on the Moon, on a train, in motion, or at rest in a laboratory. The pivotal feature in SR is that a clock ‘ticks’ differently in each moving body; each object, e.g. atom, human, man-made satellite, has its own proper time\(^3\). A clock you carry and a clock your friend carries will show different times unless you two always stay together.

Einstein was the first to realize that time ‘belonged’ to a body\(^4\). In daily experience the difference between proper time and laboratory time is too small to notice without making an effort. However, today we can observe time dilation of a clock even in a moving car.

Aside from body proper time there is the time we used in the opening paragraphs, in the discussion of the time coordinate for a meeting. There was an implicit agreement to use the time ‘ticking’ at the location of the meeting. We used that coordinate-time as our reference, this is so called ‘laboratory’ time. We

---

\(^1\)Hermann Minkowski (1864 – 1909), German mathematician who realized that time and space should be combined into a four-dimensional space.

\(^2\)H. Minkowski, address at the 80th Assembly of German Natural Scientists and Physicians (September 21, 1908); translated by the author; original text: “Die Anschauungen über Raum und Zeit, die ich Ihnen entwickeln möchte, sind auf experimentell-physikalischem Boden erwachsen. Darin liegt ihre Stärke. Ihre Tendenz ist eine radikale. Von Stund an sollen Raum für sich und Zeit für sich völlig zu Schatten herabsinken und nur noch eine Art Union der beiden soll Selbständigkeit bewahren.”

\(^3\)H. Minkowski, Ref.\(^2\) introduced the expression ‘proper time’, in his article: “Eigenzeit.”

usually choose as the laboratory the reference frame where we perform simultaneous measurements of things that happen in different locations. We will show that for another observer, not at rest with respect to the laboratory, this means that these two events do not have simultaneity, see section 8.

Causality

Causality means that the sequence of events in time cannot change. Keeping causality of events is a concern since the coordinate transformation of special relativity involves the transformation of time. We will find that, within SR, we cannot go back, e.g., to the time of our birth. The one-dimensional, unidirectional nature of time prevents us from looping motion in time. It is common to frequently revisit the same location in space. Therefore time, which is for many purposes just another coordinate tacked onto space, turns out to be, after all, profoundly different from space.

However, time ticks differently for each of us. Individuals from our past could therefore emerge from an alien spaceship to tell us how different the world used to be. And, for the same reason, if you know a friendly and technologically advanced alien, you could visit the future a thousand years from now. But once you reached the future, there is no way back; the sequence of events cannot change; for each of us the time arrow points forward. The adherence to the arrow of time for all bodies is referred to as the principle of causality.

Is there deeper understanding of ‘time’?

‘What is time?’ is perhaps the most important and yet least understood of the fundamental questions. Moreover, we now see several notions of time. We have recognized a few different ways that we can consider and measure time:
1) The coordinate time, also called laboratory time, we use;  
2) The proper time, i.e. the time proper to a body;  
3) The cosmological time is the proper time (i.e. age) of the Universe.

Within SR we cannot distinguish between clocks created to measure these different times. The difference is entirely confined to the choice of the frame of reference in which these clocks are placed. Therefore, the possible difference between them is not addressed in this book.

1.2 Measuring space and time

The question ‘what is time?’ relates directly to ‘what is space?’. Both questions require that we first establish how each is measured. In the International System of Units (SI) the unit of time, the second, and the unit of spatial size,
the meter are chosen arbitrarily. The time-unit is universally adopted in scientific and daily life since the measurement of time is rooted in thousands of years of tradition with the world being united by the wealth of astronomical observations.

Since 1967, the ‘scientific’ second is the period equal to 9,192,631,770 cycles of the radiation which corresponds to the transition between the two hyperfine interaction split levels of the ground state of the Cesium-133 atom. This definition makes the cesium oscillator (sometimes referred to as an atomic clock) the primary standard for time measurements.

The speed of light is a quantity which today has an exactly defined value and is not measured anymore. Writing the speed of light in the form\footnote{The path to the letter \( c \) becoming the symbol for the speed of light is described by: K.S. Mendelson, “The story of \( c \),” \textit{Am. J. Phys.} \textbf{74}, 995 (2006). A few key points from this article: Einstein switches from \( V \) to \( c \) about three years after inventing relativity. Why \( c \)? The latinized German writing “Constante” instead of ‘Konstante’ by W. Weber, who also may have been driven by the sequence \( a,b,c \); and the Latin word for speed ‘celeritas’, though this interpretation appears explicitly only in 1959 work of Asimov.}{c \equiv 299,792,458 \frac{m}{s},} \tag{1.1}

we recognize that the definition of the meter in effect defines the magnitude of the speed of light. Therefore, since 1983, the meter, as a unit, has been defined as the distance traveled by light in vacuum within a given fraction of a second:

\[ 1 \text{ m} \equiv c \frac{1 \text{s}}{299,792,458}. \tag{1.2} \]

By defining the meter in this way, the accuracy of distance measurement is limited only by time-keeping, which turns out to be very precise\footnote{A natural unit of laboratory distance is the light-nanosecond (\( 10^{-9} \) fraction of a light-second, \( \equiv 1 \text{ c.ns} \approx 29.97 \text{ cm} \)), which corresponds nearly to an American unit of length, the ‘foot’, (1 ft \( \approx 30.5 \text{cm} \)). It can be useful to remember that light takes a bit more than a nanosecond to travel one foot. With due deference to SI units, it is a mystery to this author why \( 1 \text{ c.ns} \approx 0.984 \text{ft} \) by lucky chance a good scale both at the human and the fundamental length scale has not been adopted as the SI-unit of length. Several nuisance factors would disappear from the SI-unit tables: for example Eq. (1.2) simplifies since \( c = 10^9 \text{ ‘ins’}/\text{s} \). Even today a unit of length based on lns could result in a more widely accepted unit system, perhaps uniting the US and European systems of measure. We will discuss SI-unit choices again introducing energy units in Insight on page 204 and addressing units associated with electromagnetic theory in Insight on page 320 yet we adhere in this book to SI-unit system as this is the one system students know.}.

In astronomy, it is common to refer to distance based on how far light travels in a year; that is, the light-year (lyr). Defined by the International Astronomical Union (IAU), 1 lyr is the distance that light travels in a vacuum in one ‘Julian’ year (Jyr) where Jyr= 3.15576 \( 10^7 \text{s} \); and a Julian light year (lyr) by definition:

\[ 1 \text{ lyr} = 9.4607304725808 \cdot 10^{15} \text{ m}. \tag{1.3} \]
Historically, the unit of length was defined along with the unit of time. Therefore the measurement of the speed of light $c$ providing the connection between these arbitrarily chosen units became necessary. Precise measurements of $c$ led to the recognition of the universal nature of $c$. We will describe how we learned that the speed of light is observed to be the same in the Universe and in the laboratory, and that $c$ is found to be independent of the state of motion of the source or of the observer. These experimental results led to the development of SR.

Natural units

The vast majority of scientists working in the fields such as relativity, astrophysics, particle physics and related areas use a natural system of units in which, as just discussed, time is used to measure distance. To see how this works imagine that at a doctor’s visit I tell the nurse I am 6 ns tall. She will take her own measurement triggering a laser pulse at my feet, bouncing the pulse from a mirror at top of my head, with a good clock measuring a time interval of 12 ns, that is after multiplication with $c$, twice my height in feet. This measurement shows clearly that an independent unit of length as such does not exist once speed of light is known. The speed of light in this system of measurement is:

$$ \text{speed of light} \equiv c = \frac{\text{distance traveled}}{\text{time}} = \frac{1 \text{ light} \cdot s}{1s} = 1 \text{ no units.} \quad (1.4) $$

It is usual to omit the $c$ from many equations. In unifying the measurement of time and space:

a) we abandon the distinction between unit of time and unit of length;

b) we eliminate from use the arbitrary factor 299,792,458 connecting the unit of time to an arbitrary unit of distance, arising solely from historical context;

c) as much as we can use time to measure distance, we can use distance to measure time – this is commonplace when, for example, the distance traveled at speed of light between events matters.

Caution: we can measure distance using time, provided that the speed of light is the same everywhere, at any time in all inertial frames of reference. The principle that the speed of light is the same allows a universal relation between space and time. Since time and space are one and the same, we can proceed all the way in terms of units as well to omit, as Eq. (1.4) shows, the letter $c$ from all equations. Even so, in this book $c$ appears in all equations; in other texts the same equations may be written without the symbol $c$ included explicitly for the reason described.
1.3 Speed of light and the æther

Astronomy and the speed of light

The speed of light has fascinated scientists for millennia – but only as recently as the 17th century did the invention of the telescope allow the speed of light to enter the realm of measurable quantities. Around 1679, a Danish astronomer (later statesman) Ole Roemer, working at the Paris Observatory, was puzzled by the orbital period of Jupiter’s moon Io which seemed to depend on the distance between Earth and Jupiter: an orbital period must be always the same.

Roemer believed that the terrestrial observation of what happens near Jupiter depends on the speed of light. Christiaan Huygens (1629-1695) analyzed Roemer’s Io orbit data, taking into account the effect of the orbiting motion of both the Earth and Jupiter around the Sun to obtain $c \approx 220,000 \text{km/s}$ (note that ‘km’ was not yet defined; the value reported here is converted to modern language). The difference from the true value is attributed to the quality of available data and not to omissions in the theoretical analysis.

The speed of light was first precisely obtained by James Bradley\(^7\), who worked with Samuel Molyneux\(^8\) until Molyneux’s death in 1728. They intended to measure the parallax of Gamma Draconis, or 33 Draconis, a star in the zenith of London. The zenith is an imaginary point in the sky directly opposite the direction of the ‘vertical’, the apparent gravitational force at that location. Instead, they discovered another effect, the aberration of light, allowing the measurement of the speed of light with unprecedented precision. We will address stellar aberration in technical detail in section 7.2.

Parallax is the difference in the apparent position of an object viewed by an observer at rest from two different positions, along two different lines of sight. The parallax method is widely used in determining distance. However, Molyneux and Bradley did not observe a parallax for the simple reason that the method does not work well for far distant objects. Gamma Draconis is known today to be at a distance of $R = 154.3 \pm 0.7 \text{ lyr}$. Compared to the diameter $D \simeq 1000 \text{ ls}$ of the Earth solar orbit we have $R/D \simeq 0.5 \cdot 10^7$. Thus the parallax is $P = 2\pi/0.5 \cdot 10^7 = 1.3 \mu\text{rad} = 0.27''$. The parallax effect would be buried in the measurement error of the Molyneux telescope, the most precise astronomical instrument of the epoch. Since Molyneux and Bradley did not realize that Gamma Draconis was beyond

---

\(^7\)James Bradley (1693 – 1762), English astronomer, Astronomer Royal from 1742. Appointed to the Savilian chair of astronomy at Oxford in 1721, Bradley’s work provided the first direct evidence for the movement of the Earth around the Sun as well as a precise measurement of the speed of light based on the newly discovered aberration effect.

\(^8\)Samuel Molyneux (1689 – 1728) member of the British parliament and an amateur astronomer, Fellow of the Royal Society in 1712. Molyneux commissioned precise telescopes and engaged James Bradley.
the parallax distance measurement technique, they hoped for a good measurement of the distance.

Bradley and Molyneux observed an unexplained relatively large (±20") circular ‘motion’ of Gamma Draconis around the vertical defined on Earth. Since Gamma Draconis was in apparent motion synchronized with Earth motion around the Sun, this new ‘aberration’ effect had to be due to the motion of Earth around the Sun, as Bradley realized shortly after Molyneux’s death.

The aberration of light can be understood by considering an observer in a car driving through a snowstorm. Driving into the falling snowflakes, the angle of impact onto windshield tilts apparently to be more horizontal. In an extreme case the snowflakes seem not to fall when the windscreen hits them at high speed, thus instead of falling from the top the aberration due to motion makes them appear to ‘fall’ from the front.

Knowing the orbital speed of the Earth around the Sun (close to 30 km/s, follows from Kepler’s laws), Bradley determined the speed of light from the aberration angle measurement of a zenith star using a non-relativistic version of the method described in section 7.2. Bradley’s result \( c = 301,000 \text{ km/s} \) remained the best value until the terrestrial measurement of 1862 by Foucault that we discuss in the following section.

**Terrestrial measurement of the speed of light**

The first measurement of the speed of light near the surface of the Earth was made by Louis Fizeau\(^9\) using two fixed mirrors, with one partially obscured by a rotating cogwheel. Fizeau’s 1849 value for the speed of light was 5% too high. A year later Fizeau and Foucault\(^10\) used a much improved apparatus based on light reflecting off a rotating mirror toward a stationary concave mirror about 35 kilometers away (see figure [1-2]). The mirror, rotating at an angular velocity \( \omega \), moves slightly in the time it takes for the light to bounce off and return from the stationary mirror. The returning light is therefore deflected a small angle \( \phi' \) from the original beam. A more detailed discussion of the Fizeau-Foucault experiment is found in exercise I–1.

By 1862 Foucault achieved a precision similar to Bradley, reporting \( c = 298,000 \pm 500 \text{ km/s} \). This result was close enough to Bradley’s to nurture the

---

\(^9\)(Armand- Hippolyte-) Louis Fizeau (1819 – 1896), French physicist. In 1849 he published the first ‘terrestrial’ measurement of the speed of light, and in 1850 with Foucault, this measurement was considerably refined. In 1851 he demonstrated the Fresnel drag (see exercise III–10 on page 94).

\(^{10}\)(Jean Bernard) Léon Foucault (1819 – 1868), French physicist, best known for the invention of the Foucault pendulum in 1851. In 1850, together with (Armand-Hippolyte-)Louis Fizeau, Foucault made a precise terrestrial measurement of the speed of light.
Figure 1-2: The principle of the Fizeau-Foucault apparatus to measure the speed of light. Top: a beam reflects off the rotating mirror and travels a distance $l$ to a concave mirror, where it is reflected back traveling along nearly the same path to the rotating mirror. The light makes an angle $\phi'$ with the original beam after reflecting back off the rotating mirror. Bottom: two stages (a) and (b) of the rotating mirror geometry, see exercise I–1.

notion of the universal nature of speed of light\[11\]. Light travels at the same speed on the surface of the Earth and between stars. Moreover, it was also known that light travels through airless space, unlike sound, which does not travel at all in

\[11\]An editorial review by *Nature* of May 13, 1886 entitled “The Velocity of Light,” pp 29-32 presents in historical context a period-contemporary, æther embedded, review of the efforts to measure the speed of light.
space emptied of air.

### Exercise I–1: The Fizeau-Foucault experiment

Obtain the deflection angle, $\phi'$, in the Fizeau-Foucault experiment (figure 1–2), as a function of the angular velocity $\omega$ and the distance $l$ to the concave reflection mirror. Calculate the expected deflection $\phi'$ for a separation between the rotating mirror and the concave mirror of $l = 35\text{km}$ and an angular velocity of $\omega = 10 \text{ revolutions/s} = 20\pi \text{ rad/s}$. Discuss sources of error in this measurement of light speed.

### Solution

The light initially bounces off the rotating mirror when the mirror’s normal makes an angle $\theta$ with the horizontal, as shown in bottom detail (a) of figure 1–2. Because the light is incident along the horizontal, the angle of incidence is $\theta_i = \theta$. By the law of reflection:

1. $\theta_r = \theta_i = \theta$

Therefore, the angle $\phi$ of the reflected beam is:

2. $\phi = 2\theta$

When the light hits the concave mirror, it reflects directly back at the same angle $\phi$ to the horizontal. The time $\Delta t$ it takes the light to travel from the rotating mirror to the concave mirror and back is:

3. $\Delta t = \frac{2l}{c}$

In this time, the mirror has rotated by an angle $\Delta \theta$ to have a new angle $\theta'$ with the horizontal, as in bottom detail (a) of figure 1–2. In terms of the angular velocity $\omega$ of the mirror we have

4. $\Delta \theta = \omega \Delta t = \frac{2l\omega}{c}, \quad \theta' = \theta - \Delta \theta$

From the geometry of figure 1–2, we see that the incident angle of the returning light is:

5. $\theta'_i = \phi - \theta'$

Again by the law of reflection, $\theta'_i = \theta'_r$, and thus we can find the final deflection of the light from the horizontal, $\phi'$:

6. $\phi' = \theta'_r - \theta' = \phi - 2\theta' = 2\theta - 2\theta'$
\[ \phi' = 2\Delta \theta = \frac{4l\omega}{c} \]

With \( l = 35 \text{ km} \) and \( \omega = 20\pi \text{ rad/s} \) we expect to observe a deflection of:

\[ \phi' \approx \frac{\pi}{100} \text{ rads} = 1.7^\circ. \]

Conversely, a precise measurement of \( \phi' \), for a prescribed distance \( l \) and a rotation frequency \( \omega \) yields the speed of light

\[ c = \frac{4l\omega}{\phi'}, \]

with an error \( \delta c \)

\[ \delta c \equiv \delta \phi' \frac{\partial c}{\partial \phi'} = -\delta \phi' \frac{4l\omega}{\phi'^2}. \]

This can be simplified to read

\[ \frac{\delta c}{c} = -\frac{\delta \phi' \phi'}{\phi'^2}. \]

Beside the measurement error in \( \phi' \), the experimental set-up requires a precise determination of the parameter \( l \), and the independent determination of the rotation frequency \( \omega \). These normally can be done to a much greater precision than the measurement error \( \delta \phi' / \phi' \).

---

Maxwell, light, and Einstein’s æther

Upon ‘improvement’ of the prior formulation of the connection between electricity and magnetism, James Clerk Maxwell recognized\(^{12}\) that his new equations allowed for the presence of wave solutions. He evaluated the speed at which electromagnetic waves would move. He showed that this speed can be expressed in terms of the ratio of the strengths of the electric and magnetic forces. Maxwell found for the velocity of the EM-waves, within the measurement error, the speed of light \( c = 310,740,000 \text{ m/s} \). He wrote, “…light and magnetism are affections of the same substance, and light is an electromagnetic disturbance propagated...”\(^{12}\)

\(^{12}\)J. Clerk Maxwell, “A Dynamical Theory of the Electromagnetic Field,” *Phil. Trans. R. Soc. Lond.*) **155** 459-512 (1 January 1865)
through the field according to electromagnetic laws.” To amplify his discovery of the unification of light and his laws of electromagnetism, Maxwell adds “The only use made of light in the experiment (to measure the speed of light, JR) was to see the instruments.” (see page 499 in Ref.[12].) In 1886 and 1887 Heinrich Hertz\textsuperscript{13} confirmed Maxwell’s waves. By the time of Einstein’s special relativity, Maxwell’s wave velocity had been measured\textsuperscript{14} to be \( c = 299,710 \pm 30 \text{ km/s} \).

**James Clerk Maxwell** *British-Scottish physicist, 1831-1879*

Maxwell began his studies at the age of 16 at University of Edinburgh. He moved to Cambridge in 1850 and graduated from Trinity in 1854 with a degree in mathematics. In 1855 Maxwell was made a fellow of Trinity College, before returning to Scotland a year later as professor of natural philosophy at Marischal College, Aberdeen. He was laid off following an academic reorganization and moved in 1860 to King’s College, London. In 1871 he moved to the newly endowed Cavendish Laboratory at Cambridge. “From a long view of the history of mankind–seen from, say, ten thousand years from now–there can be little doubt that the most significant event of the 19th century will be judged as Maxwell’s discovery of the laws of electrodynamics.” (Richard P. Feynman, Lectures on Physics, volume 2)

Maxwell’s unified theory answered the question where the electromagnetic energy resides. Previous theories had assumed that the energy was located at, or on, magnets and/or electrically charged bodies. In Maxwell’s theory, electromagnetic energy could be stored in the deformation described by a ‘field’ of the carrier of the waves. Maxwell devoted a great amount of effort to the understanding of this medium, the “carrier” of electromagnetic waves. Given the common speed of both light and Maxwell’s waves, æther as a medium was also required for light propagation, and hence arose the name “luminiferous” (meaning light-bearing) æther. Because Maxwell’s waves oscillated transverse to the direction of prop-

\textsuperscript{13}Heinrich Rudolf Hertz, (1857 — 1894), German physicist, proved the existence and propagation in space of electromagnetic waves. The frequency unit ‘Hz’ is named after him. He has presented the Maxwell’s equations found in every book today. Thus in literature from before 1933 one often sees these equations called Maxwell-Hertz equations.

agation, the models developed assumed that the æther was incompressible, not allowing density compression waves.

The question “what is æther” remained unresolved and was perhaps the paramount challenge facing Albert Einstein a few decades later. In his initial work on relativity of 1905, Einstein dispensed with æther; a substance which in the new relativistic world is unobservable and physically non-existent. However, 15 years later, after the formulation of the general theory of relativity, Einstein revised his position. In a letter to Lorentz he says: “It would have been more correct if I had limited myself, in my earlier publications, to emphasizing only the non-existence of an æther velocity, instead of arguing the total non-existence of the æther, for I can see that with the word æther we say nothing else than that space has to be viewed as a carrier of physical qualities.”

Explaining his 1905 position writing in 1920 (see Preamble, Ref. 10) Einstein says: However, considered alone in the context of special theory of relativity (used in the meaning of ‘theory of electromagnetism’, JR), the æther hypothesis is a hypothesis devoid of content. In the field equations of electromagnetism, in addition to charge densities, only the electric and magnetic field strengths enter. It appears that in a vacuum, these equations consistently determine the outcome of all electromagnetic processes uninfluenced by any other physical quantities. The electromagnetic fields appear as the ultimate, irreducible realities. Therefore, the consideration of electromagnetic fields as states of the hypothetical homogeneous, isotropic æther-medium appears on a first inspection superfluous.

In this discourse of 1920 prepared to honor Lorentz and given in the presence of Lorentz, Einstein makes it clear that his æther differs from proposals made by others including Lorentz: any æther must satisfy the principles of special relativity. Therefore æther cannot be usual ponderable medium (see Einstein’s remarks in the introduction to this volume). He further argues that æther is a necessary ingredient, being the carrier of all dimensioned variables.

Einstein deduced the reality of the æther by inspecting how the other foundational theory, general relativity, relates to electromagnetic field equations. He argues that the general theory of relativity requires and induces modifications of the æther. Therefore, there should also be a modification within the theory of electromagnetism, accounting for the deformation of the æther resulting from

\footnote{Letter to H.A. Lorentz of November 15, 1919, see page 2 in Einstein and the Æther, L. Kostro, Apeiron, Montreal (2000).}
\footnote{Translated by the author from the original: “Allerdings erscheint die Ätherhypothese vom Standpunkte der speziellen Relativitätstheorie zunächst als eine leere Hypothese. In den elektromagnetischen Feldgleichungen treten außer den elektrischen Ladungsdichten nur die Feldstärken auf. Der Ablauf der elektromagnetischen Vorgänge im Vakuum scheint durch jenes innere Gesetz völlig bestimmt zu sein, unbeeinflußt durch andere physikalische Größen. Die elektromagnetischen Felder erscheinen als letzte, nicht weiter zurückführbare Realitäten, und es erscheint zunächst überflüssig, ein homogenes, intropes Äthermedium zu postulieren, als dessen Zustände jene Felder aufzufassen wären.”}
the gravity of matter. In Einstein’s words: *The principal difference between the æther of general theory of relativity and the case of Lorentz-æther is that in the first case the state of the æther at any point is determined by the presence of matter in this neighborhood. This is so since laws of physics have the form of differential equations. On the other hand, the state of the Lorentzian æther in the absence of electromagnetic fields is determined by nothing other than itself, and the state is everywhere the same.*

We see in this quote that in the opening Einstein reminds that the field equations of general relativity allow us to evaluate everywhere the response of the space-time metric to the presence of matter. By implication in general relativity the concept of æther is meaningful. However, in the Lorentz-Maxwell theory of electromagnetism in domains where the field is absent the æther must have self-determining properties, thus the æther must be everywhere the same.

In the theory of special relativity, the question of what is the origin of energy stored in the mass of a particle is of paramount relevance. Einstein pursues the hypothesis that elementary particles are condensations of electromagnetic energy.

*Considering that the elementary constituents of matter are in their nature nothing else but condensations of the electromagnetic fields, . . . , a point to which we return at the end of this book, see section [28.2]. We must therefore read the arguments about æther understanding that for Einstein, material particles are a part of electromagnetic theory framework, carrying in their mass a locked-in energy content. The contemporary theories of mass do not contradict Einstein’s deductive arguments about reality of the æther. In fact the present day association of mass with a ‘Higgs field’ could be understood as a specific implementation of Einstein’s remarks about the æther providing “standards of space and time (measuring-rods and clocks)”, and thus also of energy and equivalently, mass.*

The quantum structure vacuum (alias “the æther”) is yet more complex. The phenomenon of quark confinement is believed to be caused by transport properties of the quantum structure present everywhere: the “vacuum”; that is, Einstein’s quantum æther, abhors free quarks. This structure dissolves, liberating quarks

---

17 Translated by the author from the original, see Preamble, Ref. 10: “Das prinzipiell Neuartige des ¨Athers der allgemeinen Relativit¨atstheorie gegen¨uber dem Lorentzschen ¨Ather besteht darin, daß der Zustand des ersteren an jeder Stelle bestimmt ist durch gesetzliche Zusammenh¨ange mit der Materie und mit den ¨Atherzust¨anden in benachbarten Stellen in Gestalt von Differentialgleichungen, wahrend der Zustand des Lorentzschen ¨Athers bei Abwesenheit von elektromagnetischen Feldern durch nichts außer ihm bedingt und ¨uberall der gleiche ist.”

18 See Preamble, Ref. 10: “Da nach unseren heutigen Auffassungen auch die Elementarteilchen der Materie ihrem Wesen nach nichts anderes sind als Verdichtungen des elektromagnetischen Feldes, . . .”

within the localized volume of a quark-gluon plasma at temperatures achieved in relativistic heavy ion (nuclei) collision experiments.

Discussion I-1 – Is there an æther?

Topic: We kick off conversation by two students and the professor choosing as topic the æther[20]

Simplicius: Everybody says there is no æther.

Student: Surely they mean no ponderable material æther. In any case, ‘what everybody says,’ is not a scientific argument. By the way, who is everybody?


Professor: Both Lorentz and Einstein would disagree with the Wikipedia writer. I believe that Einstein claimed the absence of æther not due to simplicity of argument i.e. Occam’s Razor, but the issue of physical reality in a theory where results did not depend on æther presence. After Einstein included gravity in the relativity framework, he needed æther.

Student: To illustrate the æther problem, let me ask a provocative question, do you think the Earth is flat?

Simplicius: I’m sorry? What?

Student: You see, once upon a time, practically ‘everybody’ believed it! Do you believe that the Earth moves around the Sun?

Simplicius: That is evident: Copernicus, Galileo...

Student: ...it seems to me you trust a condemned scientist like Galileo?

Simplicius: Justice is not always well served. I trust Galileo; I trust Einstein.

Student: Which Einstein, the one at age 26, or the other at age 40?

Simplicius: I would think the older one knows more.

Student: But did you know that by the time Einstein turned 55, he was a refugee? Can you really trust a homeless scientist?

Simplicius: Why did he become homeless?

Student: Here we are, let me exaggerate – it all started with the æther!

Simplicius: How?

Professor: A colleague of Einstein’s performed an experiment and found a divisible

---

[20] This and many other conversations that follow actually took place, but we present generally an extended and/or dramatized versions. Here ‘student’ is a more seasoned helper to the Professor, and Simplicius is a novice.

[21] William of Ockham’s also written as ‘Occam’ (c. 1285-1349) razor is a principle urging, in the words of Ptolemy (c. AD 90-c. AD 168): “We consider it a good principle to explain the phenomena by the simplest hypothesis possible.”
æther, then Einstein pointed out an experimental mistake; æther was indivisible. The experimentalist enlisted with the Nazis and introduced the idea of German science – applying Occam’s Razor argument that Einstein, being a Jew, had to be wrong. When the Nazis seized power, Einstein was by chance abroad, and he became a stateless, homeless refugee.

Simplicius: I don’t believe that the Nazis came to power over the question of æther.

Professor: The æther controversy divided scientists living in a highly polarized society.

Simplicius: If æther was so important once, I am at a loss to see why we for the most part ignore it today.

Professor: The subtle nuance - there is æther but it is not made of usual matter - was very hard to grasp in the 1920s, and it was too hot to handle just after WWII. Instead everyone focused on the simpler ‘no æther’ perspective presented by Einstein in the ‘original’ 1905 SR paper.

Student: And in doing so one can teach relativity following Einstein’s footsteps made at age 26 without need for the very complex Einstein’s æther ideas that he presented at the age of 40.

Professor: However, we cannot rely on the 1905 position today, there is another subtle issue recognized by Mach: We must distinguish inertial and accelerated reference frames. That is where Einstein’s æther of 1920 is helpful.

Simplicius: So why do I see only the ‘young’ Einstein point of view written about several generations later?

Professor: In part this is so since today the word æther has been replaced by ‘quantum vacuum’.

Simplicius: ...what a funny combination of words: ‘a quantum of nothing’!

Student: Checking Wikipedia on this general topic I read: “...the Higgs field, arising from spontaneous symmetry breaking, is the mechanism by which the other fields in the theory acquire mass.”

Simplicius: Can you please translate?

Professor: The Higgs field provides the measuring stick for the masses of particles. That is just like Einstein’s æther.

Student: In 1919/20 Einstein wrote that æther is “...space is endowed with physical qualities.” Now the Higgs field endows particles everywhere with their inertial mass.

Simplicius: So the Higgs field brings the æther back...

Professor: ...in its quantum vacuum dynamical reincarnation. The properties of the quantum æther are just what Einstein imagined about the permitted properties of the æther before quantum mechanics was invented. Quantum mechanics brings in additional dynamical features absent in classical theory.

Simplicius: I am returning to Wikipedia’s remark about “Occam’s Razor” evidence against æther; why in the end is it wrong?
Professor: Occam’s Razor advises you, if you do not know better, to take the simplest point of view. It is not a scientific argument, and clearly we know better today about the æther and the quantum vacuum properties. The adjective ‘simple’ is often used synonymously with ‘known’ in science, so the Occam’s Razor argument essentially suggests avoiding the invention of new theories. However, if the known ‘simple’ has contradictions, we must propose new ideas which need time to find their simple form.

Simplicius: Can you give me an example?

Student: Have you tried reading Newton’s ‘Principia’ or Maxwell’s ‘Electromagnetism’? Good Luck! According to Occam’s Razor, they must be wrong.

Simplicius: But I have read Einstein’s 1905 ‘Relativity’. I could understand it.

Professor: Because Einstein did not cite anyone in that revolutionary work. Being entirely on his own, with no position, students, grants, reputation, etc., Einstein could ignore the complexities of all past misunderstandings.

Simplicius: In his 1905 work Einstein disposes of the æther.

Professor: He simply pushes it aside as unobservable. In 1919/20 Einstein explains that he went too far with his 1905 existential argument. In the following 15 years he understood that to assure the consistency of relativity there could not be an æther wind. An immovable imponderable æther of 1920 creates consistency with the general theory of relativity. Today we need the æther also in the context of quantum theory. We have introduced Einstein’s non-material æther, renaming it the quantum vacuum.

2 The Michelson-Morley Experiment

2.1 Earth’s motion and the æther

In 1881 Michelson conducted an experiment in an attempt to measure a) the movement of the Earth relative to the material æther, and b) the effect of the movement of the Earth on the speed of light arising from the æther wind. In 1887 an improved experimental effort by Michelson with Morley\textsuperscript{22} followed.

The Michelson-Morley (MM) apparatus consisted of a two-armed interferometer with the two light paths as depicted in figure 2-1. The apparatus is made to rotate around an axis. We assume a geometry such that at one time, an arm of the MM apparatus will be aligned parallel to, and at another time, perpendicular to the velocity of the Earth $\vec{v}$. While the light wave is traveling from Q, one of

\textsuperscript{22}Edward Williams Morley (1838 – 1923), Professor of Chemistry from 1869 to 1906 at what is now Case Western Reserve University.
2. The Michelson-Morley Experiment

**Albert Abraham Michelson** *American physicist, 1852 - 1931*

Michelson arrived as a child in the United States in 1855. From 1869 to 1881 he served in the marines, eventually as an instructor in physics. He undertook his postgraduate study in Berlin and Paris. Starting in 1880, and collaborating later with E.W. Morley, Michelson developed precise experiments to investigate the effect of the Earth’s velocity on the speed of light. In 1893 Michelson joined the University of Chicago, creating its acclaimed Department of Physics. In 1907 he was the first American to receive the Nobel prize for physics “for his optical precision instruments and the spectroscopic and meteorological investigations carried out with their aid.”

The mirrors (here either $M_1$ or the silver coated surface of $P$ which reflects beam towards $M_2$) either approaches or recedes from the other. There should be a difference in the time it takes to travel from $Q$ and $T$ for the two optical paths. The interference fringe-shift in the detector $T$ as the experiment is rotated should allow the observation of the velocity of the Michelson interferometer with respect to the material æther.

Michelson’s objective was to push the precision of speed of light measurement to below the Earth’s orbital speed of 30 km/s. Today we know that the peculiar velocity of the Earth with respect to the Cosmic Microwave Background is about 12 times greater than the orbital velocity. The three main components of Earth’s velocity vector are illustrated qualitatively in figure 2-2: the smallest is orbital speed around the Sun; the orbital speed in the Galaxy is 10 times larger; and the velocity of our Galaxy with respect to the Cosmic Microwave Background (CMB) is yet about twice larger. The net velocity with respect to CMB is $369 \pm 1$ km/s.

The detailed mathematical description of the light and mirror motion is inherent in our study of the light clock in section 4.2, and we defer a detailed description of the optical paths to this discussion. Here we note that the optical paths were defined by mirrors attached to a common material body. Any changes

---

Figure 2-1: The Michelson-Morley Interferometer. The light follows two partially overlapping paths: one QPM$_2$T shown in blue and the other QM$_1$PT shown in green. The glass plate G compensates the greater path length of the QM$_1$PT beam which is reflected on the silver coating placed on the upper side of the mirror P. The two beams produce variable interference in the observer’s T-device depending on the orientation with respect to $\vec{v}$.

Figure 2-2: The velocity vectors of the Earth around the Sun (speed $30.2 \pm 0.2$ km/s), Sun around the center of the Milky Way (speed $\approx 250–350$ km/s) and Milky Way in the Universe (speed $\approx 550–650$ km/s) define the net motion of $369 \pm 1$ km/s against the Cosmic Microwave Background.

of this body as it travels through the æther thus influences the outcome of the experiment as well.
Michelson and Morley’s experimental null result, a result of unprecedented precision at that time, at the level of $2.5 \cdot 10^{-5}$ of the speed of light, was a sensation. Neither the motion of the apparatus nor any influence on the light speed was detected at the upper bound of an 8 km/s shift relative to a stationary material æther in which light propagates at the speed $c \simeq 300,000$km/s.

2.2 Principle of Relativity

Inertial observers

Galileo put forward the principle that the laws of physics are the same in any inertial reference system that moves at a constant speed in a straight line, regardless of its particular speed or direction. Hence, there is no absolute motion, and thus no absolute rest and therefore no ‘center’ of the Universe.

This principle provided the basic framework for Newton’s laws of motion, with the first law of motion we present as: Every body perseveres in its state of rest or of uniform motion in a straight line, unless it is compelled to change that state by forces impressed upon it. This is also presented in the form “Unless acted upon by an external force, an object at rest tends to stay at rest and an object in motion tends to stay in motion with the same speed and in the same direction.” This is also referred to as the principle of inertial motion, or simply the principle of inertia.

An inertial observer is an observer for whom Newton’s first law is true.

We will discuss rotating bodies in Part [XI] of this book, see in particular section [29.2].

Galilean Transformation

Consider the conventional nonrelativistic relation between two inertial observers, $A$ and $A'$, with $A'$ moving at velocity $\vec{v}$ relative to $A$. When inertial observer $A$ measures the velocity $\vec{u}(t)$ of a body, this will differ at all times from the measurements made by observer $A'$ by the relative velocity:

$$\vec{u}'(t) = \vec{u}(t) + \vec{v}.$$ (2.1)

Since the velocity of a body is the rate of change in time of the position vector, the Galilean transformations of the coordinates of a body from $A$ to $A'$ consistent with Eq. (2.1) must be:

$$t' = t,$$

$$x' = x + v_x t, \quad y' = y + v_y t, \quad z' = z + v_z t.$$ (2.2)
Galileo Galilei Tuscan (Italian) Physicist, 1564 – 1642

Called the father of modern science by Einstein, Galileo pioneered scientific reductionism, and insisted on the use of quantitative and repeatable experiments, allowing results to be analyzed with precision.

Galileo reduced the complexity of the real world by seeking to recognize key governing factors. He knew that many subdominant effects had yet to be included into each and every consideration, and that imprecision of measurement also hindered experimental agreement with models considered.

His adherence to experimental results and rejection of allegiance to all other authority in matters of science ushered in the development of modern world.

The Vatican’s ban on reprinting Galileo’s works was partially lifted in 1718 and in full 100 years after his death.

Principle of relativity as used in this book

The Principle of Relativity requires the physical equivalence of all inertial observers: that is, two observers, who differ only in that one is moving at some fixed finite velocity relative to the other, are equivalent. This statement defines a class of inertial observers. From now on, an ‘inertial observer’ is any member of the class of all inertial observers. The laws of physics are the same for any inertial observer.

Most important is to understand how the Principle of Relativity modifies previous understanding. At the time of the heliocentric Universe, the Sun is at the center, and at rest. Now:

a) The Principle of Relativity in any context forbids a preferred point of origin; all places in the Universe are equivalent.

In his 1905 paper Einstein speaks of the “unsuccessful attempts to discover any motion of the Earth relative to the ‘light medium’...” and carries on to conjecture that the laws of physics possess “no properties corresponding to the idea of absolute rest.” He finishes raising this conjecture to the level of a principle he calls “Principle of Relativity”.

---

"From Galileo Galilei portrait of 1636 by J. Sustermans, Hayden Planetarium, NY"
b) In the context of special relativity, any and all laws of physics do not refer to a preferred frame of reference; Einstein declines the possibility of an “absolute rest frame.”

While the laws of physics according to the Principle of Relativity do not refer to any preferred frame of reference, a further condition is needed in order to define non-inertial, i.e. accelerated motion and this must introduce a preferred observer, Mach’s cosmological rest frame, section 29.2. The theory of special relativity is usually presented in the limit that all forces causing acceleration are arbitrarily weak (what weak means is explained in section 29.3). However, interpretation of SR will require that we know which body is accelerated and which is not, irrespective of the magnitude of the acting force, see for example section 12.3 and the following exercise [V–4] on page 165.

A survey of professional web pages which address the Principle of Relativity reveals quite a few different ways to argue. As an example of one such argument, where the summarized claims are in italics (here the sequence is changed):

All experiments run the same in all inertial frames of reference.  
(#2 in original list) This statement is paraphrasing our above discussion.

No experiment can reveal the absolute motion of the observer.  
(#3 in original list) However, any experiment that explores vastness of the Universe can reveal motion with respect to Mach’s cosmological rest frame.

Absolute motion cannot appear in any law of physics.  
(#1 in original list) This claim is restatement of a stronger claim listed previous to it, i.e. (3), since in our view laws of physics follow from experimental reality.

Body motion and the speed of light

Within the corpuscular view of light, a moving emitter is ‘throwing’ the light-particles. Therefore, the Galilean view of light velocity follows from Eq. (2.1); the source and light velocities add vectorially. On the other hand, we know that the speed of sound is a property of the medium in which sound propagates (air, water, etc.), and is not dependent on the motion of the emitter. However, the motion of material, such as air, can change the speed of sound:

\[ c'_s = c_s + v, \]

(2.3)

where \( v \) is the ‘wind’ speed, and not the velocity of the source.

The speed of Maxwell waves representing light was at first understood in analogy to the speed of sound. Maxwell considered a medium, a material æther, necessary for his waves to propagate. Since the speed of light was the property of the material medium, only a modification of the state of the material æther, and in particular ‘æther wind’ could modify the observed velocity of light.
The first insight about the universal nature of the speed of light comes from Maxwell’s study of the speed of electro-magnetic wave propagation. The form of Maxwell’s equations made $c$ independent of the velocity of the wave source, and independent of the wavelength of the wave.

There remains, however, the possibility that $c$ depends on the state of the æther. The opinion in the late 19th century was that Maxwell’s equations were valid only with respect to the æther at rest. Given the large magnitude of the speed of light, it was thought that it would take elaborate experimentation to discover the limits of validity of Maxwell’s equations inherent in the motion of the æther.

Since the material æther was seen to be at rest in some specific reference frame, one could proceed to measure Earth’s velocity vector with respect to material æther. Furthermore, observation of changes in the speed of light could tell us about the properties of the material æther. Such experiments were naturally of great interest.

However, before we proceed, let us remember that, paraphrasing Einstein’s words, there cannot be an æther velocity, and that only relative velocities play a role in SR. We will learn (see section 27) that Maxwell’s equations can be cast into a form valid for any inertial observer and as long as this form is valid, there is always a universal speed of light, valid for all inertial observers in all reference frames

$$c' = c. \quad (2.4)$$

Equation (2.4), the universality of the speed of light, is arguably the key input into Einstein’s formulation of SR.

### 2.3 Cosmic microwave background frame of reference

All scales of distance expand as the Universe ages. By comparing the wavelength of the quantum of light, a photon, emitted by a distant atom, to the expected wavelength emitted by an atom in laboratory, we can tell how long this photon has traveled before being observed. This effect is the cosmological redshift.

The cosmic microwave background radiation (CMB) are the ashes of the ‘big-bang’ in the form of radiation dating back to the hot Universe era during which atoms were formed. This radiation discovered in 1964\textsuperscript{25} fills the entire present

\textsuperscript{25}Arno Penzias and Robert Woodrow Wilson were awarded the Nobel Prize for Physics 1978 for the discovery of CMB. CMB radiation was predicted in 1946, albeit at $T=50K$, by Georg Gamov, (1904 – 1968) a Russian-American theoretical physicist, student of A. Friedman of cosmological FLRW model fame, best known for the explanation of nuclear alpha decay via quantum tunneling, and his work on star evolution and the early Universe, also the author of “Mr Tompkins’ adventures” series of popular-scientific books.
day universe with a thermal $T_{\text{CMB}} = 2.7255(6) \text{K}$ microwave (cm-size wavelength) black body spectrum. In essence we can say that we see everywhere the primordial CMB photons which were already present when atoms were formed.

Cooling of the Universe due to the expansion allowed for ion-electron binding at about 372,000 years after the Big Bang, and the atomic Universe became transparent to radiation. Therefore the ambient temperature today is much lower. The relatively low $T_{\text{CMB}}$ is thus due to a 1000-fold cosmological redshift by the expansion of the Universe. These CMB photons were originally formed at energies corresponding to temperatures $T \simeq 2,970 \text{K}$ when ions and electrons filling the early Universe recombined. In the absence of free electrons the Universe became transparent to radiation.

In the above considerations we implicitly made the assumption that the laws of physics and thus also atomic emission lines were the same cons ago as they are today. Attempts to find time variation of natural constants continue. The limit on relative variation of the fine-structure constant is

$$\alpha \equiv \frac{e^2}{4\pi \varepsilon_0 \hbar c} = \frac{1}{137.035399074(44)} , \quad \frac{1}{\alpha} \frac{d\alpha}{dt} = \frac{(0.20 \pm 0.20) \times 10^{-16}}{\text{yr}} , \quad (2.5)$$

and for the proton to electron mass ratio

$$\mu \equiv \frac{m_p}{m_e} = 1836.15267245(75) , \quad \frac{1}{\mu} \frac{d\mu}{dt} = \frac{-(0.5 \pm 1.6) \times 10^{-16}}{\text{yr}} , \quad (2.6)$$

both obtained assuming a constant rate of change during the lifespan of the Universe. This shows that we can proceed assuming that natural constants are constant, and consider properties of the early Universe using the physics laws determined today.

The CMB radiation background thus provides a ‘natural’ frame of reference which can be universally recognized. A moving observer sees a Doppler-deformed CMB radiation spectrum. This means that one can recognize relative motion in the Universe with respect to the CMB rest-frame of reference. We keep in mind that the equivalence of all observers inertial with respect to CMB rest frame and our knowledge of which observer is at rest with respect to CMB does not violate the principle of relativity: equally well we could imagine measuring velocities with respect to any other inertial ‘beacon-observer’ in the Universe. The CMB is just a very convenient ‘beacon’ we can refer to.

---


3 Material Bodies in Special Relativity

3.1 Time dilation, body contraction

With the advent of Maxwell’s electromagnetism, a way was sought to transcribe the Maxwell equations from one reference frame to another. Larmor realized that time could not remain unchanged. A physical reason for this is that the synchronization of clocks located at different locations takes ‘time’ to complete. Einstein realized that each body has its own ‘body time’. Minkowski called later this body time ‘proper time’. To realize the importance of this new concept let us compare $t$, the coordinate time, to the clock time on a train departing from a station and thus measuring train proper time $\tau$ of which an increment is

$$d\tau = dt \sqrt{1 - v^2 / c^2}.$$  \hfill (3.1)

Since $d\tau < dt$, we speak of time dilation. Furthermore, after the train returns to the station, the cumulative time dilation effect is shown by the train clock. Thus time dilation is hard to misinterpret. Time dilation is as inherent to point-sized bodies as it is to large bodies.

The first to derive time dilation was Sir Joseph Larmor. Larmor’s two-step Lorentz transformation formula we present in Eq. (6.27) on page 80 is evidently the same as Eq. (3.1) provided that the clock is immobile in the moving frame of reference, $x' = 0$. This shows that the proper time of a body is a concept compatible with the contents of the Lorentz relativistic coordinate transformation. In the last phrase of the letter of John Bell on page iv we see that Bell believes that Larmor was fully aware of the physical meaning (time dilation) of his formula Eq. (6.27). Bell writes: “I (Bell) disagree with him (Rindler) only on pedagogy and history (his remark about Larmor not understanding time-dilation is astonishing to me).”.

A body traveling through a resisting material medium such as material æther could be subject to deformation. When the MM experimental news hit the scientific world in the pre-relativity context, this explanation was recognized as necessary. The concept of the body contraction was proposed independently by the Irish physicist George F. FitzGerald and by Hendrik A. Lorentz. Both

---

28Sir Joseph Larmor (1857 –1942), Irish theoretical physicist, author of Aether and Matter, credited with discovery of Lorentz transformation, see Ref. 29.


proposed a shortening of the physical length of the interferometer along the direction of motion such that

\[ L(v) = L_0 \sqrt{1 - v^2/c^2}. \]  

Note that this relation is ‘inverse’ compared to Eq. (3.1) in that the proper length \( L_0 \) is on right hand side and the laboratory length \( L(v) \) on left. Body length in orthogonal directions with respect to the direction of motion remains unchanged.

The unobservability of laboratory speed \( v \), and for that matter any speed including the speed of the æther in the MM experiment requires the presence of both time dilation Eq. (3.1) and Lorentz-FitzGerald body contraction Eq. (3.2). The introduction of the Lorentz-FitzGerald body contraction explained why there is no fringe shift in the MM experiment. However, this did little to answer questions about the medium (æther) in which light propagates as Einstein clarified in his writings after 1920 (see Introduction).

Once the existence of both the body contraction and time dilation is recognized, there is a glaring conflict with the Galilean coordinate transformation. For example, two inertial observers, one at rest in the frame of the æther, and another at rest in the frame of the Michelson-Morley experiment, would report different body lengths of the apparatus. Body contraction means that either these two inertial observers are no longer equivalent, or that the Galilean coordinate transformation needs to be adapted to this new situation. In section 6 we therefore introduce the Lorentz transformations which improve the Galilean coordinate transformations, reconciling the inconsistency of the coordinate transformation with the effect of the Lorentz-FitzGerald contraction and time dilation.

After the introduction of the Lorentz-FitzGerald body contraction hypothesis to explain the MM experiment, it was widely believed that a smaller MM interference effect should still arise. Namely, if the presence of the material æther leads to body contraction, then conversely, æther wind must be created by the passage of the body through the æther; \textit{i.e.}, the body drag of the æther. The Lorentz-FitzGerald body contraction makes a more precise MM type experiment necessary.

The æther might be dragged along by the Earth, just like water is dragged along by a ship, or air is dragged along by a moving car. The drag-effect, the æther wind, would have a velocity smaller than that of the Earth, considering that the æther would have to be in some sense a rigid solid to be able to sustain transverse waves such as the electromagnetic waves predicted by Maxwell. Depending on how one models the matter-æther interaction, the expected æther-drag-effect on the velocity of light might be well below the current best experimental limits, see figure [14-1] which reached a precision at the \( 10^{-18} \) level, corresponding to the absence of an æther wind with velocity greater than 3Å/s.
The Lorentz-FitzGerald body contraction is a pivotal feature of relativity and its proper comprehension is essential. There is some confusion about the meaning of ‘Lorentz contraction’ in contemporary literature:

- Some introductory physics and/or even relativity textbooks, and unfortunately many popular books claim ‘Lorentz contraction’ is space contraction; that is certainly not the case, special relativity phenomena do not alter properties of space-time.
- A few books speak of distance contraction, without clarifying what the word means. We will show in section 9 in what sense two independent events can ‘appear to have contracted distance’: note the use of the word ‘appear’ here, justified as we will arrange the measurement in such a way that we find a contracted spatial separation between two events. This is done to show the consistency of Lorentz coordinate transformations with the Lorentz-FitzGerald body contraction.

Given the need to sort such claims into right, wrong, and conditionally right, and to explain the associated misunderstandings, the Lorentz-FitzGerald body contraction will be a recurrent topic in the first half of this book. The reality of the body contraction effect, *i.e.* the ability to measure it, will be addressed in depth. We will indeed show in section 10.3 that it is, in principle, possible to build a ‘length clock’ that both measures instantaneous contraction, and retains the information about the accumulated effect of the Lorentz-FitzGerald contraction, something that any ordinary clock does for time dilation. However, this is only possible if we start a body and allow it to acquire speed, before it returns to base, the laboratory observer. In this way the body contraction clock acts just like any clock traveling with the body which scores time dilation.

### 3.2 Reality of the Lorentz-FitzGerald body contraction

In the MM experiment the Lorentz-FitzGerald contraction affects the laboratory table, a material body. This compensates the change of the length of the light path modified by the displacement of the mirrors while light travels between them. This change leaves in place the effect of time dilation common to all optical paths. This explains the null result of MM experiment as long as the velocity of light is the same in all directions. Given the Lorentz-FitzGerald body contraction, the MM experiment shows the absence of the æther drag effect; that is, the universality of the speed of light, and given the present day experiments we described earlier, to a very good precision.

The effects a body experiences have to be consistent with the new coordinate transformation of relativity, the Lorentz coordinate transformations, which we will study in section 6. This new transformation will allow us to interpret consistently the same experiment in different inertial frames. However, using
this transformation we can ‘undo’ the body contraction by choosing to perform a measurement in a frame of reference in relative rest. This means that, depending on which observer evaluated the situation, body contraction is present, or it is not. Is the body contraction real then, or merely a mathematical trick?

To better understand how that may work, we consider the Galilean world, where a transformation of an observer into the comoving frame of a body allows her to report that she observes no kinetic energy or momentum of another body. We became very accustomed to that fact. Similarly, in SR we need to become used to the fact that, aside from energy and momentum, the size of a body and the magnitude of the time unit of a clock in a body depend both on the velocity of the observer, and, what will be new, the method of measurement. Relevant in the method is the choice about simultaneity of measurements of the body length, as we discuss in section 9.

We prove an effect is real by considering an experiment where the outcome demonstrates the phenomenon. For example, kinetic energy is ‘real’ even if we can identify a frame of reference wherein it vanishes. In the same way, the Lorentz-FitzGerald body contraction is a real phenomenon if we can present an experimental setup where the outcome depends on the change in body length. The existence of a reference frame, or better said, of an inertial observer who cannot measure body contraction is of no consequence in determining if body contraction is real.
Let us consider the example of a train that enters a mountain tunnel. Even if the train is longer than the tunnel, due to the Lorentz-FitzGerald body contraction, when the train is sufficiently fast, the contracted train will vanish from the sight of an observer at rest with respect to the mountain. This observer could be at that time taking a movie and with luck some of the frames will show only the mountain, with the train being entirely inside the tunnel.

To create a good experiment and considering that the train is moving fast, care must be taken to place the camera at the same distance from both ends of the tunnel. In this case the light from both tunnel ends takes the same time to reach the ‘train station’ camera. The measurement of the train in the tunnel is **simultaneous** by an observer at rest with respect to the mountain. If we had the opportunity to measure previously the train length at rest in the mountain reference frame, the sequence of movie pictures shows the reality of the contraction that the train experiences subject to gentle acceleration before entering the tunnel.

Lorentz always had the explicit opinion that the Lorentz-FitzGerald body contraction is “real.” The views of Einstein were more subtle, to the point of a misunderstanding he needed to clear up in 1911. The key statement reads[^32]: “The claim by the author (Varičak) of a difference between Lorentz’s view and that of mine with regard to physical properties (of Lorentz-FitzGerald contraction) is not correct. The question as to whether the Lorentz-FitzGerald contraction is a physical phenomenon or not can lead to a misunderstanding. For a comoving observer it is not present and as such it is not observable, however it is real and in principle observable by physical means by any non-comoving observer.” The discussion we presented above about the ‘reality’ of the Lorentz-FitzGerald body contraction is exactly what we read in Einstein’s clarification.

In this short paragraph Einstein confirms Lorentz’s view that the train in the tunnel is objectively shorter. He makes sure to say that this cannot be observed by anyone riding in the train, which has produced many claims of a ‘paradox’. This book will resolve such claims eliminating these unnecessary mystifications. On this note let us also mention that in the context of measuring time the general folklore is indeed that there is nothing “apparent” about the difference in time reported by two very precise clocks that do not remain always together. However, the correct understanding allowing evaluation of the ages of different moving

bodies is equally complex. Even so, the general folklore is that twins will in general age differently; the one who went away to study the stars is younger upon return. We will need to use the tools of special relativity correctly to answer the question by how much the traveling twin is younger. The twins differ as only one traveling was subject to an acceleration. We will see that traveling twin stays younger.

The change of the body dimension, and of the body proper time unit which was present in the past cannot be made irrelevant today, since we needed to introduce acceleration, however gentle and/or short-lived. Paraphrasing Einstein, to observe the body contraction in the direction of motion, we must consult a non comoving observer. In the Michelson-Morley experiment this is done by rotating the interferometer, so as to compare the distance between mirrors attached to a body in direction of motion with those orthogonal to it.

The reality of relativistic physical phenomena depends on the availability of instruments that can measure the effects. John S. Bell, known for his interest in the clarification of the difference between classical and quantum reality, has further developed Lorentz and Einstein’s remarks on reality of Lorentz-FitzGerald contraction. He refined a thought (‘Gedanken’) experiment that allows an observer to record the effect of the Lorentz-FitzGerald contraction. The moving body is set in motion while the experiment is progressing. This allows the recording of the magnitude of the Lorentz-FitzGerald contraction at any speed of the body as long as this speed is reached due to acceleration: the Bell-instrument compares length scale prior to acceleration being applied with a length scale in the body that is now speeded up. We return to this topic in section 10.

Some of my colleagues may wonder why in this book the reality of the Lorentz-FitzGerald body contraction is given much attention. In the effort to explain the context, consider the collision of two relativistic nuclei ‘heavy ions’ as achieved at several accelerator facilities, e.g. at CERN near Geneva, Switzerland or at the Brookhaven National Laboratory, Long Island, for the purpose of forming the new state of matter, quark-gluon plasma. Each of the nuclei comes from the two opposite directions at the same speed; the situation is analog to two Lorentz-contracted ‘trains’ colliding, each compressed by a very large factor. Given the original ball-like geometry of atomic nuclei, an observer standing in the laboratory will see two thin pancakes colliding as shown on left in figure 3-1. As the nuclei separate again on the right, a lot of new matter emerges, and observation of these particles helps us understand the processes that transpire, and help us to unravel the secrets of the early Universe.

But what does an observer comoving with one of the nuclei see? A large, non-contraction nucleus, is hit by an ultra thin up to million-fold contracted one. This incoming pancake of nuclear matter takes even at the speed of light up to 1000 times longer, as compared to the previously described case of an observer in laboratory, to cross the non-contraction nucleus. The clock of the comoving
Figure 3-1: On the left two Lorentz-FitzGerald contracted nuclei approach each other, in the middle these ‘heavy ions’ crossed and there is a new phase of quark-gluon matter forming between them, to the right the remnants separate, with many new matter particles produced.

observer actually ticks much slower compared to the laboratory clock. The change in body contraction seems to have been compensated by the time dilation effect; at first sight all seems in order. However, there are many time dependent processes, so which time is relevant to the understanding of the collision process, the one ticking in laboratory, or the one ticking in one of the colliding nuclei? Among the things we learn is that we must keep track of events in terms of the proper time of the observed bodies.

Another related issue is, can we treat the colliding nuclei as a rigid body, or is it a heap of independent particles? At ‘low’ energy the target acts as a strongly bound rigid body. For a very high energy particle arriving, the target body appears as an unbound assembly of particles, and there is no Lorentz-FitzGerald body contraction. To read more about this see conversation IV-2 on page 141.

There is no ‘one right way’ to think about the ultra-relativistic heavy ion collisions. This book prepares the reader to figure out, case-by-case, experiment-by-experiment the best way to interpret this and other complex situations.

Discussion I-2 – Train in a tunnel and the principle of relativity

**Topic:** We discuss why a train accelerated to a finite speed is contracted.

*Simplicius:* There cannot be any reality to the Lorentz-FitzGerald body contraction, right? Given the principle of relativity all a measurement does is compare lengths and give the relative change; we do not know if the train or the tunnel is shorter.

*Professor:* Let us set up a thought experiment to decide this matter. To begin with, the train is at rest with respect to the station and the mountain, and we establish that it is longer than the tunnel by a measurement in the common rest frame. This can be
done *e.g.* by taking a photo of the train parked at the station in front of the mountain. After that measurement, we ask the train driver to gently, very gently accelerate so that any effect that acceleration may have, is negligible. Once the train is fast enough, we let it coast inertially into the tunnel. We take a movie of this with a very fast camera. Inspecting the frames we see that the fast train can fit into the tunnel. To finish the experiment, we ask the driver to slow and stop the train, which we observe to have again its original length.

*Simplicius:* As long as nobody on the train has any memory of being ‘shorter’—I could claim it was the mountain that was longer.

*Student:* I disagree, since we started with a train that was at rest and later was accelerated we can include on the train a device to keep score of both momentary relative size, and the accumulated contraction effect, based on ideas made popular by John S. Bell, see section 10. We can measure the momentary length of the train from within the train.

*Simplicius:* However, by virtue of the principle of relativity I could place Bell’s device in the mountain and claim that the mountain is being accelerated towards the train. I would expect to demonstrate that it is the tunnel that grew longer instead of train becoming shorter.

*Professor:* The principle of relativity does not apply to non-inertial observers; we are accelerating the train, not the mountain. Acceleration of the train makes it different and, more to the point of your remark, it makes Bell’s device report a change in body size of the train. The second device left with the mountain will report no change in body length properties.

*Simplicius:* But if only the train and the tunnel are present in a Universe, how can you tell who is subject to acceleration?

*Professor:* Both Einstein and Poincaré appreciated Mach’s ideas about ‘who is accelerated’ and I believe that one of the reasons Einstein reintroduced relativistically invariant æther was to be able to recognize who is accelerated, who is subject to a force.

*Simplicius:* Could you give an example how force distinguishes a non-inertial observer from an inertial one?

*Professor:* Consider two stars in orbit around each other. To agree there is force . . .

*Student:* . . . you will need an observer in Mach’s reference frame, the rest frame of all mass in the Universe. He can report that these stars rotate around each other, experience acceleration, and thus can radiate.

*Professor:* Indeed, shortly before his æther essay, Einstein published a study of such a two star system predicting it would radiate gravitational energy; that is, these stars must experience a true force of acceleration. Observation of this effect earned a Nobel Prize to Russell Hulse and Joe Taylor in 1993, we will return to this interesting situation in section 29.2, see also Ref. 19 on page 442.

*Student:* As stars rotate they radiate, the Universe is characterized by an inertial ob-
server, who sees how the rotating stars are accelerated.

Professor: Both for gravitational and electromagnetic forces, the emission of radiation by bodies that are truly accelerated with respect to Mach’s Universe inertial frame, and the associated problem of radiation reaction, remains one of the research frontiers which we describe at the end of this book in section 29. However, we recognize here and now that the accelerated train is distinct from the inertial mountain, and that we can measure how the train contracts using devices located on and off the train. The change of the body property we observe can be made consistent with reference frame choices as we discuss in section 8.

Discussion I-3 – What and Who Is in Motion in the Universe

Topic: We discuss the consistency of knowing about the motion with reference to the preferred inertial frame of the Universe.

Simplicius: Everything in the Universe is in motion today, thus subject to some Lorentz-FitzGerald body contraction. Since everything is contracted, I can undo this contraction, so there must be ‘lengthening’. I am also troubled by the fact that in general this is non-inertial motion. For example fixed stars rotate around in their galaxy, galaxies around centers of their clusters and so on.

Student: I heard that in Einstein’s General Relativity, all motion due to gravity by ‘point’ particles is a free fall. What we think is motion subject to a force, turns out to be free fall following curvature in non-flat space. Thus most matter in the Universe is perhaps coasting in free fall.

Simplicius: But we needed acceleration to observe the effect of the Lorentz-FitzGerald contraction.

Student: There must be some acceleration since there are forces other than gravity. We can walk on the Earth since there are forces preventing us from falling freely towards its center. In the same way, we can impart on a body a recoil motion by ejecting matter. That is how the solar system was assembled, by accreting matter. Sometime in the distant past practically all material bodies were subject to acceleration and are today coasting at some speed.

Simplicius: How can I today tell what is contracted more, and what less?

Student: Eons ago, the Universe became transparent to light. We still can see this light which by now, due to Universe expansion, has microwave wavelength. We call this background light the Cosmic Microwave Background (CMB) radiation. Studying this distribution, we realize that it defines a frame of reference in which all visible matter was on average at rest. We determine the solar system’s peculiar velocity vector against this CMB reference frame as a byproduct of this investigation. Look again at figure 2-2 on page 21.
Simplicius: How can you tell that it is not the Universe that started to move against the solar system? And, does the existence of a preferred cosmological rest-frame of reference not contradict the principle of relativity?

Professor: The reference to the Universe as a whole and the proper observer for this system is consistent with the principle of relativity as long as there is no preferred spatial point, as there would be, had we claimed that the Earth is at rest and in the center of the Universe. It is best we agree from now on to relate motion in the Universe to the frame of reference in which the matter in the Universe was at rest 13.8 billion years ago.

Simplicius: What remains of the principle of relativity?

Professor: Everything. That principle applies to individual inertial observers and bodies in relative motion.

Simplicius: Which body is contracted?

Student: We know that Earth’s, or better, the solar system’s, peculiar velocity arose in the process of solar system formation. Therefore, there is a Lorentz-FitzGerald contraction of material bodies in the direction of the peculiar velocity vector with respect to the universal rest-frame.

Professor: We must keep in mind that the body contraction can be also reported by an observer at rest with respect to the Universe frame. Therefore sending a rocket to this Machian reference frame is similar to the case of the moving train. As the train slows down, for the observer in the frame of reference at rest with respect to the Universe, it expands back to its truly natural length. However with respect to the moving point of origin on Earth, if it remains in motion, there is body contraction. This can be so since both measurements are different, carried out by different observers under different conditions we will study in section 9.

Discussion I-4 – What is right and who is wrong explaining SR

Topic: Is there really an issue with books on relativity? And if yes, how can we find a good book?

Professor: After these many arguments on how to think about special relativity embedded into the wider context of things like æther and Mach’s reference frame, it seems to me that we need to clarify how to differentiate good from wanting relativity texts.

Student: Well, I found quite a few books where contraction of space is described. These I think we can safely send back to publishers with a refund request.

Professor: Agreed. However, the situation is rarely that clear; misconceptions can be deeply hidden. We saw how Einstein himself needed to straighten how he is to be understood. I also have a letter from John Bell telling me to beware of following Einstein
in my presentation of relativity, see page iv. So naturally, to know if someone really
got distracted when writing is not that simple. On the other hand, we can recognize
easily when there is a misunderstanding of principles. I once had to be a modern
physics lecture substitute, and as is customary in such a situation I received notes with
instructions to not deviate from. The lecturer whom I helped did not know of my
interest in ‘how to teach relativity’. The fact is that I could not teach the students that
particles travel from the upper atmosphere to the surface of the Earth because “space
is contracted, proving time dilation.”

Simplicius: This is evidently a multiply wrong argument: space is not contracted, and
when a particle travels in free space, there is not a body that is subject to Lorentz-
FitzGerald body contraction. Finally, time dilation seems to me to be distinct from
body contraction so how can one effect prove the other? Did the lecture-note writing
lecturer invent all this?

Professor: No, I checked the textbook, and it was precisely the source of these hand
written class notes.

Student: Do you have an excuse for the author of the textbook?

Professor: None. But I did convince a few colleagues that this textbook was a poor
choice for the classroom. And, we found quite good other texts, both clear and correct.
So even in such a subtle topic like relativity it is not too difficult to find excellent
introductory texts. However, when I turned around to do other things, the bad book
was back in use. Old habits are strong.

Simplicius: So what did you do next?

Professor: I created for myself a list of bad books, and I hope all those who study this
book will do the same. Please beware of the fact that as books are reissued in later
eDITIONS relativity mistakes are often added in. In one case, the path to 3rd edition
was really a full turn from perfect relativity introductory text to a version that teaches
‘s’pace contraction proving time dilation’ which is as bad as one can find – and I cite
verbatim here. I cannot believe that the renown formal book author saw what a ghost
writer was doing.

Student: I now understand why there is so much about body contraction in this book.

Professor: Because we have a problem: as you were making this true comment I opened
a well selling popular text, randomly on page 116, and in the third line from the top
I read: “I referred to the shrinking of the distance between Earth and the star... as
‘length contraction.’” I then realized that the section has the title “Star trips and
squeezed space” and jumping to the end of this section I see it ends with “The rela-
tivistically correct interpretation of length contraction is that measures of space itself
differ in different reference frames...”.

Simplicius: I browsed the reviews of the book I see you hold in hand: there are good
students out there who note the problem – here is a review I think you will like: “I read
all the reviews...bought the book...but on page 137 (nice page number, JR) I finally
threw the book down in disgust. It does not clarify relativity. It is totally confusing
3. Material Bodies in Special Relativity

as to which frame of reference he’s using. Its use of diagrams as examples for clarity is laughable and not helpful. I made it through medical school so I can’t think that I’m unteachable...I am utterly at a loss, using what I learned from this book, to explain relativity to someone. I cannot recommend this book. Those that have read this book, I defy you to explain length contraction to me clearly. I’m sending it back for a refund.”

Professor: I sincerely hope this reviewer has something good to say about this book.

3.3 Path towards Lorentz coordinate transformations

In classical mechanics the Galilean transformations of coordinates allow us to infer the values of momentum and velocity a body has with reference to an observer in motion. There is a consistency between the coordinate transformation and the measurement of momentum and velocity of the body.

The relativistic modification of the properties of a body in motion, and specifically the Lorentz-FitzGerald body contraction and the moving body time dilation, are grossly inconsistent with the Galilean coordinate transformation. Namely, the measurements of e.g. body length made by two different Galilean observers will always produce the same result. This is inconsistent with the Lorentz-FitzGerald body contraction, and, there is no time dilation, since for Galilean transformations time-scale does not change.

We conclude that there must be a new set of coordinate transformations, which probably impact aside from the three space coordinates also time on account of time dilation. This means that we are looking for a transformation of the coordinates of an event $t, x, y, z$ observed in a frame of reference $S$ into the coordinates $t', x', y', z'$, of the same event observed in the frame $S'$ moving relative to $S$ at a velocity $\vec{v}$.

The transformation we seek is called ‘the Lorentz transformation’. In several attempts Lorentz sought a coordinate transformation consistent with Maxwell’s equations, the latest effort[33] does not present the ‘Lorentz’ transformation in the relativistic present day format. The name for the new coordinate transformation was created by Sir Joseph Larmor, see Ref.29, and by Henri Poincaré in the here relevant short report describing Lorentz’s latest work[34] Presenting the ‘Lorentz transformation’, Poincaré claims that he is adding little if anything new;

he checked the results of Lorentz and ‘insignificantly’ amended these – there is no detail or derivation in this presentation of the correct transformation we call today following Larmor, Poincaré, and possibly others, as ‘Lorentz transformation’.

A comprehensive derivation and discussion of the ‘Lorentz’ transformation is seen for the first time in the much longer, and scientifically complete work of Einstein (see Ref. 7 on page xiv) written in German, which Poincaré read and understood fluently. Annalen der Physik (Berlin) dates Einstein’s paper as received on June 30, 1905. Poincaré’s competing paper is received at the publishers July 23, 1905. This Poincaré-Einstein priority question is without merit as the actual credit for the first publication of the correct form of the Lorentz-transformation is due to Sir Joseph Larmor, see Ref. 29.

In this circumstance Einstein holds priority in presenting a derivation of the relativistic coordinate transformation from first principles, opening with his work the door to a new domain of physics. Einstein’s paper which has not a single citation, was written in his capacity as a junior patent clerk without a doctoral degree, and was published in arguably the most prestigious physics research journal of the

---

epoch, and without review, by the editor Prof. Max Planck\textsuperscript{36}. For clarification of historical time-line: Einstein is awarded a doctorate by the University of Zurich in January 1906 and in April 1906 is promoted to be a second-class technical expert at the Swiss Patent Office, in January 1908 he obtains habilitation (venia docendi) at University Bern, a procedure delayed for a few months by rules and regulations governing the process; at time of his habilitation Einstein is still 28 years old.

Einstein obtained the Lorentz transformations from the following general principles:
1. The isotropy of space and the homogeneity of space-time;
2. The principle of relativity, i.e. the equivalence of inertial observers;
3. A universal velocity of light $c$.
There are no further tacit or contextual assumptions required. This means in particular that the medium in which Maxwell waves propagate is undetectable by an inertial observer, a point made by Einstein in his 1905 work.

However, as soon as forces are introduced, questions about æther reappear. That is why upon completion of the theory of gravity and seeing need for a further, more comprehensive unified theory that includes electromagnetism, Einstein reintroduced the non-ponderable (we would say today, non-material) æther.

### 3.4 Highlights: How did relativity ‘happen’?

We briefly recall several developments leading up to Einstein’s formulation of the Special Theory of Relativity which we have now introduced:

1. Within Maxwell theory, light is unified with electric and magnetic dynamics: electromagnetic waves propagate at the same speed as light; light is an electromagnetic wave.
2. The velocity of light is universal; it is the same when measured in a laboratory on Earth or in the interstellar medium and at any wavelength.
3. Since wave velocity is universal, the Maxwell theory cannot be invariant under Galilean transformations.
4. The Michelson-Morley experiment is interpreted in terms of the Lorentz-FitzGerald body contraction hypothesis.
5. Galilean transformations cannot be reconciled with the new revolutionary insights: the universality of the velocity of light, the non-universality of body size, and the proper time of each body.

\textsuperscript{36}Max Planck (1858-1947) a renown German theoretical physicist is said to have made two breakthrough discoveries: that of Planck constant, \textit{i.e.} quantum of action $\hbar$, and of Albert Einstein.
We list now in *chronological* order the scientific work that is most relevant to these key developments. Most of these works have already been cited: exceptions are the first entry on ‘Fresnel drag’, and the related work by von Laue, which we will address further below in exercise [III–10] on page [94]; and Larmor’s priority to the form of Lorentz transformation, see Part III, Ref. 29.

<table>
<thead>
<tr>
<th>Year</th>
<th>Author</th>
<th>Title</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1851</td>
<td>Fizeau</td>
<td>On the hypotheses relating to the luminiferous æther, and on an experiment which seems to demonstrate that the motion of a body changes the speed with which light propagates through its interior</td>
<td></td>
</tr>
<tr>
<td>1856-73</td>
<td>Maxwell</td>
<td>A Dynamical Theory of the Electromagnetic Field</td>
<td>Maxwell, James Clerk, <em>Philosophical Transactions of the Royal Society of London</em> 155 (1865) 459-512. This article accompanied a December 8, 1864 presentation by Maxwell to the Royal Society. To be seen together with two publications, of 1856 and 1861, and the 1873 <em>A Treatise on Electricity and Magnetism</em>.</td>
</tr>
<tr>
<td>1897-1900</td>
<td>Larmor</td>
<td>(1) On a Dynamical Theory of the Electric and Luminiferous Medium; (2) Aether and Matter</td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Author(s)</td>
<td>Titles</td>
<td>Details</td>
</tr>
<tr>
<td>------</td>
<td>---------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
Part II

Time Dilation, and Lorentz Contraction
Introductory remarks to Part II; outlook to Parts III & IV

The reconciliation in the understanding of physics laws creating the present day SR context was achieved by Einstein in 1905 with the introduction of the (proper) body time, that is a clock ticking differently in each and every moving body. Therefore, a convenient starting point in the formal development of SR is a discussion of the proper time, and, of a clock made of a light beam bouncing forth and back between two mirrors, the light-clock.

The study of the light-clock in motion will lead us to the quantitative recognition of the two independent effects that relativistic acquired motion has on a material body: a) the time dilation, and b) the Lorentz-FitzGerald body contraction. The understanding of the light-clock is equivalent to the understanding of why the MM experiment has ‘null’ outcome.

As a body is taken from laboratory reference frame to a different speed, its body-clock measures the proper time that is different from the laboratory time. This time dilation effect is naturally accumulated by a body proper clock set in motion. The time elapsed shown by a clock depends on history of the motion with reference to the base inertial observer, allowing clock comparison in what is often called a ‘twin’ experiment.

The effect of time dilation can be observed by clocks placed on planes and satellites, and in the study of the lifespan of moving, naturally decaying particles, such as muons. All unstable elementary particles have an internal clock that defines their mean proper lifespan. Depending on the status of the observer the actual observed lifespan can differ from the proper lifespan. We return to this topic repeatedly in the book, showing different ways of arriving at the same result.

The relativistic ‘Lorentz-FitzGerald’ body contraction is recognized (akin to the MM experiment) considering the need to construct a light-clock that functions reliably: the time recorded must be independent of the orientation of the mirrors with respect to the direction of motion. Such a generalized light-clock functions just like MM experiment: in the MM measurement a directional difference in optical path is searched using an interference method, while rotating mirror clocks help establish that time measurement is independent of the the light-clock axis orientation.

The need to consider body contraction to assure a consistent light-clock reading implies that if time dilation is a ‘real’ effect, then so must be the Lorentz-FitzGerald body contraction. We return to this point in section 10 in Part IV where we present a device allowing to measure the Lorentz-FitzGerald body contraction and/or the memory of this effect recorded. This discussion is postponed in order to be able to argue, and to show, that the idealized Lorentz-FitzGerald body contraction ‘clock’ is functioning as expected from the view point of different inertial observers. To show this we must use the relativistic coordinate transformations obtained in the following Part III.
4 Time Dilation

4.1 Proper time of a traveler

Applying the principle of relativity comparing two inertial observers when we only know the relative velocity, it seems that one should not be able to tell which of the two observers’ clocks is fast and which is slow. Any of the two observers feels equal and could make just the same comment about the condition of the other. However, in order to compare the time readings of two clocks at one location, it will be necessary in general for at least one of these two clocks to cease to be inertial in order to intersect with the world line of the ‘laboratory’ clock. That requirement of accelerated motion now distinguishes these clocks, where one follows the non-inertial world line of the traveler. We keep in mind that the principle of relativity does not apply to accelerated motion, the traveler cannot claim that it is the laboratory that accelerates to approach her.

By telling which of the clocks to be compared is subject to an acceleration we have defined a measurement prescription. The clock of the accelerated body remains always at the same location according to the body, while the measurement in the laboratory must take into account the fact that the body clock is in motion and thus each time-tick occurs at a different location. To be specific, for \( \vec{v} = d\vec{x}/dt \), and \(|\vec{v}| < c\) the time dilation Eq. (3.1) can be stated in the format

\[
(d\tau)^2 = (dt)^2 - \left(\frac{d\vec{x}}{c}\right)^2, \quad (dt)^2 = \left(\frac{d\vec{x}}{c}\right)^2 + (d\tau)^2 \rightarrow (dt)^2 \geq (d\tau)^2. \quad (4.1)
\]

We see that the increment of proper time \((d\tau)^2\) is always smaller compared to the increment of laboratory time \((dt)^2\); both are equal only if the traveler is at rest with respect to the laboratory. The time interval \((dt)^2\) of a clock that moves by \((d\vec{x}/c)^2\) is always larger compared to the proper time interval \((d\tau)^2\) which is measured in the body at the same location.

This view of the proper time measurement resolves the argument that once the relative motion is established (that is, after the acceleration phase is over), both the moving clock observer and the laboratory observer could claim to be time dilated; the difference that remains is in how the time is measured, \(i.e.\) which of the two observers measures at the location of the clock. Thinking in this way is necessary in order for the time dilation result to reappear correctly when using the relativistic (Lorentz) coordinate transformations, see Part \[IV\].

Integrating Eq. (3.1) we obtain a relation between the proper time of the traveler with her world line \(\vec{x}(t)\) observed in the laboratory

\[
\tau_2 - \tau_1 = \int_1^2 dt \sqrt{1 - \left(\frac{d\vec{x}}{c dt}\right)^2}, \quad (4.2)
\]

and we see again that always \(\tau_2 - \tau_1 \leq t_2 - t_1\). The time recorded by a clock at rest with a body is always less compared to the time recorded by the clock that remains at rest in the laboratory which observes the body in motion.
4. Time Dilation

Figure 4-1: One clock travels in an airplane; the other remains on the Earth.

In the following exercise exercise II–1, we consider the magnitude of time dilation that a direct clock time measurement must be sensitive to. Considering only the special relativity time dilation, both the plane and the satellite clocks stay ‘younger’ compared to the laboratory clock. We find a relatively large effect. Experiments have confirmed the phenomenon of time dilation to high precision, see section 14.1 on page 184. Experimental observation of the effect establishes time dilation as a measured effect around which we can develope in the following the SR framework.

Exercise II–1: Time dilation in airplanes (and satellites)

Compare the time dilation measured in a passenger plane (traveling 1000km/h), and in a near-Earth orbiting satellite (making an orbit every 1.5hrs) to the time measured by a clock at rest on Earth.

Solution

Here we compare clocks of which one is on Earth scoring the time $t$ and another flies on either a plane (or a satellite), as depicted in figure 4-1, scoring proper time $\tau$.

The clocks are synchronized before the plane or satellite takes off, $\tau_1 = t_1 = 0$, and we ignore the short take-off time in our evaluation. A large passenger plane flies at a typical speed near 1000 km/h. This corresponds to $v/c = 1/(300 \cdot 3600) \approx 10^{-6}$. We assume that the speed of the plane is constant. From Eq. (4.2) we find

$$\frac{1}{\tau_2} = \frac{1}{\sqrt{1-(v/c)^2}} \approx 1 + \frac{1}{2} \left(\frac{v}{c}\right)^2$$

and with this

$$2 \frac{t_2}{\tau_2} = 1 + \frac{1}{2} \cdot 10^{-12}.$$
We see that after 6 hours of flight between London and New Yorn one can expect a
mismatch of clocks of $3600 \times 6/2s \times 10^{-12} = 11 \text{ ns}$. This effect is not difficult to measure
in the 21st century.

A near-Earth satellite orbits the Earth in about 1.5h and thus it travels at 40 000
km/1.5h which is 27 times faster compared to the plane, and accordingly the effect of
time dilatation would be $27^2 \simeq 700$ greater. Therefore the reference clock on such a
satellite is out of synchronization with the Earth clock by a factor that is about 700
times greater compared to the result seen in Eq. 2, that is $3.5 \cdot 10^{-10}$ fraction of the
elapsed time. Thus after each satellite orbit which for the near-Earth objects takes
about 100 minutes, the orbiting clock would be short by $2 \mu s$ compared to Earth clock.
Such a relatively big time dilation effect impacts, for example, the GPS navigation
system, which is adjusted accordingly; GPS signal adjustment procedure is by itself a
test of SR, see Ref. 20 on page 186.

End II–1: Time dilation in airplanes (and satellites)

Discussion II-1 – Clock on the relativistic train

**Topic:** We discuss time dilation observed on a ‘train’ moving at a relativistic speed.

*Student:* I see that you have a watch; did you compare the clock time on the train and
at the station before and after the trip?

*Simplicius:* Indeed, I see the train-bound clock is slow.

*Student:* That is the effect of time dilation. The traveler on the train aged less than
the observer at the station did.

*Simplicius:* But how can it be that it is the train that is younger, and not the station
that is older? How did you know that?

*Student:* We measure time at ‘equal space’; that is, the clock in the train is always at
the same spot in that train-frame of reference, and the watch on your hand is where
you are. Therefore, to compare time, we refer to time measurements performed in an
inertial frame of reference at the same location.

*Simplicius:* How does this explain that the train clock is slow?

*Student:* The train was first at rest with respect to the station reference frame; we
synchronized the clocks at that time. After, the train-bound clock moved away, later
it returned with the train; it measured time moving along with the train. The train
differs from the station in that it cannot coast along in inertial motion. That is the
condition which distinguishes the clock on the train assuring that less time has passed
on the watch on your hand.

*Simplicius:* Are you sure? If I sat in the train I could claim that it is you who went
away and came back. In this case I get the opposite result; from this perspective the
station clock seems younger. How can this ever be resolved?
Student: The point is that for two clocks flying apart and coming together later, at least one of the clocks needed to undergo acceleration to change direction of motion. At least one of the flying clocks has to turn around.

Professor: Indeed, the relativity principle does not extend to an accelerated observer. The on-board observer is non-inertial, is set in relative motion and as the train accelerates away, all its clocks measuring proper time slow down.

Simplicius: What if we there were two different trains?

Professor: The fun part is of course to be able to tell the outcome when all clocks were (gently) accelerated while their carriers were exploring the Universe. Looking at several travelers, the one who traveled farthest will come back youngest. We will return to this point again, see exercise V–4 on page 165 and exercise V–6 on page 171.

---

Discussion II-2 – How to measure time dilation

**Topic:** Let us talk about how best to design an experiment to measure the special relativity time dilation effect.

Simplicius: Having read these pages I see that all I need is a precise clock, a plane ticket, and we are done.

Student: Not really! You must have heard that light is deflected by gravity.

Simplicius: What does this have to do with my experiment?

Student: Our time clock depended on the light wave going straight to the mirror. If the path is curved by gravity, the path changes and time dilation changes.

Simplicius: That cannot be a significant effect on Earth; maybe you are thinking of a black hole.

Student: Not true: the escape speed from the Earth is a qualitative measure of the strength of the gravitational potential. Since the speed we can employ flying a plane is significantly smaller than escape speed, the effect of gravity could be bigger.

Simplicius: So how can we measure the motion related SR time dilation effect on Earth?

Student: People came up with a clever idea. Note that when the effects of gravity and special relativity are small the time dilation effect is the sum of two small numbers: that is, the effects of gravity and special relativity are additive. In order to eliminate the gravity effect we form a difference of two time dilations obtained such that the velocity of motion is different, but the effect of gravity is the same. This was accomplished by sending two clocks, one West around the Earth and the other East around the Earth.

Simplicius: How can it matter which direction around the Earth the planes flew?

Student: Assume for simplicity that the planes fly around the equator. The total velocity of each plane that went on the trip is \( v_\pm = v_r \pm v_p \) where \( v_r \) is Earth’s rotation
speed \( v_r = 40\,000\text{km/24h} \), and \( v_p \) the flight speed against the ground. Both planes have otherwise the same history, of take-off and landing, of flight height, same as much as possible. Since the Earth rotates pretty fast the total speed of the two planes as compared to an inertial observer floating in space is not at all the same.

*Simplicius:* That is neat, the Earth rotation speed is \( v_r \approx 1\,700\text{km/h} \) and that is faster compared to the speed of the planes. Thus SR difference effect containing \( v_r^2 - v_p^2 = 4v_r v_p \) is enhanced by the speed of the Earth’s rotation.

*Student:* Moreover, in the difference, the effect of gravity and acceleration drops out.

*Simplicius:* Did it work?

*Student:* Yes, this measurement confirmed special relativity, see Ref. [17] on page [184]. In fact the effects of both special relativity and gravity were measured. The GPS signal also allows a test along this line, see Ref. [20] on page [186].

### 4.2 Relativistic light-clock

Our next objective is to understand better the process of measuring time. To this end we consider a generic ‘light’ clock comprised of two mirrors firmly attached to a solid body. This assures that, using our device, we study time ticking within this body. The tick of time is the reflection of light in the mirror. Our objective is to compare the body clock times of two or more light-clocks. For the purpose of our discussion, a gentle acceleration can be present acting on one of the light-clocks, but has to be always as small as needed so that we can argue the case without a theoretical framework adapted to non-vanishing acceleration.

In a first step, we set up the light-clock such that the effect of the Lorentz-FitzGerald contraction is not present; that is, light travels between body mirrors normal to the direction of motion. In a second step to which we return in section 5.2 we show that, as long as we allow for the body contraction, the outcome of time measurement is the same for a clock where the light travels in direction of motion. We return to show that irrespective of the orientation of the light path with respect to the motion of the light-clock the clock scores the same time in exercise II–5.

Two light-clocks are shown in figure 4-2; on the left we see one that is at rest with respect to the laboratory observer, and on the right the clock moves with speed \( v \) nearly normal to light path. In all practical configurations, the angle that we see the light path inclined against \( v \) is very, very small, rivaling the precision with which the mirrors can be made normal to the direction of light propagation.

We see for the stationary clock on the left in figure 4-2 that light travels between two mirrors, \( M_1 \) and \( M_2 \). We describe the clock based on the observations
made by an observer in the same inertial frame as the clock, that is, there is no relative velocity between the observer and the clock. Measurements made in the rest frame of the body of interest are called ‘proper’. Thus the ‘proper time’ of a body is the time measured by a clock at rest with respect to the body. To avoid misunderstanding, the proper time of a body is a property of that body. A frame of reference is introduced here to show that proper time can be measured in any frame of reference that is at rest with respect to the body.

In figure 4-2a we further see that the two mirrors attached to one supporting body are separated by the proper body distance $l$. We keep in mind that measurement of $l$ is made at the same time in the rest frame of the body; that is, we use a rigid meter-stick which we apply to measure the mirror separation $l$. The unit of body-clock proper time is

$$I_{(a)} = \frac{2l}{c}.$$ \hfill (4.3)

The time $\tau$ elapsing in the body is measured as number of units $I_{(a)}$. We can think of this as number of ‘$I_{(a)}$-seconds’.

In the laboratory frame the traveler’s clock ‘(b)’ has a nonvanishing velocity $\vec{v}$. For simplicity, we first assume that clock ‘(b)’ moves at a constant velocity $\vec{v}$ relative to the laboratory observer, and in this discussion, perpendicular to the mirror normals, as is depicted on right in figure 4-2b. The time that the light needs to travel from $M_1$ to $M'_2$ and back to $M''_1$ is

$$I_{(b)} = \frac{2\sqrt{l^2 + y^2}}{c}.$$
We can also write this time in terms of how long it takes the mirror to move the distance \( y \)

\[ I_{(b)} = \frac{2y}{v}. \]

Solving these two equations for \( y \) yields

\[ y = l \frac{v/c}{\sqrt{1 - (v/c)^2}}. \tag{4.4} \]

The period of the moving clock is

\[ I_{(b)} = \frac{2y}{v} = \frac{2l}{c} \frac{1}{\sqrt{1 - (v/c)^2}}, \quad \Rightarrow I_{(b)} = I_{(a)} \frac{1}{\sqrt{1 - (v/c)^2}}. \tag{4.5} \]

This result is easily understood by considering the longer optical path seen in figure 4-2b as compared to figure 4-2a which requires \( I_{(b)} > I_{(a)} \): for the laboratory observer the moving clock has a longer optical path compared to the identical clock at rest in the lab.

To understand who stays ‘younger’ imagine that the speed achieved by the traveler is so close to velocity of light that the traveling clock has only a chance to ring a few ‘\( I_{(b)} \)-seconds’ before reaching destination. Clearly the traveler is in that case much younger when compared to his twin for whom the laboratory clock at rest ticks many ‘\( I_{(a)} \)-seconds’. As we increase the speed \( v \) the number of ‘\( I_{(b)} \)-seconds’ decreases and in the limit of \( v \to c \) the traveler like a photon will not age; the traveler catches-up with the flow of time.

Discussion II-3 – **Light speed is very special**

**Topic**: What is special about the speed of light?

**Simplicius**: Why is there such a difference between sound and light? Both propagate with velocities that are properties of the ‘medium’ yet I never heard that I needed to learn a new transformation to understand sound.

**Student**: That is so since all matter around us can move faster than sound, and we can take air along for the ride.

**Simplicius**: Can you not pack the æther and run?

**Student**: No, that is the point Einstein made clearly. If light velocity cannot be changed, than you cannot talk about taking æther along as then the light velocity could be changed. Conversely, for light velocity to be universal the æther must be indivisible, and the concept of velocity cannot apply to æther.

**Simplicius**: Still, I do not see why I need a new relativistic form of coordinate transformations?

**Student**: Because you want the light-clock to tick correctly when described (albeit in
very lengthy calculations) by Maxwell equations.

_Simplicius:_ OK, could I say that this transformation explains time dilation and Lorentz-FitzGerald contraction?

_Student:_ Actually, this new transformation is consistent with time dilation and Lorentz-FitzGerald contraction under well defined measurement prescription. Moreover, this transformation allows us to recognize that the speed of light is always the upper speed limit of all matter from which you can make mirrors.

_Simplicius:_ How does this last point arise?

_Student:_ If the mirrors of the light-clock were to move with \( v > c \), they could outrun the light in the clock – time could stand still or even tick backwards.

_Simplicius:_ OK, I see, for the mirror not to outrun the light \( c \) must be the upper speed limit of all matter that can interact with light.

_Student:_ All matter that interacts with light has yet more to do with light; just recall the most famous formula \( E_0 = mc^2 \). Note \( c^2 \) here.

_Simplicius:_ I never thought about \( E_0 = mc^2 \) like that: matter has locked rest energy which is the mass at rest multiplied with the square of speed of light. This means that whatever interacts with light is in some sense made of light.

_Student:_ We therefore call it ‘visible’ matter.

_Simplicius:_ And dark matter?

_Professor:_ That is the big conceptual question. We know observing the effects of gravity that there is around a lot of invisible, thus ‘dark’ matter. The question could ‘dark’ be gray is crucial. ‘Gray’ means visible, but with more effort. Then, the gray (not dark) matter must have the same mass-energy relation \( E_0 = mc^2 \), same limiting speed \( c \). Truly dark means whatever this stuff is, it does not know there is light. I wonder how could such dark stuff know speed of light? Truly dark matter could be a very interesting new physics situation. Interesting physics may be lurking in the Universe and we, literally, need to learn to ‘see’.

---

**Discussion II-4 – Riding the photon wave, seeking tachyons**

**Topic:** We discuss the consequences of the time clock results

_Simplicius:_ This root \( \sqrt{1 - (v/c)^2} \) troubles me. What if \( v > c \)?

_Student:_ First things first: note that when \( v \to c \) the light-clock stops ticking.

_Simplicius:_ Stops?

_Student:_ Yes, as \( v \to c \) one tick of the moving clock is eternity elsewhere. The clock can go anywhere in the Universe and not even tick once...

_Simplicius:_ ...running with the light wave.
Student: Yes, that is an important insight. A light-clock that moves almost as fast as the speed of light will hardly make a tick, yet it can go anywhere.

Simplicius: I see! Finally I understood – riding a photon the rider does not age and goes all over the Universe.

Student: Yes, I would love to catch and ride the light wave!

Simplicius: So again, what about $v > c$?

Student: Nothing much. The mirrors would outrun the light, and the light-clock makes no sense to me.

Simplicius: You seem to imply that there could not be matter moving faster than light just because your clock will not work?

Student: Well, the clock is just one example how and why matter that can interact with light must move with a speed that is limited by speed of light $c$. That is what we clearly have assumed when ‘building’ the clock.

Simplicius: However, I did come across tachyons[1], someone created such particles moving faster than light.

Student: You mean as an idea using a pen to write the word ‘tachyon’ on paper?

Professor: Let us not reject new ideas without proper scrutiny: we see tachyons in the scientific literature of mid-20th century[2] at the time invented to help handle in quantum field theory the situation in which the sign of the ‘mass squared’ needed to be changed. However a very different idea had at the time already resolved the problem[3], the recently discovered Higgs particle was predicted by virtue of these considerations.

Simplicius: Thank you! So what are tachyons today?

Student: A word in science fiction and movie shows characterizing material bodies that can be faster than light.

Professor: Even if I fully agree with this characterization, we must check to see if such particles somehow could make sense. Tachyons are beastly things that also were getting faster while losing energy. Anything that could move faster than the speed of light, and accelerate losing energy will always run far away.

Simplicius: But we could make such particles in laboratory, and if so, they could be made by cosmic rays elsewhere and come our way as a cosmic ray.

Professor: The search is on for new cosmic particles, tachyons, and many others. A discovery of a new particle, even one that is not really new in the sense that a tachyon would be, is a ticket to fame. So far nothing that moves faster than light has been found.

Student: We already noted that matter and light are related given $E_0 = mc^2$. So how

---

[1] Greek: ταχύς (tachyus), meaning “swift, quick, fast, rapid”


is it possible to look for matter with $v > c$?

_Simplicius:_ Why not try $E_d = mc^2_d$.

_Professor:_ That is not what was ordered for tachyons. These peculiar creatures were invented to be dependent on, and related to the speed of light, and thus, importantly in our present conversation, quite visible. Only their physical properties, energy and momentum, varied oddly as a function of their speed. That is why in the laboratory, our experimental colleagues could not fail to discover them, given their odd properties.

_Simplicius:_ How would they appear to us?

_Student:_ Faster than light but interacting with light means Mach cone shock waves of light. We see this happen when ultra relativistic particles travel in materials. Speed of light in matter is reduced, thus we have an effective model of tachyonic motion with all the unusual radiative properties.

_Professor:_ To my regret Tachyons are only words on paper. This also agrees with the findings of a few good people who question this theoretical idea in terms of mathematical consistency. But to be open minded we should always open our eyes. We have not seen anything tachyonic. We will return to discuss this topic in conversation $V-2$ on page 155. In the interim I recommend that if you are interested, a search of literature beginning with an early review by Bilaniuk and Sudarshan.$^4$

---

**Exercise II–2: Travel range of muons before decay**

We introduce particles called muons in the Insight on page 55: muons observed on Earth’s surface originate in the upper layers of Earth’s atmosphere from collisions of high energy cosmic ray particles with the atomic nuclei of Nitrogen and Oxygen as is illustrated in figure 4-3. These muons move with a high velocity. How great must the speed of a muon be for it to travel a distance of 13000 m before it decays?

**Solution**

Muons are created in cosmic ray processes and have a finite lifespan, thus their birth and death are distinct events. Considering their natural decay lifespan, each muon has to be viewed as equipped with a built-in light-clock which measures the muon’s proper time.

According to Eq. (4.2), in the Earth’s rest frame $E$, the lifespan of a muon is measured to be longer than the proper lifespan,

$$\frac{1}{\sqrt{1 - \left(\frac{v}{c}\right)^2}}$$

$^4$O.M. P. Bilaniuk, E. C. G. Sudarshan, “Particles Beyond the Light Barrier,” _Physics Today_, 43 (5) (1969), among their insights we read: “Proper mass of a superluminal particle is not an observable physical quantity; it is a parameter devoid of any immediate physical significance.”
Insight: Elementary Matter and the Muon

All matter around us is made of electrons $e$, and nucleons (protons $p$ and neutrons $n$). Electrons are bound by the Coulomb force in atomic orbits around atomic nuclei, which are much more localized on account of the stronger mutual interactions between nucleons, and $\simeq 1840$ times higher mass comparing a nucleon to the electron. Exploration of nucleon properties demonstrates that these particle have further constituents, the lighter and fractionally charged $u$ and $d$ quarks. The weakly interacting partner of the electron, the neutrino $\nu_e$ rounds off the ‘family’ of 4 distinct stable elementary particles we ‘see’ around us. However, we have further discovered two repetitions of these 4 fundamental elementary particles, for a total of three matter particle ‘families’. Among the 8 additional elementary matter particles is the muon. It is a 206.77 times heavier sibling of the electron. Both electrons and muons are subject to exactly the same electromagnetic and weak interactions. However, it is the heavier muon that by emitting neutrinos can decay to the lighter electron by a weak-interaction process. The proper mean lifespan of the muon is $\tau_{\mu} = 2197\text{ns}$, $c\tau_{\mu} = 658.654\text{m}$. We also note that the half-life of a muon; that is, the time it takes for half of the muons to decay is $\tau_{\mu}/2 = \tau_{\mu} \ln 2 = 1523\text{ns}$

Muons are produced most abundantly in the decay of charged pions $\pi^\pm$. These particles have a lifespan $\tau_\pi = 26\text{ns}$, $c\tau_\pi = 7.805\text{m}$. Pions are the most frequently produced particles in high energy collisions of strongly interacting nucleons as they are the lightest and also strongly interacting particles. Pions decay after a few 10’s meters and create a significant flux of muons in the cosmic rays that impact the surface of the Earth, see figure 4-3.

In this dilated lifetime $\tau_E$ a muon travels the mean distance $S_E = v\tau_E$. Therefore,

$$2 \quad \tau_E = \frac{S_E}{v}$$

We insert this result into the equation above and obtain

$$3 \quad S_E = \frac{\tau v}{\sqrt{1-(v/c)^2}}.$$  

Note that Eq.3 can be used to measure the lifespan of the particle by evaluating the distance traveled in a detector at measured velocity $v$. For a muon $c\tau = 658.654\text{m} \equiv S$, and in this example, $\gamma = 1/\sqrt{1-(v/c)^2} = 19.74$.

We solve for $v$ and obtain, using binomial approximation

$$4 \quad v = \frac{c}{\sqrt{1+S^2/S_E^2}} \simeq 1 - 0.5 \frac{S^2}{S_E^2}$$
Substituting in $S_E = 13\text{km}$, as well as $c = 2.998 \times 10^5 \text{km/s}$ yields

$$5 \ v = 0.9987 \ c = c - 385 \text{ km/s}$$

Thus a fast moving muon, just 385 km/s shy of the speed of light, can reach the Earth’s surface due to the large time dilation effect.

The muon survives the trip to the Earth’s surface because of the effect of time dilation, distinctly different from any other effect such as the Lorentz-FitzGerald body contraction. In addition, though we were discussing travel distance as occurring in the atmosphere, there is no material rigid body involved here and thus no body contraction. Certainly, there is no space contraction, or for that matter ‘distance contraction’. The reader must beware of ‘explanations’ that invoke any contraction effect in this context and recognize these as being incompatible with SR. We deepen this matter in the following conversation, and we return to this topic in conversation [III-1] on page [84] and exercise [VIII–11] on page [301].
Discussion II-5 – Why muons reach the surface of the Earth

**Topic:** All agree that muons can cross 10's of kilometers, the distance from their origin in upper atmosphere to the Earth’s surface, see the last exercise [II–2](#). However, a light ray could only travel 660 meters far during the muon mean lifespan \( \tau_\mu = 2.2 \mu s \). This situation is arguably most misunderstood SR contents. This conversation aims to clarify the circumstance.

*Simplicius:* Was the two-plane experiment (see conversation [II–2](#) on page 48 and section 14.1) the first measurement of time dilation?

*Student:* The effect of time dilation was recognized to govern particle decay and their ability to travel far.

*Simplicius:* I heard about muons living longer than naively expected. However, in another book I saw that this is due to ‘space contraction’.

*Student:* Within the scope of special relativity, space cannot be affected by the motion of a body. Was this a science fiction book?

*Simplicius:* No, it was a book on Modern Physics; it costs $200 and all physics students have to have it. To my friends on first glance, the explanation sounds good; the book even claims that the ability to explain muon impact on Earth either as time dilation or as the Lorentz- ‘space’ or ‘distance’ contraction proves the ‘consistency’ of special relativity. To help prove this claim they place a tall mountain in the book figure that will also contract.

*Student:* I am confused about the logic of this argument. If that story about space or distance contraction was right without the mountain, why do they need a mountain? I would think that the mountain in this argument means that they interpret material body contraction as being due to some other ‘space or distance’ contraction. This must mean that they insist that body contractions are driven by space contraction.

*Simplicius:* Yes, it looks odd to me now. What happens if I have two muons traveling together but at slightly different speeds: is the space contracted differently for each – what kind of physics is this?

*Student:* Let us remember that we are trying to figure out if muons travel far because of time dilation or because of something else. To see the issues more clearly, let us talk next about an observer in the train that is going through the mountain tunnel of our train & tunnel conversations, conversation [II–1](#) on page 47.

*Simplicius:* As I recall, the train clock was synchronized before the train left the station.

*Student:* Therefore we know the train clock will be slow once the train comes back after traveling through the tunnel.

*Simplicius:* OK, using that other book I can say that the tunnel was contracted and I needed less time to travel through it.

*Student:* On the other hand if and when the length of the tunnel is compared to the
contracted meter stick found on the train one would think more time is needed to cross the tunnel.

Simplicius: Agreed: I learned that the train that is sent on the trip is seen by an external observer to be shorter while traveling at high speed, and clocks on board tick more slowly.

Professor: Both effects, body contraction and time dilation occur together and relate to the body that is making the trip. The presence of both effects is necessary in order to have consistent reality; for example, the body contraction assures that the clock time does not depend on clock orientation. However, please remember that time dilation does not ‘explain’ Lorentz-FitzGerald body contraction and vice-versa. These two effects are different and relate to how we measure body proper time and body size.

Student: Creating the muon at a speed $v$ we assure that its clock, measuring muon proper time is slow compared to a laboratory observer clock that sees the muon moving at speed $v$, and, had the muon a size, it would be also contracted in direction of motion – that is the only contraction effect we can expect. Since muon is point-like in this consideration only time dilation enters.

Simplicius: But just hypothetically, could we not argue that as seen by an observer comoving with the muon the Earth has a speed $v$ and is instead contracted?

Student: If we could without making any other adjustment to the process of measurement reverse the situation as you propose, all people on Earth would also age more slowly. Wouldn’t it be great if we could slow our time by simply accelerating muons, or any other unstable particle in a device we construct? Clearly such contemplations are nonsensical.

Simplicius: While I would like to construct such a muon-driven-age-dilation machine, I also realize now how wrong this line of thought must be: just because there is a muon flying in space my aging would slow, or perhaps I would shrink, this clearly is a bad idea. To avoid more confusion, we need to explain the muon case in full.

Professor: We return several times to the challenge posed by the muon traveling in the Universe. We re-discuss the topic in conversation III–1 on page 84. How muon time dilation works is presented in section 7.3 on page 104. We look at a similar problem in exercise III–15 on page 106 and from a more elaborate perspective in exercise VIII–11 on page 301. There, we will also show that it is possible to argue that it is the Earth that is inertially moving towards the muon, the outcome is the same because the process of measurement is different: the time simultaneity is altered. But to clarify this point we need to move on reading the book together.
5 Lorentz-FitzGerald Body Contraction

5.1 Universality of time measurement

In section 4.1 we derived time dilation using only the principle that light propagates at a constant speed independent of the motion of the mirrors. We have considered the case of one particular orientation with respect to the direction of motion: the mirrors were mounted on a body with the mirror normal being exactly orthogonal to the direction of the velocity vector $\vec{v}$.

However, the principle of relativity requires that time measurement cannot depend on a particular orientation of the clock. Two inertial observers at rest with respect to each other must measure the same proper time, irrespective of the orientation of the mirror clock they are using. To see why this must be, take two identical but differently oriented pairs of light-clocks. One pair of clocks is in the laboratory, and the other pair goes with the traveler.

A laboratory observer sees her clock pair ticking synchronously with one another. The traveler in motion notes that his two clocks tick differently from each other. If this indeed could happen, the traveler would be measuring his velocity vector even during the period that he is coasting inertially. This violates the principle of relativity, and thus the moving clocks must remain synchronized irrespective of their orientation; in other words, the measurement of time is universal.

In order to achieve this result, it is necessary to introduce the Lorentz-FitzGerald contraction of a body. In essence when we discuss the situation below, we retrace in our own language the arguments that led to this discovery following the recognition that the Michelson-Morley interferometer did not reveal the motion of the Earth. Both the light-clock and the null outcome of the Michelson-Morley experiment rely, in the end, on the requirement that the optical path length is independent of the orientation of the mirrors.

5.2 Parallel light-clock

We consider a light-clock moving at constant velocity relative to the observer, but this time with the motion in the same direction as the light beam, as depicted in figure 5-1.

We now determine the time $t_{m1}$ light needs to travel from the left mirror $M_1$ to the right mirror at a new location denoted $M'_2$ in figure 5-1. We see that by the time the light reaches the right mirror, it is an extra distance $x$ further away. To find the time light travels we note that it is the time it takes the mirror to move the distance $x$ at velocity $v$:

$$t_{m1} = \frac{x}{v}.$$
This is also the flight time of the light from the left mirror to the new position of the right mirror:

\[ t_{m1} = \frac{l' + x}{c}. \]

Here in anticipation of the contraction of the body we have called the separation between the two mirrors \( l' \). Setting both expressions equal to each other

\[ \frac{x}{v} = \frac{l' + x}{c}, \]

allows us to determine \( x \)

\[ x = \frac{l' \, v/c}{1 - v/c}. \] (5.1)

On the return path of light, the left mirror moves to right by \( x' \) reducing the amount of time needed by light to reach the mirror, thus respectively,

\[ t_{m2} = \frac{x'}{v}, \quad t_{m2} = \frac{l' - x'}{c}, \]

and from these relations we obtain,

\[ x' = \frac{l' \, v/c}{1 + v/c}. \] (5.2)

The entire distance covered by the light beam is

\[ S = (l' + x) + (l' - x'), \]

or, when one substitutes in \( x \) and \( x' \) from equations 5.1 and 5.2,

\[ S = 2l' + \frac{l' \, v/c}{1 - v/c} - \frac{l' \, v/c}{1 + v/c}. \]
After rearranging to give a common denominator we find

\[ S = \frac{2l'}{1 - (v/c)^2}. \quad (5.3) \]

The period of the clock oriented in parallel to the direction of motion is then:

\[ T_{\text{parallel}}^B = \frac{S}{c} = \frac{2l'}{c} \frac{1}{1 - (v/c)^2}. \quad (5.4) \]

We recall for comparison the result for the rotated clock, Eq. (4.5)

\[ T_{\text{orthogonal}}^B = \frac{2l}{c} \frac{1}{\sqrt{1 - (v/c)^2}}. \quad (5.5) \]

### 5.3 Body contraction

We now have two different expressions for the clock period, and if \( l = l' \) it is impossible that in general Eq. (5.4) and Eq. (5.5) are equal – other than for \( c \to \infty \). It is natural to require that our light-clock should measure time independent of the orientation of its mirrors. If this is not the case we could, for example, measure absolute velocity by comparing the time shown by two differently oriented clocks. Moreover, it would really not seem possible to measure time correctly. Therefore we now require that the clock time be independent of orientation. The condition that the two clocks score the same time is obtained by setting Eq. (5.4) equal to Eq. (5.5),

\[ \frac{2l'}{c} \frac{1}{1 - (v/c)^2} = \frac{2l}{c} \frac{1}{\sqrt{1 - (v/c)^2}} \]

\[ (5.6) \]

We find that to maintain the same period, the length \( l' \neq l \). The change in length is a function of the velocity

\[ l' = l\sqrt{1 - (v/c)^2}. \quad (5.7) \]

This equation says that the body on which the mirrors are mounted undergoes compression in the direction of motion. FitzGerald in 1889 and Lorentz in 1892 independently proposed the hypothesis that a body would be contracted in the direction of its motion by the factor \( \sqrt{1 - (v/c)^2} \). This effect is known today as the Lorentz-FitzGerald body contraction.

Henceforth to shorten the notation we will use the dimensionless velocity

\[ \beta = \frac{v}{c}. \quad (5.8) \]
and the Lorentz factor

\[ \gamma = \frac{1}{\sqrt{1-(v/c)^2}} = \frac{1}{\sqrt{1-\beta^2}} \]  

(5.9)

The Lorentz-FitzGerald body contraction can be written

\[ l' = \frac{l}{\gamma} \]  

(5.10)

and time dilation becomes

\[ t' = \gamma t \]  

(5.11)

Compression of a body traveling through æther in the direction of motion seems to be a natural hypothesis to reconcile the two expressions seen Eq. (5.6). However this also means that the clock will tick the same, independent of its orientation. When such magic occurs, one usually expects a more complete theory to follow. Within such a theory the Lorentz-FitzGerald body contraction is a necessary element; it is required by virtue of both the universality of the speed of light and the principle of relativity.

Note that the comparison we have performed between the parallel and orthogonal to motion orientations of the light-clock is just the same as the comparison of the optical paths for the two arms of the Michelson-Morley interferometer, see figure 2-1 on page 21. Thus the result that the clock will function independent of its orientation is equivalent to the explanation of why the Michelson-Morley experiment will not show a fringe shift when the interferometer is rotated.

Historically, the Lorentz-FitzGerald contraction was the first step leading towards the concept of time dilation, for it is easy for us to accept that a body’s shape changes, and far more difficult to agree that we age at different, personal, rates. Einstein was 13 years old at the time the Lorentz contraction entered the mainstream of scientific discussion. At age 26 he realized that a consistent theory requires a change in the way we think about time, proposing that each body must have its proper time.

Lorentz recognized that his contraction should be derived directly from Maxwell’s theory of electromagnetic interactions. However, his efforts to show this were hindered by introduction of material æther. Without material æther Lorentz’s idea worked. Today we can expand these considerations to all known forces and recognize that the Lorentz-FitzGerald body contraction is a universal phenomenon applying to all visible material bodies.

To summarize: consider the Lorentz-Fitzgerald contraction of a metal rod, shown in figure 5-2. When it is oriented perpendicular to its direction of motion, it has the length \( l = l_0 \); when it is turned so that it is parallel to its direction of motion, it will have the length

\[ l = l_0 \sqrt{1-(v/c)^2} \]  

(5.12)
The principle of relativity assures that no physical observable can depend on the value of $v$. For example, a meter stick attached to the body also contracts, so measurement of the contraction effect by the comoving observer is impossible; it would amount to measurement of velocity, violating equivalence of all inertial observers and allowing for the existence of some absolute reference body. However, we can measure the effect of the Lorentz contraction if we set up an experimental circumstance in which we change the body velocity. We will return in section 10 to discuss how one can then measure the reality of the Lorentz-FitzGerald body contraction.

**Discussion II-6 – Relativistic train entering a tunnel**

**Topic:** We discuss the Lorentz-FitzGerald contraction of a train at relativistic speed. This leads us to the question, how we can tell who is accelerated in the Universe?

**Professor:** We need to perform a measurement to determine that the tunnel is indeed shorter than the train. This must be done in the frame of reference for which the mountain and the train are at rest, e.g. the train station. After this, we start the train engine and slowly pick up speed.

**Simplicius:** I ride on the train. I measure the train always to have the same length irrespective of its speed. Thus I know it will be too long to fit in the tunnel.

**Student:** If the train should be contracted, your measuring rod will be also contracted. Using a contracted rod, you cannot realize the train has become shorter.

**Simplicius:** How can I measure differently?

**Professor:** Sitting on the station you measure the moving train differently compared to when you are riding the train. This is so because you always measure the size of an
object at equal time in your ‘proper’ frame in which you are at rest. These two measurements (station and train) are different in that what is simultaneous to one inertial observer is not to another moving with a finite relative velocity.

*Simplicius:* Please confirm: inertial observers do not agree what is simultaneous?

*Student:* As long as the speed of light is finite this is so.

*Simplicius:* If so, ‘my’ frame of reference must be special: all experiments, measurements, and observations refer to my frame of reference.

*Student:* Yes, that is your ‘proper’ frame of reference, you have your ‘proper’ time, and your ‘proper’ standard of length.

*Simplicius:* When I ride the train, in my proper frame I will always measure the same train length, the same ‘proper’ length. How can I understand the body contraction if I am a contracted passenger on the train?

*Professor:* Imagine the tunnel has two end-gates which can be shut locking in the moving train. We drop the gates simultaneously in the mountain frame of reference. For the rider on the train, the drops in front and in the back occur at two different times. This is so since what is simultaneous for an observer on the mountain is not simultaneous in the moving train frame of reference. The difference is just enough for the moving train not to be hit by dropping gates.

*Simplicius:* Is there another way to explain this?

*Student:* No, there are only these two ways: a) Observing from the station frame of reference, when the train trip started you conclude the moving train is contracted, this is the Lorentz body contraction; b) You sit on the train and do not realize you are contracted but you see the two gates, one at front, and one in the back lock in the train at different times even if you were told by someone at the station that the locking will be done at the same time.

*Professor:* By realizing that events simultaneous to one inertial observer cannot be simultaneous to another (unless they have zero relative velocity), Einstein resolved observational contradictions that the Lorentz-FitzGerald body contraction can generate. That is so since a measurement requires a statement of how time is treated. Moreover, we see that the Lorentz contraction is not reversible; that is, you cannot claim that sitting on a train you see a contracted mountain approach you. When making this statement you changed the measurement’s ‘proper’ frame. Given that the train and the mountain were compared at the station where the ride started, the comparative body length measurement that makes sense is the one carried out from the station reference frame. All this will be shown in technical detail in section 9.

Discussion II-7 – *Is the Lorentz-FitzGerald contraction ‘natural’?*

*Topic:* The realization of the Lorentz-FitzGerald contraction was a pivotal step on the
winding path to relativity theory. We discuss how and why, given the MM experimental result, this step had to happen. The topic expands as we try to resolve many questions the reader will have at this time.

Simplicius: With Maxwell’s recognition that electromagnetic waves and light propagate at the same ‘light’ speed, even I could argue that, just like sound in air, light and Maxwell waves propagate in some new medium, the ‘Æther’.

Student: I agree, it is hard not to see it differently given what was known at that time. Even today, everything we know has a maximum speed $c$ provided by this obscure new medium according to Einstein, just like the air provides the sound speed. Thus for Maxwell, there had to be an Æther filling the Universe. The MM interferometry experiment was devised to observe how the æther responds to the motion of the Earth orbiting the Sun against the stationary æther. The outcome of the experiment was all but that.

Simplicius: Why was this such an important question to pose?

Student: If there is a material æther, why is matter not hindered in motion and yet Maxwell’s waves are transverse vibrations with respect to the direction of wave propagation? Maxwell understood that transverse waves arise in a hard and rigid medium. Sound in air is a longitudinal compression wave; in general this means that the medium is relatively easy to penetrate, and indeed air can be tossed around. Thus it was not understood how material bodies could move at all in the æther without an observed resistance. How could we not see an effect on the orbit of the Earth around the Sun?

Simplicius: I see, that is why no effect of motion reported by MM added to the enigma surrounding the æther.

Professor: In this context the body contraction hypothesis seemed to offer an escape, postponing to a later stage (see below) the conflicts with notion of material æther.

Simplicius: To me it seems that there are other interpretations. I could argue that the null result of MM experiment and the stability of Earth orbit simply mean that the Earth, and the æther, are comoving together around the Sun.

Student: Local motion of the carrier of light would affect the path of star light, alter the path of light and the value of the speed of light. The agreement of terrestrial and astrophysical methods of light speed measurements exclude this interpretation.

Simplicius: I see, but what you say about these experiments seems to be excluding any chance of MM observing an effect.

Student: Foremost, that comment clearly relates to experimental precision. A big æther drag effect on light can be excluded. But some small drag seemed possible and the MM experiment was designed to measure a directional difference in the speed of light rather than the absolute speed of light.

Simplicius: How could all this lead to a body contraction hypothesis? Let me take a ruler, and rotate it in space. While I do this my ruler is compressed and decompressed and yet I do not notice much happening to it.
Professor: The magnitude of the effect expected by Lorentz, and independently by FitzGerald was about 5 parts in a billion, so the body change was undetectable in the experiment you propose, but fixed the problem of the absence of any MM interference fringe shift.

Student: This is much smaller than the typical thermal expansion of the ruler when temperature changes by one degree, an effect that is always present when you pick up anything with your hands.

Professor: This discussion also shows the need for precise control of temperature and many other environmental conditions in order to achieve the required experimental precision in modern MM experiments, see section 14.2.

Student: The Lorentz-FitzGerald body contraction was the departure point of the search Lorentz initiated for the relativistic coordinate transformations that leave the form of Maxwell equations invariant.

Simplicius: Trying to understand a body contraction I come to think about resistance to motion by a material æther.

Student: We will return to discuss this topic, see conversation IV-2 on page 141. We will recognize that the electromagnetic force field emanating from the moving charged nuclei compresses atoms, and hence a chain of cohesively bound atoms, and more generally, any rigid material body, producing the Lorentz-FitzGerald body contraction. Thus it is not the direct reaction from the æther, but the effect of Maxwell’s equations that cause body contraction.

Simplicius: You are now explaining body contraction as being due to the nature of electromagnetic forces, yet we seem to find all these effects also without any reference to forces, but from rather general considerations.

Professor: Understanding body contraction is similar to the understanding of energy conservation: at first we learn this as a principle, and later we study the outcome in terms of microscopic interactions that must be consistent with this principle.

Simplicius: Except that there seems to be nothing conserved regarding the body size. We turn a ruler into the direction of motion and it contracts, ever so slightly.

Professor: This works because all laws of physics respect the rules of Lorentz coordinate transformations. This assures that we can look at the effects of special relativity as being either an outcome of dynamics; i.e., forces, or as an outcome of coordinate transformations.

Simplicius: So all physical properties of a body depend on its velocity – we have discussed energy, momentum, size, – is there anything that does not change?

Student: Yes, there are invariant properties of a physical body which all observers will measure to be the same, for example the inertial mass \( m \). But kinetic energy, or here the observed body length, are clearly not among these invariant properties.

Simplicius: Yet people speak of proper time.

Professor: I agree, the proper time is the time measured by the clock attached to the
body. An unstable particle has a proper time lifespan that is invariant.

Exercise II–3: Example of magnitude of body contraction

A rod moving in the x-direction is found to have a length in the x direction of
\( \ell = \frac{2}{5} \ell_0 \), where \( \ell_0 \) is its proper length. At what velocity \( v \) is the rod traveling?

Solution

The Lorentz contraction of the rod is given by

\[ \ell = \ell_0 \sqrt{1 - \frac{v^2}{c^2}}. \]

Solving for \( v \) yields

\[ v = c \sqrt{1 - \frac{\ell^2}{\ell_0^2}} = \frac{4c}{5}. \]

Exercise II–4: Body orientation

A rod, as shown in figure 5-3 is at an angle \( \theta^0 \) with respect to the horizontal, starts to move along the horizontal and reaches the speed \( v \) relative to the laboratory frame. Obtain the angle \( \theta \) the rod now makes with the horizontal.

Solution

In its rest frame, the rod has length parallel to its motion of \( \ell_x^0 \) and perpendicular length \( \ell_y^0 \), related to the angle \( \theta^0 \) by:

\[ \tan \theta^0 = \frac{\ell_y^0}{\ell_x^0}. \]

Using Eq. (5.12), we find that in the laboratory frame, the component of length parallel to the motion of the rod is contracted by

\[ l_x = \frac{\ell_x^0}{\ell_x} \sqrt{1 - \left(\frac{v}{c}\right)^2}, \]
while the component perpendicular to the motion is unchanged:

\[ l_y = l_y^0. \]

We now relate these quantities to the angle \( \theta \) that the rod makes with the horizontal in the lab frame

\[ \tan \theta = \frac{l_y}{l_x} = \frac{l_y^0}{l_x \sqrt{1 - (v/c)^2}}. \]

Substituting in Eq. 1, we have

\[ \tan \theta = \frac{\tan \theta^0}{\sqrt{1 - (v/c)^2}}, \]

which has the solution

\[ \theta = \tan^{-1} (\gamma \tan \theta^0) . \]

---

**Exercise II–5: Light clock in arbitrary orientation**

Earlier, our examination of the light-clock oriented perpendicular to its motion revealed time dilation. The principle of relativity required that time dilation affects all processes equally in any given inertial frame. Because light travels at the same speed for any observer, we were obliged to invoke the Lorentz contraction hypothesis
Figure 5-4: An arbitrarily oriented light-clock: on left at rest and on right in motion at velocity $\vec{v}$, see exercise II–5.

to explain how a light-clock oriented parallel to its motion could experience the same time dilation as the one oriented perpendicular. The task now is to show that when the Lorentz-FitzGerald contraction is applied, a light-clock in any orientation relative to its motion will be subject to the same universal time dilation.

Solution

We describe our light-clock in its rest frame with a displacement between the mirrors of $a$ in the direction parallel to the velocity and $b$ in the direction perpendicular, as shown in figure 5-4. We have only the restriction that the distance between the mirrors in the rest frame is constant in any orientation:

1. $a^2 + b^2 = l^2$

From our previous discussion (see Eq. (4.3)), we know that the period of this clock in the rest frame is

2. $I_A = \frac{2l}{c}$.

Now we must consider its period $I_B$ as observed when in motion at velocity $\vec{v}$, as shown in figure 5-4. The total period is given by the time taken to travel along $S_1$ and $S_2$

3. $I_B = t_1 + t_2$.

Analogous to our earlier discussion of the light-clocks, we find two expressions for the time of travel along each section of the path

4. $t_1 = \frac{S_1}{c}$,
and

5. \[ t_1 = \frac{y_1}{v} = \frac{y_1}{c\beta}. \]

Setting these quantities equal and squaring we obtain

6. \[ y_1^2 = \beta^2 S_1^2. \]

To find the path length followed by the light, we must consider also that the light-clock is contracted in the direction of its motion; the displacement between the mirrors in the direction of the motion becomes \( \frac{a}{\gamma} \). From the geometry of the problem, see figure 5-4, we now see that the distance traveled by the light is

7. \[ S_1^2 = \left( \frac{a}{\gamma} + y_1 \right)^2 + b^2. \]

Substituting in to the previous equation and expanding we obtain

8. \[ y_1^2 = \beta^2 \left( \frac{a^2}{\gamma^2} + \frac{2a}{\gamma} + y_1^2 + b^2 \right)^2. \]

Rearranging terms and recognizing that \( 1 - \beta^2 = \gamma^{-2} \) yields

9. \[ \gamma^{-2} y_1^2 - \frac{2a\beta^2}{\gamma} y_1 - \left( \frac{a^2}{\gamma^2} + b^2 \right) \beta^2 = 0. \]

We now solve the quadratic equation for \( y_1 \)

10. \[ y_1 = \frac{1}{2} \gamma^2 \left( \frac{2a\beta^2}{\gamma} \pm \sqrt{\frac{4a^2\beta^4}{\gamma^2} + \frac{4\beta^2}{\gamma^2} \left( \frac{a^2}{\gamma^2} + b^2 \right)} \right) \]

which simplifies to

11. \[ y_1 = \gamma \beta \left( l + a \beta \right), \]

where we have chosen a positive sign for \( l \) because we have defined the term \( y_1 \) as positive, and \( a \beta < l \). We can apply the same method to obtain \( y_2 \), but we have to use negative sign for \( l \).

12. \[ y_2 = \gamma \beta \left( l - a \beta \right). \]

The total time traveled is given by

13. \[ I_B = t_1 + t_2 = \frac{y_1}{c \beta} + \frac{y_2}{c \beta}. \]
Substituting in our results we find

$$I_B = \frac{2\gamma l}{c}.$$  \tag{14}

This shows that the time the light takes for a return trip between mirrors is independent of the orientation of the clock. As a consequence, the time dilation effect is universal as is required by the principle of relativity.

This exercise demonstrates that a moving body subject to Lorentz-FitzGerald contraction can provide the material body for a light-clock that registers time dilation. This demonstrates that the time dilation and Lorentz-FitzGerald contraction are two different effects which, when combined together, lead to a consistent theoretical framework.

Note that this is exactly the computation needed to show that as the Michelson-Morley interferometer is rotated (see figure 2-1 on page 21), there is no expected change in the optical paths of the light, independent of the orientation, and thus as long as the speed of light $c$ is a constant, we cannot observe any effect of body motion.

End II–5: Light clock in arbitrary orientation
5. Lorentz-FitzGerald Body Contraction
Part III

The Lorentz Transformation
Introductory remarks to Part III

In Part III our objective is to understand and characterize a change in reference frame from one inertial to another inertial observer in the context of SR. The relativistic coordinate transformation was named by Poincaré after Lorentz, who was the first to systematically attempt to resolve the incompatibility of Maxwell’s electromagnetism with the Galilean transformations\(^{1}\). However, the transformations were derived from first principles by Einstein, and, in a different manner, by Poincaré nearly at the same time in Spring-Summer 1905. Generalizing the Galilean coordinate transformations to SR we follow arguments presented by Einstein.

We begin clarifying that Lorentz coordinate transformations are passive, the body is still in the same state of motion as before; it is the observer changing her frame of reference. We show how the three physics inputs determine the form of Lorentz transformation: the isotropy and homogeneity of space, the principle of relativity discussed in Part I, and the universality of the speed of light.

Classic results of relativity are derived in this Part III: the addition theorem of velocities is presented for the simple colinear, and general case of two arbitrary velocities. The invariance of proper time under Lorentz transformations is demonstrated twice, using two different methods. The rapidity, replacing body speed, is introduced. The merits of rapidity are demonstrated by showing the additivity property for colinear motion.

Among physics examples that have impacted the development of SR, the Fresnel light drag by a fluid in motion is shown to emerge from velocity addition theorem. Another important application of Lorentz transformations is the aberration of light. Aberration is the modification of the perception of direction of incoming light emitted by a source in relative motion with the observer. We study the aberration effect for a focused light-ray emitted by a star. Our discussion is based on Lorentz coordinate transformations without a reference to any wave phenomena such as Doppler shift, to be addressed in section 13 in Part V. The vector character of the aberration effect is discussed, showing that only transverse motion relative to the line of sight contributes.

\(^{1}\)Lorentz transformation should better be called Larmor-Lorentz-Einstein-Poincaré LLEP-transformation, to honor also (see section 3.3): a) the actual first presentation of the correct Lorentz transformation by Sir Joseph Larmor, the SR inventor Einstein, and Poincaré, who recognized the group properties of relativistic coordinate transformations. This correct naming could have avoided the confounding name-connection between the relativistic Lorentz coordinate transformation and the Lorentz-FitzGerald body contraction— which many simply call the Lorentz contraction. The name similarity is linking (incorrectly) in minds of many the space-time coordinate transforms to the body contraction property.
6 Relativistic Coordinate Transformation

6.1 Derivation of the form of the Lorentz transformation

Passive and active transformations

There are, in principle, two ways to address coordinate transformations:

1. We can transform actively; that is, we can move by the transformation the coordinates of the world line traced by a particle and retain the coordinate system in place.

2. We can transform passively; that is, we change the coordinate system and we leave the world line unchanged.

Anyone who mastered this detail about active and passive transformation in the Galilean world can verify her answer by considering the limit \( v/c \to 0 \) in SR. Our derivation of the Lorentz transformation will address a passive transformation; that is, a true coordinate transformation in which we evaluate how the coordinates of a body change when comparing different observers. If, however, we want to consider an active transformation operating on a physical body, e.g. such that a body observed by comoving observer as being at rest is made to be moving with \( \vec{v} = v \hat{i} \), then to achieve this we must transform the comoving frame of reference with \( \vec{v} = -v \hat{i} \).

The reason we choose to focus on passive transformations when addressing SR is, that despite the seeming equivalence in the context of any change in motion can be distinguished by the measurement process. Here we must, for example, always remember that the clock that is time dilated is attached to the twin who made the trip away to the stars, and returned.

The situation is illustrated in figure [6-1] where we see two observers \( S \) and \( S' \) recording the motion of the body \( K \) with velocity \( \vec{u} \) on left and \( \vec{u}' \) on right. There are two coordinate systems \( x, y, z, t \) and \( x', y', z', t' \) attached to, and at rest with respect to these two \( S \) and \( S' \) observers; in figure [6-1] we only show \( x, y \) and \( x', y' \) coordinates.

In the left frame we see observer \( S \) at rest and observer \( S' \) in motion to right with a speed \( v \), while in the right frame we see observer \( S \) in motion to the left also with a speed \( v \) while \( S' \) is at rest. Thus the relative velocity of the two observers is the same in both cases, while the velocity of the body as measured by either of the two observers is different in both cases. While \( \vec{u} \) is recorded by \( S \) on the left, \( \vec{u}' \) is measured by \( S' \) on the right.

Under the Galilean coordinate transformation we read off figure [6-1], the location of the body: with time the origin of \( S' \) approaches the body so we must have \( \vec{x}' = \vec{x} - t\vec{v} \) which when differentiated with respect to time produces the velocity
Figure 6-1: On left: a body $K$ is moving with velocity $\vec{u}$ with respect to the observer $S$, while another observer $S'$ is in motion with velocity $\vec{v}$. On right: the same body $K$ is moving with velocity $\vec{u}'$ with respect to the observer $S'$, while the observer $S$ is in motion with $-\vec{v}$. See text for further discussion.

addition theorem $\vec{u}' = \vec{u} - \vec{v}$. Exchanging the role of $S$ with that of $S'$ we obtain the inverse transformation $\vec{x} = \vec{x}' - t\vec{v}' = \vec{x}' + t\vec{v}$ and hence $\vec{u} = \vec{u}' - \vec{v}' = \vec{u}' + \vec{v}$. These relations form those seen in section 2.2, see Eq. (2.1) and Eq. (2.2) in that we consider here the passive transformation.

Using the isotropy and homogeneity of space

The isotropy of space means that there is no preferred direction. Assuming isotropy it suffices to consider one arbitrary direction of motion and the results we find are valid generally. Thus we can choose the relative velocity $\vec{v}$ to be parallel to one of the coordinate axes, say the $x$-axis of the coordinate systems $S$. We consider coordinate transformation such that the transformed coordinate system $S'$ has its $x'$-axis parallel to the $x$-axis. By making these two axes $x$ and $x'$ parallel we exclude from current consideration rotation transformations superposed on top of the Lorentz transformation along the same axis, which often in professional slang are called ‘(Lorentz) boost(s)’.

The homogeneity of space means that the two observers $S$ and $S'$, with their coordinate systems $x, y, z, t$ and $x', y', z', t'$, respectively, both recognize the forceless motion of a body as being linear and uniform. We call this body motion ‘inertial’.

Thus, if a body $K$ in system $S$ moves with constant velocity $\vec{u}$, its velocity $\vec{u}'$ in system $S'$ should likewise be constant. This requirement can only be fulfilled when the coordinate transformation involving both time and space is a linear transformation.

\[
x' = a_{11}x + a_{12}t + k_1, \tag{6.1}
\]

\[
t' = a_{21}x + a_{22}t + k_2, \tag{6.2}
\]

\[
y' = y. \tag{6.3}
\]
\[ z' = z. \] (6.4)

We now choose coordinate transformations between \( S \) and \( S' \) such that at a prescribed instant in time \( t'_0 = t_0 = 0 \) we also have \( x'_0 = x_0 = 0 \). We must have

\[ k_1 = k_2 = 0. \] (6.5)

This means that for both coordinate systems the origins coincide. By making this choice we do not consider here transformations which are called translation in space (for \( k_1 \neq 0 \)) or/and translation in time (for \( k_2 \neq 0 \)).

As shown in figure 6-1, system \( S' \) moves relative to \( S \) with velocity \( \vec{v} = v\hat{i} \). We want to consider the movement in \( S \) of the coordinate origin of \( S' \). This corresponds to the tracking of \( x' = 0 \). Note that we are thus considering how the coordinate system transforms, which is so-called ‘passive’ transformation.

For the origin of \( S' \) moving with velocity \( v \) we obtain the equation

\[ a_{11}x + a_{12}t = 0, \] (6.6)

For a small increment in space \( dx \) and time \( dt \) we thus have

\[ -\frac{a_{12}}{a_{11}} = \frac{dx}{dt} \equiv v. \] (6.7)

The last equality is the definition of \( v \) given that \( x' = 0 \) and our choice to consider motion along the \( x \) axis only. We use this result in Eq. (6.1) and obtain

\[ x' = a_{11}(v)(x - vt). \] (6.8)

Here the notation \( a_{11}(v) \) reminds us that the coefficient \( a_{11} \) is in general a function of \( v \), as is inherent in Eq. (6.7).

**Using the principle of relativity**

As we have already presented in figure 6-1, we can consider the motion in frame \( S' \); then \( S \) moves relative to \( S' \) at \( \vec{v}' = -\vec{v} \). By the principle of relativity, both points of view are equally valid. The transformation from \( S' \) to \( S \) must then take on the same form as Eq. (6.8):

\[ x = a_{11}(v')(x' - v't'). \] (6.9)

The velocity \( v' \) in Eq. (6.9) is the velocity of \( S \) relative to \( S' \), whereas \( v \) in Eq. (6.8) is the velocity of \( S' \) relative to \( S \), therefore \( v' = -v \). We thus have

\[ x = a_{11}(-v)(x' + vt'). \] (6.10)
We could have chosen the coordinates such that both the $x$- and $x'$-axis pointed in opposite directions. In this case we modify relations Eq. (6.8) and Eq. (6.10) by transformations $x \rightarrow -x$, $x' \rightarrow -x'$ and $v \rightarrow -v$. After imposing this transformation Eq. (6.8) and Eq. (6.10) read

$$x' = a_{11}(-v)(x - vt) \quad x = a_{11}(v)(x' + vt'). \quad (6.11)$$

To assure that the transformation we are seeking is not dependent on the choice of the direction of the coordinate axis, the coefficient $a_{11}$ must not be dependent on the sign of the velocity; that is, on the direction of $\vec{v}$. Therefore $a_{11}(v) = a_{11}(-v)$ must be true.

We implement this by writing $a_{11}(v^2)$. We then have, restating Eq. (6.11):

$$x' = a_{11}(v^2)(x - vt) \quad x = a_{11}(v^2)(x' + vt'). \quad (6.12)$$

Using the universality of speed of light.

In order to determine the coefficient $a_{11}(v^2)$, we consider the light ray coordinates in $S$, see figure 6-2. At time $t = t' = 0$ the origins of the two systems coincide. At this point in time a flash of light is sent out from the origin. In system $S$ at time $t$ the flash of light (light cone) reaches the position

$$x = ct, \quad (6.13)$$

shown in figure 6-2.

Similarly in $S'$

$$x' = c't', \quad (6.14)$$

Figure 6-2: Space-time path of a flash of light in system $S$
The speed of light should be the same for all inertial observers; this is a direct experimental input into the reasoning about the new coordinate transformation we are seeking. To justify this hypothesis recall that we have found in section 1.3 that $c$ was the same if measured on Earth or in interstellar space, and we have also described there the result that Maxwell’s equations provide for waves that propagate with light velocity.

We thus demand that the speed of light is observed to be the same in all inertial systems in any state of motion,

$$c = c' \quad (6.15)$$

Therefore in system $S'$ at time $t'$ the light cone reaches the position

$$x' = ct'. \quad (6.16)$$

We substitute these two conditions, i.e. $t = x/c$ and $t' = x'/c$ into Eq. (6.12) and obtain how the flash of light observed in one system is observed in the other. This results in two transformation constraints

$$x' = a_{11}(v^2)x \left(1 - \frac{v}{c}\right), \quad x = a_{11}(v^2)x' \left(1 + \frac{v}{c}\right). \quad (6.17)$$

which applies only to the light cone coordinates, but which must be consistent.

Multiplying the left hand side and right hand side of these two equations with each other and canceling the common factor $x'x$ we obtain

$$1 = a_{11}(v^2)a_{11}(v^2) \left(1 - \frac{v}{c}\right) \left(1 + \frac{v}{c}\right), \quad (6.18)$$

and thus

$$a_{11}^2(v^2) = \frac{1}{1 - (v/c)^2}. \quad (6.19)$$

In consideration of the limiting case $v/c \rightarrow 0$ which must yield the Galilean coordinate transformation, we choose the positive root\(^2\) and obtain:

$$a_{11} = \frac{1}{\sqrt{1 - (v/c)^2}} \equiv \gamma. \quad (6.20)$$

With Eq. (6.7) we have also

$$a_{12} = \frac{-v}{\sqrt{1 - (v/c)^2}} \equiv -\beta c \gamma. \quad (6.21)$$

\(^2\)Transformations which reverse the direction of time (time reversal) and/or of space (parity) which would require the negative root are called “improper LT”. Such transformations are part of the Poincaré group of all space-time transformations.
6.2 Explicit form of the Lorentz transformation

With Eq. (6.21) and Eq. (6.20) we can write for the spatial part of the coordinate transformation:

\[
x' = \frac{x - vt}{\sqrt{1 - (v/c)^2}} \quad \text{or} \quad x' = \gamma(x - \beta ct).
\]  

(6.22)

Next we determine the coefficients \(a_{21}\) and \(a_{22}\) required in the transformation equation for the time coordinate \(t'\), Eq. (6.2). \(t'\) is seen also in Eq. (6.12) on the right hand side. We solve Eq. (6.12) for \(t'\)

\[
t' = \frac{x}{va_{11}} - \frac{x'}{v};
\]

here \(a_{11}\) is given in Eq. (6.20) and \(x'\) in Eq. (6.22).

\[
t' = \frac{x}{v\gamma} - \frac{\gamma(x - vt)}{v} = \gamma \left( t - x \left( \frac{1}{v} - \frac{1}{v\gamma^2} \right) \right)
\]

and we obtain

\[
t' = \frac{x}{v\gamma} - \frac{\gamma(x - vt)}{v} = \gamma(t - (v/c^2)x).
\]  

(6.23)

We thus have

\[
t' = \frac{t - (v/c^2)x}{\sqrt{1 - (v/c)^2}} \quad \text{or} \quad ct' = \gamma(ct - \beta x).
\]  

(6.24)

We can now read off Eq. (6.24) the coefficients of the transformation equation 6.1

\[
a_{21} = -\frac{v}{c^2} \frac{1}{\sqrt{1 - (v/c)^2}} = -\frac{1}{c^2} \beta \gamma, \quad a_{22} = \frac{1}{\sqrt{1 - (v/c)^2}} = \gamma.
\]  

(6.25)

The two transverse coordinates are not modified as we excluded rotations of the coordinate system from present consideration;

\[
y' = y, \quad z' = z
\]  

(6.26)

We can also write for Eq. (6.24) the modified exact form

\[
t' = t \sqrt{1 - (v/c)^2} - \frac{x'v}{c^2}.
\]  

(6.27)

One easily checks the equivalence using \(x'\) as given in Eq. (6.22). Thus Eq. (6.27) describes time dilation for coordinate transformations for which \(x' = 0\), thus when \(t'\) is the proper time of a body, see section 4.1 and Eq. (4.2).
Eq. (6.27), and Eq. (6.22), Eq. (6.26) are the form of relativistic coordinate transformation proposed before year 1900 by Sir Joseph Larmor, see Ref. [29] on page 27, result obtained following the lead of Lorentz in his study of the invariance transformations of Maxwell equations, as Larmor says: “...to second order in $v/c$.”—Larmor who is improving work of Lorentz therefore calls these coordinate transformations ‘Lorentz transformations’.

This name sticks. The relations Eq. (6.22), Eq. (6.24), Eq. (6.26) are the set of Lorentz transformations” in $x$-direction or “$x$-boost” in more colloquial language. We recall that spatial isotropy assures that our considerations apply to boosts along any direction in space. Therefore, the other two orthogonal boosts in $y$, and, respectively $z$ directions follow by renaming $x \leftrightarrow y$, and respectively, $x \leftrightarrow z$.

Exercise III–1: The universal speed of light

Consider light traveling in frame $S$ parallel to the $x$-axis with a wavefront at position $x = ct$. Confirm that in frame $S'$, moving at an arbitrary velocity $v$ relative to $S$, the speed of light remains unchanged. Since we used $c = c'$ this exercise is a cross check of our work.

Solution

In the frame of reference system $S$, if $x = 0$ at $t = 0$, the light expands such that

1. $x = ct$

In the moving frame of reference, where the speed of light is $c'$, we have analogously

2. $x' = c't'$,

and with this and the transformation equations Eq. (6.22) and Eq. (6.24),

3. $c' = \frac{x'}{t'} = \frac{x - vt}{t - (v/c^2)x}$.

Inserting Eq. 1, we find as expected

4. $c' = \frac{c - v}{1 - v/c} = c$.

Thus we have checked that the Lorentz transformation indeed contains the condition that the light velocity is the same for all observers.

End III–1: The universal speed of light
The derivation of the Lorentz transformation has one characteristic hypothesis that makes no sense in consideration of the Galilean transformations, namely \( c' = c \). One can wonder what happens to the derivation of the Lorentz coordinate transformation for \( c' = c - v \), a relation one could write for a corpuscular light theory.

**Solution**

We allow for two different velocities \( c \) and \( c' \)

1. \( c' = c - v \)

where the minus sign is appropriate for the passive transformation. Note that it is the coordinate origin of \( S' \) that moves with velocity \( v \) and hence a ‘tennis ball’ light has a smaller velocity in the \( x' \)-direction in \( S' \).

Considering Eq. (6.17) but allowing for \( c' \) we find

\[
x' = a_{11}(v^2)x \left( 1 - \frac{v}{c} \right),
\]

2. \( x = a_{11}(v^2)x \left( 1 + \frac{v}{c'} \right) = a_{11}(v^2)x' \left( 1 + \frac{v}{c' - v} \right) = a_{11}(v^2)x' \frac{x'}{1 - \frac{v}{c}}.
\]

which makes sense only if

3. \( a_{11}(v^2) = 1 \).

We now seek to determine the coefficients \( a_{21} \) and \( a_{22} \) required in the transformation equation for the time coordinate \( t' \), Eq. (6.2). We solve for \( t' \) seen in Eq. (6.12):

4. \( t' = \frac{x}{va_{11}} - \frac{x'}{v}. \)

We use Eq. 3 and obtain \( x' \) proceeding in analogy to derivation of Eq. (6.22).

5. \( t' = \frac{x}{v} - \left( \frac{x - vt}{v} \right) = \left( t - x \left( \frac{1}{v} - \frac{1}{v} \right) \right), \)

and obtain

6. \( t' = t \).

Hence

7. \( a_{21} = 0, \quad a_{22} = 1. \)
Considering Eq. 3 and Eq. 6 and inspecting Eq. (6.12) we recognize that we have demonstrated Galilean transformation. This shows that as long as the speed $c$ is transforming as in usual Galilean world, we recover the Galilean transformation of coordinates from the relativity arguments combined with isotropy and homogeneity of space-time.

Einstein’s postulate $c' = c$ is the key hypothesis that changes everything.

---

**End III–2: What causes time transformation?**

---

**Exercise III–3: Lorentz transformation of $dx, dt$**

We consider the transformation of coordinate differences $dx = x_2 - x_1$, $dt = t_2 - t_1$ to those observed in another coordinate system $S'$ moving with velocity $v$ along the $x$-axis.

**Solution**

We will study the Lorentz transformation of two different events $(t_1, \vec{x}_1)$ and $(t_2, \vec{x}_2)$

1. $x'_1 = \gamma(x_1 - \beta ct_1), \quad ct'_1 = \gamma(ct_1 - \beta x_1),$

2. $x'_2 = \gamma(x_2 - \beta ct_2), \quad ct'_2 = \gamma(ct_2 - \beta x_2).$

The difference between coordinates of these two events and their respective Lorentz transformations thus are

3. $x'_2 - x'_1 = \gamma(x_2 - x_1 - \beta(c(t_2 - t_1))), \quad ct'_2 - ct'_1 = \gamma(ct_2 - ct_1 - \beta(x_2 - x_1)).$

The transverse directions are not transformed

4. $y'_2 - y'_1 = y_2 - y_1, \quad z'_2 - z'_1 = z_2 - z_1.$

This shows that event differences are subject to the same Lorentz transformation as any single event. Taking the difference between events to be small we find as a corollary the often used form

5. $dx' = \gamma(dx - \beta cd t), \quad dy' = dy, \quad dz' = dz, \quad cd t' = \gamma(cd t - \beta dx) .$

This shows that since the Lorentz-transformation is linear it applies to linear combination of coordinates in the same way as it does to one coordinate.

---

**End III–3: Lorentz transformation of $dx, dt$**
Discussion III-1 – Why muons reach the surface of the Earth II

**Topic:** We return to discuss the trip that a muon makes from the upper atmosphere to Earth’s surface using this time the Lorentz transformation as a means of observing the circumstance in Earth’s rest frame.

_Simplicius:_ I have learned that the Lorentz transformation is not the same as the Lorentz-FitzGerald body contraction.

_Student:_ Yes. Using the Lorentz transformation we evaluate how events change when the observer changes from one inertial reference frame to another.

_Simplicius:_ Is the Lorentz-FitzGerald contraction a part of the Lorentz transformation?

_Professor:_ Not directly. Using the Lorentz transformation we compute what a set of coordinate events in one frame of reference means to another observer. The Lorentz-FitzGerald contraction refers to the contraction in size of a moving material body. Since the Lorentz transformation must be consistent with the Lorentz-FitzGerald body contraction, we can in principle find the contraction using the coordinate transformation approach.

_Simplicius:_ In another physics book I read that a muon produced at a height of 10km by (secondary) cosmic particles reaches the surface of the Earth because space contracts. Since I learned from you that space does not contract, how do you explain that the muon reaches the surface of the Earth? In its lifespan of \( \tau_\mu = 2.2 \mu s \), it can only travel 660 meters at speed of light. That is 1/15 of the distance.

_Student:_ In the muon rest frame – said, more accurately, in the frame of a muon-comoving observer – the time between two events, birth and death of the muon, is \( \Delta t = t_2 - t_1 \equiv \tau_\mu \), i.e. the muon lifespan. The distance traveled during that time is \( \Delta x = x_2 - x_1 = 0 \), i.e. for the comoving observer, the muon does not change position and lives its lifespan.

_Simplicius:_ I see, in comoving frame of reference a muon lives its lifespan and does not move at all. But how long does it live and how far does it travel when observed from the surface of the Earth?

_Student:_ Let me answer by employing the Lorentz transformation. For an observer in any other reference frame who sees the muon at a velocity \( v = \Delta x'/\Delta t' \), where the prime coordinates are those of the other (e.g. an observer on Earth), these coordinates become, according to the Lorentz transformation:

\[
\Delta t' = \gamma(\tau_\mu + v(\Delta x = 0)/c^2), \quad \Delta x' = \gamma((\Delta x = 0) + v\tau_\mu),
\]

and we indeed find that in this other reference frame the muon lives \( \Delta t' = \gamma \tau_\mu \) and during that time it travels the distance \( \Delta x' = \gamma v\tau_\mu \). Note also that \( \Delta x'/\Delta t' = v \).

_Simplicius:_ I see. For an observer on Earth the muon lifespan is stretched by \( \gamma \). We have derived time dilation using the Lorentz transformation. This is the other explanation I read about, that the muon reaches the Earth since it lives effectively longer when observed by an Earthbound observer.

_Professor:_ The time dilation effect is introduced here by employing the Lorentz coordinate transformation and implementing a measurement prescription, that the proper
lifespan of the muon $\tau_\mu$ is measured where $\Delta x = 0$. To find time dilation we compare the clock time measured at rest $\Delta x = 0$ with respect to the muon with a clock that is manifestly different: an inertial clock where both $t', x'$ are changing.

Simplicius: How about the argument that space contraction confirms relativity since it provides an alternate explanation of muon time dilation and explains how the muon reaches the Earth?

Professor: There is no `space contraction'; explanation in terms of a contracted spatial distance traveled by the muon is an incorrect application of the principles of SR.

Simplicius: However, I do not like the time dilation explanation. The reason is that in the theory of relativity I was told I can reverse the point of view: I can choose to be the observer approaching the muon with velocity vector $-\vec{v}$. How does your explanation do?

Professor: The situation is that any observer moving with the same speed $v$ will, within her proper time $\tau_\mu$, travel as far as the muon.

Simplicius: OK, I see that my argument was circular; we already computed $\Delta x'$ measured by the observer. Still, let me explain better the reason I do not like the time dilation explanation. Once I reverse the point of view, and I look at the clock attached to the observer, once it is this observer that is moving towards the muon, the time dilation reverses, right? In this case the observer would see a muon decay in his time $\tau_m u/\gamma$ and thus the distance traveled would not be 10km but $660m/15=44m$.

Professor: There is only one frame of reference in which the muon is at rest and that is the frame in which the muon lives $2.2 \mu s$, which is its proper time lifespan. For all other moving observers the observed lifespan of the muon is longer. Any other observer who moves with respect to the muon can tell the time it takes in her frame for the muon to decay. This will be described in section [7.3] on page [104]. In exercise [VIII–11] on page [301] we evaluate the distance traveled as reported by Earth-bound and muon-bound observer. The result is the same because the clock determining the muon proper lifespan lives within the muon. The process of creation and decay of the muon singles out the muon as the traveler whose intrinsic clock determines the lifespan – all other observers will be `older', i.e. age more when compared to the muon.

6.3 The non-relativistic Galilean limit

A wealth of daily experience shows that Galilean coordinate transformation (GT) is correct in the non-relativistic limit in which the speed of light is so large that it plays no physical role, that is $c \to \infty$. Any coordinate transformation replacing the GT must also agree with this experience, and thus must contain the GT in the non-relativistic limit.
To obtain the non-relativistic limit of the Lorentz transformation, we expand relations of interest to us choosing as the small expansion parameter \( \frac{v}{c} \ll 1 \). We have

\[
\gamma = 1 + \frac{1}{2} \left( \frac{v}{c} \right)^2 + \frac{3}{8} \left( \frac{v}{c} \right)^4 + \frac{5}{16} \left( \frac{v}{c} \right)^6 + \ldots \quad (6.28)
\]

Using this expansion in Eq. (6.22) yields

\[
x' = x - vt + vt \left( \frac{x}{2vt} - \frac{1}{2} \right) \left( \frac{v}{c} \right)^2 + \ldots, \quad (6.29)
\]

while for Eq. (6.24) we obtain

\[
t' = t - t \left( \frac{x}{vt} - \frac{1}{2} \right) \left( \frac{v}{c} \right)^2 + \ldots, \quad (6.30)
\]

Note that we have introduced above \( x/vt \), comparing the coordinate \( x \) to distance traveled at speed \( v \).

The correction terms in both Eq. (6.29) and Eq. (6.30) have a similar format, proportional to a term quadratic in \( \frac{v}{c} \). This means that even for body speeds that reach 10% of light speed, the relativistic corrections are not overwhelming. While in daily life such speeds are rarely attained, they do occur frequently in atomic physics, and dominate particle and nuclear physics. This also means that in order to observe experimentally the effect of relativity, we must study body speeds that are closer to \( c \).

### 6.4 The inverse Lorentz transformation

The Lorentz transformation allows an observer to determine the coordinates of an event, if these are known to another inertial observer. The coordinates \((t, x, y, z)\) of an event in \( S \) can be transformed into \((t', x', y', z')\) in \( S' \). With the velocity \( v \) of \( S' \) relative to \( S \) pointing by our choice in the \( x \) direction we have,

\[
\begin{align*}
ct' &= \frac{ct - (v/c)x}{\sqrt{1 - (v/c)^2}} = \gamma (ct - \beta x), \\
x' &= \frac{x - vt}{\sqrt{1 - (v/c)^2}} = \gamma (x - \beta ct), \\
y' &= y, \quad z' = z, \quad (6.31)
\end{align*}
\]
Seen from $S'$ the motion is in the negative $x'$ direction and thus the inverse Lorentz transformation must be:

\[
\begin{align*}
ct &= \frac{ct' + (v/c)x'}{\sqrt{1 - (v/c)^2}} = \gamma(ct' + \beta x'), \\
x' &= \frac{x' + vt'}{\sqrt{1 - (v/c)^2}} = \gamma(x' + \beta ct'), \\
y &= y', \quad z = z' .
\end{align*}
\]

(6.32)

The four equations of (6.32) are the inverse Lorentz transformation. We have obtained them from Eq. (6.31) through the application of the principle of relativity, whereby one exchanges the primed coordinates with those that are not primed and substitutes $-v$ for $v$. As one would expect, it is also possible to obtain the same inverse transformation by solving for $x'$ and $t'$ in Eq. (6.31).

---

**Exercise III–4: Inverting Lorentz transformation**

Solve the Lorentz transformation Eq. (6.31) for $x, ct, y, z$ assuming that the primed coordinates are given.

**Solution**

We see that $y = y'$ and $z = z'$. We look at the Lorentz transformation

1. $x' = \gamma(x - \beta ct), \quad ct' = \gamma(ct - \beta x),$

and we form two linear combinations;

2. $x' + \beta ct' = \gamma x(1 - \beta^2) - \gamma \beta ct + \gamma \beta ct = \gamma x(1 - \beta^2),$

and

3. $ct' + \beta x' = \gamma \beta x - \gamma \beta x + \gamma ct(1 - \beta^2) = \gamma ct(1 - \beta^2).$

Since $(1 - \beta^2) = 1/\gamma^2$ we multiply by $\gamma$ to obtain

4. $x = \gamma(x' + \beta ct'), \quad ct = \gamma(ct' + \beta x').$

Thus the back-transformation is achieved by taking $\beta \to -\beta$. This is also the inverse transformation in the sense that transforming $S \to S' \to S'' = S$ is achieved by transforming first with $\beta$ and next with $-\beta$.

---

**End III–4: Inverting Lorentz transformation**
7. Properties of the Lorentz Transformation

7.1 Relativistic addition of velocities

Case of two parallel velocities

An observer in $S'$ passes with velocity $\vec{v}$ in the $x$-direction with respect to another observer located in the frame $S$. An object is seen by the observer in frame $S$ to move at a constant velocity $\vec{u}$, as depicted in figure 7-1. What velocity is measured by the observer in $S'$? In the non-relativistic case, the answer is known from the GT:

$$\vec{u}' = \vec{u} - \vec{v} \quad (7.1)$$

However, such a Galilean law of addition of velocities does not respect the requirement that a material body velocity cannot be greater than light velocity.

To obtain the relativistic form of velocity addition we need to remember how velocity relates to position $\vec{r}$ and time $t$, in frame $S$ we have

$$\vec{u} = \frac{d\vec{r}}{dt} \quad (7.2)$$

It is important to remember that both space and time coordinates are subject to the Lorentz transformation, giving us $\vec{u}' = \frac{d\vec{r}'}{dt'}$ rather than $\frac{d\vec{r}'}{dt}$.

In what follows we orient the coordinate system $S'$ such that the velocity vector $\vec{v}$, Eq. (7.1) always points along the $x$-axis,

$$\vec{v} \equiv v\hat{i} \quad (7.3)$$
Given this choice the body velocity vector $\vec{u}$ cannot be constrained, it can point in any arbitrary direction.

We first consider the special case of $\vec{u}$ being parallel to $\vec{v}$. To obtain the transformed velocity we employ Lorentz transformation in the infinitesimal form

$$u'_x = \frac{dx'}{dt'} = \frac{dx - v dt}{dt - \frac{v}{c^2} dx} = \frac{dx}{dt} - \frac{v}{1 - \frac{v}{c^2}} dx$$

We recognize $\frac{dx}{dt}$ as $u_x$ resulting in the relativistic velocity addition relation

$$u'_x = \frac{u_x - v}{1 - u_x v / c^2}, \quad u'_y = u'_z = 0.$$  \hfill (7.5)

The Galilean relation Eq. (7.1) is corrected by the denominator which enforces the limit on maximum allowed speed.

---

**Exercise III–5: Checking relativistic velocity addition, case $\vec{u} \parallel \vec{v}$**

Consider the value of $u'_x$ when $u = 0.9c$ and $v = \pm 0.8c$.

**Solution**

We will be using the SR result Eq. (7.5) and compare to the GT Eq. (7.1). For $v = -0.8$ the GT predicts a speed that is 70% higher than the speed of light. However the SR result is

$$u'_x = \frac{0.9c + 0.8c}{1 + 0.9 \cdot 0.8} = \frac{1.7c}{1.72} = 0.9884c,$$

which comes close to speed of light but clearly respects the limit.

For the case $v = 0.8$ the GT result Eq. (7.1) is that the particle has become nearly nonrelativistic moving with 10% of $c$. However, situation is different for the SR case Eq. (7.5), we find

$$u'_x = \frac{0.9c - 0.8c}{1 - 0.9 \cdot 0.8} = \frac{0.1c}{0.28} = 0.36c$$

Thus the motion remains closer to relativistic as compared to the GT case.

---

**Exercise III–6: Relativistic relative approach speed**

An observer on Earth sees two rockets approach each other with equal and opposite velocities $u_{\pm} = \pm |u| = \pm 0.6c$. For this observer, the distance between the rockets
diminishes according to relative speed \( u_+ - u_- = 1.2c \). Thus the distance between the ships diminishes with a speed that exceeds the speed of light! Does this situation violate relativity? Consider an observer riding along in one of the rockets. What velocity of the other rocket does he report?

**Solution**

Relativity only requires that \( c \) be the maximum speed for light and for physical bodies. The relative speed at which coordinate separation between the rockets diminishes as recorded by a third arbitrary observer has nothing to do with the relative speed of two bodies measured by an observer riding one of the rockets. The whole point of the (special) theory of relativity is that there is no physical relevance to such an arbitrary third observer.

For an observer riding in one of the rockets, the relative velocity corresponds to the velocity of the other rocket as observed from this rocket. To find this velocity we must appropriately add individual velocities. For simplicity we consider the observer traveling at \( u_- \). The transformation of the velocity \( u_+ \) of the other rocket into this frame is

\[
\frac{1}{1 - \frac{u_+u_-}{c^2}} = \frac{2|u|}{1 + |u|^2/c^2}.
\]

With \( |u| = 0.6c \), we find that the relative velocity given by Eq. 1 is less than the speed of light: \( u'_+ = \frac{15c}{17} = 0.88c \).

---

**Exercise III–7: Shuttle craft rescue**

Star Wars scene: The Insurrection base reports to its shuttle traveling with \( u_s = 0.3c \) that it is being chased by a rocket following it with \( u_r = 0.6c \), and their rescue spaceship is approaching from the opposite direction traveling with the velocity \( u_S = -0.6c \); all velocities are parallel. What relative speeds are registered in the shuttle craft? Are these consistent with exercise III–6?

**Solution**

Let us denote by subscript ‘S’ = quantities in the spaceship frame and by subscript ‘s’ = quantities in the shuttle frame. We boost to the shuttle frame. The spaceship approaches the shuttle with relative velocity

\[
v_{SS} = \frac{u_S - u_s}{1 - u_Su_s/c^2} = \frac{-0.9c}{1 + 0.18} = -0.763c.
\]

However, the rocket is gaining on the shuttle with

\[
v_{rs} = \frac{u_r - u_s}{1 - u_ru_s/c^2} = \frac{0.3c}{1 - 0.18} = 0.366c,
\]
so rescue is possible but not assured. We check if the shuttle craft crew computations are correct by boosting with \( v = v_{rs} = -0.366c \) from shuttle frame to the rocket frame of reference. This gives the rocket observed velocity of the spaceship, independent of what the shuttle is doing:

\[
3 \quad v_{Sr} = \frac{u_{Ss} - v_{rs}}{1 - u_{Ss}v_{rs}/c^2} = \frac{-0.763c - 0.366c}{1 + 0.763 \cdot 0.366} = \frac{-1.129c}{1 + 1.28} = -0.88c .
\]

This is the expected result

\[
4 \quad v_{Sr} = \frac{u_S - u_r}{1 - u_xu_s/c^2} = \frac{-0.6c - 0.6c}{1 + 0.36} = -0.88c ,
\]
as we already evaluated in exercise [III–6]. It is rather complicated to show the above result analytically. This becomes much easier once additional tools are developed and we return therefore to this problem once more in exercise [III–18] on page 113.

End III–7: Shuttle craft rescue

Case of two arbitrary velocities

In the more general case where the direction of \( \vec{u} \) is arbitrary it is helpful to decompose \( \vec{u} \) into its component \( u_x \) parallel to the relative velocity \( \vec{v} \), Eq. (7.3), and the orthogonal components \( u_y \) and \( u_z \):

\[
\vec{r} = x\hat{i} + y\hat{j} + z\hat{k}
\]

\[
\vec{u} = \frac{dx}{dt}\hat{i} + \frac{dy}{dt}\hat{j} + \frac{dz}{dt}\hat{k} = u_x\hat{i} + u_y\hat{j} + u_z\hat{k} . \tag{7.6}
\]

Likewise we decompose \( \vec{u}' \) observed in frame \( S' \), remembering that the \( x' \)- and \( x \)-axes are parallel:

\[
\vec{u}' = \frac{dx'}{dt'}\hat{i} + \frac{dy'}{dt'}\hat{j} + \frac{dz'}{dt'}\hat{k} = u'_{x}\hat{i} + u'_{y}\hat{j} + u'_{z}\hat{k} . \tag{7.7}
\]

Even though \( dy' = dy \) and \( dz' = dz \), we see \( dt' \) in the denominator and thus all three velocity components transform; i.e., both the parallel and orthogonal components of the velocity are modified when we transform coordinates. Carrying this out we note that the first form of Eq. (7.5) remains valid while for \( u'_y \) and \( u'_z \) we find

\[
u'_y = \frac{dy'}{dt'} = \frac{dy\sqrt{1 - (v/c)^2}}{dt - \frac{v}{c^2}dx} = \frac{u_y\sqrt{1 - (v/c)^2}}{1 - \frac{v}{c^2}dx} = \frac{u_y\sqrt{1 - (v/c)^2}}{1 - \frac{vu_x}{c^2}}, \tag{7.8}
\]

\[
u'_z = \frac{dz'}{dt'} = \frac{u_z\sqrt{1 - (v/c)^2}}{1 - \frac{vu_x}{c^2}} .
\]
In the last step we are recognizing $dx/dt = u_x$.

To summarize, for $\vec{v}$ along $x$-axis, see Eq. (7.3), the relativistic velocity addition equations are for parallel and orthogonal addition, respectively

$$u'_x = \frac{u_x - v}{1 - u_x v/c^2}, \quad (7.9a)$$

$$u'_y = \frac{u_y}{1 - u_x v/c^2} \sqrt{1 - v^2/c^2}, \quad u'_z = \frac{u_z}{1 - u_x v/c^2} \sqrt{1 - v^2/c^2}. \quad (7.9b)$$

**Exercise III–8: Relative rocket motion**

An observer on Earth sees two rockets traveling with the velocities $\vec{u}^\pm$. What is the relative speed of the rockets?

**Solution**

We employ the result Eq. (7.9a). We align the coordinate $x$-axis with vector $\vec{u}^-$ and position the observer on this rocket ‘-’. Upon the Lorentz transformation into rocket ‘-‘, that is $v = u_x$ Eq. (7.9a) we find the relative velocity vector is $\vec{u}^r = \vec{u}^+-$

$$u^r_x = \frac{u^+_x - u^-_x}{1 - u^+_x u^-_x / c^2},$$

$$1 \quad u^r_y = \frac{u^+_y}{1 - u^+_x u^-_x / c^2} \sqrt{1 - (u^-_x/c)^2},$$

$$u^r_z = \frac{u^+_z}{1 - u^+_x u^-_x / c^2} \sqrt{1 - (u^-_x/c)^2}.$$

This is the relative velocity vector expressed in a coordinate system of which the orientation of the $x$-axis is parallel to the laboratory rocket ‘-‘ velocity vector. Thus $u^r_x$ is what the observer in rocket ‘-‘ sees as the velocity of ‘+‘ rocket in approach, while $u^r_y$ and $u^r_z$ are components that are normal to the approach axis.

**End III–8: Relative rocket motion**

**Exercise III–9: Relativistic forward projection**

We consider two sources of photons $S$ and $S'$. $S$ is at rest with respect to the laboratory observer while $S'$ moves with speed $v' = 0.6c$ relative to $S$ along the $x$-axis. In the rest frame of both systems one measures an angle of emission $\vartheta = 60^0$
between the direction of movement of the photon and the direction of movement of $S'$ as depicted in figure 7-2. What is the angle of emission that the laboratory observer reports for the photon originating in the moving source $S'$?

**Solution**

Geometry: the source $S'$ moves along the $x$-axis, which is chosen to be parallel to the $x'$-axis; the movement of the photon is confined to the $xy$-plane, figure 7-2. The $y$ and $y'$ axes are also parallel.

The velocity of the photon in system $S$ is therefore

$$\vec{v} = c \cos 60^\circ \hat{i} + c \sin 60^\circ \hat{j} = u_x \hat{i} + u_y \hat{j};$$

that is $u_x = 0.5c$, $u_y = 0.866c$.

We calculate the component $u'_x$ for the photon emitted by moving source using the addition of velocity theorem,

$$u'_x = \frac{u_x + v'}{1 - u_x v'/c^2}.$$

Inserting $u_x = 0.5c$ and $v' = 0.6c$, one obtains $u'_x = 0.846c$.

To obtain $u'_y$ we consider the requirement

$$c = \sqrt{u'^2_x + u'^2_y},$$

see exercise III-11 on page 96. Therefore

$$u'_y = \sqrt{c^2 - u'^2_x} = \sqrt{1 - 0.846^2} c = 0.533c.$$

We recall that this transverse to motion component of velocity of light is changed solely due to the transformation of laboratory time.
For the angle $\vartheta'$ between the direction of the photon and the $x$-axis of the observer at rest in system $S'$, we then have

$$\tan \vartheta' = \frac{u'_y}{u'_x} = \frac{0.533c}{0.846c} = 0.630.$$ 

and therefore: $\vartheta' = 32.2^\circ$.

We performed this computation for photons but it applies equally to the emission of all relativistic particles.

This exercise shows an important phenomenon often observed in particle physics experiments, namely that particles are relativistically focused pointing forward along the direction of motion of the source. Such a situation arises for example when a fast cosmic particle hits the upper atmosphere. Particles are produced in a rest frame intrinsic to the reaction which moves at a high speed in the same direction as the primary cosmic ray. Our computation explains why the secondary particles are focused in this direction towards the Earth. The study of light aberration leads to a more general result, see exercise III–13 on page 103. A more systematic discussion of this effect is presented in exercise VI–8 on page 222.

---

### Exercise III–10: Fresnel drag coefficient

The speed of light in a medium with a refraction index $n > 1$ is known to be $\tilde{c} < c$. Furthermore, $\tilde{c}$ is dependent on the velocity $v$ of the medium relative to a (laboratory) observer. Obtain the lowest correction in $v$: $\tilde{c}(v) = \tilde{c}_0 + vf(n)$ as measured by the laboratory observer, where $f$ is the Fresnel drag coefficient (Fresnel$^3$, 1818).

**Solution**

The speed of light is reduced by the index of refraction $n$ in a stationary medium. It is known that the Maxwell equations allow one to deduce the relation

$$\tilde{c}_0 = \frac{c}{n}.$$

We wish to describe what happens when the medium is set in motion where $v$ is medium velocity relative to the lab frame in the same direction, or against, the beam of light as is shown in figure 7-3.

There is a fluid comoving observer $S(v)$ moving with $\pm v$ with respect to the laboratory observer $S$ for whom the light can be traveling through a stationary medium in

---

$^3$Augustin-Jean Fresnel (1788–1827), luminary French physicist and founder of the theory of wave optics.
Figure 7-3: Illustration of a table-top interference experiment to measure Fresnel’s drag coefficient: one observes the change in the interference between light carried with and against the (changing) flow velocity $v$ of a fluid, see exercise III–10.

Each of the arms of interferometer seen in figure 7-3. To find the speed of light $\tilde{c}(v)$ as measured by the observer in the lab frame, we can thus use the addition of velocities theorem, Eq. (7.9a):

$$\tilde{c}(v) = \frac{\tilde{c}_0 + v}{1 + \tilde{c}_0 v/c^2} = \frac{c/n + v}{1 + v/(nc)} = \left(\frac{c}{n}\right) \left(\frac{1 + nv/c}{1 + v/(nc)}\right).$$

Any velocity that can be achieved in a laboratory satisfies $v \ll c$, thus we can expand:

$$\tilde{c}(v) \simeq \frac{c}{n} \left(1 + \frac{nv}{c}\right) \left(1 - \frac{v}{nc} \ldots\right) \simeq \frac{c}{n} \left(1 + \frac{nv}{c} \left(1 - \frac{1}{n^2}\right)\right)$$

$$\simeq \frac{c}{n} + v \left(1 - \frac{1}{n^2}\right).$$

We see that Eq. 3 applies to both cases, when $v$ is positive and negative.

The Fresnel drag coefficient is therefore

$$f = 1 - \frac{1}{n^2}.$$

For the special case of a vacuum ($n = 1$), we find that $f = 0$ confirming the drag limit in the vacuum where the speed of light is independent of the motion of the observer.

Further reading: The first experimental demonstration of Fresnel drag was carried out by Fizeau 1851, and much experimental and theoretical work followed in the second half of 19th century. Fresnel drag was one of cornerstone precursor to the development of SR; the concept of æther motion does not enter present discussion, a topic that has
been much a part of Fresnel drag experiment prior to Einstein’s development of SR. The solution of Fresnel drag using SR principles is due to Max von Laue\footnote{Max von Laue, “Die Mitführung des Lichtes durch bewegte Körper nach dem Relativitätsprinzip.” (The Entrainment of Light by Moving Bodies according to Principle of Relativity) \textit{Annalen der Physik} \textbf{328} 989–990 (1907). Max von Laue (1879–1960) won the Nobel prize in 1914 “for his discovery of the diffraction of X-rays by crystals”; active opponent of the Nazification of German physics.}

End III–10: Fresnel drag coefficient

Exercise III–11: Maximum speed for arbitrary $\vec{u}$.

An object moves at a velocity of magnitude $u \ (u < c)$ in an arbitrary direction as observed in frame $S$. Show that the velocity $u'$ observed in frame $S'$ moving relative to $S$ at velocity $v$ along the $x$-axis is less than the speed of light, $u' < c$.

Solution

We want to prove the relationship:

1. $u'^2 < c^2$.

We can rewrite $u'^2$ in components

2. $u'^2 = u'_x^2 + u'_y^2 + u'_z^2$,

where $u'_x$ is the component of the velocity parallel to the $x$- and $x'$-axes, and the $u'_y$ and $u'_z$ are the two transverse components. Using Eq. (7.9a) to express these components in terms of $u$ and $v$ we obtain, after some simple algebraic manipulations

3. $u'^2 = \frac{(u_x - v)^2 + (u_y^2 + u_z^2)(1 - v^2/c^2)}{(1 - vu_x/c^2)^2} = \frac{u^2 + v^2 - 2u_xv - (u_y^2 + u_z^2)v^2/c^2}{(1 - vu_x/c^2)^2}$

   $= \frac{u^2 - c^2}{(1 - v^2/c^2)} + \frac{c^2(1 - v^2/c^2)}{(1 - vu_x/c^2)^2} - 2u_xv + v^2 + u_x^2v^2/c^2$,

where $u^2 = u_x^2 + u_y^2 + u_z^2$. The $v^2$ terms cancel and we obtain

4. $u'^2 = c^2 \left(1 - \frac{(1 - u^2/c^2)(1 - v^2/c^2)}{(1 - vu_x/c^2)^2}\right)$.

Inspecting the large brackets in Eq.\footnote{Max von Laue, “Die Mitführung des Lichtes durch bewegte Körper nach dem Relativitätsprinzip.” (The Entrainment of Light by Moving Bodies according to Principle of Relativity) \textit{Annalen der Physik} \textbf{328} 989–990 (1907). Max von Laue (1879–1960) won the Nobel prize in 1914 “for his discovery of the diffraction of X-rays by crystals”; active opponent of the Nazification of German physics.} we see that a manifestly positive term is always subtracted from unity. This term becomes arbitrarily small when either of the speeds, $u, v$ approaches speed of light, yet always

5. $u'^2 < c^2$. 
Thus we confirm that irrespective of the direction of motion the addition theorem of velocities derived from the Lorentz transformation is consistent with $c$ being the universal and highest speed which cannot be exceeded.

It is of interest to consider the special case of a light ray with $u = c$ observed in the moving frame $S'$. We find

\[ u'^2 = c^2, \quad \text{for} \quad u = c. \]

This result is of use in section 7.2.

Another cross check of interest is verification of the result Eq. (7.5) and Eq. (7.9b) corresponding to $u_y = u_z = 0$, that is $u^2 = u_x^2$. We form common denominator in Eq. 4 to obtain

\[
\begin{align*}
6 & \quad u'^2 = c^2 \frac{1 - 2vu_x/c^2 + v^2u_x^2/c^4 - 1 - v^2u_x^2/c^4 + u_x^2/c^2 + v^2/c^2}{(1 - vu_x/c^2)^2}, \\
7 & \quad = \frac{(u_x - v)^2}{(1 - vu_x/c^2)^2},
\end{align*}
\]

which we recognize as the square of Eq. (7.5).

End III–11: Maximum speed for arbitrary $\vec{u}$.

Exercise III–12: Relativistic relative motion - vector notation

State the magnitude of the relative velocity seen in Eq. 4 in exercise III–11 in coordinate axis independent vector notation which focuses on the difference in the velocity vectors.

Solution

We recall that in Eq. 4 in exercise III–11 the velocity $v$ has a single component with $\vec{\beta}_2 = (v/c, 0, 0)$. We call now $\vec{\beta}_1 = (u_x/c, u_y/c, u_z/c)$. In this notation Eq. 4 in exercise III–11 reads

\[ \beta^2 \equiv u'^2/c^2 = \left( 1 - \frac{(1 - \vec{\beta}_1^2)(1 - \vec{\beta}_2^2)}{(1 - \vec{\beta}_1 \cdot \vec{\beta}_2)^2} \right). \]

We rewrite Eq. 1 with common denominator to obtain

\[ \beta^2 = \frac{1 - 2\vec{\beta}_1 \cdot \vec{\beta}_2 + (\vec{\beta}_1 \cdot \vec{\beta}_2)^2 - 1 + \vec{\beta}_1^2 + \vec{\beta}_2^2 - \vec{\beta}_1^2 \vec{\beta}_2^2}{(1 - \vec{\beta}_1 \cdot \vec{\beta}_2)^2}. \]

Noting $(\vec{\beta}_1 \times \vec{\beta}_2)^2 = \vec{\beta}_1^2 \vec{\beta}_2^2 - (\vec{\beta}_1 \cdot \vec{\beta}_2)^2$ in the numerator we find

\[ \beta^2 = \frac{(\vec{\beta}_1 - \vec{\beta}_2)^2 - (\vec{\beta}_1 \times \vec{\beta}_2)^2}{(1 - \vec{\beta}_1 \cdot \vec{\beta}_2)^2}. \]
As a cross-check we note agreement with Eq. (7) in exercise III–11, of which Eq. (3) is a generalization.

Our result Eq. (3) exhibits the relativistic relative speed comprising a counter intuitive contribution arising from a component normal to both velocity vectors. This last term in the numerator of Eq. (3) vanishes when \( \vec{\beta}_1 \parallel \vec{\beta}_2 \).

---

End III–12: Relativistic relative motion - vector notation

### 7.2 Aberration of light

We consider a light ray originating in a distant star, see figure 7-4, where for the observer marked with \( \vec{v} \) the light arrives more horizontally as discussed before in section 1.3. Thus the inferred star location has moved by the aberration angle \( \alpha \) from its position.

The following discussion is addressing the observation of well focused light-rays and not (spherical) plane wave light. This approach side-steps arguments about modification of aberration effect by waves. This picture of the experimental situation is accurate since the light emitted by a star consists of an incoherent flux of photons produced in independent atomic processes. When we look at a star we see individual photons emitted across the entire source that make it at the same time, given the material such as telescope optics, into our eye. The accidental coherence of these photons is difficult to discern, it is of use in evaluating the size of the random light source using the so called HBT\(^5\) method.

In a light-source rest frame \( S \) we choose the coordinate axis pointing away from the Earth, hence the velocity components of the light ray are negative pointing ‘down’:

\[
\begin{align*}
    u_x &= -c \sin \theta \cos \phi, \\
    u_y &= -c \sin \theta \sin \phi, \\
    u_z &= -c \cos \theta .
\end{align*}
\] (7.10)

The angle \( \theta \) spans from the coordinate \( z \)-axis to the line connecting to the observer on Earth. The orientation of the Earthbound observer’s (moving) \( S' \) coordinate system can be for convenience and without loss of generality chosen such that the relative motion velocity vector \( \vec{v} \) points along one of the coordinate axis, and the natural choice is the \( z' \)-direction which we align with the \( z \) direction in \( S \).

We apply the velocity addition theorem, Eq. (7.9a), where the boost is now along the \( z \)- and not \( x \)-axis. The light ray velocity in the Earth’s rest frame of \( S \) is

Figure 7-4: Aberration angle \( \alpha \) of light from a distant star: Comparison of the line of sight for an observer at rest and for another observer traveling at a constant velocity \( \vec{v} \) relative to the line of sight.

The observer \( S' \) becomes:

\[
\begin{align*}
\alpha' &= \phi' - \theta' \\
\end{align*}
\]

Comparing the two first expressions in Eq. (7.11) we recognize that the azimuthal aberration vanishes,

\[
\phi = \phi'.
\]  

However, according to the last expression in Eq. (7.11) the altitude aberration

\[
\alpha \equiv \theta - \theta'.
\]
\[
\cos \theta' \equiv \frac{\cos \theta + (v/c)}{1 + (v/c) \cos \theta} \simeq \cos \theta + \frac{v}{c} \sin^2 \theta - \left(\frac{v}{c}\right)^2 \cos^2 \theta \left(1 - \cos \theta\right) + \ldots .
\]

We note that since by definition \( v > 0 \), \( \cos \theta' > \cos \theta \) and thus \( \theta' < \theta \) for any altitude angle \( \theta \in \{0, \pi/2\}; \) see figure 7.4 on page 99. For \( \theta \to 0 \) the altitude aberration vanishes; this is the case of pure ‘radial’ motion, that is motion along the line of sight. We next check the relativistic consistency of Eq. (7.15). Inverting Eq. (7.15) to obtain \( \cos \theta \) we find

\[
\cos \theta = \frac{\cos \theta' - (v/c)}{1 - (v/c) \cos \theta'} .
\]

The altitude aberration expressions Eq. (7.14), Eq. (7.15), and Eq. (7.16) show the equivalence between the two observers \( S \) and \( S' \) with respect to \( \alpha \leftrightarrow -\alpha, \ v \leftrightarrow -v, \ \theta \leftrightarrow \theta' \).

Eq. (7.16) demonstrates that the effect of aberration for an observer \( S \) is analogous to that for the observer \( S' \). The principles of special relativity are fully accounted for. We can go one step further: by the nature of the problem the speed of light must enter into the answer, and by dimensional considerations the effect has thus to involve the ratio of the relative speed of the light source with the Earth, \( v \), with the speed of light \( c \). Since the Galilean limit has to arise, we knew and expected that the aberration effect begins with linear power \( v/c \). In order to assure symmetry of the result between observers on Earth and on the Star, as described by Eq. (7.17) the functional dependence we found applying the Lorentz transformation, Eq. (7.15), is uniquely defined.

We now discuss the magnitude of the effects we expect. Considering the range of values \( v/c < \mathcal{O}(10^{-3}) \) we can neglect quadratic terms in the non-relativistic expansion shown in Eq. (7.15). In the small aberration angle limit we use

\[
\cos \theta' = \cos(\theta - \alpha) = \cos \alpha \cos \theta + \sin \alpha \sin \theta \simeq \cos \theta + \alpha \sin \theta .
\]

The linear term in non-relativistic expansion shown in Eq. (7.15) provides

\[
\alpha = \frac{v_{\perp}}{c}, \quad v_{\perp} = v \sin \theta .
\]

This is the final aberration result: the relative speed transverse to the line of sight fixes aberration angle \( \alpha \). The radial (along line of sight) speed \( v_r = v \cos \theta \) does not influence the magnitude of aberration. Naturally, the transverse speed can
contain aside of the orbital motion of the Earth around the Sun also a transverse component inherent in the star motion.

For the Earthbound observer $S'$ the velocity vector $\vec{v}$ includes the Sun’s relative motion $\vec{v}_\odot$ with respect to the observed star, modulated by the annual rhythm of Earth’s orbital velocity $\vec{v}_\oplus$ around the Sun, and daily rotation velocity $\vec{v}_{\text{rot}}$:

$$\vec{v} = \vec{v}_\odot + \vec{v}_\oplus + \vec{v}_{\text{rot}}.$$ \hfill (7.20)

We drop $\vec{v}_{\text{rot}}$ ($v_{\text{rot}} \simeq 0.46 \text{ km/s}$) from any further discussion; this is possible considering that Earth’s orbital speed $\vec{v}_\oplus \simeq 30 \text{ km/s}$ is 67 times greater. Relative star speeds $v_\odot$ range from essentially vanishing to a few times greater than the orbital speed. However, the nearby stars that are easily observed move along with the Earth around the galactic center and thus in general the dominant transverse component in the motion is the orbital velocity $\vec{v}_\oplus$.

Today we know that the London zenith star, Gamma Draconis, is approaching Earth with the radial speed $28.19 \pm 0.36 \text{ km/s}$, nearly the same as $v_\oplus = 30 \text{ km/s}$. However, the proper motion (angle expressed transverse speed) of Gamma Draconis is $-8.48 \text{ mas/yr}$ (milli-arc-sec/year). Therefore the observed aberration motion $\pm 20 \text{ as/yr}$ of Gamma Draconis, see section 1.3, is an effect 2000 times larger, must be thus dominated by the Earth motion around the Sun.

We evaluate $\alpha$ according to Eq. (7.19) assuming that orbital speed is $v_\perp$, hence $v_\perp/c = \pm 10^{-4}$. While the Earth makes a full orbit around the Sun, with orbital speed $v_\oplus = 30 \text{ km/s}$, the Earth velocity vector $\vec{v}$ produces an image of Earth’s motion in the alteration angle $^6\delta \alpha(t) \in (-10^{-4}, 10^{-4}) = (-20'', +20'')$, which is the apparent zenith star circular motion observed by Molyneux and Bradley and interpreted in this way by Bradley.

Discussion III-2 – Light aberration and æther drag

**Topic:** Aberration is a topic recurrent in the context of æther drag theories and the “relativity must be wrong” argument: i.e., it is easy to argue that the aberration effect is a measurement of the absolute speed of the æther near the Earth. Is that really the case?

**Simplicius:** I saw a book where the author argues that since one measures absolute speed of Earth by observing fixed star periodic aberration one actually observes the æther motion around the Sun.

**Student:** Why should the measurement of Earth orbital motion create this æther drag problem? Lots of good people measure relative speed of the Earth to this and that. The emphasis is on ‘relative’, here relative to the Sun.

**Simplicius:** But the point the author makes is that by looking at many fixed stars distributed uniformly in the sky we measure exclusively the Earth’s orbital speed around

---

$^6$We recall that 1 arc min = $[(2\pi)/(60 \times 360)] \text{ rad}$.
the Sun. The author concludes that the measurement is that of the motion of the material æther that accompanies the Earth in the motion around the Sun.

**Student:** I read on Wikipedia that the aberration effect had big influence on the development of the understanding leading on to the invention of relativity. Wikipedia points out that the hypothesis of the motion of the material æther could not account for the effect of aberration – there is even a graphic simulation to explain why. But it was a windy road so it is likely that some of these nearly 200 year old ideas are rediscovered - without the arguments that lead to their demise.

**Simplicius:** I also looked at the Wikipedia; that is a very theoretical web page on “aberration of light”. What I see in the book I mentioned is lots of experimental results addressing aberration of fixed stars placed in all directions in the sky. When the aberration data is analyzed, the author finds the orbital speed of the Earth in each case, so he attributes the effect to the motion of the dragged æther. I am more inclined to believe the experiment. A theory could be wrong.

**Professor:** The stars we can observe easily are in general located nearby in our Galaxy - at most a few 100 ly away. These nearby stars move along with us in the Milky Way, and do this in a way that renders the aberration effect to be often mainly caused the the motion of the Earth around the Sun with the speed of 30 km/c. Stated in technical terms, for many nearby stars the motion transverse to the line of sight is dominated by the Earth’s orbital motion around the Sun.

**Simplicius:** What you say also means that there should be some stars for which the analysis would not give the Earth speed as 30 km/c . . . .

**Professor:** . . . and you wonder why this book does not mention that? I see also in my own research domain that good people are inclined to reject measurements that disagree with their view of the World.

**Student:** I see this all the time when grading labs: my students only retain results they know fit what they should find.

**Professor:** There are many people who look at experimental data selectively. Both by picking from the results they obtain, and by making the experimental data they do not like have a large error. I see where these habits come from: I wonder what would happen if a student in a lab report presented a result suggesting Newton or Hook are wrong; can s/he expect a good grade? I imagine students ‘learn’ in their introductory labs to pick out ‘good’ data.

**Simplicius:** I would never do that, would I . . . ? So the table I saw was selected data? And the measurements that did not fit were ignored? . . . I can see now how to begin to argue with my friend who loaned me that book. What other advise do you have?

**Student:** I would focus on experimental issues to avoid protracted argument and doubt. Argue one should consider a series of precise annual rhythm aberration measurements choosing stars randomly. Soon one finds cases that contradict æther drag. Today lots of data is out there. So this can be done without any additional experimental effort.

**Simplicius:** The book shows so many data points . . .
Student: ... you do not need to find an equal number of opposite examples. One case that strongly contradicts a hypothesis is enough.

Exercise III–13: Ultrarelativistic aberration of light

Calculate the ultrarelativistic aberration of the observation angle \( \theta \) of a ‘star’ for an observer moving with velocity \( \vec{v} \) with respect to the fixed star, see section 7.2. Interpret the result for the case of a relativistic source emitting radiation observed in laboratory.

Solution

The aberration formula Eq. (7.15) is not suitable for evaluating the aberration when within measurement error we have \( v/c \to 1 \). The result suggests that all such objects appear in the zenith. This is clearly not right. The reason is that we lost ‘relativistic’ sensitivity when deriving Eq. (7.15).

An equivalent form of Eq. (7.15) is obtained using Eq. (7.13) in one of the two first expressions in Eq. (7.11)

\[
\sin \theta' = \frac{\sin \theta}{\gamma(1 + (v/c) \cos \theta)}.
\]

Adding in squares of the left hand sides of Eq. (7.15) and Eq. 1 we show \( \sin^2 \theta' + \cos^2 \theta' = 1 \). This verifies the consistency of Eq. (7.15) with Eq. 1. Taking ratio of Eq. (7.15) with Eq. 1 we obtain

\[
\cot \theta' = \frac{\gamma(\cos \theta + (v/c))}{\sin \theta} = \gamma \cot \theta + \frac{\gamma v/c}{\sin \theta}.
\]

Both Eq. 1 and Eq. 2 are useful for the case \( v/c \to 1 \), since we now have explicit value of \( \gamma > 1 \) available.

Even though we do not have a relativistic shooting star on our horizon, the results presented can be adapted to describe the case of light emitted by relativistic charged particles. However, it is the particle and not the observer (the laboratory in which the experiment is performed) that is moving. We thus need to change to the emitting frame of reference: we perform the transformation shown in Eq. (7.17): \( \theta \leftrightarrow \theta' \) and \( v \rightarrow -v \).

Now \( \theta' \) is the angle between emitted radiation and direction of motion in the frame of reference of the charged particle moving with speed \( v \) with respect to laboratory, and \( \theta \) is the observation angle in the rest frame of the laboratory. The two equivalent results are

\[
\sin \theta = \frac{\sin \theta'}{\gamma(1 - (v/c) \cos \theta')}, \quad \cot \theta = \gamma \cot \theta' - \frac{\gamma v/c}{\sin \theta'}.
\]
To see how this relation works, take $\theta' = 45^\circ$ such that $\sin \theta' = \cos \theta' = 1/\sqrt{2}$ and assume that the speed $v$ is sufficiently ultrarelativistic to justify use of $v/c = 1$. Hence we find $\sin \theta \simeq \theta = 2.4/\gamma$, thus for $\gamma = 100$ the angle of emission in the lab is $\theta = 0.024 \times 360^\circ/(2\pi) = 1.4^0$. We recognize that the radiation is observed ‘focused’ into a forward cone (small $\theta$) for large $\gamma$, even if in the emitter frame of reference it is emitted more uniformly in all directions i.e. uniformly as function of $\theta'$.

**End III–13: Ultrarelativistic aberration of light**

### 7.3 Invariance of proper time

We compare

$$s^2 = c^2t^2 - x^2 - y^2 - z^2,$$

with

$$\Delta s^2 = c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2,$$

and equivalently

$$s'^2 = c^2t'^2 - x'^2 - y'^2 - z'^2,$$

with

$$\Delta s'^2 = c^2(t'_2 - t'_1)^2 - (x'_2 - x'_1)^2 - (y'_2 - y'_1)^2 - (z'_2 - z'_1)^2.$$

It suffices to consider only the LT in the $x$-direction. If other directions need to be transformed, we can always re-orient our coordinate system so that the $x$-axis points in the direction of the LT transformation. Therefore in what follows $y^2 + z^2 = y'^2 + z'^2$ in Eq. (7.21), and similarly for the difference between two events, Eq. (7.22).

We employ the LT Eq. (6.22) and Eq. (6.24) to find

$$s'^2 = c^2t'^2 - x'^2 - y'^2 - z'^2,$$

$$= \gamma^2(c^2t^2 - \beta^2x^2) - y'^2 - z'^2,$$

$$= \gamma^2(\beta^2x^2 - x^2 - \beta^2c^2t^2) - y'^2 - z'^2.$$

Recognizing $\gamma^2 = (1 - \beta^2)^{-1}$ we have

$$s'^2 = \gamma^2(1 - \beta^2)(c^2t^2 - x^2) - y'^2 - z'^2 = s^2.$$
The same transformation property follows for the difference between two events

\[
\Delta s^2 = c^2(t'_2 - t'_1)^2 - (x'_2 - x'_1)^2 - (y'_2 - y'_1)^2 - (z'_2 - z'_1)^2,
\]

\[
= \gamma^2 [c^2(t_2 - t_1)^2 + (\beta^2 - 1)(x_2 - x_1)^2 - \beta^2 c^2(t_2 - t_1)^2] - (y_2 - y_1)^2 - (z_2 - z_1)^2,
\]

\[
= c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2 = \Delta s^2.
\]

(7.25)

and thus for a small separation between events we have

\[
ds^2 = c^2 dt'^2 - dx'^2 - dy'^2 - dz'^2
\]

\[
= c^2 dt^2 - dx^2 - dy^2 - dz^2
\]

\[
= ds^2.
\]

(7.26)

One says that \(s^2\) and, equivalently, the increment \(ds^2\), are Lorentz invariant.

Dividing by \(c\) we see that \(ds/c = d\tau\) and

\[
d\tau'^2 = \frac{ds^2}{c^2}
\]

\[
= dt'^2 - \frac{dx'^2 + dy'^2 + dz'^2}{c^2}
\]

\[
= dt^2 - \frac{dx^2 + dy^2 + dz^2}{c^2}
\]

\[
= \frac{ds^2}{c^2} = d\tau^2
\]

(7.27)

The meaning of \(\tau\) becomes clear when we consider an observer for whom there is no change of position \(dx = 0\). We see that \(dt = d\tau\). This observer’s clock which rests in the body frame of reference measures the proper body time, \(\tau\).

We found by deriving Eq. (7.26) that the increment of proper time is an invariant. That means that all observers \(S(v)\) having some speed \(v\) will agree to how fast a clock will tick inside a body,

\[
d\tau = \sqrt{dt^2 - dx^2/c^2} = dt\sqrt{1 - dx^2/(cdt)^2}
\]

\[
= dt\sqrt{1 - v^2/c^2}
\]

(7.28)

where \(v\) is velocity of a body measured by the observer \(S(v)\) and this velocity can depend on time \(t\). We see that the proper time increment \(d\tau\) is according to Eq. (7.28) shorter compared to the time of any observer \(S(v)\).
Important: for all observers the proper time \( d\tau \) of each and every body is the same, it is the observer’s \( S(v) \) clock that measures a different and longer time \( dt \) depending on how fast the observer measures the speed of the body. More generally we can say that each and every body has its proper time which is an invariant quantity.

**Exercise III–14: Proper time of interstellar probe**

The star Alpha Centauri is located 4.4 light years from our solar system. A probe leaves the solar system, traveling at a constant speed, and reaches Alpha Centauri six years later, according to observers on Earth. How much time passes on a clock traveling with the probe?

**Solution**

The use of the invariant \( \tau \) leads us to a efficient solution of this seemingly complex problem. We evaluate the Lorentz invariant proper time given by:

\[
1 \quad c^2 \Delta \tau^2 = c^2 \Delta t^2 - \Delta x^2.
\]

With \( \Delta x = 4.4 \) light years and \( \Delta t = 6 \) years we find

\[
2 \quad c^2 \Delta \tau^2 = (6c)^2 - (4.4c)^2 = c^2(4.1 \text{ y})^2.
\]

Thus \( \Delta \tau = 4.1 \) years. Note that with increasing (average) speed \( \Delta x/\Delta t \equiv v \to c \) of the probe we have \( c\Delta t \to \Delta x \) and hence \( \Delta \tau \to 0 \). A probe moving at ultra relativistic speed ages very little.

End III–14: Proper time of interstellar probe

**Exercise III–15: Positronium annihilation**

A metastable configuration of positronium (usually denoted by symbol P\(_s\)), the bound state of an electron and its antiparticle the positron, has a mean proper lifetime of \( \tau = 142 \) ns before it annihilates into three photons (this is the ortho-positronium\(^7\) where the particle spins align forming spin 1 state). If the mean lifetime observed in the laboratory of the metastable P\(_s\) in a mono-energetic beam (that is all P\(_s\) at a constant

\(^7\)The other equally bound state of positronium, that annihilates into two photons, the ‘para-positronium’ with spin-0, has a lifespan of \( \tau = 125 \) picoseconds, that is more than 1000 times shorter than for the ortho-P\(_s\) considered in the exercise.
and prescribed speed) in the lab frame is $\Delta t = 300$ ns, what is the mean lab frame distance traveled before annihilation by the Ps 'particle' in the beam?

**Solution**

Again, as in exercise III–14 we use the invariant,

$$c^2 \tau^2 = c^2 \Delta t^2 - \Delta x^2.$$  

If $\tau$ is the mean proper lifetime and $\Delta t$ the mean laboratory lifetime, then the mean distance Ps traveled from the source is

$$\Delta x = c\sqrt{\Delta t^2 - \tau^2} \approx 264 \, c \cdot \text{ns} = 260 \, \text{ft} = 79.2 \, \text{m}.$$  

We used feet as a reminder that the order of magnitude estimate is always $1\,\text{ns}$ at speed $c \approx 1 \, \text{ft}$.  

We note that the reported mean observed lifespan $\Delta t = 300 \, \text{ns} > 142 \, \text{ns}$ exceeds the proper lifespan due to time dilation. Without the effect of time dilation the maximum average travel distance must be less than $142 \, \text{ft}$ (corresponding to $142 \, \text{ns}$ mean lifespan and a speed which must be below light velocity). In this somewhat different context this exercise is a redo of the muon traveling from the upper atmosphere to the surface of the Earth, exercise II–2 on page 54, and conversation III-1 on page 84.

---

**End III–15: Positronium annihilation**

### 7.4 Two Lorentz transformations in a sequence

We consider two subsequent Lorentz transformations: a) the event $(t, x, y, z)$ in $S$ transformed into $(t', x', y', z')$ in $S'$ using $v_1$, and b) $(t', x', y', z')$ in $S'$ transformed into $(t'', x'', y'', z'')$ in $S''$ using $v_2$. We consider the case that both transformations are in the $x$-direction. We now show that the outcome is as if there was one transformation from $S$ to $S''$ by $v_x$ which arises from the relativistic addition of the two velocities.

We begin with

$$x'' = \gamma_2(x' - \beta_2 ct'), \quad ct'' = \gamma_2(ct' - \beta_2 x'),$$  

(7.29)

where we insert

$$x' = \gamma_1(x - \beta_1 ct), \quad ct' = \gamma_1(ct - \beta_1 x),$$  

(7.30)

which results in

$$x'' = \gamma_2 \gamma_1(x - \beta_1 ct - \beta_2 (ct - \beta_1 x)), \quad ct'' = \gamma_2 \gamma_1(ct - \beta_1 x - \beta_2(x - \beta_1 ct)).$$  

(7.31)
A reorganization of terms produces
\[ x'' = \gamma_2 \gamma_1 (x + \beta_2 \beta_1) - (\beta_2 + \beta_1) ct, \quad ct'' = \gamma_2 \gamma_1 (ct + \beta_2 \beta_1) - (\beta_2 + \beta_1) x. \] (7.32)

If this is to be a new Lorentz transformation again from \( S \rightarrow S'' \), then we must have
\[ \gamma = \gamma_2 \gamma_1 (1 + \beta_1 \beta_2), \quad \gamma \beta = \gamma_2 \gamma_1 (\beta_1 + \beta_2). \] (7.33)

Dividing these two equations by each other we obtain again the velocity addition theorem
\[ \beta = \frac{\beta_1 + \beta_2}{1 + \beta_1 \beta_2}, \quad \text{that is} \quad v_x = \frac{v_1 + v_2}{1 + v_1 v_2/c^2}, \] (7.34)

where \( v_x = c \beta \). We see that two Lorentz transformations in sequence in the \( x \)-direction are described by a transformation with the velocity obtained according to the relativistic addition theorem for the two velocities of these transformations.

Comparing to our earlier result Eq. (7.5) and Eq. (7.9b) we note a change in sign: in the prior result the velocity \( v \) was that of a moving body. In the present consideration \( \beta_2 \) describes change of coordinate system. Thus we have shown the equivalence of both passive and active relativistic velocity addition theorems.

We next check if \( \beta^2 < 1 \). This is best done evaluating
\[ 1/\gamma^2 \equiv 1 - \beta^2 = \frac{(1 - \beta_1^2)(1 - \beta_2^2)}{(1 + \beta_1 \beta_2)^2} \] (7.35)

which is positive and smaller than unity for any \( \beta_1^2, \beta_2^2 < 1 \).

For \( \beta_1 = -\beta_2 \) we find from Eq. (7.34) as well as from Eq. (7.35) that with the 2nd transformation we transformed back to the original system \( S'' = S \) since \( \beta = 0, \gamma = 1 \).

**Exercise III–16: Two Lorentz transformations in different directions.**

Find the transformation describing the consequence of two Lorentz transformations carried out in sequence, the first one in \( x \)-direction with \( v_{1x} \) and the second one in \( y \)-direction with \( v_{2y} \).

**Solution**

Since \( z = z' = z'' \) we will ignore this coordinate. The first transformation is from system \( S \) to \( S' \) along \( x \)-axis

1. \[ x' = \gamma_1 x - \beta_1 x ct, \quad y' = y, \quad ct' = \gamma_1 (ct - \beta_1 x), \]
and the second from $S'$ to $S''$ along $y$-axis

\[ \begin{align*}
2 \quad x'' &= x', \quad y'' = \gamma_{2y}(y' - \beta_{2y}ct'), \quad ct'' = \gamma_{2y}(ct' - \beta_{2y}y'), \\
3 \quad y'' &= \gamma_{2y}(y - \beta_{2y}\gamma_{1x}(ct - \beta_{1x}x)) = \gamma_{2y}(y + \gamma_{1x}\beta_{2y}\beta_{1x}x) - \gamma_{1x}\gamma_{2y}\beta_{2y}ct, \\
&\quad ct'' = \gamma_{2y}(\gamma_{1x}(ct - \beta_{1x}x) - \beta_{2y}y) = \gamma_{2y}\gamma_{1x}ct - \gamma_{2y}\gamma_{1x}\beta_{1x}x - \gamma_{2y}\beta_{2y}y
\end{align*} \]

By substitution we obtain the double primed coordinates expressed in terms of the not primed coordinates

\[ 
\begin{align*}
x'' &= \gamma_{1x}(x - \beta_{1x}ct), \\
y'' &= \gamma_{2y}(y - \beta_{2y}\gamma_{1x}(ct - \beta_{1x}x)) = \gamma_{2y}(y + \gamma_{1x}\beta_{2y}\beta_{1x}x) - \gamma_{1x}\gamma_{2y}\beta_{2y}ct \\
ct'' &= \gamma_{2y}(\gamma_{1x}(ct - \beta_{1x}x) - \beta_{2y}y) = \gamma_{2y}\gamma_{1x}ct - \gamma_{2y}\gamma_{1x}\beta_{1x}x - \gamma_{2y}\beta_{2y}y
\end{align*} \]

We see that this format is very different from the usual Lorentz transformation.

An interested reader can now check that for two non-parallel Lorentz transformations the sequence in which they are carried out matters. One says that these transformations do not commute. This means is that if we were first to consider the transformation in the $y$-direction with $v_{2y}$, followed by a transformation in $x$-direction with $v_{1x}$, the result we obtain would be different from the result we presented. This is further developed in exercise VIII–2 on page 283.

---

### 7.5 Rapidity

We now introduce a new way to characterize speed within the Lorentz transformation through a function\footnote{It is very inconvenient at this point to introduce the rapidity using the conventional symbol $y$ which can be confounded in this book with a coordinate. We thus use $y_r$ which is a cumbersome compromise.} $y_r(\beta)$. We are motivated by the desire to find a new variable which, unlike $v$, is not bounded by $c$. Therefore, it can more accurately characterize motion at ultra relativistic speeds. We also want that in nonrelativistic limit

\[ y_r = \beta \quad \text{for} \; \beta << 1. \]  

(7.36)

This relation suggests that we call $y_r$ something that relates to speed, and the name rapidity\footnote{The more common symbol in use for rapidity is $y$. Our use of coordinates $x$, $y$, $z$ could lead to confusion. The use of $y_r$ as symbol for rapidity, however, can be confounded with another variable, pseudo rapidity, introduced in particle physics. In this book $y_r$ is the best symbol choice.} is in common use.

We further recall that in the nonrelativistic limit two velocities add vectorially. If these velocity vectors are parallel, they add as numbers in nonrelativistic limit;
that is, the two speeds add. Among many functions \( y_r(\beta) \) we choose a unique one such that for parallel relativistic motion the rapidities sum just like speeds do for nonrelativistic motion. How this could work is easily recognized by recalling the addition theorem

\[
\tanh(y_{r1} + y_{r2}) = \frac{\tanh y_{r1} + \tanh y_{r2}}{1 + \tanh y_{r1} \tanh y_{r2}},
\]

which we compare with the addition theorem of relativistic parallel speeds

\[
\beta_{12} = \frac{\beta_1 + \beta_2}{1 + \beta_1 \beta_2}.
\]

Thus we explore a new variable \( y_r \) that satisfies

\[
\beta = \tanh y_r = \frac{e^{y_r} - e^{-y_r}}{e^{y_r} + e^{-y_r}} < 1.
\]

We find the useful relation

\[
e^{y_r} = \sqrt{\frac{1 + \beta}{1 - \beta}} = \gamma (1 + \beta), \quad e^{-y_r} = \sqrt{\frac{1 - \beta}{1 + \beta}} = \gamma (1 - \beta),
\]

and hence

\[
y_r = \ln \sqrt{\frac{1 + \beta}{1 - \beta}} = \frac{1}{2} \ln \left( \frac{1 + \beta}{1 - \beta} \right) \equiv \tanh^{-1}(\beta).
\]

Considering that hyperbolic exponential functions can be written in terms of the exponential function

\[
cosh y_r = \frac{e^{y_r} + e^{-y_r}}{2}, \quad \sinh y_r = \frac{e^{y_r} - e^{-y_r}}{2},
\]

a short computation shows

\[
cosh y_r = \gamma, \quad \sinh y_r = \gamma \beta.
\]

We cross check our computation

\[
\cosh^2 y_r - \sinh^2 y_r = \gamma^2 (1 - \beta^2) = 1,
\]

and furthermore

\[
\tanh y_r = \frac{\sinh y_r}{\cosh y_r} = \frac{\gamma \beta}{\gamma} = \beta,
\]

as expected, see Eq. (7.39).
Expanding Eq. (7.45) in a power series we see that

\[
\beta = \tanh y_r = y_r - \frac{1}{3} y_r^3 + \frac{2}{15} y_r^5 - \ldots ,
\]

(7.46)

and from Eq. (7.41)

\[
y_r = \frac{1}{2} \ln \left( 1 + \beta \right) - \ln \left( 1 - \beta \right)
= \frac{1}{2} \left[ \left( \beta - \frac{\beta^2}{2} + \frac{\beta^3}{3} - \ldots \right) - \left( -\beta - \frac{\beta^2}{2} - \frac{\beta^3}{3} - \ldots \right) \right]
= \beta + \frac{1}{3} \beta^3 + \frac{1}{5} \beta^5 + \ldots .
\]

(7.47)

On the other hand for \( \beta \to 1 \) we use a Taylor series for Eq. (7.39) to write

\[
1 - \beta = 2 \left( e^{-2y_r} - e^{-4y_r} + e^{-6y_r} - \ldots \right) .
\]

(7.48)

It is interesting to note how rapidity describes the approach to the speed of light: \( y_r = 10 \) corresponds to a fractional deviation from the speed of light by \( 4 \cdot 10^{-9} \), corresponding to 1.24 m/s.

We now write the Lorentz transformation using rapidity in the form

\[
x' = \frac{x - vt}{\sqrt{1 - \left( \frac{v}{c} \right)^2}} = \gamma (x - \beta ct) = x \cosh y_r - ct \sinh y_r ,
\]

\[
c't = \frac{ct - \left( \frac{v}{c} \right) x}{\sqrt{1 - \left( \frac{v}{c} \right)^2}} = \gamma (ct - \beta x) = ct \cosh y_r - x \sinh y_r .
\]

(7.49)

The form of Eq. (7.49) is similar to a rotation around the z-axis:

\[
x' = x \cos \phi + y \sin \phi ,
\]

\[
y' = y \cos \phi - x \sin \phi .
\]

(7.50)

but with one sign being different and a hyperbolic angle of rotation. This takes into account the different properties of space-time: while usual rotation e.g. around the z-axis leaves \( \rho^2 = x^2 + y^2 \) invariant, the Lorentz transformation leaves invariant \( s^2 = (ct)^2 - \vec{x}^2 \). We return to this point in section 20.1

With the help of rapidity we thus recognize the Lorentz transformation to be a new form of rotation in what one calls Minkowski space-time. LT is characterized by an angle which is hyperbolic and not trigonometric. This difference arises from the unbounded character of \( y_r \) compared to \( 0 \leq \phi \leq 2\pi \), where \( \phi \) is a regular rotation angle.
We have constructed the rapidity variable to be additive which means that when we consider two LT in sequence that the rapidities add just like two angles add in case of regular rotations
\[ y_r = y_{r_1} + y_{r_2} \]  \hspace{1cm} (7.51)
This will be the content of the exercise [III-17] below. It is important to remember that Eq. (7.51) is only true for two Lorentz transformations in same spatial direction. Even so, Eq. (7.51) is a pivotal property of rapidity, which behaves just like nonrelativistic addition of velocity does. This makes rapidity an extraordinarily important tool in the study of many problems in physics.

For example, the rapidity formulation of LT Eq. (7.43) allows us to show the invariance of proper time just as we show that length of a vector remains the same under rotations:
\[
s'^2 = c^2 t'^2 - x'^2,
\]
\[= (c t \cosh y_r - x \sinh y_r)^2 - (x \cosh y_r - c t \sinh y_r)^2 \]  \hspace{1cm} (7.52)
\[= (\cosh^2 y_r - \sinh^2 y_r)c^2 t^2 + (\sinh^2 y_r - \cosh^2 y_r)x^2
\]
\[= c^2 t^2 - x^2 = s^2 ,\]
which is the same result as obtained in section [7.3]

Exercise III-17: Addition of rapidity

Demonstrate that rapidities for two Lorentz transformations carried out in sequence add akin to the situation with rotation angles.

Solution
We now carry out two Lorentz transformations in sequence, see section [7.4] employing rapidity format. We have
\[ x' = x \cosh y_{r_1} - c t \sinh y_{r_1} , \quad ct' = c t \cosh y_{r_1} - x \sinh y_{r_1} , \]
\[ x'' = x' \cosh y_{r_2} - c t' \sinh y_{r_2} , \quad ct'' = c t' \cosh y_{r_2} - x' \sinh y_{r_2} . \]
We insert the first transformation into the second
\[ x'' = (x \cosh y_{r_1} - c t \sinh y_{r_1}) \cosh y_{r_2} - (c t \cosh y_{r_1} - x \sinh y_{r_1}) \sinh y_{r_2} , \]
\[ct'' = (c t \cosh y_{r_1} - x \sinh y_{r_1}) \cosh y_{r_2} - (x \cosh y_{r_1} - c t \sinh y_{r_1}) \sinh y_{r_2} . \]
Reordering terms we find

\[ x'' = x \left( \cosh y_r \cosh y_r' + \sinh y_r' \sinh y_r - ct \left( \cosh y_r \sinh y_r' + \sinh y_r' \cosh y_r \right) \right) \]
\[ ct'' = ct \left( \cosh y_r \cosh y_r' + \sinh y_r' \sinh y_r \right) - x \left( \cosh y_r \sinh y_r' + \sinh y_r' \cosh y_r \right) \]

It is well known that

\[ \cosh(y_r + y_r') = \cosh y_r \cosh y_r' + \sinh y_r' \sinh y_r \]
\[ \sinh(y_r + y_r') = \cosh y_r \sinh y_r' + \sinh y_r' \cosh y_r \]

which can be also checked using

\[ \cosh y_r = \frac{e^{y_r} + e^{-y_r}}{2} \quad \sinh y_r = \frac{e^{y_r} - e^{-y_r}}{2} \]

The combined LT transformations thus have the form

\[ x'' = x \cosh y_r - ct \sinh y_r, \quad ct'' = ct \cosh y_r - x \sinh y_r. \]

where

\[ y_r = y_r' + y_r. \]

The rapidities add under Lorentz transformations carried out in the same direction, just like is the case with speeds of two Galilean transformations carried out in sequence in the direction of two parallel velocities.

End III–17: Addition of rapidity

Exercise III–18: Shuttle craft rescue: rapidity method

We return here to the exercise exercise [III–7] on page [90] employing instead of speeds/velocities the associated rapidity. We recall that in the Star Wars scene the insurrection base reports to its shuttle traveling with \( u_{sh} = 0.3c \) that it is being chased by a rocket following it with \( u_r = 0.6c \), and their rescue spaceship is approaching from the opposite direction traveling with the velocity \( u_S = -0.6c \), all velocities are parallel. What relative speeds are registered in the shuttle craft? Are these consistent with exercise [III–6]?

Solution
We first establish the appropriate rapidities using Eq. (7.41). We find:

1. \[ y_{rs} = 0.5 \ln \frac{1.3}{0.7} = 0.310, \quad y_{rt} = 0.5 \ln \frac{1.6}{0.4} = 0.693, \quad y_{rs} = 0.5 \ln \frac{0.4}{1.6} = -0.693 \]

We boost to the shuttle frame. The spaceship approaches the shuttle with rapidity

2. \[ y_{rss} = y_{rr} - y_{rs} = -1.003, \quad \rightarrow v_{ss} = 0.763, \]

where we use Eq. (7.39) to compute velocity \( v \). However, the rocket is gaining on the shuttle with rapidity

3. \[ y_{r sr} = y_{rt} - y_{rs} = 0.383, \quad \rightarrow v_{sr} = 0.365, \]

so rescue is possible but not assured. We check if shuttle craft crew computations are correct by boosting with \( v = v_{rs} = -0.366c \) from the shuttle frame to the rocket frame of reference. This gives the rocket observed rapidity of the spaceship, independent of what the shuttle is doing:

4. \[ y_{r sr} = y_{rs} - y_{r s} = -1.003 - 0.383 = -1.386, \quad \rightarrow v_{sr} = 0.88 \]

This is the expected result

5. \[ y_{r sr} = y_{rs} - y_{r s} = -0.693 - 0.693 = -1.386 \]

since trivially

6. \[ y_{r sr} = y_{rs} - y_{r s} = y_{rs} - y_{rs} - (y_{rt} - y_{rs}) = y_{rs} - y_{rt}. \]

The advantage of the use of rapidity in Star Wars context and for that matter all science fiction conforming to Einstein’s relativity is now evident, but this simple variable has yet to be discovered by movie makers.

---

Exercise III–19: Lorentz transformations in a sequence: factor \( \gamma_{12} \)

Obtain using rapidity the Lorentz factor \( \gamma_{12} \) for two Lorentz transformations carried out in sequence along the same axis.

Solution
We take advantage of the fact that rapidities of these two transformations add

1. \( \gamma_{12} = \cosh(y_{r1} + y_{r2}) \)

We further have in general

2. \( \cosh(a + b) = \cosh a \cosh b + \sinh a \sinh b \)

and thus

3. \( \gamma_{12} = \cosh y_1 \cosh y_2 (1 + \tanh y_1 \tanh y_2) \)

We use Eq. (7.39) and Eq. (7.43) to obtain

4. \( \gamma_{12} = \gamma_1 \gamma_2 (1 + \beta_1 \beta_2) \)

Naturally, this result follows in a more cumbersome computation using velocity addition and evaluating \( \gamma_{12} \) explicitly, as we have shown deriving Eq. (7.33). However, the use of rapidity simplified the evaluation considerably.

End III–19: Lorentz transformations in a sequence: factor \( \gamma_{12} \)
Part IV

Measurement
Introductory remarks to Part IV

We explore the consistency of the Lorentz coordinate transformation with the relativistic body properties: time dilation and Lorentz-FitzGerald body contraction. The objective is to introduce the definition of spatial separation measurement which, when combined with the coordinate transformation, produces time dilation and the Lorentz-FitzGerald body contraction assuring consistency of body properties with coordinate transformations.

We introduce a graphic method to characterize the relation of Lorentz transformation to the process of the measurement of body properties. The equal time definition of coordinate events, corresponding for example to the ends of a moving body, produces within the context of Lorentz coordinate transformation a consistent interpretation of the Lorentz-FitzGerald body contraction. Similarly, we show that ‘equal location in space’ measurement of time produces in the context of the Lorentz coordinate transformation the time dilation effect when a body is observed from another frame of reference.

A clock attached to a body scores the time always at the one and the same body location. This clock measures accumulated body proper time that is dilated by any motion imparted on the body, which makes the time dilation effect ‘real’. Since a body does not on its own retain the record of the Lorentz-FitzGerald body contraction, and nobody has build a body-contraction ‘clock’ some think body contraction is not real.

John S. Bell adopted an example of two independently propelled identical rockets which maintain their spatial separation distance as measured at their start. Tracking two such beacon rockets floating within a vacuum chamber within the body it is possible to preserve the original unit of length which is non-material and thus not subject to the Lorentz-FitzGerald body contraction. This observation allows the construction of a device that records the Lorentz-FitzGerald body contraction as it occurs in the moving body. These two beacon rockets could also be bouncing a light signal between them thus allowing to compare the proper time in the rocket to a preserved standard unit of time, allowing to determine the time dilation as it occurs.

The measurement of the Lorentz-FitzGerald body contraction and/or time dilation as it occurs means that only body property changes caused by a (gentle) acceleration can be observed. Any other reference velocity effect remains unobservable. However, we do not need to track in time the acceleration process to acknowledge that it has existed in the past. For example there is a universal preferred inertial CMB frame in which the matter we are made of was once at rest, and against which we can determine our velocity and evaluate the change in the body properties.
8 Body Properties and Lorentz Transformations

8.1 Graphic representation of Lorentz transformation

We consider an observer at rest in the frame of reference $S$, in which a clock moves in the direction of the $x$-axis with $\vec{v} = \vec{v}_i$. This movement is depicted by the line $\ell$ in the space-time diagram figure 8-1. $\ell$ is the world line of the clock.

The world line $\ell$ creates an angle $\vartheta$ with the $x$-axis, and we have

$$\cot \vartheta = \frac{x - x_1}{ct} = \frac{v}{c}$$

and therefore

$$\vartheta = \arccot \frac{v}{c}.$$  \hspace{1cm} (8.1)

Since $ct > vt$, the world line of any object must have a slope steeper than $45^\circ$. With $90^\circ$ representing a clock at rest, the range for motion from left to right is $135^\circ > \vartheta > 45^\circ$.

Let $S_0$ be the proper coordinate system of the clock, in which the clock is at rest. $S_0$ has the velocity $+\vec{v}$ with respect to $S$. We now find the new coordinates $ct_0$ and $x_0$ and draw these along with the coordinate axes $ct$ and $x$. The proper coordinates of the clock in $S$ are $ct_0$, $x_0$.

In order to incorporate this new coordinate system in our original diagram figure 8-1, we draw the $t_0$-axis parallel to $\ell$, as the clock measures the time $t_0$, see figure 8-2. The $x_0$-axis will be where the coordinate $t_0 = 0$. The LT relation, which describes the coordinate time $t_0$ in terms of $ct$ and $x$ is

$$ct_0 = \frac{ct - (v/c)x}{\sqrt{1 - (v/c)^2}} = \gamma (ct - \beta x).$$  \hspace{1cm} (8.2)
8. Body Properties and Lorentz Transformations

Figure 8-2: The $t_0$-axis of the rest frame of the clock $S_0$ is parallel to $\ell$, the $x_0$-axis has inclines with reference to $x$-axis 'inverse' to that of $t_0$-axis with respect to the $t$-axis, see text.

Setting $t_0 = 0$ in Eq. (8.2) yields the line

$$ct = \frac{v}{c} x,$$

(8.3)

which describes the $x_0$-axis in the $(ct, x)$ diagram depicted in figure 8-2. The $x_0$-axis and the $x$-axis make an angle $\varphi$, which is given by

$$\varphi = \arctan \frac{v}{c}.$$

(8.4)

The axes $ct_0$ and $x_0$ of the system $S_0$ lie within the axes $ct$, $x$ of the system $S$. The $ct_0$ axis is always above the 45º line ‘light cone’ (dashed), and the $x_0$ below, see figure 8-2.

We now consider an event $E_1$ located on the world line $\ell$ of the clock; for example, the striking of the hour. We obtain the coordinates of this event in both systems by projecting parallel to the axes of the frame of reference as shown in figure 8-3a: in system $S$ the event has coordinates $x_{E_1}$ and $t_{E_1}$; in $S_0$ the event has the coordinates $x_{0E_1}$ and $t_{0E_1}$. To clarify, the projection of the position onto the $x_0$ axis occurs in the direction of the $ct_0$ axis, and the position must be constant in $S_0$ where the clock is at rest.

8.2 Simultaneity and time dilation

We consider the same clock and let it strike twice creating the events $E_1$ and $E_2$ shown in figure 8-3b. In $S$ the clock has moved between the two events and
Lesson 1: The time difference between events is recognized to be dependent on the observer’s frame of reference.

Lesson 2: Temporal simultaneity in a reference frame is a unique measurement prescription of two events.
observations are assembled and compared, the time difference between the events can be determined in both $S$ and $S_0$.

We obtain the result of this comparison using the inverse Lorentz transformation Eq. (6.32), where $S$ moves with velocity $-v$ relative to $S_0$:

$$t_{E_1} = \frac{t_{0E_1} + (v/c^2)x_{0E}}{\sqrt{1-(v/c)^2}},$$

$$t_{E_2} = \frac{t_{0E_2} + (v/c^2)x_{0E}}{\sqrt{1-(v/c)^2}}.$$  \hspace{1cm} (8.5)

Because the two events occur at the same location $x_0$ in $S_0$, the relation between the two time differences as measured in $S$ and $S_0$ is

$$t_{E_2} - t_{E_1} = \frac{t_{0E_2} - t_{0E_1}}{\sqrt{1-(v/c)^2}} = \gamma(t_{0E_2} - t_{0E_1}).$$  \hspace{1cm} (8.6)

This result shows that the proper (at rest in a body) time interval $t_{0E_2} - t_{0E_1}$ is always shorter than the time interval measured in the laboratory frame $S$ where the body is seen to be moving with speed $v$. The proper time is measured at the same point in frame $S_0$ by a clock at rest. The laboratory time of the clock is measured at two different locations in frame $S$. This is the key difference which makes the laboratory time larger than the proper time. We thus found the time
dilation that we have discussed earlier by analysis of events employing the relativistic coordinate transformation, the Lorentz transformation.

**Lesson 3:** The clock ‘at rest’ (not moving in space) ticks slower compared to any clock in motion; the effect is called time dilation.

Given two observers it is the full travel history of both observers after they synchronized their watches which determines who will be younger when they meet again. We return to this matter in greater detail, see section 12.3.

9 Different Methods of Measuring Spatial Separation

The process of measuring the spatial separation interval $\Delta x$ between two events is considerably more involved than time measurement and we need to establish clear operational approach at the outset. There are two motivations for this circumstance:

1) When we discuss spatial separation of events, time separation must be considered and the temporal simultaneity of a measurement will be identified as a defining constraint.

2) We will not introduce a ‘length-clock’ informing us about Lorentz-FitzGerald body contraction as we want to show, based on our findings here and in the following section 10, that such a device can be constructed.

We now assume that the observer is at rest in the system $S$. The observed body is at rest in system $S_0$. Our objective is to determine the magnitude of the spatial separation between events characterizing the two ends of observed body. We discuss the following three different possible methods of measurement:

(a) The body’s ends are lit permanently and we photograph them with a camera located in the system $S$. This means that the observation is simultaneous in $S$, but not in $S_0$.

(b) The ends of the body light up for a brief instant in the body rest frame, and their picture is taken by a camera that has a permanently open objective. The measurement of any two points on the body is therefore simultaneous in $S_0$ but not simultaneous in $S$.

(c) We place mirrors at the two body ends and illuminate the body for a short instant by a flash of lightning. The light is then scattered at different times, and returns from the two body ends to a camera located in $S$ with a permanently open objective with the signal from the two ends arriving at different times.
9. Different Methods of Measuring Spatial Separation

9.1 Spatial separation measurement with the signal synchronized in the rest frame of the observer \((S)\)

Case (a) is depicted in figure 9-1. The camera’s aperture opens and closes instantaneously at time \(t_E \equiv t_{E1} = t_{E2}\) in system \(S\). The incoming observed light is therefore sent out at different times \(t_{0E1}\) and \(t_{0E2}\) in system \(S_0\). We define the spatial separation in the two frames of reference

\[
\begin{align*}
l_0 &= x_{0E2} - x_{0E1}, \\
l_{\text{sim}} &= x_{E2} - x_{E1}
\end{align*}
\]

where the superscript ‘sim’ reminds us in which frame of reference the measurement is simultaneous.

The Lorentz transformation Eq. (6.31) yields

\[
\begin{align*}
x_{0E1} &= \frac{x_{E1} - vt_E}{\sqrt{1 - (v/c)^2}}, \\
x_{0E2} &= \frac{x_{E2} - vt_E}{\sqrt{1 - (v/c)^2}}.
\end{align*}
\]
Figure 9-2: Measurement, Case (b): two events simultaneous in the moving system of the body, $S_0$.

Taking the difference of the two equations we obtain

$$l_0 = \frac{l_{\text{sim}}}{\sqrt{1 - (v/c)^2}} , \quad \text{or} \quad l_{\text{sim}} = l_0 \sqrt{1 - \left(\frac{v}{c}\right)^2}.$$  \hfill (9.1)

The measured distance $l_{\text{sim}}$ is shorter than the distance $l_0$.

**Lesson 4a:** The spatial separation $l_{\text{sim}}$ of two events evaluated in the frame $(S)$ in which the two events are simultaneous is shorter by the factor $\sqrt{1 - (v/c)^2} = \gamma^{-1}$ when compared to a non-simultaneous spatial separation measurement carried out in another frame $(S_0)$ moving with speed $v$.

### 9.2 Spatial separation measurement with the signal synchronized in the rest frame of a body $(S_0)$

We consider the case (b) illustrated in figure 9-2. The events $E_1$ and $E_2$ are occurring at the two ends of a body $K$, emitting a light flash simultaneously in body rest frame $S_0$ at the same time $t_{0,E} \equiv t_{0,E_1} = t_{0,E_2}$. The two events are not simultaneous in $S$. 


The distance between the two events is again

\[ l = x_{E2} - x_{E1}, \]

\[ l_{sim}^0 = x_{0E2} - x_{0E1}. \]

We determine \( x_{E2} \) and \( x_{E1} \) using the inverse Lorentz transformation Eq. (6.32):

\[ x_{E1} = \frac{x_{0E1} + vt_{0E}}{\sqrt{1 - (v/c)^2}}, \]

\[ x_{E2} = \frac{x_{0E2} + vt_{0E}}{\sqrt{1 - (v/c)^2}}. \]

By taking the difference we now obtain

\[ l = \frac{l_{sim}^0}{\sqrt{1 - (v/c)^2}}, \] (9.2)

or

\[ l_{sim}^0 = l \sqrt{1 - \left(\frac{v}{c}\right)^2}. \] (9.3)

**Lesson 4b:** The spatial separation \( l_{sim}^0 \) of two events evaluated in the frame in which the two events are simultaneous is shorter by the factor \( \sqrt{1 - (v/c)^2} = \gamma^{-1} \) when compared to a non-simultaneous spatial separation measurement carried out in another frame moving with speed \( v \).

**Lesson 5:** No matter what we call \( l_0 \) or \( l \), what matters is the frame of reference in which the spatial separation between two events describing the body length is simultaneous as denoted by superscript ‘sim’ for simultaneous. The spatial distance \( l_{sim} \) is always shorter compared to any other distance measurement between events when measured at unequal times. The textbook statement \( l = l_0 \sqrt{1 - (v/c)^2} \) must be replaced by

\[ l_{sim} = l \sqrt{1 - \left(\frac{v}{c}\right)^2}. \] (9.4)

Thus there is no inconsistency. This resolves the apparent inconsistency in regard to Lorentz contractions as measured by observers in different frames by making it clear that i) absolute speed is not measurable, but only the speed with respect to a singular observer ‘sim’, and, ii) any two observers claiming that the other is Lorentz contracted are performing two very different measurements in regard to how time is treated.

**Exercise IV–1: Special character of the equal time observer**
We have introduced in section 7.3 the invariance of \((\Delta s)^2\). Use this to demonstrate that the ‘sim’ observer always will see a contracted event separation.

Solution

The invariance under Lorentz transformation, Eq. (7.25), reversing the sign means 
\[-(\Delta s)^2 = -(\Delta s')^2,\]
that is,

\[
1 \quad (\Delta x)^2 - (\Delta ct)^2 = (\Delta x')^2 - (\Delta ct')^2,
\]
where we ignore transverse to motion directions \(y\) and \(z\), and reference frames \(S\) and \(S'\) are associated with the coordinate sets. The two event separations are \(\Delta x\), \(\Delta ct\) and \(\Delta x'\), \(\Delta ct'\) respectively.

Let \(S \rightarrow S^{\text{sim}}\) be the equal time observer; that is \(\Delta ct = 0\). We then have the identity

\[
2 \quad (\Delta x^{\text{sim}})^2 = (\Delta x')^2 - (\Delta ct')^2.
\]

We have now demonstrated unequivocally the special character of the equal time observer ‘sim’: \(\Delta x^{\text{sim}} < \Delta x'\) where \(S'\) is any other observer for whom \(\Delta ct' \neq 0\).

We can also write

\[
3 \quad (\Delta x^{\text{sim}})^2 = (\Delta x')^2 \left(1 - \frac{(\Delta ct')^2}{(\Delta x')^2}\right).
\]

The space time intervals that appear are not related to motion as is evident in \(\Delta ct' < \Delta x'\) and additional Lorentz transformation algebra is needed to obtain the explicit form of the Lorentz contraction Eq. (9.4).

End IV–1: Special character of the equal time observer

9.3 Spatial separation measurement due to illumination with light emitted in the rest frame of the observer

Finally we consider the more complex case (c), as illustrated in figure 9-3. Note that the body is at rest in \(S_0\) as the body ends project onto values \(x_{0E_1}\) and \(x_{0E_2}\) for all time \(ct_0\). The flash of lightning occurs at \(t = 0\) and \(x = 0\) in \(S\) and propagates along the path \(x = ct\), illuminating the moving body at different times and positions. In both frames of reference the observation of the body is made at unequal time, that is \(\Delta t = t_{E_1} - t_{E_2} \neq 0\) and \(\Delta t_0 = t_{0E_1} - t_{0E_2} \neq 0\), see figure 9-3.
The light reaches the ends of the body at events $E_1$ and $E_2$ clearly at different times in both frame $S_0$ and $S$. However, $x_{E_1} = ct_{E_1}$ and $x_{E_2} = ct_{E_2}$. We insert these constraints into the Lorentz transformation (6.31) and obtain

$$x_{0E_1} = \frac{x_{E_1}(1 - v/c)}{\sqrt{1 - (v/c)^2}},$$
$$x_{0E_2} = \frac{x_{E_2}(1 - v/c)}{\sqrt{1 - (v/c)^2}},$$

which leads to a relation between the event separation $l_0$ and the laboratory observed event separation $l$

$$l_0 = l \frac{(1 - v/c)}{\sqrt{1 - (v/c)^2}} = l \sqrt{\frac{1 - v/c}{1 + v/c}}.$$

This third measurement case is thus reminiscent of the relativistic Doppler wavelength shift, an important phenomenon we describe further in section 13.

Further reading: J. Terrell\(^1\) writes in abstract: “Observers photographing the meter stick simultaneously from the same position will obtain precisely the same

Figure 9-4: The rectangular prism as observed from above and the paths of light hitting the photo plate. The corners of the body, the thick dots at 1,2,3,4 represent position lights. See exercise [IV–2].

picture, except for a change in scale given by the Doppler shift ratio”.

Exercise IV–2: Quasi-rotation of a moving body for a special observer

A rectangular prism, with the edge lengths $a_0$, $b_0$, and $c_0$, moves with a velocity $v$ parallel to the edge $a_0$ past an observer $S$ at rest. A light shines from each corner of the prism, and the observer $S$ takes a snapshot of it as it flies past. What picture does the observer obtain and how can she interpret it?

Solution

The light from each corner must arrive simultaneously in $S$ at the camera, and thus must be sent from the different corners of the prism at different times in order to be recorded simultaneously. Inspection of figure 9-4 shows that the light which emanates from edge 1 must travel the extra stretch $b_0$ compared to the light from edge 4.

Light from edge 1 must be sent out

1 \[ \Delta t = \frac{b_0}{c} \]

earlier as compared to edge 4. During this time the rectangular prism moves an additional distance

2 \[ \Delta x = v \Delta t = \frac{v b_0}{c} \]
Thus in order to arrive at the same time on the photo plate, emission occurs from points 1 and 4' and similarly 2 and 3'. We note also that a Lorentz-FitzGerald body contraction applies only to the side $a_0$.

A possible interpretation of the photo plate picture is shown in figure 9-5. The edges of the moving rectangular prism appear at locations which are the same as would be seen considering a projection of a prism rotated out of the photo plate plane around an axis perpendicular to the normal of the photo plate and the velocity $\vec{v}$ by an angle $\vartheta$, where

$$\sin \vartheta = \frac{v}{c} \Rightarrow \sqrt{1 - \left(\frac{v}{c}\right)^2} = \cos \vartheta.$$ 

The observed effect is a quasi-rotation attributed to both the effect of Lorentz-FitzGerald body contraction and the time difference required to see signals simultaneously in $S$ originating from the back and the front of the body.

What we learn is that the Lorentz-FitzGerald body contraction is an input into any body measurement, irrespective of the inspection method which always needs to account for the effect of the finite speed of light. What we observe depends on how we
perform the measurement. The quasi-rotation interpretation with angle given in Eq. 3 is specific to the body motion normal to the line of observation. This result is valid during one particular observational instant and the effect is found in a specific method of inspection of the body. Naturally there will be some generalization of this result (see below).

Further reading: This effect has had some advocates. The article by J. Terrell Ref.1 (loc.cit) stimulated considerable discussion and several further presentations, for most readable see contributions by Victor F. Weisskopf2 and by Mary L. Boas3. In our view the ability to use measurement methods to change the picture image of the object observed is consistent with our presentation and is not impacting in any way the Lorentz-FitzGerald body contraction.

End IV–2: Quasi-rotation of a moving body for a special observer

9.4 Train in the tunnel: is the tunnel contracted?

We have considered three different measurements of the distance between coordinate events and obtained three different results. The first two cases describe the Lorentz-FitzGerald body contraction using the Lorentz transformation and show that the observer for whom the time interval between the two coordinate events vanishes observes the shortest, Lorentz contracted separation of the events. The third case (c) leads to a different result, analogous to the Doppler phenomenon, described in section 13: here the measurement events are not simultaneous in the source frame of reference, nor in the observer frame of reference.

We now reconsider the example of the train traveling through a gated tunnel. The situation is that a long train starts at a station and passes at a very high speed through a tunnel which is shorter than the length of the train as measured at the station. This situation can looked at in both the context of the Lorentz-FitzGerald contraction and of the coordinate transformation:

- Lorentz-FitzGerald contraction advocates note that the length of the moving train is contracted and the train will fit easily in the tunnel. However, they start to worry when Simplicius points out a seemingly equally valid view of the situation: the principle of relativity should mean that considering a frame of reference where the train is at rest and the tunnel is moving, the tunnel would be Lorentz-FitzGerald contracted and the train does not fit. Fortunately, we have just seen that there is no trivial reciprocity; it is important to define the measurement process and remember that time in the context of measurement is relevant.

• Coordinate transformation experts note, much to the relief of the concerned Lorentz-FitzGerald contraction advocates, that in the rest frame of the train (rather than the tunnel) the front and rear doors no longer drop down at the same time, and thus when the front of the train reaches the tunnel exit, at the rear end there is still time to escape the slam of that second door.

As we have demonstrated, in either frame of reference – the train or the mountain – the coordinate transformation experts determine exactly the same Lorentz-FitzGerald contracted length for the train since the measurement process determines which body is contracted. The Lorentz transformation is therefore consistent with the Lorentz-FitzGerald body contraction, and the principle of relativity is confirmed. Those who claim that in the rest frame of the train the tunnel is observed to be contracted perform a different experiment which has nothing to do with the train traveling through a tunnel.

Recall here again the Michelson-Morley experiment, in which we compare two lengths, one contracted (parallel to motion), and the other not (perpendicular to motion). The contraction in the direction of motion is needed to assure that the optical path is equal in all directions irrespective of the relative velocity vector with the æther, the carrier of light waves. Thus no matter what relative velocity there is between the experimental table with the MM interferometer and the æther, the outcome of the experiment is null; the magnitude of the relative velocity is unmeasurable. However, it is not the æther which is rendered nonexistent, it is the velocity of the æther that has been shown to be inaccessible to observation, a crucial point Einstein explained in 1920, as we discussed in the Introduction.

This seems to contradict the assertion that we can observe Lorentz-FitzGerald contracted body length. This is not the case; we can directly measure the Lorentz-FitzGerald contraction as it occurs, e.g., for the train that starts from a station, accelerates and later fits into the mountain tunnel. In the next section we will demonstrate that the classic Lorentz-FitzGerald contraction of a body that starts to move is indeed a real effect, in that an experimental arrangement can be built to measure the effect of contraction and keep it in memory even after the motion has stopped and the body is no longer contracted.

Discussion IV-1 – Body contraction and coordinate transformation

**Topic:** We take a second look at how one connects Lorentz coordinate transformation with Lorentz-FitzGerald body contraction.

*Simplicius:* Without doubt there must be a way to recognize the Lorentz-FitzGerald body contraction using the Lorentz transformation.

*Professor:* It is unfortunate that ‘Lorentz’ appears twice in this statement as it often creates confusion. Lorentz was seeking a transformation that would be consistent
with his explanation of the MM experiment in terms of the Lorentz-FitzGerald body contraction, yet remain consistent with Maxwell electromagnetism. Ultimately this problem was solved by others but being first to pose the question he was awarded the naming credit. It would have been far better to avoid the ‘Lorentz’ confusion seen that ‘contraction’ and ‘transformation’ are two different concepts.

**Student:** Which, however, must be consistent, as Lorentz already recognized.

**Simplicius:** So how does that work for the train in the tunnel?

**Student:** We measure at the station the train length \( L_0 = x_2 - x_1 \) with a measuring stick, which means we made simultaneous observations in the station frame of reference of both train ends, measuring at \( t_2 - t_1 = 0 \). For the observer in the train once the train departs the station the train length remains the same since the measurement process is the same. Another way to say this is that both the stick and the train change in same way.

**Professor:** The important point here is that this measurement is carried out at \( t_2 - t_1 = 0 \) in the train reference frame. To notice something happened to the moving train length we must measure insisting on equal time measurement in the station frame of reference. The result we find is that a train is observed to be contracted by an observer who sees it moving and measures at equal time in her frame, e.g. observing from the station frame of reference, which is the same as that of the tunnel.

**Simplicius:** If the contraction depends on the process of measurement, it seems to me that it is not ‘real’.

**Professor:** No, the point is that while the contraction is real, how we understand the consistency with the Lorentz transformation properties requires extra thought about time difference involved. Two observers will, unless they are comoving obtain different results.

**Student:** This is just like with kinetic energy: two different Galilean observers see two different values of the body kinetic energy.

**Professor:** Einstein derived the Lorentz transformation from basic principles. The connection and interpretation of a measurement process leading to an understanding of the the Lorentz-FitzGerald body contraction in terms of a coordinate transformation Einstein discovered is another step.

**Simplicius:** I find it very odd that a young clerk in a patent office came across the solution of such a fundamental riddle that a prize-winning Lorentz could not solve.

**Professor:** Lorentz became a great admirer and friend of Einstein’s. Why was the solution of the Lorentz-transformation riddle found by Einstein? I would argue that Einstein was in a unique and privileged circumstance. On one hand he had to evaluate numerous patents written in the wake of Maxwell, Hertz, Edison, Tesla, and others, writings which were probably not always arguing their case correctly. Confronted with many misunderstandings, Einstein formulated a new way to think about electromagnetism, perhaps out of necessity to reconcile the differing points of view to do his patent clerk job effectively. In fact, in his first 1905 relativity paper Einstein acknowledges his
lifelong friend Michelangelo Besso, an engineer, whom he had known since 1896, and with whom he walked nearly everyday to work in Bern. I believe that this acknowledgment is not only of the person but also of the technical-engineering environment that was Einstein’s intellectual home.

Exercise IV–3: Compounding Lorentz contraction: two moving frames

We look at how the observed length of the rod aligned with the direction of motion changes for two different observers moving along the same direction – the measurement is made at equal time in each observer’s frame. We are interested in expressing the length $L_2$ seen by observer $S_2$ given the observed length $L_1$ where we know both the velocity $v_2$ of $S_2$ with respect to the body and $v_{12}$ the relative velocity of $S_2$ with respect to $S_1$.

**Solution**

Perhaps a pictorial example is illuminating here: let the observer $S_1$ be at rest in the train station from which a train starts accelerating towards the tunnel, achieving speed $\beta_1$. Let $S_2$ be an observer that is on a second train that follows the first, moving slower $\beta_2 < \beta_1$. The relative speed between $S_1$, the station observer and $S_2$ the second train observer is according to Eq. (7.34)

$$\beta_{12} = \beta_1 - \beta_2, \quad y_{r12} = y_{r1} - y_{r2}. \tag{1}$$

We show relative velocity and relative rapidity relations. We will need below

$$\tanh y_{r2} = \beta_2, \quad \cosh(y_{r1} - y_{r2}) = \gamma_{12}, \quad \sinh(y_{r1} - y_{r2}) = \beta_{12} \gamma_{12}. \tag{2}$$

For each of two observers $S_i$ measuring in their respective rest frame at their equal time the rod (train) is contracted to the length $L_i$. Using rapidity $y_{r_i}$ of the observer with respect to body rest frame

$$L_0 = \cosh y_{r_1} L_1, \quad L_0 = \cosh y_{r_2} L_2, \quad \Rightarrow \quad \cosh y_{r_2} L_2 = \cosh y_{r_1} L_1. \tag{3}$$

The length measured by observer $S_2$ can now be written

$$L_2 = \frac{\cosh(y_{r1} - y_{r2} + y_{r2})}{\cosh y_{r2}} L_1, \tag{4}$$

where we extended trivially the argument of cosh in the nominator. Proceeding as in exercise III–19 we obtain

$$L_2 = [\cosh(y_{r1} - y_{r2}) + \tanh y_{r2} \sinh(y_{r1} - y_{r2})] L_1. \tag{5}$$
Using Eq. \ref{eq:2} we obtain

\[ L_2 = \gamma_{12}(1 + \beta_{12}\beta_2)L_1. \]

This expression is easily verified as it can be also written in the format

\[ \gamma^{-1}_{2} = \gamma^{-1}_{1}\gamma_{12}(1 + \beta_{12}\beta_2) \quad \Rightarrow \quad \gamma_1 = \gamma_2 \gamma_{12}(1 + \beta_{12}\beta_2), \]

which shows the \( \gamma_1 \)-factor as result of the transformation of \( \gamma_2 \) introducing the transformation between \( S_1 \) and \( S_2 \). Thus we have no doubt about Eq. \ref{eq:6}.

We restate our result

\[ \sqrt{1 - \beta_{12}^2}L_2 = (1 + \beta_{12}\beta_2)L_1, \quad \frac{\sqrt{1 - \beta_{12}^2}\sqrt{1 - \beta_2^2}}{1 + \beta_{12}\beta_2} = \sqrt{1 - \beta_1^2}. \]

This is not the naively expected product of Lorentz contraction factors.

We obtained the result Eq. \ref{eq:8} assuming: a) the relativistic addition of velocities, Eq. \ref{eq:1}, and b) both observers agree on the proper (rest frame) length of the body (train standing at station), Eq. \ref{eq:3}. This shows that contractions do not accumulate multiplicatively.

In retrospect this is almost a trivial result in view of Eq. \ref{eq:7}. The reason is that each observer performs a different measurement, at equal time in her respective reference frame.

End IV–3: Compounding Lorentz contraction: two moving frames

10 The Bell Rockets

10.1 Rockets connected by a thread

Imagine a physicist in a relativistic rocket who wants to measure the Lorentz-FitzGerald body contraction. S/he would find that all body lengths remain the same, as all measuring instruments would also be subject to the Lorentz-FitzGerald body contraction. Because of this difficulty, it would seem that we cannot prepare and maintain the standard of length between two different reference frames. On the face of this one very simple argument, many conclude that the measurement of Lorentz-FitzGerald contractions is, in principle, impossible.

To remedy this pessimistic view, we recall that the Lorentz-FitzGerald contraction was discovered to explain the Michelson-Morley experiment; it shows no fringe-shift because the body of the experiment is contracted in the direction of
motion. MM is a laboratory demonstration of the Lorentz-FitzGerald contraction of an unknown magnitude. The magnitude is unknown since an absolute velocity is in principle non-measurable within the context of Einstein’s theory of relativity.

We thus conclude from the two simple arguments that a measurement of the Lorentz-FitzGerald contraction requires comparison of the standards of length between two reference systems with a known relative velocity. We proceed to present the method devised by John S. Bell which involves an apparatus that will keep a ‘memory’ of the history of the contraction. J.S. Bell’s article, *How to teach special relativity* posed a question based on these concepts which has come to be known as the “Bell Rocket Paradox”.

To explain the method we consider the example depicted in Figure 10-1 of two point particles initially separated in space by a distance $D = x_2 - x_1 = D_0$ and starting to move from rest in the laboratory frame $S$. The particles are accelerated uniformly and identically for all time $t > 0$, leading to the equations of motion describing each world line and in particular their difference:

$$m \frac{d^2}{dt^2} (\vec{x}_1 - \vec{x}_2) = 0,$$

$$\frac{d}{dt} (\vec{x}_1 - \vec{x}_2) = \Delta \vec{V},$$

$$(\vec{x}_1 - \vec{x}_2) = \Delta \vec{V}t + \Delta \vec{x}_0.$$

We see that if we start our particles simultaneously from rest, *i.e.*, with $\Delta \vec{V} = 0$, their spatial separation $D = D_0 \equiv |\Delta \vec{x}|$ remains constant. This is so since the spatial separation between two particles is not subject to Lorentz-FitzGerald body

---

4John Stewart Bell (1928-1990), was a British-Irish physicist who is famous particularly for his work on quantum theory. For thirty years, from 1960 until his untimely death, he was a pivotal figure in the CERN Theory Division in Geneva. Bell is most known for his study of Einstein Podolsky Rosen quantum paradox and the development of Bell inequality. For appreciation article see R. Jackiw and A. Shimony, “The Depth and breadth of John Bell’s physics,” Phys. Perspect. 4, 78 (2004) doi:10.1007/s00016-002-8359-3 [physics/0105046 [physics.hist-ph]] and special issue of *Europhysics News*, 22 No.4 pp65-80 (April 1991)


Figure 10-1: Two point particles separated by distance $D = x_2 - x_1 = D_0$. (a) at rest, and in case (b) moving at velocity $\vec{v}$ acquired at a later time.

contraction. Therefore we recognize that the independent but identical motion of two bodies provides the standard of spatial distance parallel to motion which can be maintained between two reference frames: the initial rest frame and the second rest frame at any finite relative velocity.

If we consider along with J.S. Bell two rockets instead of two point particles, as in Figure 10-2, we find that the centers of mass of the rockets also maintain a constant separation $D = D_0$. However, because the rockets themselves are rigid bodies, the measured length of either rocket will surely contract about its center of mass by a factor of $\gamma$. The Lorentz-FitzGerald contraction has an effect only on the physical length of each rocket, but not on the spatial separation between the two centers of mass of the rigid bodies.

Next we add an arbitrarily weak, thin thread of length $l$, just long enough when the rockets are at rest to span the gap connecting the two rockets by their centers of mass, as is shown in Figure 10-3.

10.2 The thread breaks

Bell’s question is, what happens to this thread as the rockets gently accelerate? We have already considered the separation in space of the two independent rockets which must remain constant, since the rockets certainly accelerate independently. We have already discussed several times that any object, such as a train in a tunnel or here a thread will be subject to Lorentz-FitzGerald body contraction in the conditions here described

$$l = \frac{D_0}{\gamma} \rightarrow l < D_0 , \quad D(t) = D_0 .$$

(10.2)
10. The Bell Rockets

Figure 10-2: Two rockets of length $h$ separated by distance $D = x_2 - x_1 = D_0$. (a) at rest, and in case (b) moving at velocity $\vec{v}$ acquired at a later time.

Figure 10-3: Two rockets separated by distance $D = x_2 - x_1 = D_0$ and connected by a thin thread of (a) at rest, and in case (b) moving at velocity $\vec{v}$ acquired at a later time.

The obvious and correct answer, that the Lorentz-FitzGerald contraction causes the string to break, is often met with skepticism. Indeed one can see that if the thread was sufficiently strong it could pull the rockets together despite the presence of two independent rocket engines, and the entire system should contract as one rigid body. However, one can always choose a thread that is as weak as needed to allow the rockets to move independently and such a string must break.
10.3 Lorentz-FitzGerald contraction measured

We now refine Bell’s ‘two rockets connected by a thread experiment’ by providing two ratcheted spools, one with thread allowed to unroll on one rocket, and another roll-up thread spool on the other rocket. The spool with thread unwinds adding more thread when we accelerate the rockets, allowing the Lorentz-FitzGerald body contracted thread not to break but always connecting the two rockets. In this way we can read off and gauge the body contraction at any speed. When the rockets slow down, the roll-up spool picks up the slack in the connecting thread. The roll-up spool assures permanent memory of changing body contraction effects. Bell’s ‘two rockets with two ratcheted spools’ allow measurement of both transient contraction effects and cumulative body contraction effects.

Even though this Bell’s spooled thread method appears extremely cumbersome, it establishes the in-principle observability of the effect of the Lorentz-FitzGerald body contraction based on a comparison with a spatial separation distance standard. Therefore it clarifies the physics contents of one of the most contentious phenomena of special relativity.

We now return to our original question: can we transport between different reference frames the standard unit of length without contraction, and if so how? We see now that the Bell rocket example offers a method by comparing (contracted) material bodies with a (constant in time) spatial separation. One can envision a box carried on the relativistic train. The box contains two miniature rockets, each accelerating independently at the same rate, and exactly the same rate as the relativistic train. The constant separation of the floating rockets then preserves the unit distance providing all along the travel path the reference unit of length against which one determines instantaneous contraction of the relativistic train.

This instrument does not measure the absolute velocity, but the relative velocity achieved since travel started, and does this by means of the associated Lorentz-FitzGerald body contraction. Like relativistic trains and rockets, Bell’s device remains in the realm of distant technology.

Exercise IV–4: Bell’s example considered from one of the rockets

We consider how an observer $S'$, located on the rear rocket, as shown in figure 10-3 evaluates the situation with Bell’s rockets.

Solution

We call the measurements of the positions of the spaceships in $S'$, $x'_1$ and $x'_2$, and the positions of either end of the string in $S'$, $l'_1$ and $l'_2$. The separation of the ships, $D'$,
and the length of the string, \( l' \) are

\[
1 \quad D' = |x'_2 - x'_1|, \quad l' = |l'_2 - l'_1|.
\]

Measurements in \( S' \) take place at \( t'_2 - t'_1 = 0 \). Assuming the string is accelerated by the rear rocket, independently of the front rocket, we know that its length in \( S' \) will be constant: \( l' = D_0 \).

To determine the separation between the rockets, we first look at the temporal back transformation into \( S' \):

\[
2 \quad t_2 - t_1 = \gamma ((t'_2 - t'_1) + \frac{v}{c^2} (x'_2 - x'_1)) = \gamma \frac{v}{c^2} D' = t_2 - t_1 = \gamma \frac{v}{c^2} D'.
\]

Even though we have not solved for \( D' \) explicitly, it is clear that a measurement of the rockets simultaneously in \( S' \) corresponds to a measurement in \( S \) of the rear rocket’s position prior to the measurement of the front rocket’s position. The effect is analogous to the case of a train in the tunnel in the train reference frame. The front rocket will further advance during this time difference and thus the string connecting the rockets, which have moved further apart, must break.

---

End IV–4: Bell’s example considered from one of the rockets

---

Exercise IV–5: Measurement simultaneous in \( S \)

Another way to attack the Bell problem is to look again at \( S' \), located on the rear rocket, as shown in figure 10-3, but now with measurements simultaneous in laboratory frame \( S \), that is \( t_2 - t_1 = 0 \). Given this context, calculate the separation of the rockets in \( S' \), \( D' = x'_2 - x'_1 \).

Solution

A measurement simultaneous in \( S \) corresponds to a measurement with time difference in \( S' \):

\[
1 \quad t'_2 - t'_1 = \gamma ((t_2 - t_1) - \frac{v}{c^2} (x_2 - x_1)) = -\gamma \frac{v}{c^2} (x_2 - x_1) = -\gamma \frac{v}{c^2} D_0
\]

We now put this time difference, eq. [1] into the spatial inverse transformation from \( S' \) to \( S \):

\[
2 \quad x_2 - x_1 = \gamma ((x'_2 - x'_1) + v(t'_2 - t'_1)) = \gamma ((x'_2 - x'_1) - \frac{v^2}{c^2} (x_2 - x_1))
\]

Solving for \( x'_2 - x'_1 \), which is the definition of \( D' \) from eq. [1], yields:

\[
3 \quad D' = (x'_2 - x'_1)|_{t_2=t_1} = \frac{D_0}{\sqrt{1 - v^2/c^2}} > D_0.
\]
The length of the string is constant in $S'$, $l' = D_0$, no matter what the time difference between measurements. Measured at $t_1 = t_2$, the separation of the two rockets in $S'$ is greater than the length of the string; the string must break in $S'$. Not surprisingly, the ratio of these two quantities is:

$$\frac{4D'}{l'} = \gamma$$

---

Discussion IV-2 – Rigid bodies, relativity and ‘length clock’

**Topic:** ‘The thread breaks only if two rockets connected by a thread do not constitute a single rigid body. But when is a body not rigid? Is the theory of relativity able to describe the behavior of all extended material bodies? Many of these and similar questions can be debated – our conversation includes today a colleague ‘Iwo’ who will intervene only to keep the debate ‘honest’. This discussion picks up the argument we embarked on in conversation II-7 on page 65.

**Iwo:** The detailed analysis of the ‘two rockets riddle’ does not contain an essential ingredient: the discussion of the concept of a rigid body in the framework of relativity theory.

**Simplicius:** Right, the whole discussion we are having relies on the physical properties of the connecting thread. We expect that the thread length changes, and if the thread is strained it must be able to fail.

**Student:** Actually, I think this topic is moot, for two reasons: i) We introduced a spool of thread which means we can and indeed should use a thread that will not break or stretch but which unwinds. ii) Any study of Lorentz-FitzGerald body contraction must consider extended objects; there is nothing new to this case.

**Simplicius:** (Addressing the Professor) The question of principle raised by Iwo is interesting and I would like to learn more. Could you please explain how the Lorentz-FitzGerald body contraction, here of the thread, arises from the physical structure of the thread?

**Professor:** Very well. Let us begin by noting that a solid material object implies the presence of quantum physics. The birth of quantum theory was in part due to many ‘in principle’ problems of classical matter. I believe that material bodies of finite extent exist solely due to quantum physics.

**Simplicius:** So how exactly does the use of quantum physics help us to form rigid extended bodies?

**Student:** Crystals are a good example.

**Simplicius:** How does a crystal rod connecting two crystal rockets contract?
Professor: We need to determine how the electromagnetic forces that shape the quantum crystal structure, in terms of quantum physics and the electromagnetic interactions, change for different observers.

Simplicius: But Maxwell’s electromagnetism is a classical theory and inherently relativistic!

Professor: True or false, it is OK to think in this way since what matters is how the quantum electrons in the crystal adjust to the fact that the charged atomic nuclei we can think of as being classical move.

Simplicius: Moving charges make currents.

Student: And in turn, electrical currents induce magnetic fields. Thus, the moving atomic nucleus is a source of both electric and magnetic fields.

Professor: True, and in the next step we use the fields of moving nuclei to find by solving the quantum dynamical equations the needed deformation of the electron orbitals. This done, we find that the crystal will be contracted.

Simplicius: Can you get big contractions?

Professor: The computation is difficult if we want to make a big contraction in one step. Instead, we make very many small contraction steps, and use addition theorems. Thus all we need to show is that for a very small speed we obtain the expected tiny contraction of the crystal.

Simplicius: Again, how do you get a big contraction?

Professor: I can make a 2nd step with a second tiny velocity and so on. We have shown that the effects of two consecutive Lorentz transformations add according to the relativistic theorem of velocity addition. Similarly we can compute the expected summed up contraction comprising many small contractions. After accounting for the effect of all the small steps, we must find the expected result, a contraction of any magnitude, depending on the total velocity.

Simplicius: Even if crystals contract this does not mean that a piece of wood, liquids, or a heap of sand will do so.

Professor: Sand is a heap of individual crystalline silicon corns, and we actually only expect contraction of each crystalline particle. So in the absence of other forces in microgravity a heap of sand will not contract as this would mean contraction of space. This is like having many rockets chase each other, not only two as considered in Bell’s example. Each sand-rocket is contracted but the space between them is not. However, if you connect rockets strongly while the rocket engines are very weak, the entire system could contract.

Simplicius: What about a wooden stick?

Professor: Wood is made of complex macro molecules and therefore it is entirely intractable on my computer. However, as electron orbitals will contract and compress the long molecules just like the case of a crystal, I am convinced a piece of wood contracts just like a crystalline rigid body.
Simplicius: There are easily deformable substances such as water. I doubt that the body contraction is precise in this intermediate case.

Professor: The question is if ‘soft bodies’ experience a change in ‘packing’ in the context of body contraction. After some thought I am prepared to defend the following point of view: any bound body; that is, a body where removal of atoms or molecules costs noticeable energy, and that includes water and all liquids but few if any gases, will be subject to one and the same Lorentz-FitzGerald contraction. Only an unbound pile of matter will not contract. In between there are the very, very weakly bound piles which are difficult to understand.

Student: I still have one issue to resolve: In many relativity texts I see that there is a causality problem considering rigid bodies. Does this create another problem for Bell’s rockets?

Simplicius: I also remember reading that more generally, quantum physics has conflicts with causality.

Professor: The problem that these books note is that an infinitely rigid body could transmit a signal at superluminal speed: you push at one rod end, and the other end of a long stick moves at the same instant even if it is far away.

Student: I was taught that if I push at one end of the quantum crystal, the action propagates in the crystal in causal fashion and the other end moves away later. The action propagates no faster than the crystal’s light velocity which is known to be lower than light velocity in free space.

Professor: This is so since crystals actually are not infinitely rigid; they can fail when forced, and they support wave propagation, compression, density oscillations and all that is needed to assure causality and Lorentz-FitzGerald body contraction.

Simplicius: Yet some claim quantum physics can be acausal; is this possible?

Professor: This is loose talk. Quantum electrodynamics, the relativistic quantum field theory of charged particles achieves causality.

Iwo: All this is nice and good but again, accelerated extended rigid bodies have no place in relativity!

Professor: Agreed as it concerns considerations developed in the first years of special relativity before quantum physics was invented. Since then, we have learned about relativistic quantum physics. Moreover, in our world acceleration is so weak that it is practically irrelevant in that it is unable to perturb the arguments we presented. In fact there are 20 orders of magnitude that separate the ‘strong’ acceleration from usual laboratory experimental realm.

Iwo: I think SR describes very well the acceleration of particles under extreme conditions: do you claim that forces acting in the LHC\textsuperscript{7} are insignificant?

\textsuperscript{7}The Large Hadron Collider (LHC), the world’s largest and highest-energy particle accelerator 27 kilometers in circumference, lies in a tunnel as deep as 175 meters beneath the Franco-Swiss border near the city of Geneva.
Professor: The electromagnetic forces keeping particles in a LHC orbit, which while large compared to daily experience, are still extremely weak on natural scale. Were it not so, the charged particles kept by the magnetic field in the accelerator would radiate and lose energy upon traveling quantum scale distance! This is the meaning of ‘strong’ acceleration which breaches the physics discussed in this book. I will return to the topic in the book’s last chapter.
Part V

Time
Introductory remarks to Part V

The emphasis of this part of the book is on time and consistency of theory. This connects to pivotal questions which we all weight in our mind, such as: Why is superluminal speed of travel not natural? What does causality mean? Are we sure causality always prevails? Is travel in time possible? Of course everything that follows here is based on principles of SR and thus by assumption we have adopted the speed of light as the highest speed in the Universe. Still, there is much to discuss as the questions about consistency of the approach, and consequences abound.

In terms of new concepts and vocabulary this part of the book is very rich. We begin with an introduction to timelike, and spacelike event separation, the future and the past light cone, and light cone coordinates. We discuss causality and even tachyons. These concepts are laying foundations for several topics that need clarification.

Causality and quantum physics are subject of an essay that follows, extending and connecting the ideas of special relativity into the vibrantly growing field of quantum non-locality. We describe the current stage of the field and argue that there is no conflict with special relativity. What happens is prescribed by a causal process, but it is clear that we cannot resolve the quantum non-locality questions in context of local realism of special relativity.

We next explain how and why the time dilation effect is not reversible. Time dilation appears in common discussion in the context of rocket space travel (that is, the ‘traveling twin’ stays younger). We argue that by sending one twin on a space trip we assure by process of rocket acceleration that the path of the twin will reduce its proper time below that recorded by the inertial observer (the ‘twin at rest at base’). We extend the argument to include the case of triplets and argue which of two travelers will be younger compared to the base reference time, and explain why. We solve the example of spaceship travel.

When we turn to discuss the Doppler shift we see how the process of measurement determines which side of the shift equation gets the Lorentz $\gamma$ factor. Resolving this conundrum can only mean that time dilation has actually nothing to do with Doppler frequency shift. The proper measurement definition of Doppler shift requires understanding of the aberration effect (see section 7.2 in Part [II]), and from application of Lorentz coordinate transformation connecting observer coordinates to the coordinates of the light source.

To close we look in which way special relativity is subject to precision tests. We present the different experimental efforts based on Doppler effect, Michelson-Moreley experiment, and discuss a few theoretical considerations that seek to find limits to the SR framework presented in this book.
11 The Light Cone

11.1 The future

We know that between any two events in a vacuum a signal can propagate at a maximum speed of $c$. We place ourselves at the event point $x = y = z = 0$, at $t = 0$ and emit a light signal. Let us first consider a highly focused laser pulse propagating along the $x$-axis only. The signal reaches all event points that lie along the lines $x = ct$, and also $-x = ct$, should we point a second laser in the opposite direction. These two lines are shown in figure [11-1]. Similarly, we can accelerate material particles, shooting these with speeds $v < c$ along direction $\pm x$. Their paths will fill the region above these two lines in figure [11-1]. The unshaded domain for $ct > 0$ cannot be reached by particles originating at $x = y = z = 0$ and moving with $v < c$.

A signal transmitted from the origin, $x = y = z = 0$, could be emitted in any direction in the three spatial dimensions. If the signal were a pulse of light, it could reach the points described by

$$\vec{x} = (ct)\hat{n},$$

where $\hat{n}$ is a unit vector pointing in any direction. We can square this equation

$$\vec{x}^2 = (ct)^2.$$  \hspace{1cm} (11.2)

We write out Eq. (11.2) showing all coordinates

$$(ct)^2 = x^2 + y^2 + z^2,$$

and now we recognize that for this case of a flash in all directions, light traces not only the two boundary lines as shown in figure [11-1], but a much larger
domain. Setting $z = 0$ for illustration; that is, allowing vector $\hat{n}$ to point in the plane spanned by $x, y \neq 0$, we obtain a cone in the space-time-manifold shown in figure 11-2,

$$(ct)^2 = x^2 + y^2.$$ (11.4)

In space-time for $z = 0$ as Eq. (11.4) and figure 11-2 shows, the light rays appear on the boundary of the 'light-cone’. Generally we call the shapes shown in figure 11-1 (one spatial dimension) and in figure 11-2 (two spatial dimensions) the light-cone.

To understand better where the spherical light flash can go, consider again the spherical flash of light emitted at space-time point $x = y = z = 0, t = 0$. Allowing now $z \neq 0$ we write

$$(ct)^2 - z^2 = x^2 + y^2.$$ (11.5)

Allowing $z \neq 0$ we fill all points within the cone. For example, the event at $x/2, y/2, ct = \sqrt{x^2 + y^2}$ requires according to Eq. (11.5) $z^2 = 3(x^2 + y^2)/4$. A flash of spherical light emitted in past, will therefore reach, in the future, every point in the three-dimensional space given sufficient time as equation Eq. (11.3) describes. In the light-cone representation figure 11-2 this means the light goes to all points within the cone. The boundary of the light-cone is defined by Eq. (11.2) for all points where the Lorentz invariant is zero.

$$s^2 = (ct)^2 - \vec{x}^2 = 0, \quad t > 0.$$ (11.6)

$s^2$ does not change under a Lorentz transformation, see section 7.3. $s^2$ has the same value for any inertial observer. This is true regardless of the value of $s^2$, i.e.,
$s^2$ is invariant on the boundary of the light cone, as well as inside or outside the light cone. For all inertial observers connected by a proper (LT which preserves the sign of $t$) Lorentz transformation the constraints Eq. (11.6) are preserved. This means that the domain within the future light-cone is not affected by a proper Lorentz transformation change of coordinates.

The meaning of ‘future’ is thus a Lorentz invariant concept. However, we are restricted to ‘proper’ Lorentz transformations which e.g. preserve the sequence of events. As we noted briefly in the discussion following Eq. (6.19), the inclusion of ‘improper’ LT is necessary when considering the full set of space-time transformations called the Poincaré group.

If the observer at $x = y = z = 0$ were to shoot material particles in all directions, these, with fixed energy and mass have a fixed speed $v < c$. For a signal velocity smaller than $c$ the light-cone narrows, becoming a vertical line for particles at rest. For a material signal transmission, the future domain is the region inside the light cone defined where the Lorentz invariant is positive, so we have

$$s^2 = c^2t^2 - \vec{x}^2 > 0, \quad t > 0.$$  \hspace{1cm} (11.7)

The light-limit $s^2 = 0$ is approached asymptotically as the energy of material particles is increased. Since speed $v > c$ is not allowed, the space-time diagram such as figure [11-2] excludes for $t > 0$ much of the domain outside the future light-cone due to finite value of $c$. We are unable to influence observers outside of our light-cone.

Without giving technical details we note that in the context of quantum physics material particles explore a very, very small domain outside the light-cone, the range is exponentially small with characteristic quantum distance, the Compton wavelength $\lambda_C = h/mc$, $h$ being Planck constant.

The quantity $s^2$ is closely related to the proper time of a body, see section 7.3. By considering an observer always at rest at the origin of a reference frame we have

$$\sqrt{s^2/c^2} = \sqrt{t^2 - (\vec{x} = 0)^2/c^2} = \tau,$$  \hspace{1cm} (11.8)

where the use of $\tau$ denotes the fact that time measured by this observer at rest is the proper time in this reference frame. For another observer the values of $t$, and $\vec{x}$ change according to the Lorentz transformation, however the value of $\tau$ does not change, see Eq. (7.27).

### 11.2 The past

The same line of argument developed above for the future light-cone, case $t > 0$, also applies to the ‘past’ light-cone, case $t < 0$, the region shown in
Figure 11-3: The light-cone: the region of the past.

We can receive messages from all event points in the region bounded by the past light-cone; that is to say, the present can be influenced by any event in the shaded region in figure 11-3. This region contains all points that could be in our past.

When we look at the stars we see light from within the past light-cone, since each star shines in all directions. However the starlight we see can be of very different ages, determined by the distance light needs to travel to be recorded today. The domain of the universe we can observe is larger and larger as we look at older and older light. We will address this situation in section 12.1.

Exercise V–1: Light cone coordinates

Time and space coordinates are not the optimal coordinates to describe physical phenomena involving speed close to that of light. We explore here Lorentz transformation properties of light cone coordinates.

Solution

Ultrarelativistic motion occurs near to the light cone and can be described by light cone coordinates

\[ x_- = ct - x, \quad x_+ = ct + x, \]

while coordinates \( y, z \) remain unchanged. Depending on which direction a particle moves one of the coordinates \( x_\pm \) is very small, and the other is very large.
We consider the invariant
\[ s^2 = (ct)^2 - x^2 - y^2 - z^2 = (ct - x)(ct + x) - y^2 - z^2 = x_- x_+ - y^2 - z^2 , \]
and recognize the product \( x_- x_+ \). This shows that under a Lorentz transformation the two new coordinates Eq. 1 must transform inversely to each other

\[ x'_+ = e^{-y_r} x_+ , \quad x'_- = e^{y_r} x_- . \]

In the following we show that the exponent is the rapidity \( y_r \).

Using the explicit form of the (passive) Lorentz transformation we obtain

\[ x'_\pm = \gamma (ct - \beta x \pm \beta ct) = (ct \pm x) \gamma (1 \mp \beta) . \]

A straightforward computation yields

\[ \gamma (1 \mp \beta) = \frac{1 \mp \beta}{\sqrt{(1 \pm \beta)(1 \mp \beta)}} = \sqrt{\frac{1 \mp \beta}{1 \pm \beta}} = \exp \left\{ \ln \sqrt{\frac{1 \mp \beta}{1 \pm \beta}} \right\} = e^{\mp y_r} , \]

proving Eq. 3; compare here with Eq. (7.40). To boost to a high rapidity a particle moving to right such that \( x \simeq ct \), we make an active transformation and find light cone coordinates

\[ x'_- = ct' - x' = e^{-y_r} (ct - x) , \]

which is relatively small for positive \( y_r \), and

\[ x'_+ = ct' + x' = e^{y_r} (ct + x) , \]

which is relatively large for positive \( y_r \).

---

12 Causality, Twins

12.1 Timelike and spacelike event separation

We divide the entire space-time as seen from the origin of the coordinate system into two event regions. All events in the region

\[ s^2 > 0, \quad c^2 t^2 - \vec{x}^2 > 0 \quad \text{“timelike” events} \quad (12.1) \]
are called timelike. This includes the regions within the past and future light-cones. In contrast, the remainder of space-time, a region where

\[ s^2 < 0, \quad c^2 t^2 - \vec{x}^2 < 0 \]  \quad \text{“spacelike” events} \quad (12.2)

is called the spacelike region. An observer at the origin of the coordinate system cannot influence events in the spacelike region. Moreover, causal action originating in the spacelike region cannot be felt by this observer. The naming of the two regions follows the larger of coordinate values; that is, for \( ct \geq |\vec{x}| \) the region is timelike and where \( |\vec{x}| > ct \) it is spacelike.

In order to further generalize this discussion of the invariant quantity \( s^2 \), we now release the observer from the origin of the coordinate system. The more general definition of the space-time interval between two events \( E_1 \) and \( E_2 \) is

\[ s^2 = c^2(t_1 - t_2)^2 - (\vec{x}_1 - \vec{x}_2)^2. \quad (12.3) \]

This is equivalent to our previous definition of \( s \) if we shift our origin to one of the event points. Equation (12.3) thus defines the invariant ‘distance’ between two events; it is invariant in the sense that it does not change when events are subject to LT. For \( s^2 \geq 0 \), the distance is called timelike and the event with smaller \( t \) can due to causality influence the event with larger \( t \). On the other hand, for \( s^2 < 0 \), the two events are separated by a spacelike distance and they cannot communicate in a causal way.

At first glance it would seem that there are large domains of the Universe which not only are ‘out of sight’ but also will forever remain so, and thus could be subject to conditions that we cannot influence or ever observe. However, this, in general, need not be the case, since we can recognize two cases where the influence of the past could exceed the limits imposed by the light cone definition:

1) It was possible in the distant past \( t \to -\infty \), to furnish a large future region with measuring sticks and clocks synchronized by a light signal, starting from \( x = 0 \) as indicated in the larger light-cone in figure [12-1]. The future light-cone of
this initial synchronization signal is far larger and includes much more space-time than is within our future cone today – this means that an observer in the distant past could make sure that the Universe is the same on a scale vastly larger than accessible to us today;  
2) The light-cone of an event originating outside of our light-cone will ultimately intersect our cone. Observers far outside of our future light-cone and at rest or in motion with speed $v < c$ with respect to us today move into our light-cone domain at some later time.

**Discussion V-1 – Homogeneous Universe**

**Topic:** We now discuss a beginning in time, the Big-Bang, and how this can be consistent with a homogeneous Universe.

_Simplicius:_ I have a problem with the ideas presented here: I heard that there was a beginning in time of the Universe, a Big Bang. So nobody can go back further, and thus not all of Universe is in the future light-cone, ever.

_Professor:_ Good remark. The Universe is about 13.8 billion years old. We cannot go further back.

_Student:_ I have a problem with that. Today I can look to the right and to the left and see stars from far and away. If they are older than $13.8/2 = 6.9$ billion years and are seen in the opposite direction, I believe these stars are so far apart, more than 13.8 billion years, that they could not have been causally connected by a signal emitted after the Big Bang.

_Simplicius:_ That is indeed my problem: since there was a beginning, you cannot go back far enough in time to synchronize the behavior across the vast visible Universe that existed in the distant past. This is unsettling: the Universe to the ‘right’ and to the ‘left’ could be very different yet we see it to be the same in all properties we know how to measure today.

_Professor:_ The situation is more complex: the Big Bang implies that there is, and was, spatial expansion of the Universe. The expansion means that the light-cone is opening wider. Conversely, this means that a small spatial domain from the past created the present day vast Universe. Specifically, if the Universe spatial expansion were at a constant rate, we could argue that the stars we see back 6.9 billion years on left were not 13.8 but only 6.9 billion light years apart from those on right. Thus they could be synchronized in their properties at the time of the big-bang that is yet another 6.9 billion years earlier. Therefore, the Universe we look back at was not 27.6 billion light years across. Allowing for spatial expansion it must have been much, much smaller. Even so, the modern cosmological model finds it to be sufficiently large to leave us with a big problem, the so called ‘horizon problem’: why does the Universe today appear very much the same in every direction that we can look?

_Simplicius:_ This sounds just like what I feared; the Big Bang model of the Universe one way or another conflicts with causality, and probably special relativity.
**Professor:** The challenge is to reconcile the idea of a beginning in time, a Big Bang, with synchronization of properties of the Universe across a still rather large spatial domain. The solution presently adopted is that at the birth of the Universe, the Universe ‘hyper-inflated’, that is expanded rapidly, and this process assures that a large domain of the Universe we observe is connected and hence can be the same when we look at the horizon.

**Simplicius:** I heard that ‘rapidly’ means actually superluminal.

**Student:** I do not see how change in spatial scale can be affecting sequence of events; even in presence of superluminal spatial expansion it should not be possible to find an observer for whom the sequence of events reverses.

**Professor:** I also do not see how I could loop in time against the cosmic time arrow. Surely, our views about the first instances of a Big Bang will continue to evolve, and the topic we discussed will be revisited many times.

---

**Exercise V–2: Faster than light signals and causality?**

An inertial observer $S$ originates a ‘tachyonic’ signal that allegedly propagates faster than the speed of light (i.e., superluminal); that is, $\frac{\Delta x}{\Delta t} = v > c$. Considering another inertial observer, $S'$, moving relative to $S$ with the velocity $v'$, determine the range of $v'$ for which the causality is violated.

**Solution**

The world line in $S$ of the emitted signal has allegedly $v > c$. The observer $S'$ moves relative to $S$ with velocity $v'$. This observer is at rest in his system, and he measures the time interval $\Delta t'$, which can easily be determined considering the Lorentz-transformation from $S$ to $S'$:

$$\Delta t' = \frac{\Delta t - v'\Delta x/c^2}{\sqrt{1 - v'^2/c^2}}$$

$$= \Delta t \frac{1 - (\Delta x/\Delta t)(v'/c^2)}{\sqrt{1 - v'^2/c^2}}$$

We are looking for the condition such that $\Delta t'$ reverses sign compared to $\Delta t$; that is, where $vv'/c^2 > 1$. This is normally impossible, but by assumption we are considering $v > c$. We see that an observer $S'$ moving below speed of light can see the sequence of events reversed subject to condition $c > v' > c^2/v$.

Thus when a signal travels at $v > c$, for all observers $S'$ traveling at

$$v' > c \left(\frac{c}{v}\right),$$
causality, i.e. the sequence of events, is violated. A superluminal signal could reach observer \( S' \) before this observer could possibly observe (in his system) the cause of the signal.

Note that although the final answer Eq. 2 is symmetric between \( v' \) and \( v \) we assume that observer \( S' \) has \( v' < c \) (and therefore the signal has \( v > c \)) since we could not write Eq. 1.

End V–2: Faster than light signals and causality?

Discussion V-2 – Tachyons?

**Topic:** Can particles that travel with superluminal velocities exist?

**Simplicius:** Star Trek frequently mentions particles that are faster than light, called tachyons. There is even a book “Physics of Star Trek”\(^1\). Can tachyons exist?

**Student:** I checked the book. No tachyons there. The author, a well regarded theorist points out large effects of time dilation would ruin the show, so the warp drive was introduced to allow travel over galactic distances without need to approach light velocity and to battle time dilation.

**Simplicius:** But traveling faster than light over a large distance will for sure cause problems with causality, right?

**Student:** Any signal that travels faster than \( c \) will spell problems for relativity, and the easiest to recognize are the challenges to causality as we just considered. On the other hand, we can speculate how we could sidestep the problems that warp drive introduces. I think we will be in trouble unless there is another synchronization speed, perhaps within a different dimensional context (read ‘sub-space’), which is billions of times greater than that of light.

**Professor:** If so, everything we learn in this book about special relativity is applicable only to the realm of our space-time and matter we are made of, i.e., matter made of ‘light’, in the sense that its rest energy is \( mc^2 \) (note the \( c \)). Should there be another realm where everything can move much faster, I guess the ‘others’ could pick us up and take us quickly to a different location without a breach of causality.

**Simplicius:** This does not really answer the question: Do you believe tachyons could exist?

**Professor:** Science differs from science fiction in that it is based on observation of phenomena present around us or on proposals which, however difficult, can be decided by experiment. Your question reaches outside of the science paradigm\(^2\) of present day.

---


\(^2\) Thomas Kuhn, *The Structure of Scientific Revolutions* (University of Chicago Press, 1962) characterized a paradigm shift (or revolutionary science) as a change in the basic assumptions ruling science.
This does not mean that the answer to your question is ‘impossible’. It is unavailable based on all that we know about the world around us. A scientific revolution can occur in the future, a paradigm shift, which, based on a new experimental observation and/or compelling theoretical framework, will allow us to face these questions.

Simplicius: So, ‘maybe’?

Student: No, our Professor clearly said, neither yes, nor maybe, nor no. The answer is ‘nobody knows’.

Professor: Indeed, well put. We already agreed tachyons are not around, see conversation II-4 on page 52. Even so this is a good moment to tell an anecdote. A few years ago there was some commotion about the possibility that neutrinos could be just a little faster than the speed of light. It began with scientific publications proposing the neutrino as a tachyon and when the speed of neutrinos was finally for the first time very precisely measured, it seemed in excess of \( c \). The appearance of such an experimental possibility would have been just such a paradigm-shifting experimental result. I and many of my colleagues dropped our research projects and we turned to think about this topic. But it soon became clear someone did not plug a wire!

Simplicius: Too bad this experiment was wrong.

Professor: Actually, it felt good to try and fail. We must keep testing relativity. However, always remember that the end of Newtonian mechanics was only in part brought about by testing Newton’s laws; just as important if not more was the study of electromagnetism, a new physics domain, which was in conflict with classical physics.

Student: Is there anything new like this we can hope to discover today?

Professor: The Universe is full of stuff we have not seen, literally. We call the material part ‘dark matter’ and there is four times more of that around than of the matter we observe. Then there is the yet more mysterious ‘dark energy’: three times the amount of matter, acting akin to Einstein’s ‘cosmological constant’. We found all this by checking the action of gravity in the Universe at large distances, so some people think the gravitational force could be wrong. There are well known problems with reconciling gravity and quantum physics, and even at the level of classical electromagnetism we do not understand the physics of large acceleration and radiation reaction. The list of mysteries around us is long.

Simplicius: Please do not forget to add tachyons to your list.

Student: No. The mysteries that the Professor mentioned are established scientific riddles originating in experimental results for which no natural explanation is yet known. Tachyons have no place on this list as nothing indicates their presence in nature.


\(^4\)“Measurement of the neutrino velocity with the OPERA detector in the CNGS beam” OPERA Collaboration (T. Adam et al.). September 2011. 24 pp. e-Print: arXiv:1109.4897, result retracted a few months later.
Insight: Causality and proper time

Space-time distance, $s$, between two event points or, equivalently, proper time $\tau$ are

\[ s^2 \equiv c^2 t^2 - \vec{x}^2, \quad s^2 = c^2 \tau^2, \]

both are relativistically invariant. When $s^2 > 0$ two events, one at origin and the other at $(ct, x)$, they are said to be ‘timelike’ since $ct$ dominates $\vec{x}$ – in fact there is an observer $S'$ for whom $\vec{x}' = 0$. Such two events can influence each other. Conversely, when $s^2 < 0$, these two events are said to be ‘spacelike’. There is an observer who can claim that the time separation between these events is absent, $t' = 0$. Such events cannot communicate with each other.

Information can propagate at a maximum speed of $c$ and that creates the light-cone $|\vec{x}| = ct$. Therefore, there is a region of space-time, the ‘future’ that we can influence with our actions, and another region the ‘past’, from which our state can be influenced. Event points outside of the light-cone are acausal (that is require faster than light communication) with respect to coordinate origin.

Causal sequence of events cannot be altered as long as all velocities $v \leq c$.

Quantum phenomena obey special relativity and in particular there is no acausal transmission of information.

Proper time of a moving particle: To address more general motion we consider the differential increment of proper time of a body given by

\[ d\tau = dt \sqrt{1 - \bar{v}^2 / c^2} = \sqrt{(dt)^2 - (dx)^2 / c^2}, \]

where $\bar{v}$ is the instantaneous velocity of the body. The proper time, or ‘age’ of the body is then for an observer at laboratory time $t$

\[ \int^t d\tau = \int^t \sqrt{1 - v^2 / c^2} dt. \]

---

Exercise V–3: Spatial event separation

Two events are observed by an inertial observer $S$ to occur at $s_1 = (ct_1, \vec{x}_1)$ and $s_2 = (ct_2, \vec{x}_2)$ and are measured to have a spacelike separation. Demonstrate that even if $t_1 \neq t_2$ there is another inertial observer $S'$ moving with velocity $v$ who observes these events to occur simultaneously. Is this also possible for timelike event separation? Obtain the spatial separation of these two events observed by $S'(v)$ and write this result solely as a function of $v$ and $x'_1 - x'_2$, eliminating any dependence on $t'_1, t'_2$. Explain
and interpret your finding for the case that the two events correspond to the ends of a physical body.

Solution

Since $s^2 = (s_1 - s_2)^2$ is an invariant under Lorentz transformations, see section 7.3, once we have $s^2 < 0$ (spacelike separation Eq. (12.2)) all other observers $S'(v)$ agree that

1. $s^2 = s'^2 = c^2(t'_1 - t'_2)^2 - (\vec{x}'_1 - \vec{x}'_2)^2 < 0$

For a timelike observer $s^2 > 0$ we cannot ever satisfy the condition $t'_1 - t'_2 = 0$ since in that case it follows, according to Eq. (1), that for observer $S'$ we must have $s'^2 < 0$ which contradicts the invariance requirement $s^2 = s'^2$.

Thus only in case of spacelike event separation can we look for a frame of reference such that $t'_1 - t'_2 = 0$

2. $0 = t'_1 - t'_2 = \gamma(t_1 - t_2 - (x_1 - x_2)v/c^2)$

where as usual we oriented the coordinate system with $x$-axis being the axis of velocity $v$.

From Eq. (2) follows what the ‘equal time observer’ $S'(v)$ thinks about the measurement made by observer $S$

3. $t_1 - t_2 = (x_1 - x_2)v/c^2$.

Solving Eq. (3) for $v$ allows us to determine, given the measurements made by observer $S$ the velocity $v$ which is defining $S'$

4. $v/c = c(t_1 - t_2)/x_1 - x_2$.

Now let us consider the observer $S'$ who is special with $t'_1 - t'_2 = 0$ as our primary reference and compare to what $S'$ thinks about the properties of the events. Observer $S'$ reports that the spatial separation between the two events is

5. $x'_1 - x'_2 = \gamma(x_1 - x_2 - v(t_1 - t_2))$

We can use condition Eq. (3) to eliminate $t_1 - t_2$ and we obtain

6. $x'_1 - x'_2 = \gamma(1 - v^2/c^2)(x_1 - x_2) = \sqrt{1 - v^2/c^2}(x_1 - x_2) = \frac{1}{\gamma}(x_1 - x_2)$

Remarkably we find that the spatial separation of the events recorded by ‘equal time observer’ $S'$ is reduced by the reciprocal of the Lorentz factor compared to the spatial separation in the frame of reference $S$ where the events are not at equal time; rather they differ in time according to Eq. (3).
The situation can be now interpreted as follows: the equal time observer $S'$ can assign the two events which occur for him simultaneously to be associated with the ends of a material body traveling at velocity $v$. We recall the train in the tunnel discussion example. For $S'$ the long train fits into the tunnel due to Eq. (6). Observer $S$ is riding the train, and also reports that the train fits in the tunnel since the gates locking the train in the tunnel are actuated at different times according to Eq. (3).

While the physical reality is the same, the discussion of spatial events has a different connotation as compared to the discussion of body length. The consistency of Lorentz transformation combined with the measurement prescription ‘equal time observer’ $S'$ with body Lorentz contraction does not mean that the Lorentz contraction is not ‘real’. What we find is that the length of the body, and proper time of the body change in a way that is required to assure consistency with coordinate transformations. This consistency is essential for the relativity theory to be right.

We are familiar with a similar situation in classical mechanics where in the comoving body reference frame the momentum of the body vanishes. Naturally, this does not mean that momentum is not a ‘real’ property of a moving body.

---

12.2 Essay: Quantum entanglement and causality

This section is a digressive essay outside the main context of this book, as knowledge of and about special topics in quantum physics is expected. It is presented here in response to the frequent misunderstanding that there is a problem with causality considering long distance quantum entanglement.

The statement that the cause must precede the outcome by a time it takes a message to travel with speed of light has emerged in the context of quantum physics as not self-evident. The problem arises because ‘quantum nonlocality’ poses a challenge to classical local realism on which our conceptual view of the world is built and on which this book relies. To clarify what follows, proof of the ‘violation of the Bell inequality’ means that quantum non-locality prevails over classical local realism.

We begin by citing a claim from an abstract of a prominently published research paper that there can be superluminal signaling, read in plain language, an exchange of information:

"The experimental violation of Bell inequalities using spacelike separated measurements precludes the explanation of quantum correlations through causal influences propagating at subluminal speed. . . . assuming the

---

5Named after the same John S. Bell who contributed the ‘Bell rocket’ example extensively used in this book.

impossibility of using non-local correlations for superluminal communication, we exclude any possible explanation of quantum correlations in terms of influences propagating at any finite speed.

Let us consider \( S_1 \) as the originator of two quantum signals, and \( S_2, S_3 \) the two measuring observers. Above the authors believe that it is the process of measurement at location \( s_2 \) by the observer \( S_2 \) that defines the state observed by \( S_3 \) at \( s_3 \) at a spacelike separation

\[
0 > (s_2 - s_3)^2 = c^2(t_2 - t_3)^2 - (\vec{x}_2 - \vec{x}_3)^2. \tag{12.4}
\]

The spatial quantum nonlocality connecting spacelike distances in view of the authors require infinite speed. This topic clearly begs for further discussion in these pages.

Is the instantaneous sharing of quantum information across large spacelike distances in conflict with causality? We argue now why this is not the case. Note that \( S_3 \) does not know if \( S_2 \) measures at event \( s_2 \) first while \( S_2 \) does not know if \( S_3 \) measures at event \( s_3 \) first – a priori the observer closer to the source of the signal measures first. However, spacelike separated observers cannot communicate their location.

Who measures first, \( S_2 \) or \( S_3 \), is a moot question in the context of relativity given Eq. (12.4). This is so since for spacelike separated events we can always find an observer \( S' \) for whom the sign of \( \Delta t' \) is reversed as compared to \( \Delta t = t_2 - t_3 \)

\[
\frac{t'_2 - t'_3}{t_2 - t_3} = \gamma \left( 1 - \frac{v_x}{c^2} \frac{x_2 - x_3}{t_2 - t_3} \right) < 0 \quad \text{if} \quad \frac{v_x}{c} > c \frac{t_2 - t_3}{x_2 - x_3}. \tag{12.5}
\]

The event sequence \( t_2, t_3 \) does not matter for another reason as well: the measurement outcome was set up by observer \( S_1 \) who created the entangled quantum-non-local state and thus a table of probabilities for the measurement outcome. Thus the issue is that both \( S_2 \) and \( S_3 \) must be in causal sequence with \( S_1 \), the causal sequence of \( S_2 \) and \( S_3 \) does not matter at all.

Having clarified that there is no conflict with relativity that quantum nonlocality entails, we could now close the discussion, especially so since there are no known conflicts between quantum physics and special relativity. However, there are a few related issues worth further discussion, for example the statement (loc.cit.) “...we exclude any possible explanation of quantum correlations in terms of influences propagating at any finite speed”.

The laws of quantum physics at the microscopic scale are symmetric with regard to time arrow reversal, the so called T-symmetry. In general, when one studies the time evolution of a quantum system to assure causal, i.e., ‘arrow in time forward’ evolution we choose to consider the forward time arrow. The
fact that we discovered direct and very weak T-breaking in behavior of certain elementary processes has no immediate bearing on this argument.

A consequence of analyticity (word used in the mathematical sense describing the behavior of complex functions) of quantum amplitudes is that a quantum particle tries to tunnel into causally forbidden domain outside the light cone, but can only go the distance consistent with quantum uncertainty provided by the Compton wavelength \( \lambda_C = \frac{h}{mc} \) of the particle (for an electron, \( \lambda_{eC} = 2.4 \times 10^{-6} \mu m \), and for all other particles this is a much shorter distance). This microscopic quantum non-locality is a widely accepted feature of quantum physics, and addresses the behavior of one particle only; the macroscopic quantum non-locality requires two particles.

Let us next make the concept of quantum entanglement more precise: a source-observer \( S_1 \) sends out two particles which, by process of their creation, are ‘connected’ even though traveling in opposite directions. We speak of a fully correlated state when measurement of the spin polarization of one particle predicts with certainty the behavior of the other, potentially at a very distant location. Entanglement means the non local macroscopic presence of a quantum system, demonstrated by the fact that measurement performed ‘here’ implies an outcome ‘there, far and away’. The conundrum is, how can a measurement outcome here influence without time delay the other location?

Given this apparent conflict of quantum entanglement of particles with classical local realism it would help if quantum physics non-locality was wrong – revealed by the Bell inequality not being violated. However, we have seen definitive proof of the validity of quantum physics in a loophole proof demonstration of the violation of the Bell inequality by Hensen and colleagues using entangled electron spins and detectors separated by 1.3 kilometers. Two more such reconfirms of quantum physics followed. A neutrino based Bell-type experiment shows quantum entanglement over the distance of 735 km.

We are thus for sure facing the logical challenge: how a highly delocalized quantum system coordinates its behavior over spacelike distance. We keep in

---


8The situation with neutrinos is evolving: the three neutrinos may have masses that are much smaller – at the level of 0.01–0.1 eV which implies a quantum scale which is \( 5 \times 10^6 \) times larger – at the scale of 10 \( \mu m \). Moreover, we are with near certainty emerged in a sea of cosmic neutrinos, see for example: J. Birrell, and J. Rafelski, “Proposal for Resonant Detection of Relic Massive Neutrinos,” *Eur.Phys.J. C* 75 91 (2015).


mind that there cannot be any active communication between $S_2$ and $S_3$. All we know is that if either of the two observers $S_2$, $S_3$ measures one outcome, the other observer will measure a predictable, indeed prescribed outcome. This does not establish faster than light communication, as we read the expected outcome off the code table provided by $S_1$ who has set up the quantum-non-locality. It is the joint action of observers $S_1$ and $S_2$ that signals $S_3$, located at timelike distance from $S_1$. Without the action of $S_1$ there is nothing to measure at $S_3$. Causality is assured. However, measurement by $S_2$ collapses the many outcomes defined by $S_1$ to one particular ‘world’. This argument can be repeated considering how observers $S_1$ and $S_3$ signal $S_2$, located at timelike distance from $S_1$.

As it seems, there are virtual alternative worlds evolving, and the measurement at $s_2$ (or, respectively, $s_3$) makes one of them real, creating the same reality for $s_3$ (or, respectively, $s_2$). Resolving the communication-causality conflict we create the challenge of understanding how we collapse many possible future worlds into one reality valid everywhere. The solution of these questions awaits. At the first sight, this seems not to require any modification of SR.

Even if the above discussion addressed a quantum phenomenon, it bears resemblance to a classical physics context we now develop. This sharpens our understanding of the non-locality problem: An observer $S_1$ can predetermine what happens at the distant event $s_2$. For example, we ($S_1$) need to pre-program what a Mars rover at $s_2$ will do to assure that it can autonomously explore the distant surface for the period that signals take to make the return trip. We also call the rover observer $S_2$. More generally, $S_1$ can predetermine two events $s_2$ and $s_3$ associated with two observers $S_2$ and $S_3$ present in the future light cone of the observer $S_1$, and being separated by spacelike distance $(s_2 - s_3)^2 < 0$. This is in particular precluding according to special relativity that $S_2$ and $S_3$ can communicate with each other.

$S_1$ can predetermine what happens to $S_2$ and $S_3$ in a way that is ‘hidden’, so that the two spacelike observers $S_2$ and $S_3$ are not necessarily aware of having been synchronized with respect to their behavior. The outcome could fool $S_2$ and $S_3$ into arguing that a superluminal information exchange has occurred between them. Here we keep in mind that Bell inequality was invented to differentiate the quantum world from one with such a hidden variable classical context. Yet, let us proceed.

We note that additional complications arise when such a synchronization is carried out by $S_1$ without self-awareness of what s/he is up to. Now everybody involved, $S_1$, $S_2$, and $S_3$, can argue that faster than light communication is possible. How this happens we show by example. Keep in mind we are discussing classical physics processes.

Consider that $S_1$ makes a radio broadcast. When it is received simultaneously at two ends of your town, continent, planet, stellar system, etc, with receiving observers $S_2$ and $S_3$ at spacelike separation $(s_2 - s_3)^2 < 0$, this does not break
causality. Considering that everyone can hear the transmission of a radio station, the question of whether you turn your receiver on or not is the pivotal information content: while one listener knows what the other listener could hear, she does not know if the other observer was listening or not.

Following this line of thought, imagine the radio message instructs the space-like separated observers $S_2$ and $S_3$ to message each other, these are events $s_2$ and $s_3$. Let there be another observer $S_4$ for whom the events $s_2$ and $s_3$ are in the past light cone. $S_4$ reports that $S_2$ and $S_3$ were exchanging messages with each other as if they were instantaneously communicating. However, we know that the observers $S_2$ and $S_3$ were only executing commands introduced by $S_1$ to them. All of this is just like the radio example, even if the observers $S_2$ and $S_3$ do something upon receiving the signal from $S_1$ and appear to $S_4$ to exchange information instantaneously over spacelike distance $(s_2 - s_3)^2 < 0$.

In the final example $S_1$ informs both spacelike separated observers $S_2$ and $S_3$ how to incorporate a random amount of variance from certainty into the predetermined messages being exchanged, so that the messages $S_2$ and $S_3$ send each other are not fully predetermined but are the outcome of a probability-based game. Given the complexity of the game the observer $S_4$ is even more convinced that $S_2$ and $S_3$ are both aware of each other, and are messaging by superluminal communication.

To make the classical situation even more akin to the present day quantum entanglement experiments, let us imagine that the synchronizing communication that $S_1$ transmits to $S_2$ and $S_3$ was arranged by someone else (let us call this observer $S'_1$) at the transmission station so that the actual $S_1$ is not even aware that the synchronization message with action instructions is sent out. To make things worse let $S_4$ be also $S_1$ but observing at a later time. Clearly, $S_4 = S_1(t + \Delta t)$ could claim she has observed superluminal communication between $S_2$ and $S_3$. She would be fooled into doing this by $S'_1$ who may be now ‘dead’, and cannot rectify the claim made by $S_4$.

To sum up, the recognition of a possible predetermination of the measurement outcome removes in classical thinking the paradox of instantaneous communication in an example that mimics, but is not equal, to quantum entanglement. Yet this example naturally motivates in the quantum context the discussion of how $S_1$ by producing the entangled particles can establish a predetermined synchronization between $S_2$ and $S_3$.

In a popular rendering of the recent results of Hensen and colleagues (loc. cit.) the science writer of *Economist*\(^{12}\) takes the matter much further: “Just maybe . . . all these counter-intuitive findings . . . were all predetermined at the Universe’s birth, and all these experiments are playing out just as predetermined.” However, a) A ‘grand’ universal determinism is not needed, as our example clarified; b)

---

\(^{12}\)Economist of October 24, 2015, p77, “Hidden no more.”
classical determinism in small or big is not the explanation, for reason of Bell inequality. To date we have not found how to resolve the challenge that two particle quantum non-locality presents.

12.3 Time dilation revisited

Time dilation and the “twin paradox” continue to challenge many students. Imagine that an astronaut $S_A$ undertakes a trip to a distant star and back in a spaceship reaching $v \approx c$, while her/his twin remains on ‘the Earth base’, an inertial coasting observer $S_E$. Considering the effect of time dilation, the time in the spaceship $t_A$ should pass at a significantly slower pace than the time in the inertial base $t_E$. The traveling twin will therefore age more slowly than the base twin.

However, considering the principle of relativity, one could argue (wrongly) that the Earth’s base inertial system $S_E$, which we have assumed to provide the reference frame for the relative time measurement is not any different from the system at rest of the spaceship. From the viewpoint of the twin $S_A$ on the spaceship, the spaceship is at rest, and the Earth is traveling at a large velocity, carrying the Earth’s base twin $S_E$ away. Setting up the problem in this way, how can we tell which twin is younger when they reunite on Earth’s base? Furthermore, we could consider another observer $S_B$ following a different world line for whom, both the Earth’s base $S_E$, and the other spaceship, are in motion. Clearly, it cannot be that the aging of twins depends on the presence of this observer $S_B$.

The difference between the traveling twins is established by noting the length of their world line. The twin who has traveled away $S_A$, and later comes back to meet the twin $S_E$ on Earth’s inertial base, has a longer world line path in space-time. Because $S_A$ was subject to some acceleration, however small, it is not possible to reverse the situation claiming that the Earth twin $S_E$ traveled away. All inertial observers will agree that the accelerated body has a longer world line. The last statement is verified by evaluating

$$dl_A^2 - dl_E^2 = (dt^2 + dx^2) - dt^2 = dx^2,$$  \hspace{1cm} (12.6)

where $dl_A$ is the incremental worldline length of accelerated twin, and $dl_E$ that of laboratory twin. Under LT we find $dx^2 \rightarrow dx'^2 = [\gamma(dx' - dt'v)]^2 \geq 0$, where the increment vanishes for an observer who at a given instant in time is comoving with the accelerated clock, $dx'/dt' = v$. Except for this accidental instantaneous match in general $dx'^2 > 0$. We conclude that the integral over the entire world line difference Eq. (12.6) is always positive.

Although it is the acceleration of the traveling twin $S_A$ that changes his space-time path and distinguishes him from the other Earth’s base twin $S_E$, we need not
to improve on special relativity and/or invoke general relativity to find out who is younger, since the acceleration that distinguishes the world lines can be arbitrarily small. Thus we can make the determination of the time dilation entirely using the principles of special relativity. However, for purposes of clock comparison, special relativity necessitates that all travelers must meet at a common space-time point. This means that considering two inertial twins, one can remain inertial while (at least) the other \( S_A \) must accelerate or decelerate to meet \( S_E \) who will be the oldest traveler. By how much we find considering space-time paths.

**Exercise V–4: Which triplet is youngest?**

“Triplets” A, B, and E follow space-time paths that all start at the origin as shown in figure 12-2.

Triplet A travels at constant velocity \( v_A < c \), starting at \( t = 0 \) and ending at \( t = t_f \) (where \( t \) is measured by an inertial observer defining the spatial origin of the reference frame, i.e. triplet E below).

Triplet B travels at constant velocity \( v_B < c \) away from the origin until \( t = t_f / 2 \), at which point she turns around and heads back to the origin.

Triplet E remains at rest at the origin (on Earth’s inertial base).

Calculate and compare the proper time elapsed for each traveler between \( t = 0 \) and \( t = t_f \). In this exercise consider the period during which there is a change in velocity, i.e. when the travelers are subject to acceleration, as being negligible when compared to other travel periods.

**Solution**

For all of the triplets we obtain the proper time in terms of the velocity of each triplet measured by the laboratory observer.
For triplet E, \( u = 0 \), we have simply

\[ \tau_E^f = t_E^f. \]

For triplet A, its speed \( v_A \) is constant and

\[ \tau_A^f = t_A^f \sqrt{1 - \frac{v_A^2}{c^2}} = t_A^f / \gamma_A \]

which describes the usual time dilation, but the triplet is still far and away from Earth. For triplet B, \( v_B \) is again constant, only the direction of the velocity changes. Thus we have again

\[ \tau_B^f = t_B^f \sqrt{1 - \frac{v_B^2}{c^2}} = t_B^f / \gamma_B \]

For \( u_A = u_B \) we have \( \tau_A = \tau_B \) because the lengths of the paths in figure 12-2 are the same. However, while triplet B is already back, triplet A needs to take the road back to join B and E (not shown in figure 12-2). No matter how triplet A is going to come back she has to travel at finite speed, which we can choose for simplicity to be again \( v_A \) and this part of the trip will add to time dilation of A compared to the Earth observer E and compared to the observer B who is now at rest. We see that the result for two of the triplets who traveled at equal speed is that the one that went further away must be youngest when all triplets reunite.

We can explain this more precisely: the incremental change in the proper time is

\[ d\tau^2 = dt^2 - dx^2 / c^2 = d\tau_E^2 - dx^2 / c^2 \]

where we inserted the relationship that the laboratory time is also the proper time of the triplet at rest. Thus for any infinitesimal change \( dx > 0 \) the result is

\[ d\tau^2 < d\tau_E^2 \]

where the equal sign appears for \( dx = 0 \). Any addition to the distance traveled results in corresponding decrease in the traveler’s proper time compared to the situation before the addition.

For unequal \( v_A \neq v_B \) nothing changes in the above argument. The traveler who goes farthest away will in the end be youngest. Of course, it is easy to complicate our argument in that one of the travelers makes many turns, so that the distance traveled away will be small. What matters in this case is of course the sum of the distances traveled. To operationally resolve this situation one can evaluate the path-length of the world lines considered.

Laying a thread on the world path we see that longest thread is associated with the most traveled, youngest triplet, since any local extension of the thread results according
to Eq. 5 in a further reduction in the proper time of the traveler. The geometric length of the world line drawn in the \((t,x)\)-plane is a convenient measure allowing a comparison of any two observers traveling along different world lines.

---

End V–4: Which triplet is youngest?

12.4 Spaceship travel in the Milky Way

In this section we discuss the possibility of exploration of the Milky Way (our galaxy). Though this seems today impossible, this is not in contradiction with the basic laws of physics. Consider as an example a spaceship which takes off from the Earth and flies with a constant and quite comfortable acceleration of \(a = 1g = 9.81 \text{ m/s}^2\). It is perhaps worth noting that the characteristic time constant for accelerating at \(a = g\), is \(T = c/g = 353.8 \text{ days} \approx 1 \text{ year}\). This shows that after accelerating for one year with \(a = g\), the explorer spaceship should be close to the speed of light.

As result, an expedition subject to a constant acceleration will be able to travel extraordinary distances during a human lifespan. Should one day we find a way (think here about \(E = mc^2\)) to provide for the energy needs of this expedition, exploration of the Milky Way, and even a greater region of the Universe appears to be a possibility. However, as result of time dilation we expect a considerable difference between the spaceship time and the Earth time. Thus for any expedition this will be a one-way street, not so much because there is no way to come back, but because the time dilation effect separates explorers from the home base in time. Thus we consider here the duration of the voyage as seen both from the Earth and from the spaceship, given the desired distance we want to travel.

The coordinate values with primes will refer to the frame of reference moving along with the spaceship, and those without primes refer to the base on Earth. The spaceship’s ‘constant’ acceleration is due to processes in its rest frame, therefore it has in its frame of reference a constant differential velocity increase of

\[
dv' = a'dt' .
\]  

(12.7)

As this equation defines the meaning of \(a'\), there can be no question posed regarding its validity.

For the observers remaining on Earth, the velocity increase \(dv\) must be calculated using the addition theorem for velocities Eq. (7.9a):

\[
v + dv = \frac{v + dv'}{1 + vdv'/c^2} = \frac{v + a'dt'}{1 + va'dt'/c^2} ,
\]  

(12.8)
$a'dt'$ is naturally very small compared to $c$, so we expand terms, taking $(a'^2 dt'^2) \to 0$ to obtain

$$v + dv \simeq v + a'dt' \left(1 - \frac{v^2}{c^2}\right),$$

(12.9)

therefore

$$\frac{dv}{1 - v^2/c^2} = a'dt'.$$

(12.10)

This important equation relates the acceleration present in the spaceship’s rest frame to the differential change in velocity of the spaceship as observed from the Earth.

We integrate Eq. (12.10) and obtain

$$\text{arctanh} \frac{v}{c} = \frac{a't'}{c}.$$  \hspace{1cm} (12.11)

We have set the integration constant to zero since we require $v = 0$ at $t = t' = 0$. The velocity of the spaceship as a function of the ship’s proper time follows directly from Eq. (12.11):

$$v = c \tanh \frac{a't'}{c}.$$  \hspace{1cm} (12.12)

At this point we recognize a relation to the rapidity of the spaceship $y_s$, see section 7.5. Considering Eq. (7.45) we have

$$y_s = \frac{a't'}{c}.$$  \hspace{1cm} (12.13)

We could have obtained this equation directly by considering the incremental change in $y_s$ observed from the base due to the intrinsic acceleration process

$$dy_s = \frac{dv'}{c} = \frac{a'}{c} dt'.$$

(12.14)

The equality applies since the change in rapidity is very small, see Eq. (7.47). Since rapidity is additive, see exercise III–17, the effect accumulates, leading to Eq. (12.13). For immediate use we note that

$$\frac{1}{\sqrt{1 - (v/c)^2}} = \gamma = \cosh y_s = \cosh \frac{a't'}{c},$$

(12.15)

$$\frac{v/c}{\sqrt{1 - (v/c)^2}} = \frac{v}{c} \gamma = \sinh y_s = \sinh \frac{a't'}{c}.$$  \hspace{1cm} (12.16)

The rocket time $t'$ is related to Earth time $t$ by the Lorentz transformation. Consider a short increment of time

$$dt' = \frac{dt - (v/c^2)dx}{\sqrt{1 - (v/c)^2}}.$$  \hspace{1cm} (12.17)
We rearrange terms, and after squaring we obtain:

$$dt^2 (1 - (v/c)^2) = \left(1 - (v/c^2) \frac{dx}{dt} \right)^2 dt^2. \quad (12.18)$$

But we know that $v = \frac{dx}{dt}$, therefore:

$$dt' = \sqrt{1 - (v/c)^2} dt, \quad dt = \frac{dt'}{\sqrt{1 - (v/c)^2}}. \quad (12.19)$$

This is the usual time dilation formula: the time on the spaceship $t'$ advances slower than time on Earth $t$. Using Eq. (12.15)

$$dt = dt' \cosh \frac{a't'}{c} \rightarrow t = \frac{c}{a'} \sinh \frac{a't'}{c} = \frac{c}{a'} \frac{v/c}{\sqrt{1 - (v/c)^2}}, \quad (12.20)$$

where the last relation follows using Eq. (12.16). Solving this algebraic equation for $v(t)$ results in

$$v = \frac{a't}{\sqrt{1 + \left(\frac{a't}{c^2}\right)^2}}. \quad (12.21)$$

In terms of Earth time $t$ the distance traveled by the ship is

$$D = \int_0^x dx = \int_0^t vdt = \frac{c^2}{a'} \left(\sqrt{1 + \left(\frac{a't}{c}\right)^2} - 1\right). \quad (12.22)$$

We can find a relation $x(t')$ inserting into Eq. (12.22) the first form of Eq. (12.20)

$$D = \frac{c^2}{a'} \left(\cosh \frac{a't'}{c} - 1\right). \quad (12.23)$$

The two relations Eq. (12.22) and Eq. (12.23) express the same world path $x$, one parametrized in Earth time ($t$) and one in spaceship time ($t'$). We set $x = D$ which is the target distance and solve for $t$ (Earth time) and $t'$ (spaceship time), respectively

$$t = \frac{c}{a'} \sqrt{\left(\frac{a'D}{c^2} + 1\right)^2 - 1}, \quad t' = \frac{c}{a'} \text{arcosh} \left(\frac{a'D}{c^2} + 1\right). \quad (12.24)$$

To complete this analysis we evaluate the limit $D \to \infty$. We find, respectively

$$t \to \frac{D}{c}, \quad t' \to \frac{c}{a'} \ln \left(\frac{2a'D}{c^2}\right). \quad (12.25)$$

For the observer on Earth the travel time to a distant location is just about what the light would take, so a target 1000 ly away will be reached after 1000
years. However, the time in the spaceship advances slowly. Since $c/a' \simeq 1\text{y}$, the travel time would be $t' \simeq \ln 2000$ years; that is about 8 years. By the time the expedition returns to home base, two millennia have passed, while in the spaceship time practically stood still. This becomes even more extreme considering the time needed to cross the Milky Way: in the reference frame of Earth this will take about 30 000 years, however only 11 years will elapse for the traveler.

**Exercise V–5: Exploration of Vega**

Consider the duration of the trip to Vega, the brightest star in the constellation Lyra located 25 light years away from the Earth. In order to arrive in minimum time with maximum comfort we assume that the spaceship accelerates at $a = 1g$, and decelerates at the same rate after the mid-point of the trip. We allow for six months of research at Vega, and assume that the spaceship returns to Earth using the same sequence of acceleration and deceleration.

**Solution**

The ship’s space-time path from the perspective of the Earth-bound observer, is given by Eq. (12.24) and is shown in figure 12-3. The spaceship travels to Vega in two stages of $D = 12.5$ light years. We insert therefore the value 12.5 light years for $D$ and we use
1 light year = \( c \cdot 3.16 \times 10^7 \) seconds:

\[
t = \frac{299,800\text{km/s}}{9.81\text{m/s}^2} \sqrt{\left(\frac{9.81\text{m/s}^2 \cdot 12.5 \cdot 3.16 \cdot 10^7 \text{s} \cdot c}{299,800\text{km/s} \cdot c} + 1\right)^2 - 1} = 13.4 \text{ years },
\]

\[
t' = \frac{299,800\text{km/s}}{9.81\text{m/s}^2} \operatorname{arcosh}\left(\frac{9.81\text{m/s}^2 \cdot 12.5 \cdot 3.16 \cdot 10^7 \text{s} \cdot c}{299,800\text{km/s} \cdot c} + 1\right) = 3.22 \text{ years }. \]

Including the half-year of research time we obtain the total round trip time. We find:

(\text{Earth}) \quad t_E = 4t + 0.5 \text{ years} = 54 \text{ years}, 2 \text{ months}, 22 \text{ days},

(\text{traveler}) \quad t'_R = 4t' + 0.5 \text{ years} = 13 \text{ years}, 4 \text{ months}, 11 \text{ days}.

The main point of this exercise is that from the perspective of the spaceship crew, Vega, a very interesting (scientifically) star is reachable within a reasonable expedition schedule. The ‘only’ problem is that everyone at the base has aged by additional 40 years compared to members of the expedition. We do not need to worry too much about that as the technology of a constant 1g-acceleration is not on the present-day technology horizon.

At this point a reversed condition question arises, Fermi’s famous question: “Where is everybody?” referring to absence of a contact with more advanced civilizations. This question was posed during a luncheon conversation with Emil Konopinski, Edward Teller, and Herbert York in the summer of 1950. Fermi’s companions on that day have provided accounts of the incident\(^{[3]}\). This author likes E. Teller’s answer (loc.cit.) “…as far as our galaxy is concerned, we are living somewhere in the sticks, far removed from the metropolitan area of the galactic center.

---

**End V–5: Exploration of Vega**

**Exercise V–6: Aging in relativistic rockets**

Two spaceships leave Earth with the assignment of investigating habitable planets at distant stars and then returning to Earth 100 years later Earth time. One ship flies to star \( x_1 \) and then returns to Earth; the other ship flies in the opposite direction to stars \( x_2 \) and then \( x_3 \) before returning. In this problem we do not assume the velocity is a constant. a) Give a space-time diagram of the expedition. b) How does the spaceship crew determine the Earth year? c) How can we interpret the relative aging of the crews?

**Solution**

Figure 12-4: Space-time diagram of voyages to stars at $x_1$, $x_2$, and $x_3$. See exercise V–6.

a) The expeditions start in the year $t = 2050$, as is depicted in figure 12-4. During each acceleration phase the spaceships almost reach the speed of light, so the world lines continue only a little more steeply than the cone of light. Both expeditions make the prescribed stops and return to base in year 2150.

b) As in section 12.4 we assume, that the acceleration occurs in the direction of motion, hence Eq. (12.10) applies

$$\frac{dv}{1 - (v/c)^2} = a(t)d\tau,$$

in which $v = dx/dt$ is the velocity of the spaceship, $a$ is its acceleration in the spaceship rest frame and $\tau$ refers to the spaceship proper time. Integration leads to

$$\frac{v}{c} = \tanh \frac{1}{c} \int_0^\tau a(\tau')d\tau'.$$

If we use the instantaneous time dilation to link the Earth time $t$ with the spaceship time $t'$, then it follows that (using Eq. (12.15)):

$$dt = \frac{d\tau}{1 - (v/c)^2} = d\tau \cosh \frac{1}{c} \int_0^\tau a(\tau')d\tau'. $$

From this we find that the Earth time $t$ is a function of the spaceship-time-dependent acceleration:

$$t = \int_0^\tau \cosh \left[ \frac{1}{c} \int_0^{\tau'} a(\tau'')d\tau'' \right] d\tau'.$$

During the flight the Earth time can be calculated using Eq. 4. In acceleration-free phases there is a time-dilation, which follows from Eq. 3 given the attained velocity $v$
c) The spaceship crews age based on their proper time,

\[ \text{AGE} \equiv \int_0^\tau d\tau'. \]

The proper time is related to Earth time by the Lorentz transformation. We must repeat earlier consideration since now there is for much of the time an acceleration present. Let us consider a short increment of time. Holding \( v \) constant with respect to time in that instant, we have

\[ d\tau = \frac{dt - (v/c)^2 dx}{\sqrt{1 - (v/c)^2}}, \]

\[ d\tau^2 (1 - (v/c)^2) = \left(1 - (v/c)^2 \frac{dx}{dt}\right)^2 dt^2. \]

But we know that \( v = \frac{dx}{dt} \), therefore:

\[ d\tau = \sqrt{1 - (v/c)^2} dt, \quad d\tau = \sqrt{dt^2 - (dx/c)^2}. \]

We see that with regard to the differential relation between proper time and base time nothing changed even when \( v \) is not a constant as long as acceleration is small, a point we will look at more precisely in section 22.1 on page 304. Thus all earlier considerations apply and in particular the path length along the world line in figure 12-4 helps to determine the relative age of both expeditions and to compare them to the time at base.

---

End V–6: Aging in relativistic rockets

---

13 Relativistic Doppler Shift

13.1 Introducing relativistic Doppler shift

The phenomenon of motion-related change of sound frequency is today a common experience. The frequency is heard to be higher in the approach of a fast racing car as compared to when the vehicle recedes. Fast vehicles, planes or trains were not available in 1842 when Doppler\(^\text{14}\) considered the influence of binary star motion on the color of emitted light.

The classical (non-relativistic) Doppler shift was seen as being due to a change in the time it takes two wave crests to arrive in the detector, originating in the

\(^{14}\)Christian Andreas Doppler, (1803–1853), Professor of Mathematics and Practical Geometry at the Technical Institute (1841–47) (now Czech Technical University in Prague) at the time of the Doppler shift announcement in 1842/3.
Relativistic Doppler Shift
change in the separation between emitter and observer during this interval in time. It is easy to convince oneself based on the wave crest idea, that when source and observer move toward each-other, the frequency of wave crests increases, and the wavelength is shorter (blue-shift). Conversely when source and detector separate, the frequency of crests diminishes, and the wavelength is longer (red-shift). This non-relativistic shift thus does not exist if the relative motion of the source and observer is transverse to the line of sight, that is the line connecting source and observer.

After development of SR and the banishment of material æther, it must be understood that the original Doppler’s theory is more atune with the sound propagation in air. Air is a ponderable medium and thus sound is actually modified by the motion of the source with reference to the carrier of sound, the air. How we hear a moving object depends on the condition of the source with respect to the state of the carrier of the signal.

Therefore the reader should apply relativistic Doppler shift as discussed in this book to the measurement of light only. There are essential changes compared with the case of sound:

- Light travels in space in a way that must be completely independent of the dynamics of the source, there is no effect of motion of the source that can be ascribed to propagation of light.

- The relativistic Doppler wavelength shift must be relative and reciprocal.

Consider two light sources, one in laboratory and another on a rocket going on an exploration run in the Universe. We would argue that the wave-crest time ticks on the rocket would be slow as compared to those ticking in the lab. However, observers in lab and the rocket knowing only about the standarized light signal emitted by the other body should not expect a modification of the signal at the source. This means that the understanding of relativistic Doppler shift cannot possibly rely on the classical Doppler effect combined with time dilation at the source. Solving this riddle will keep us busy for a few extra pages.

Discussion V-3 – Doppler frequency formula.

**Topic:** We clarify why the light emitted by a distant observer is Doppler shifted by the effect of relative motion.

_Simplicius:_ I have been reading ahead in these fine print pages about Doppler shift and I must wonder why Professor gave this matter so much attention.

_Student:_ The Doppler shift method allows to measure the speeds of stars in our galaxy creating a map of motion and helping, for example, to disentangle the distribution of dark matter. The Doppler shift helps your GPS work properly. Do I need to continue
Insight: Relativistic Doppler and the ‘other’ shifts

There are three different mechanisms influencing observed frequency $\nu$ or, equivalently, the wavelength $\lambda$ of (star)light, where $\lambda \nu = c$. These effects are due to distinct physical phenomena:

1. The relativistic Doppler shift which we discuss in detail in this book is due to relative motion between source and observer. Relativistic Doppler shift is present within the special theory of relativity and it does not depend on action by a force, nor does it depend on how far light must travel.

2. Light emitted from within any gravitational potential well must do work to escape. This effect falls under the ‘general’ theory of relativity, but it can be also described to a very good approximation by employing energy conservation within SR. The effect has been demonstrated in a laboratory experiment\(^a\). This effect is rarely confounded with the Doppler shift as it is due to photon interaction with gravity. A photon escaping from a stellar source will do work to get away and thus is always subject to a gravitational redshift.

3. A very distant source of light is in general not speeding with respect to the Universe CMB, thus we normally should expect that such light (atomic spectral lines) when emitted should be observed to be similar if not the same as the light emitted by a similar source today. However, this light travels far through an expanding Universe and in the time it takes to reach the detector on Earth, the free-streaming photons are subject to the the expansion of the Universe is ‘stretching’ the length of the wave from crest to crest. The effect is large and grows with the time a photon takes to reach the observer producing a cosmological emission energy red-shift that grows with cosmological distance. Today it is widely agreed that this cosmological red-shift is neither due to motion nor to direct action of a force. Even so, cosmological redshift is sometimes described in popular renditions as being due to distant stars receding from us at ever increasing speed, the further they are. Such characterization is incompatible with the wealth of experimental information. The understanding of red-shift relation to source distance is highly refined today and has allowed experimental determination of accelerated expansion of the Universe\(^b\) leading to the concept of repulsive gravity due to dark energy.


\(^b\)The Nobel Prize in Physics 2011 was awarded to Saul Perlmutter, Brian P. Schmidt and Adam G. Riess “...for the discovery of the accelerating expansion of the Universe through observations of distant supernovae.”
to motivate your interest?

Simplicius: Why not write down the good old Doppler? There is nothing more nothing less, just include a Lorentz-gamma factor with what Doppler wrote down around 1842. This accounts for time dilation in the source. Finished. I think this could save some student head aches and a few pages in this book.

Student: Really? Keep in mind light is always emitted just like we know it would be emitted by a hot star. It is emitted for all to see. The atoms on the star that radiate the light cannot possibly know that they are in motion with regard to the Earth observer who is going to measure them many thousand years later. Or, for that matter, any other observer far and away. In absence of material æther the only thing that matters in the measurement of the signal is the relative state of the observer, many possible observers, each reporting a different measured frequency spectrum

Simplicius: Upps, I feel foolish! Yes, the relativistic Doppler shift is thus generated in the process of measurement. However, that is not what I read in another book.

Student: However, it is obvious: the light emitting observer sees the wave go away in his coordinate system and the light phase oscillates in space-time according to the emitter’s set frequency. However, the same phase oscillations in the other observer’s frame of reference will be interpreted in a different fashion.

Simplicius: This seems not at all like the classic Doppler shift with sound waves!

Student: Right, this is so since sound waves travel in air, a ponderable medium, and we can incorporate in the interpretation of the signals aside of relative motion between the loudspeaker and microphone also the relative motion with the air of one or both the loudspeaker and microphone.

Simplicius: I am still bothered by the fact that I can on this ‘Earth’ location discover the relative state of motion of a star out there.

Professor: This is only possible because the emitter uses a reference signal, e.g. the spectral lines e.g. in hydrogen. There is also a remaining ambiguity as to what is the shift in frequency, and what is the angle aberration shift, a very important part of the relativistic Doppler shift that is often not mentioned at all. This is the same aberration we studied in section [L.3] in the context of Bradley’s precise measurement of the speed of light.

Simplicius: When can we ignore the aberration of the observation angle?

Professor: For parallel motion only, when the line of sight is the line of motion. Let me explain why I make considerable effort to clarify Doppler effect here: when teaching relativity I and many Professors like to motivate students’understanding of the Doppler shift by introducing time dilation. Yet this situation is not at all similar to a twin going on a trip in the Galaxy. Only when aberration of the line of sight is accounted for, the relation between source and observed frequencies is reciprocal and allows consistent relativistic Doppler formulas. The lesson here is that we cannot talk about non-parallel Doppler shift without introducing the observation angle aberration shift.
13.2 Relativistic Doppler shift

Light as Maxwell realized more than 150 years ago was a wave solution of the equations we today call Maxwell equations. These are transverse waves which comprise the ‘plane-wave’ factor, see section 24.1 on page 345.

\[ W(x, t) = \Re e^{i\phi(x, t)} = \cos \left( \omega t - \vec{x} \cdot \vec{k} \right) = \cos \left( 2\pi \nu t - 2\pi \vec{x} \cdot \vec{n}/\lambda \right). \]  

(13.1)

This is the light wave emitted from a source \( S \) at rest with coordinates \( t, x \) in direction \( \vec{n} \). \( \lambda \) is the wavelength of light, and we introduce

\[ \vec{k} = 2\pi \vec{n}/\lambda \quad \omega = 2\pi \nu. \]  

(13.2)

For any wave of light the crest of the wave travels at the velocity of light:

\[ \omega^2 - c^2 \vec{k}^2 = 0, \quad |\vec{k}| = \omega/c. \]  

(13.3)

The observer \( S' \) travels in direction which we fix to be the \( z \) axis. The line of sight to the light source is characterized by a unit vector \( \vec{n} \), such that \( \vec{k} = k\vec{n} \). Typically we would need to look at spherical waves emitted by the source, but for simplicity we focus our attention on a plane wave traveling along the line of sight. This situation as shown in figure 13-1 corresponds to our earlier presentation in figure 7-4 when we discussed the aberration of light in section 7.2.

We focus our attention on the phase of the wave Eq. (13.1)

\[ \phi = \omega t - \vec{x} \cdot \vec{k}. \]  

(13.4)
The magnitude of this phase should not depend on the observer; in other words, the phase should be a Lorentz invariant. If this is not so, the wave could be used to cause observer-dependent physical phenomena. However, each observer invokes a different set of coordinates, e.g., for $S$ it is $t, \vec{x}$, and for $S'$ it is $t', \vec{x}'$. The only way we can keep $\phi$ invariant is that for each observer, $\omega, \vec{k}$ also change. Such a change is expected in consideration of the Doppler shift. We now aim to find how different relativistic observers record the change $\omega \to \omega', \vec{k} \to \vec{k}'$ such that

$$\phi = \omega t - \vec{x} \cdot \vec{k} = \phi' = \omega' t' - \vec{x}' \cdot \vec{k}' .$$

(13.5)

\(\vec{v} \parallel \vec{n}\): Motion parallel to line of sight

The alignment of $\vec{n}$ and $\vec{v}$ means that

$$k_x = 0 , \quad k_y = 0 , \quad k_z = k = \omega/c .$$

(13.6)

We introduce a transformation of coordinates $t', x'$ in Eq. (13.5) to the source coordinates $t, x$

$$\phi' = \omega' \gamma (t + zv/c^2) - k_z' \gamma (z + tv) .$$

(13.7)

We try the same transformation relation for $\omega, k_x$

$$\omega' = \gamma (\omega + k_z v) , \quad k_z' = \gamma (k_z + \omega v/c^2) , \quad k_x' = 0 , k_y' = 0 .$$

(13.8)

Note that the velocity ‘sign’ is the same in Eq. (13.8) as that used in the Lorentz boost of coordinates Eq. (13.7). There is a deeper reason for this result we discuss in Part [VIII].

We find

$$\phi' = \gamma^2 (\omega + k_z v)(t + zv/c^2) - \gamma^2 (k_z + \omega v/c^2)(z + tv)$$

$$= \gamma^2 (1 - \beta^2)(\omega t - k_z z) + \gamma^2 (k_z tv + \omega zv/c^2 - k_z tv - \omega zv/c^2)$$

$$= \omega t - k_z z = \phi ,$$

due to cancellation of the four cross-terms containing $k_z t$, or $\omega z$. Thus with the assumption Eq. (13.8) the phase Eq. (13.7) remains unchanged. It is straightforward to check consistency of Eq. (13.8): the new frequency and wave number of the wave, Eq. (13.8) satisfy the constraint Eq. (13.3)

$$\omega'^2 - c^2 k_z'^2 = \gamma^2 [(\omega + k_z v)^2 - (ck_z + \omega v/c)^2]$$

$$= \gamma^2 (1 - \beta^2)(\omega^2 - c^2 k_z^2)$$

$$= \omega^2 - c^2 k_z^2 = 0 ,$$

after cancellation of the cross terms containing $k_z \omega v$.

The notation choice of primed (moving observer) and unprimed (source) coordinates is made in order to relate our argument to the study of Lorentz transformation properties as presented in this book. However, in the final reference
formulas describing the Doppler shift we revert to introductory textbook notation: it is common to denote the frequency and the wavelength of the source by subscript zero; that is \( \nu, \lambda \to \nu_0, \lambda_0 \). Once this is done, the prime on the observed quantities can be dropped.

We obtain according to Eq. (13.8) the frequency \( \nu \) of the observer moving towards light source \( \nu_0 \) with the speed \( v \) parallel to the line of sight

\[
\nu = \nu_0 \frac{1 + v/c}{\sqrt{1 - (v/c)^2}} = \nu_0 \sqrt{\frac{1 + v/c}{1 - v/c}},
\]

that is, observed \( \nu > \nu_0 \). The relation of frequencies can be also seen inverting Eq. (13.9) where we find

\[
\frac{\nu_0}{\nu} = \sqrt{\frac{1 - v/c}{1 + v/c}} = \frac{1 - v/c}{\sqrt{1 - (v/c)^2}}
\]

which is also the functional result one obtains using \( v \to -v \) in Eq. (13.9) for observer motion away from the source.

The wavelength \( \lambda \) according to Eq. (13.2)

\[
\lambda \equiv \frac{2\pi}{k} = \lambda_0 \sqrt{\frac{1 - (v/c)^2}{1 + v/c}} = \lambda_0 \sqrt{\frac{1 - v/c}{1 + v/c}}
\]

is decreased (respectively increased for \( v \to -v \)): we do see a blue-shifted \( \lambda < \lambda_0 \) wavelength for the observer and source approaching each other. For \( v \to -v \) the relative motion reverses, and the light of a receding source is seen to be at a longer wavelength \( \lambda > \lambda_0 \), that is, red-shifted.

**General case of arbitrary \( \vec{v}, \vec{n} \).**

We now redo in part the computation of the Doppler shift relaxing the condition that the line of sight is parallel to the relative motion of the source and the observer. We want to recall first the key results of section 7.2 in regard to the observation angle transformations: the line of sight when not parallel to velocity \( \vec{v} \) of moving observer \( S' \) has different altitude angle \( \theta' \) compared to \( \theta \) measured by observer \( S \) comoving with the light source. The two key and equivalent aberration transforms obtained in Eq. (7.11) are

\[
\sin \theta' = \frac{\sin \theta}{\gamma (1 + \beta \cos \theta)}, \quad \cos \theta' = \frac{\cos \theta + \beta}{1 + \beta \cos \theta}
\]

and their inversion is obtained setting \( (v/c) = \beta \to -\beta, \ \theta \leftrightarrow \theta' \)

\[
\sin \theta = \frac{\sin \theta'}{\gamma (1 - \beta \cos \theta')}, \quad \cos \theta = \frac{\cos \theta' - \beta}{1 - \beta \cos \theta'}. \]
The azimuthal angle $\varphi$ according to Eq. (7.13) is not transformed. It is straightforward to verify Eq. (13.12) by showing $\cos^2 \theta' + \sin^2 \theta' = 1$.

For the invariant phase of the light wave we find the extension of Eq. (13.7)

$$\phi' = \omega' t' - \vec{x}' \cdot \vec{k}' = \omega' \gamma (t + zv/c^2) - k'_x \gamma (z + tv) - k'_x x - k'_y y.$$  (13.14)

Here we inserted in the last equality the Lorentz transformation from the primed coordinates of the moving observer $S'$ to the light source observer $S$.

The plane wave vector $\vec{k}'$ is decomposed as in Eq. (7.10)

$$k'_x = -\frac{\omega'}{c} \sin \theta' \cos \varphi', \quad k'_y = -\frac{\omega'}{c} \sin \theta' \sin \varphi', \quad k'_z = -\frac{\omega'}{c} \cos \theta'. \quad (13.15)$$

We insert this into Eq. (13.14) and regroup terms

$$\phi' = \omega' \left[ t \gamma (1 - \beta \cos \theta') - \frac{z}{c} \gamma (\cos \theta' + \beta) - \frac{x}{c} \sin \theta' \cos \varphi' - \frac{y}{c} \sin \theta' \sin \varphi' \right]. \quad (13.16)$$

Let us first look at transformed and non-transformed phase $\phi'$ and $\phi$ components proportional to $x$ and $y$. Remembering that the azimuthal angle $\varphi$ is not transformed according to Eq. (7.13) and given that the value of $x$ and $y$ can be arbitrary we clearly must have

$$\omega \sin \theta = \omega' \sin \theta'. \quad (13.17)$$

Applying aberration transform Eq. (13.12)

$$\omega \sin \theta = \frac{\omega' \sin \theta}{\gamma (1 + \beta \cos \theta)}, \quad \Rightarrow \quad \omega' = \omega \gamma (1 + \beta \cos \theta) \quad (13.18)$$

We next proceed in the same fashion with the coefficient of $t$:

$$\omega = \omega' \gamma (1 - \beta \cos \theta') = \omega' \gamma \left( 1 - \beta \frac{\cos \theta + \beta}{1 + \beta \cos \theta} \right) \quad (13.19)$$

where we again used the aberrated transform Eq. (13.12). Straightforward algebra leads to same result as Eq. (13.18). So this checks out. Similarly we verify the coefficient of $z$

$$\omega \cos \theta = \omega' \gamma (\cos \theta' - \beta) = \omega' \gamma \left( \frac{\cos \theta + \beta}{1 + \beta \cos \theta} - \beta \right), \quad (13.20)$$

which also leads to same condition as Eq. (13.18).

Thus we have shown that transformation of $\omega'$ seen Eq. (13.18) when combined with the aberration effect, that is the transformation of line of sight altitude angle Eq. (13.12), renders the phase invariant. This means that the result Eq. (13.9)
PART V: TIME
TEST READER VERSION AUGUST 24, 2016

holds when we include in the numerator of the first expression the projection of the relative velocity onto the line of sight. Written in textbook variables:

\[ \nu = \nu_0 \frac{1 + \beta \cos \theta_0}{\sqrt{1 - \beta^2}}, \]  

(13.21)
a result that is of course naively expected and stated often without further discussion. However, this is not as naive a result as some may think. We clarified that the angle \( \theta_0 \) shown is measured in the frame of the light source. Therefore to relate this result to the observed \( \cos \theta \) line of sight of the ‘earth-bound’ observer we must insert the inverse aberration result Eq. (13.13) and the outcome is

\[ \nu_0 = \nu \frac{1 - \beta \cos \theta}{\sqrt{1 - \beta^2}}, \]  

(13.22)
an analogue of Eq. (13.10). We have on the right both the frequency \( \nu \) and the altitude angle \( \theta \) measured by an earth-bound observer whose motion is not exactly parallel or antiparallel to \( \hat{n} \) to the observed object emitting at frequency \( \nu_0 \).

Exercise V–7: Doppler measurement of velocities

The \( H_{\alpha} \)-line of a star (\( \lambda_0 = 6563 \) Å) appears to be red-shifted by 30 Å. No lateral motion of the star is observed; thus the star must be moving along the line of sight. How fast is the star moving away from the Earth?

Solution

According to Eq. (13.9) and related discussion, when the star is moving away from the Earth the frequency \( \nu \) seen by the observer at rest appears smaller than the natural frequency \( \nu_0 \)

In particular, using Eq. (13.9) we have for a source moving away from us with speed \( v \)

\[ \frac{1}{\nu} = \frac{1 - v/c}{1 + v/c}. \]  

As \( c = \nu \lambda \), the inverse of this ratio applies for the wavelength as seen in Eq. (13.11), which we can write in the form

\[ \frac{\lambda_0}{\lambda} = \frac{1 - v/c}{1 + v/c}. \]  

Clearly, the wavelength \( \lambda > \lambda_0 \) is seen by the observer. Solving for \( v \) yields

\[ v = \frac{c - (\lambda_0/\lambda)^2}{1 + (\lambda_0/\lambda)^2}. \]
Using $\lambda_0 = 6563\text{Å}$ and $\lambda = 6593\text{Å}$ we obtain $v = 0.0046c$. The velocity of the star relative to the Earth is

$$4 \quad v = 0.0046c = 0.0046 \times 300,000 \text{ km/sec} = 1,380 \text{ km/sec}$$

This velocity is small enough for the effect to be entirely described by the nonrelativistic Doppler effect formula (omitting time dilation factor $\gamma$).

---

**Exercise V–8: Predicting the Doppler shift**

A nearby star moves away from the Earth along the line of sight with $v = 5 \cdot 10^{-4}c = 150 \text{ km/sec}$. What wavelength do we expect to observe for the star’s $H_\alpha$-line on Earth?

**Solution**

We use the second formula from exercise [V–7]

$$1 \quad \frac{\lambda_0}{\lambda} = \sqrt{\frac{1 - v/c}{1 + v/c}}$$

Substitution for $v$ yields

$$2 \quad \frac{\lambda_0}{\lambda} = 0.9995,$$

and with this

$$3 \quad \lambda = 6566.3\text{Å},$$

which is by $\Delta \lambda = 3.3\text{Å}$ larger (red shift) than $\lambda_0$.

---

**Exercise V–9: Doppler effect and periodic signal from a spaceship**

An expedition rocket (source reference system $S_0$) emits a periodic reference signal. Find the period measured by an observer on Earth (detector reference system $S$). Assume that the rocket is moving away from the Earth along the line of sight. Discuss also the corresponding wavelength change if the signal is a radio wave and the spaceship is traveling away or returning from the trip.

**Solution**
Between two signal events $E_1$ and $E_2$, the rocket traveling at a speed $v$ gains in distance, and the following signal traveling at $c$ is delayed by $\Delta t_{E_2E_1}$ before reaching the detector in $S$:

$$\Delta t_{E_2E_1} = \frac{x_{E_2} - x_{E_1}}{c}.$$ 

With the Lorentz transformation where for both events $x_{E_1} = x_{E_2} = x_E$

$$x_{E_1} = \frac{x_0 + vt_0E_1}{\sqrt{1 - (v/c)^2}},$$
$$x_{E_2} = \frac{x_0 + vt_0E_2}{\sqrt{1 - (v/c)^2}},$$

we find considering the difference and using Eq. (1)

$$\Delta t_{E_2E_1} = \frac{v}{c} \frac{t_{0E_2} - t_{0E_1}}{\sqrt{1 - (v/c)^2}} = \beta \gamma (t_{0E_2} - t_{0E_1}).$$

The total time that elapses between the arrival of the signals in $S$ can now be calculated adding the two contributions, Eq. (8.6) (time dilation) and Eq. 3. We obtain

$$\tau_{E_2E_1} = t_{E_2} - t_{E_1} + \Delta t_{E_2E_1}$$
$$= \gamma (1 + \beta) \tau_{0E_2E_1}$$
$$= \sqrt{\frac{1 + v/c}{1 - v/c}} \tau_{0E_2E_1},$$

in which we have substituted

$$\tau_{0E_2E_1} \equiv t_{0E_2} - t_{0E_1}.$$ 

We note that the effect of time dilation reinforces the nonrelativistic effect by omitting the factor $\gamma$ in Eq. 3.

Let us further interpret the result Eq. 4. We introduce the striking frequency of the clock in its rest frame $S_0$ of the expedition rocket:

$$\nu_0 = \frac{1}{\tau_{0E_2E_1}},$$

and similarly as it appears to an observer at rest in $S$:

$$\nu = \frac{1}{\tau_{E_2E_1}}.$$ 

Eq. 4 thus reads

$$\nu = \sqrt{\frac{1 - v/c}{1 + v/c}} \nu_0.$$
Above $\nu_0$ is the intrinsic proper body frequency in the rocket, $\nu$ is the frequency measured by an observer on Earth. When the source on the expedition spaceship is moving directly away with velocity $+v$: we observe $\nu < \nu_0$, that is for the corresponding wavelengths $\lambda > \lambda_0$ we see motion-induced redshift. If the expedition spaceship changes course and moves directly home, we observe $\nu > \nu_0$, $\lambda < \lambda_0$ a blue-shift.

End V–9: Doppler effect and periodic signal from a spaceship

14 Tests of Special Relativity

14.1 Tests of Lorentz transformation

We present Doppler tests and time dilation tests of relativity in one (short) account. This is in part due to the language used by some of the authors of these tests, implying that the Doppler shift tests time dilation. However, what actually is tested is the precise format of the relativistic coordinate transformation on which the Doppler shift relies, and from which time dilation can be derived. As we already explained introducing the reciprocity of the Doppler effect in section 13.2, there is no direct time dilation input into the Doppler effect.

The importance of the relativistic Doppler shift as a potential test of the relativity principle and special relativity cannot be over-emphasized. Einstein recognized the pivotal role of Doppler frequency shift\(^{15}\), and defended his theory in face of erroneous experiments that were showing results up to 10 times larger.

A Doppler effect based experiment was carried out in 1938 by Ives and Stilwell, and a discussion of the historical and physics background was given in a review by Ives about a decade later\(^{16}\). A direct time dilation experiment was carried out by Hafele and Keating in 1972. They compared clocks sent on a trip around the Earth towards East and West\(^{17}\). They obtained a result in agreement with both SR and GR. GR enters the discussion since a terrestrial time dilation measurement uses clocks under the effect of gravitation. Both SR and GR time dilation effects are relatively small. This allows to consider linear first approximation in which both effects are additive. Moreover, SR effect is speed dependent, and the GR effect is gravitational potential dependent, that it depends on the distance from


\(^{16}\)Herbert Ives, “Historical Note on the Rate of a Moving Atomic Clock,” Journal of the Optical Society of America 37 (10) 810 (1947).

Earth’s surface. By performing a difference experiment it is possible to isolate both effects.

The Ives and Stilwell experiment has recently seen renewed recent interest due to the possibility of using a highly relativistic absorber/emitter, which in contemporary experiments is kept in a particle storage ring. This allows improvement over interferometric MM type experiments which depend on the velocity of the Earth with respect to a potential preferred frame of reference. As in the nill MM experiment one seeks to cancel the effect of relativity. The method employed consists of two complementary ‘parallel’ but opposite direction Doppler effects impacting the frequency of atomic clocks. The effect cancels as we see using Eq. (13.21) with either \( \theta_0 = 0, \pi \)

\[
\nu_+ \nu_- = \nu_0 \gamma (1 + \beta) \nu_0 \gamma (1 - \beta) = \nu_0^2 .
\]  

(14.1)

When the two Doppler frequency shifts, one parallel and the other anti-parallel, are combined in a product, according to SR the result should not depend on relative velocity at all. Comparison with another such experimental arrangement running at a different speed strengthens the significance of the result.

Should the Lorentz transformation have a modified form, the cancellation in Eq. (14.1) is expected not to be satisfied in a wide range of models. This allows to explore and set limits on modifications of special relativity where the invariant proper time is written in the form

\[
d\tau = \left( 1 + \alpha_{\text{RMS}} V^2 / c^2 + \alpha_{\text{RMS,2}} V^4 / c^4 + \ldots \right) \sqrt{dt^2 - dx^2 / c^2} .
\]  

(14.2)

where \( V \) is not only the local velocity but includes some unknown absolute velocity component. The parameter \( \alpha_{\text{RMS}} \) is known as the Robertson and Mansouri, Sexl (RMS) parameter. We include a further term \( \alpha_{\text{RMS,2}} \) arising from possible velocity dependence of the RMS \( \alpha \)-parameter. More generally one can think of \( \alpha_{\text{RMS}} \) as being speed dependent.

To measure \( \alpha_{\text{RMS}} \) a comparison of the two shifts measured at different velocities \( \beta_i = v_i / c \) by the null-method was carried out. We refer to the Heidelberg\(^{18}\) time dilation experiment for details. A possible deviation from Einstein’s time dilation is extracted from the two measured frequencies obtained for the two respective values of \( \beta \)

\[
\nu_{01} = \sqrt{\nu_+^{(1)} \nu_-^{(1)}} = 546466918577 \pm 108 \text{ kHz, } \beta_1 = 0.030, \ v = 9,000 \text{ km/s} \\
\nu_{02} = \sqrt{\nu_+^{(2)} \nu_-^{(2)}} = 546466918493 \pm 98 \text{ kHz, } \beta_2 = 0.064 \ v = 19,200 \text{ km/s}
\]

which result limits the time dilation modification

\[
\alpha_{\text{RMS}} = 4.8 \pm 8.5 \times 10^{-8} .
\]  

(14.3)

Tests of Special Relativity

We see that this result is consistent with zero at a high level of precision.

Yet another experiment was carried out at a more relativistic speed $\beta = 0.338$. Li$^+$ ions were stored in the ESR storage ring at the GSI laboratory at Darmstadt. Given the value of $\beta^2 = 0.114$ employed, 30 times larger compared to the Heidelberg experiment, a stringent limit on $\alpha_{\text{RMS,2}}$ could also be set.

$$\alpha_{\text{RMS,2}} < 1.2 \times 10^{-7}.\quad (14.4)$$

More generally the result constrains the RMS parameter as a function of speed:

$$\left| \frac{\alpha_{\text{RMS}} V^2/c^2 + \alpha_{\text{RMS,2}} V^4/c^4 + \ldots}{V^2/c^2} \right| \leq 2.0 \times 10^{-8}.\quad (14.5)$$

Returning to the time dilation measurement, we already described the Hafele and Keating experiment, Ref. [17]. The GPS satellite signal offers another opportunity to test time dilation and a limit on $\alpha_{\text{RMS}}$ was found at a level that is two orders of magnitude less constraining as compared to later laboratory experiments$^{20}$ Conversely, the laboratory result of the Heidelberg and GSI Ives-Stiles experiments assures that we can rely on the GPS network for precise positioning.

14.2 Michelson-Morley experiment today

In the past 120 years we have seen highly significant improvements in the Michelson-Morley experiment. The time line is shown in figure [14-1]. At the beginning this was due to the continuing efforts of Michelson himself, resulting in improved limits shown in upper left corner in the figure.

In a recent, year 2009 incarnation (bottom right corner figure [14-1]), the MM experiment involves an apparatus that floats on a thin cushion of air above a 1.3 metric ton granite table. It comprises two optical cavities, essentially pairs of mirrors that reflect light back and forth. They are both about 8.4 cm long and at right angles to each other. Because the cavities are slightly different in length, they have slightly different resonant frequencies.

A laser beam is split into two beams, one for each cavity. The frequencies of the beams are then tuned to resonate in their respective cavities using acousto-optic modulators. If the speed of light were different in different directions, it would affect the resonant frequencies of the two cavities, which could then be

---


Figure 14-1: Signal interferometry experiments of Michelson-Morley type set limits on the variation of velocity of light $\Delta c/c$, or, signal frequency $\Delta \nu/\nu$ due to the presumed 'ether drag'. The experimental progress evolution representation follows Ref.[22].

detected as a shift in the beat frequency once the apparatus is rotated and two beams are recombined. This experiment carried over a period of 13 months shows the local isotropy of light propagation at the $10^{-17}$ level[21].

An even more recent experiment[22] comprised two sapphire cylinders with their crystal axes aligned orthogonally. Resonance modes were excited in each sapphire. The difference in frequency was recorded while the apparatus was continuously rotated with a 100 s period. Evaluation of one year of data allowed the setting of a limit on the variation of frequency $\Delta \nu/\nu$ or, equivalently light velocity $\Delta c/c$ at $9.2 \pm 10.7 \times 10^{-19}$ (95% confidence interval); this is the latest experimental result seen in bottom right in figure [14-1]. The authors expect further improvement in the precision of this new method.


The verification that light propagation is independent of the direction of motion of the laboratory in the Universe has now shown the absence of ‘æther drag’ speed at the level of 3Å/s; that is, atomic size per second.

14.3 Modified coordinate transformations

We have described present day experiments which with a very high precision test special relativity: here in particular Doppler shift measurement tests and Michelson-Moreley interferometry tests. These efforts confirm specific features with high precision. In the context of exercise III–2 on page 82 we have alluded to the possibility to modify relativity by introducing modifications of the transformation properties of the speed of light. Such modifications were proposed in the context of Planck scales: the Planck energy $E_P$, and Planck length $L_P$, both expressed in terms of the Newton’s gravitational constant $G_N$

$$E_P \equiv \sqrt{\hbar c^3/G_N} = 1.221 \times 10^{19} \text{ GeV} \quad L_P \equiv \sqrt{\hbar G_N/c^3} = 1.616 \times 10^{-35} \text{ m} , \quad (14.6)$$

(for discussion of energy unit ‘GeV’ see Insight on page 204 below). It is expected that at distance scales below $L_P$ significant modification in our understanding will be needed. To no surprise, theories in which there is not only an observer-independent maximum speed of light, but also an observer-independent minimum length scale enjoy relative popularity under the name doubly (or deformed) special relativity (DSR).

To illustrate where the current day discussion is centered, let us quote from an abstract of a comment


“...relation between three different approaches of theories with a minimal length scale:
1) a modification of the Lorentz group in the ‘deformed special relativity’,
2) theories with a ‘generalized uncertainty principle’ and those with
3) ‘modified dispersion relations’
(numbering added). It is shown that the first two are equivalent, how they can be translated into each other, and how the third can be obtained from them. An adequate theory with a minimal length scale requires all three features to be present”.

14.4 Measurement of time

Time is different from space

In special relativity theory one of the most important insights is about time, which like space, is a part of space-time, an equal partner as we have reported in
the opening of this book, see section 1.1. However, there is a subtle problem. We treat time from the beginning differently: it is a parameter that characterizes the evolution of the physical system, aside from providing the additional coordinate needed to identify world events. On the other hand time and space are united in space-time. The question thus is not if, but at which level of precision the limits of validity of special relativity will be reached when time is measured.

While we can influence where we are in space, we cannot control our the motion in time. The situation with time seems to parallel the circumstance of a tree, it moves along in space where the Earth motion takes it. In analogy, we are attached to a location on the time axis, and coast along where the evolving Universe takes us.

Last but not least, in an expanding Universe time and space differ and thus special relativity cannot apply at a macroscopic scale. The domain of application of special relativity is limited to the local ‘tangential’ flat static Minkowski space-time, in the same sense that we can use Euclidean geometry on a tangential surface on Earth.

General remarks about tests

One of the challenges we note in the question relating to the CMB cosmological rest frame is how the teaching of special relativity principles is constrained by General Relativity and Cosmology. This also explains why when we open a research paper on ‘Tests of Special Relativity’, we typically first will see a lot about General Relativity, see the recent example\textsuperscript{24} a report on a test of Lorentz symmetry based on existing observations of the orbits of planets around the Sun. This report relies on earlier theoretical work\textsuperscript{25} which opens in abstract with “The gravitational couplings of matter are studied in the presence of Lorentz and CPT violation.”

The Lorentz symmetry violation concerns modification of relativistic coordinate transformation properties. We believe that the laws of physics are also invariant under a discrete symmetry transformation, the simultaneous mirroring of space (P=parity), time (T), and particles with antiparticles (C=charge). CPT arises in the context of quantum field theory, as a further consequence of what is described in this book, and the deeper insight that in relativistic quantum mechanics both particles and antiparticles appear together and are equivalent.


An example of CPT breaking would be a difference in mass between matter and antimatter, a topic of contemporary experimental interest\textsuperscript{[20]}.

**Precision time measurement and gravity**

A full discussion of gravitational time dilation is beyond the scope of this book. The mention of a few facts is a complement reference. The principle idea of General Relativity is that gravity deforms space-time geometry. In this context we not only speak of coordinate transformations but we change the space-time by presence of material bodies. This leads directly to the recognition that the proper time ticking within a body is also influenced by the gravitational potential which slightly warps the time coordinate.

For a first rough understanding, and an estimate of the magnitude of the effect we consider a mass falling in the Earth’s gravity and use the resulting velocity to estimate what the associated time dilation before the fall may be. For $\Delta h = 10^4 \text{m}$, with $g = 9.8 \text{m/s}^2$:

$$\frac{1}{2} \left( \frac{v}{c} \right)^2 = \frac{g\Delta h}{c^2} \sim 10^{-12}.$$  

This estimate, even if small, shows that the gravity-induced time dilation for a clock traveling on a plane is observable as gravity can compete with the SR motion effect. Gravity’s effect impacts time keeping on the satellite.

The global positioning system (GPS) relies on satellites orbiting the Earth. For the GPS to function properly, one must assure that the time keeper on the GPS satellite is aware that clocks on Earth are not in free fall, and are exposed to a stronger gravitational potential. Therefore the satellite emitted GPS signal is ‘corrected’ to the clock rate prevailing in Earth’s frame of reference. The precision of the GPS requires that a full account is taken of the effect of time dilation due to gravity when comparing the clocks on Earth with clocks on a satellite, see for example Wolf and Petit, loc.cit.\textsuperscript{[20]}.  

Even directly on the Earth’s surface where we are supported by force of matter, the force of gravity significantly impacts the time keeping. Two clocks will tick differently depending on location in the gravity potential of the Earth: solving Einstein’s gravity one finds that a clock will tick faster by 1 second in $10^{18}$ for every centimeter it is raised in Earth’s gravitational field. The technology of time-keeping improves with remarkable speed, in 2008 a relative precision of

5·10⁻¹⁷ was achieved\textsuperscript{27}, while in 2016 time-keeping precision better than 5·10⁻¹⁸ is reported\textsuperscript{28}, corresponding to less than 10cm height difference sensitivity.

In an independent study involving quantum interference of atomic quantum waves the gravitational redshift sensitivity reached in 2010 a new record\textsuperscript{29} with a relative precision of 7·10⁻⁹. The virtue of this approach is the extremely high frequency of a ultra cold Cesium atom’s quantum de Broglie wave. This result supports the view that gravity is a manifestation of space-time curvature. Use of interfering matter quantum waves further suggests that to lowest order in \hbar, gravity, and quantum physics, are consistent.

**Time and space in cosmology**

Most physicists agree that our Universe is not static. Thus, we can determine how much time has passed since the Big Bang. A measurement of the cosmological time, the time since the ‘beginning’, requires that we know a frame of reference defining the cosmic ‘rest’ frame – the cosmological time is defined as the proper time of our Universe. In the early Universe all the matter was on average at rest. As the Universe expanded and cooled, the plasma of ions and electrons combined and neutral atoms formed, and thus the Universe became transparent to the cosmic background light. This light is observed today as the cosmic microwave background (CMB). In this way, the distant-in-time essentially homogeneous matter defines the natural cosmological reference frame. This cosmic frame of reference can be determined in by measurement of Earth’s motion with respect to the CMB.

The age of the Universe is determined today in a global analysis of the fluctuations of CMB photons straying away from complete spatial homogeneity. In 2012 WMAP spacecraft collaboration reported the age of the universe to be\textsuperscript{30} 13.772 ± 0.059 billion years. The Planck spacecraft collaboration reported\textsuperscript{31} in 2015: 13.813 ± 0.038 billion years, both these values are obtained within ΛCDM model of the Universe, called that way for dark energy/Einstein constant Λ, and the letter acronym for ‘Cold Dark Matter’.

We see that as a matter of principle we can tell where in time we are today. Therefore time and space are subtly different: we can pinpoint absolute cosmological time as an absolute coordinate. We have no way to introduce absolute

\textsuperscript{27}T. Rosenband et al, “Frequency Ratio of Al⁺ and Hg⁺ Single-Ion Optical Clocks; Metrology at the 17th Decimal Place,” Science \textbf{319} 1808 (2008).


\textsuperscript{29}Holger Müller, Achim Peters and Steven Chu, “A precision measurement of the gravitational redshift by the interference of matter waves,” Nature \textbf{463} 926 (2010).


spatial coordinates. There is a second difference: in SR time has no beginning and no end, it is the same forever. Since there is a beginning in time, the translation in time symmetry cannot work for arbitrarily long times. Therefore, the principles of special relativity can only be explored using a time scale that is considerably shorter than the lifespan of the Universe. We note that any experiment that is carried out to test relativity usually lasts less than one year. This is 10 orders of magnitude shorter compared to the lifespan of the Universe. Clearly, the cosmological history of the Universe is irrelevant on a one year time scale and principles of SR apply.
Part VI

Mass, Energy, Momentum
Introductory remarks to Part VI

We study the connection between the energy content $E(v = 0) = E_0$ of a body at rest, its inertial mass $m$, and the speed of light $c$. The appearance of $c$ implies a profound relation of all visible material bodies to each other and establishes a paramount common property. $E(v = 0) = mc^2$ naturally combines with the known nonrelativistic kinetic energy to a simple expression that describes the total relativistic energy of a body in motion with respect to an observer.

$E_0 = mc^2$ is demonstrated by considering the consistency of the measurement by two different observers of the energy content of a moving body after it converts into radiation. We further discuss the difference between the energy of a body that includes energy of motion (kinetic energy), and the rest energy $E_0 = mc^2$, the energy in the rest frame of the body. We strongly discourage the consideration of $E(v)/c^2$ as a speed-dependent mass.

We describe several examples to certify that a composite body with complex internal dynamics as a whole satisfies Einstein’s $E_0 = mc^2$, irrespective of what is going on inside the body in terms of motion, chemical and nuclear reactions, and even radiation. This leads to the pivotal insight that all usable energy comes from reduction of the mass of reactants that serve in energy production.

We learn that the energy comes neither from the ‘wall’, nor from the ‘power station’; it comes always from mass; to be more precise, from the reduction of the mass of the final state as compared to the initial pre-burn state, irrespective of whether the power station is conventional or nuclear. The renewable energy sources use the power provided every day by the Sun: it is on the sun that the mass reduction occurs; driven by nuclear fusion.

Having established the mass and energy as different and yet very closely related, we introduce the relativistic momentum of a body by recalling the relation between the force inducing change of momentum, and the work-energy theorem relating work, energy and velocity. Both nonrelativistic and ultra-relativistic limits of the energy-momentum-velocity relation are explored. Considering that the linearity of the Galilean velocity addition is lost, the concept of rapidity is advanced: rapidity is a variable that is additive when addition of velocity in the same direction is considered.

The momentum and energy conservation is illustrated in the ‘practical’ example of a relativistic rocket. A rocket is in relativistic motion when its kinetic energy as measured from departure base dominates the remaining rest energy of the rocket. As the propellant mass is ejected, and mass is consumed to make energy for propulsion the rocket retains considering its kinetic motion half of the original (rest) energy, but only a very tiny fraction of this energy is in the remaining rest mass of the rocket.
15 Mass and Energy

15.1 Proper body energy

Among several spectacular publications in the year 1905, including works on the photo-electric effect and Brownian motion, Einstein formulated the Special Relativity Theory (see Ref. on page xiv). Einstein’s most famous equation, the equivalence between mass and energy,

$$E = mc^2,$$  \hspace{1cm} (15.1)

was argued in a separate, very short publication (see Ref. on page xiv). Einstein employed the emission of radiation by an atom and thus a partial conversion of inertial mass into energy, to deduce that the kinetic energy $T$ of a body has the form

$$T = \frac{mc^2}{\sqrt{1 - (v/c)^2}} - mc^2.$$

An analogous more dramatic version of Einstein’s arguments is presented in the following: we consider matter-antimatter annihilation radiation, that is the case that all of the inertial mass of a particle-antiparticle pair is converted into the energy of emitted radiation. In this context the reader will note the appearance below of $\hbar$, the Planck quantum of action. This is not to be interpreted to mean that quantum mechanics is needed for the demonstration of Eq. (15.1). The main reason that $\hbar$ appears is that we make explicit what in Einstein’s two papers is implicit, the relation between the frequency of light and energy of the photon, which Einstein addressed earlier in another remarkable 1905 publication\(^1\).

Let us consider the ‘meeting’ of an electron $e^-$ with its antimatter particle, the positron $e^+$. We can choose a frame of reference to view their encounter such that each arrives from opposite side very slowly and their total momentum is zero, as is shown in figure [15-1]. A reference system in which momentum of all particles vanishes is called the center of momentum (CM) frame and will be addressed in more detail in section [18]. However, here we do not need to worry, as we make the electron and positron meet each other with relative momentum near zero.

These two particles can annihilate in this encounter, releasing usually two photons. The conservation of momentum calls for the total momentum to remain zero, thus these photons must be emitted in diametrically opposite directions

$$\vec{p}_1^0 = -\vec{p}_2^0, \hspace{1cm} (15.3)$$

\(^1\)Albert Einstein, “Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt,” (translated: “On a heuristic viewpoint concerning the production and transmutation of light,”) Annalen der Physik, 17 132 (1905) (Received at publisher March 18, 1905)
Figure 15-1: A positron and electron of equal and opposite momenta annihilating to release two photons.

as shown in figure [15-1]

The photoelectric effect establishes the particle nature of the photons, with their momenta given by

\[ |\vec{p}_1^0| = \hbar k_1 = 2\pi \hbar \nu_1 / c, \quad (15.4a) \]

\[ |\vec{p}_2^0| = \hbar k_2 = 2\pi \hbar \nu_2 / c. \quad (15.4b) \]

where \( k_1 \) and \( k_2 \) are the wave numbers and \( \nu_1 \) and \( \nu_2 \) are the frequencies of the photons. Both photons thus have equal frequency \( \nu_0 \):

\[ \nu_1^0 = \nu_2^0 \equiv \nu_0. \quad (15.5) \]

Each of these photons also carries the energy

\[ E_1^0 = E_2^0 \equiv E_0 = pc = 2\pi \hbar \nu_0. \quad (15.6) \]

Aside from momentum, the energy is also conserved; thus the energy content of the electron-positron pair must be equal to the energy found in the two emitted photons

\[ E_{e^+e^-}^0 = 2E_0. \quad (15.7) \]

Here \( E_{e^+e^-}^0 \) is the intrinsic energy of the electron positron system before annihilation.

Up to now, we have considered only the process in the reference system \( S \) in which both the electron and the positron were at rest. We now consider a system \( S' \) moving relative to \( S \) with a velocity \( \vec{v} \), chosen here to be parallel to the direction of the emission of both photons. In this frame of reference we observe a Doppler shift in which one photon’s frequency increases, and the other photon’s frequency decreases, as shown in figure [15-2]. Using the relativistic Doppler relationship, Eq. (13.9), we obtain

\[ \nu^\pm = \nu_0 \frac{1 \pm v/c}{\sqrt{1 - (v/c)^2}}. \quad (15.8) \]
Figure 15-2: The moving observer sees a red shift of one photon (on left) and a blue shift of the other.

In Eq. (15.8) we have $+v/c$ for the photon which travels anti-parallel to $\vec{v}$, and $-v/c$ for the photon which travels parallel to $\vec{v}$. The energies of the photons in the moving frame of reference $S'$ are thus also shifted and are given by

$$E^{\pm} = 2\pi \hbar \nu^{\pm}$$

$$= 2\pi \hbar \frac{1 \pm v/c}{\sqrt{1-(v/c)^2}} \nu_0.$$  \hspace{0.5cm} (15.9a)

If we substitute in the energy from the system at rest (Eq. (15.6)), we can also write

$$E^{\pm} = E_0 \frac{1 \pm v/c}{\sqrt{1-(v/c)^2}}.$$ \hspace{0.5cm} (15.9b)

By the principle of relativity, conservation of energy must also apply to the moving observer. This means that in system $S'$, the combined energy of the electron and positron prior to the annihilation process should be equal to the energy carried by the two photons after the annihilation. We call the energy of the system, as observed in $S'$, $E_{e^+e^-}(v)$, anticipating that the result must be a function of the speed $v$

$$E_{e^+e^-}(v) = E^+ + E^-.$$ \hspace{0.5cm} (15.10a)

If we insert equation (15.9b), we obtain

$$E_{e^+e^-}(v) = \frac{2E_0}{\sqrt{1-(v/c)^2}} = \frac{E_{e^+e^-}^0}{\sqrt{1-(v/c)^2}}.$$ \hspace{0.5cm} (15.10b)

where we used Eq. (15.7) to simplify the coefficient.

This result differs from the result in the system at rest by the appearance of the Lorentz factor which we now expand

$$\gamma = \frac{1}{\sqrt{1-(v/c)^2}} = 1 + \frac{1}{2} \left( \frac{v}{c} \right)^2 + \frac{3}{8} \left( \frac{v}{c} \right)^4 + \ldots$$ \hspace{0.5cm} (15.11)

Inserting this expansion into Eq. (15.10b)

$$E_{e^+e^-}(v) = E_{e^+e^-}^0 + \frac{1}{2} E_{e^+e^-}^0 \left( \frac{v}{c} \right)^2 + \mathcal{O}(v^4).$$ \hspace{0.5cm} (15.12)
Comparing the 2nd term in Eq. (15.12) with the usual expression for the nonrelativistic kinetic energy,

$$E_{\text{kin}}(v) = \frac{1}{2}mv^2, \quad (15.13)$$

we recognize a relation between the energy contained in the electron-positron system and the mass of the system

$$E_{e^+e^-}^0 = m_{e^+e^-}c^2. \quad (15.14)$$

We can drop the subscript ‘$e^+e^-$’ since these equations apply to any assembly of particles bound together and moving at speed $v$ and with an inertial mass $m$. Rest energy

$$E_0 = mc^2, \quad (15.15)$$

which is arguably the most famous 20th century equation. It states that any visible material body of inertial mass $m$ has a comoving “proper energy” content $E_0$. Because $c^2$ appears for each and every visible material particle this relation establishes that all visible material bodies are related.

### 15.2 Relativistic energy of a moving body

Using Eq. (15.14) in Eq. (15.10b), we obtain the energy in the moving frame of reference:

$$E(v) = \frac{mc^2}{\sqrt{1 - (v/c)^2}} = \gamma mc^2. \quad (15.16)$$

The important message in Eq. (15.16) is that all of the energy a particle carries is controlled by the particle inertia $m$ whereby the two energy components, the rest energy, and the kinetic energy, combine into a relatively simple expression where the Lorentz factor $\gamma$ appears. To separate these components, we restate Eq. (15.16) using Eq. (15.11)

$$E(v) = mc^2 + \frac{1}{2}mv^2 + \frac{3}{8}mv^2 \left(\frac{v}{c}\right)^2 + \ldots. \quad (15.17)$$

The first term,

$$E_0 = mc^2, \quad (15.18)$$

is the energy of a particle at rest, or so-called, “rest energy”.

The full relativistic kinetic energy is obtained by removing the rest energy in Eq. (15.16)

$$T = E(v) - mc^2 = (\gamma - 1)mc^2 \simeq \frac{1}{2}mv^2 \left[1 + \frac{3}{4} \left(\frac{v}{c}\right)^2 + \frac{5}{8} \left(\frac{v}{c}\right)^4 + \ldots\right], \quad (15.19)$$
Figure 15-3: $E(v)$ in the relativistic (Eq. (15.16)) and non-relativistic cases (Eq. (15.13)).

which expands to show as the first term the non-relativistic kinetic energy followed by corrections in $(v/c)^2$ and higher order.

We see that the mass of a body defines the energy of the body, and of particular interest is of course that there is energy in the limit $v \to 0$, Eq. (15.18), which is the pivotal new result. What we learn is best stated by a direct citation from the Einstein paper: “If a body gives off an energy $\delta E$ in form of radiation, its mass diminishes by the corresponding amount $\delta E/c^2$.” In our case all of the energy of the particle-antiparticle pair was given off, so we have the full energy equivalence between $m$ and $E$.

We compare the relativistic energy Eq. (15.16) to the non-relativistic limit: the first two terms in Eq. (15.19) are plotted in figure 15-3 as a function of speed $v$. We scale energies with $mc^2$ and the speed $v$ with $c$. Note the difference between the two results near $v = c$. Within the special theory of relativity, the energy of a body diverges as $v \to c$. Thus no matter how much work is done on a body, its speed never reaches the speed of light. In this sense, the speed of light is considered a maximum speed limit for all material bodies.
15.3 Mass of a body

In older literature, it was common to absorb into the mass the speed dependent factor and to write

\[ M = \frac{m_0}{\sqrt{1 - (v/c)^2}} \]  

(notation not used in this book)

and thus

\[ M = \frac{E(v)}{c^2} \]  

(notation not used in this book)

The above definition would require that we always attach to the true mass the subscript ‘0’ to distinguish the rest mass from the quantity which actually is not a mass but up to a factor the total energy \( E(v)/c^2 \). We will see that the true mass \( m_0 \) is a Lorentz invariant, and \( M(v) \) is not. We reserve the word ‘mass’ to refer to the Lorentz invariant quantity that describes the inertial mass of a body which has the same value for all observers.

A qualitative physical argument for the Lorentz invariance of mass \( m_0 \) is as follows: Using Eq. (15.18) we assign a rest energy to all bodies. It is a physical property of the body and, according to the principle of relativity, it should be independent of the choice of the inertial system. Thus, in addition to the proper time \( \tau \), there is a second body property that is a Lorentz-invariant, the proper energy of a body, i.e. rest mass energy \( m_0c^2 \).

The use of body energy in motion as an effective mass has been criticized as follows: Albert Einstein\(^2\): It is not good to introduce the concept of the mass \( M = m/(1 - v^2/c^2)^{1/2} \) of a moving body for which no clear definition can be given. It is better to introduce no other mass concept than the ‘rest mass’ \( m \). Lev Okun\(^3\): The concept of relativistic mass, which increases with velocity, is not compatible with the standard language of relativity theory and impedes the understanding and learning of the theory . . . . In many books when an author speaks of particle energy, the implication is \( E/c^2 \equiv \langle m(v) \rangle \), in case the dependence on \( v \) is omitted (and perhaps natural units \( c = 1 \) are used) the confusion of energy with mass arises, as is seen in e.g. the Max Born’s Relativity, or R.P. Feynman’s Lectures, but not in Landau-Lifshitz Course. In this book we will always distinguish energy and mass and keep \( c \) in all equations.

From now on we drop the subscript ‘0’ and consider the inertial mass of a body to be:

\[ m \equiv \frac{E_0}{c^2} . \]  

(15.20)

\(^2\)In letter of 19 June 1949 to Lincoln Barnett, sourced in Hebrew University of Jerusalem, Israel by Lev Okun.

Einstein’s publication on the equivalence of mass and energy ends with the phrase “If the theory relates to nature, radiation transfers inertia between the emitting and absorbing bodies.” This exercise shows how that transfer of mass happens even though the photon itself is massless: A quantum of light, a photon, or even a few coherent photons as indicated in figure 15-4 are emitted from the left end and travel toward the right end of a closed box of length $L$ and mass $M$. Considering the displacement of the box but the non-motion of the system’s center of mass, show the equivalence of mass and energy. The energy and momentum of a single photon are related by $E_\gamma = p_\gamma c$.

Solution

We study the box in its rest frame. When the photon is emitted from the left end, heading to the right end of the box, it gives the box an impulse to the left. According to the conservation of momentum we have

$$ v = - \frac{p_\gamma}{M} = - \frac{E_\gamma}{Mc} , $$

in which $v$ is the speed of the box. As shown in figure 15-4 after the time interval $\Delta t$, the emitted photon hits the right end of the box. Over this time, the box has shifted to the left by the distance

$$ \Delta x = v \Delta t , \quad (\Delta x < 0) $$
In this time, the light has covered the distance given by

\[ L + \Delta x = c\Delta t. \]

From Eq. (1) through Eq. (3) we have:

\[ \Delta x = \frac{E\gamma}{Mc} \Delta t = -\frac{E\gamma}{Mc^2} (L + \Delta x). \]

When the light is absorbed, it returns the momentum \( p_\gamma \) back to the box and brings it to a standstill. However, the box has moved by \( \Delta x \); if so, what happened to the center of mass of the box? Did it also shift by \( \Delta x \)? We know that the location of the center of mass of a closed system cannot change! For this to be true we need the light to carry from the left to the right end of the box an equivalent mass \( m_{eq} \) which compensates exactly the spatial shift of the box so that there is no shift in the location of the center of mass of the box.

Thus there are two contributions to the shift of the center of mass of the system.

The first is the box of mass \( M \) moving by \( \Delta x \), and the second is the mass \( m_{eq} \) moving by \( (L + \Delta x) \). Requiring the combined shift \( \Delta x_{cm} \) to be zero, we have

\[ \Delta x_{cm} = M\Delta x + m_{eq}(L + \Delta x) = 0. \]

With Eq. (4) we have

\[ \left( -\frac{E\gamma}{c^2} + m_{eq} \right) (L + \Delta x) = 0. \]

Since \( (L + \Delta x) = 0 \) implies \( M = 0 \) according to Eq. (5), the nontrivial solution is

\[ m_{eq} = \frac{E\gamma}{c^2}. \]

Note that the quantity \( M \) does not appear in Eq. (6) which means that what matters in Eq. (1) is that momentum and velocity vectors are proportional to each other. In section 17.1 we will discuss the precise form of the relativistic momentum.

As a final step, we note that Eq. (4) can be solved for the shift of the box to the left which is of course a small but finite effect – as we have shown the transfer of mass from left to right keeps the center of mass location unchanged.

\[ \Delta x = -L \frac{1}{1 + Mc^2/E\gamma}. \]

We found that the energy and the transported mass of the electromagnetic wave are related just by the usual Einstein form Eq. (7). This effect also explains how the photon energy contributes to the mass of the box-photon system. When the photon is emitted, this energy must come from the mass of the box, and when it is reabsorbed (through atomic and thermal excitation), this energy must contribute again to the box’s mass.

---

End VI–1: “Radiation transfers inertia”
16 Generalized Mass-Energy Equivalence

16.1 Where is energy coming from?

In the following we will show, case by case, that any energy we extract or deposit in a body relates directly to the energy locked in the rest mass of the body. This clarifies the question that is often paraphrased by the answer: energy comes from the electrical socket. No. The energy we use, in whatever format, ultimately comes from a machine which was built to convert a small fraction of mass of matter into another form of energy that can serve a useful task.

Just before the French revolution, Antoine Lavoisier\(^4\) demonstrated that the combustion of gases preserves the mass of the material, just in a new form. Today we make another step: any heat produced in combustion must in fact come from a reduction of rest mass of all final reaction products. When we ‘use’ energy, say, drive a car, the combustion of gasoline with air creates a tiny mass defect well beyond Lavoisier mass scales.

Consider as another example the conversion of energy that takes place in our Sun. Our Sun is a giant fusion reactor. It burns up some of its mass. When some of the emitted radiation reaches the Earth, it helps in biological processes to break CO\(_2\) molecules, undoing the binding of carbon and oxygen. This is temporarily adding to the mass of the Earth what we later release in our power stations burning the fossil fuel. The heat is (hopefully) radiated away, and the Earth’s mass balance is restored.

On the other hand Earthbound nuclear fission, and in the future, fusion power stations, convert elements into each other, and the resultant energy is leading ultimately to additionally produced heat, which will be radiated. The element conversion on Earth will result in a slight reduction of Earth mass.

Thus the energy we use comes from reduction of mass of the Sun and, when nuclear processes are employed, reduction of the mass of the Earth. In the following, we will make case by case our argument; in section 18.4 on page 241 we return to a more quantitative description of how a radioactive decay of a heavy particle into two lighter particles results in net energy gain.

16.2 Mass equivalence for kinetic energy in a gas

Let us consider a box containing a mono-atomic gas in thermal equilibrium at temperature \(T\). When the temperature of the gas increases, the amount of thermal energy contained therein does as well. Thus, we expect the mass of the

\(^4\)Antoine-Laurent deLavoisier (1743 - 1794); French nobleman and chemist central to the 18th-century chemical revolution.
Insight: Elementary energy units

In Eq. (14.6) on page 188 we used an energy unit GeV (giga-eV: \(= 10^9\) eV): we combined the SI-unit ‘volt’ with the elementary charge \(e\) to form an elementary unit of energy ‘eV’, \(i.e.\) the amount of kinetic energy that a particle of elementary charge \(e\) acquires when accelerated across a potential step of 1 V. The relationship with joule SI-unit of energy follows by definition: \(1\text{V}=1\text{ J}/\text{C}\); thus for the conversion we need the measured value of elementary charge \(e = 1.602 176 \times 10^{-19}\text{ C}\)

\[
eV = 1.602 177 \times 10^{-19}\text{ J} , \quad 1\text{ J} = 6.241 509 \times 10^{18}\text{ eV}
\]  

(16.1)

Particle and nuclear physicists quote masses of particles in energy equivalent units of eV/c\(^2\). Evolving both sides of the last result in Eq. (16.1)

\[
1\text{ J} = 1\text{ kg}\frac{m^2}{s^2} = 6.241 509\times 10^{18} (299 792 458)^2\frac{m^2\text{ eV}}{s^2}c^2 
\]  

\rightarrow 

(16.2)

\[
1\text{ kg} = 5.609 589 \times 10^{35}\frac{\text{eV}}{c^2} , \quad 1\text{ eV} = 1.782 662 \times 10^{-36}\text{ kg }c^2 .
\]  

(16.3)

The very large number appearing in Eq. (16.3) arises since we multiply the number of elementary charges in one coulomb with the square of speed of light in SI-units.

Elementary particles have an energy-mass equivalent at GeV scale. The case of the electron is very special, at half MeV (mega-eV = \(10^6\) eV). The kinetic energy available in most powerful elementary particle accelerator is below 10 TeV (tera-eV: \(= 10^{12}\) eV). Higher powers PeV (peta: \(= 10^{15}\) eV), EeV (exa: \(= 10^{18}\) eV), ZeV (zetta: \(= 10^{21}\) eV) appear, for example, when we deal with instantaneous power of ultra-pulsed lasers. On the other end of the scale we note that at room temperature and in chemistry the unit meV (milli-eV: \(= 10^{-3}\) eV) appears: the average kinetic energy of an atom or molecule air component at room temperature is about 40 meV. The typical molecular binding energy is a few eV, electron binding energy in heavy atoms reaches to 100 keV (kilo: \(= 10^3\) eV), the nuclear binding energy is up to 15 MeV.

Another important insight is how 1 eV energy per atom or molecule translates into Joules per mol. We multiply the eV–J relation in Eq. (16.1) with the Avogadro constant \(N_A\)

\[
N_A = 6.022 141 \times 10^{23}/\text{mol} \rightarrow \frac{\text{eV}}{\text{Atom}} \equiv 0.964 853 \times 10^5 \frac{\text{J}}{\text{mol}} .
\]  

(16.4)

The use of elementary energy unit ‘eV’ is advantageous in considerations that address properties of single elementary objects. It is widely accepted as a complement to the SI-units.
box to increase. The average energy of mono-atomic gas particles is given by the equipartition theorem,
\[
< E > = \frac{3}{2} k_b T ,
\]
(16.5) where \( k_b = 1.38 \times 10^{-23} \) J/K is the Boltzmann constant. As we have already seen, the mass of the system depends on the kinetic energy of its components, and the contribution of each of the gas atoms to the mass is proportional to
\[
\delta M = \frac{< E >}{c^2} ,
\]
(16.6) so the total mass of the gas increases as its temperature is increased, as we practice in exercise VI–2.

Exercise VI–2: Thermal Mass content of a hot Zeppelin

A Super-Zeppelin is kept afloat in the sky by 100 metric tons of helium gas providing a lift capability similar to the Boeing 747. Over the course of a flight on a sunny day, the temperature of the gas increases from \( T_1 = 270 \) K to \( T_2 = 310 \) K. By how much does the mass of the gas increase?

Solution

The atomic mass of helium is \( \mu \approx 4 \) g/mol. The number of particles in the gas is given by
\[
1 \quad N = \frac{M}{\mu} = \frac{10^8 \text{ g}}{4 \text{ g/mol}} N_A \approx 1.5 \times 10^{31} ,
\]
where \( N_A = 6.02 \times 10^{23} \text{ mol}^{-1} \) is Avogadro’s number. The change in total kinetic energy due to heating is then given by
\[
2 \quad \Delta E = E_2 - E_1 = \frac{3}{2} N k_b (T_2 - T_1) \approx 1.2 \times 10^{10} \text{ J} .
\]
This seems to be large, yet this is the power of the Sun, impacting the Super-Zeppelin within an hour. The fractional change in mass of the gas due to the change in thermal energy content, the fractional mass increase, is
\[
3 \quad \frac{\Delta M}{M} = \frac{\Delta E}{Mc^2} \approx \frac{0.14 \text{ mg}}{M} \approx 10^{-12} .
\]

The “Hindenburg” destroyed by fire in 1937 got its lift from 200,000 m³ of hydrogen; that is, 18 metric tons. The desired helium filling would have been twice as heavy, 36 tons: helium has nearly 93% of the lift of hydrogen since \( (29 - 4)/(29 - 2) = 0.926 \) comparing to the air molecular weight 29.0. Helium is believed to originate from radioactive decay in the depths of the Earth and is found mixed with natural gas, sometimes in abundance exceeding 5%. Helium is generally considered to be a strategic material: it is the second-lightest element and being a noble gas is chemically non-reactive, with a boiling point at 4.4 K it plays an essential role in a wide range of technological and scientific applications.
As the smallness of this mass-gain result suggests, the blimp is in no danger of crashing. Note that it is next to impossible to measure a fractional mass change that is so small.

16.3 Potential energy mass equivalence

Let us consider now the situation of a box containing two balls of identical mass $m$ joined by a massless spring, compressed and latched at a high potential $V$, as depicted in figure 16-1. When the spring’s latch is released, the balls are thrown off. Due to conservation of momentum, the balls must have equal and opposite speed $\pm u$ at all times as measured in the original system’s rest frame. As the extension of the spring reaches equilibrium position, all of the potential energy is converted to kinetic energy of both balls.

Any observer that does not know what is ‘inside’ the box shown in figure 16-1 will always report the same mass-energy equivalent. In other words, as the potential energy of the compressed spring is released into the kinetic energy of the two masses, there is no change in the mass equivalent of the experimental box. From the conservation of energy in the original rest frame, considering the situation before and after spring release we see that we have

$$V(x(t)) + 2\frac{mc^2}{\sqrt{1 - u(t)^2/c^2}} = \text{Const.}. \quad (16.7)$$

Once all potential energy is converted into motion of the balls we have, at this time $t_m$ a speed $u_m$ and

$$2\frac{mc^2}{\sqrt{1 - u_m^2/c^2}} = Me^2, \quad (16.8)$$

where $M$ is the mass of the box comprising the balls, setting aside the contributions which are not changing with time. This fixes the constant in Eq. (16.7).

Comparing the energy content at $t = 0$ where $u(0) = 0$ and $V(x(0)) = V_0$ with the time $t_m$ where all potential energy is converted into motion, we have, using Eq. (16.7)

$$V_0 + 2mc^2 = 2\frac{mc^2}{\sqrt{1 - u_m^2/c^2}} = Me^2. \quad (16.9)$$

The merit of this consideration is now clear. When, internal to a body, potential energy is converted to kinetic energy, externally we see no change in the total mass of the body. The potential energy stored in the compressed spring is recognized as contributing the amount $\delta M = V/c^2$ to the mass of the body.
Figure 16-1: Two balls attached to a spring in its rest frame, before and after the spring releases its potential $V$. The mass of the body comprising this device does not depend on spring compression.

Conversely, after all available potential energy is converted to motion, we have exactly this mass effect now in the internal kinetic energy of components of the body which contribute to the rest mass accordingly.

Exercise VI–3: Potential energy mass defect

All massive bodies are bound by gravity to Earth’s surface. Ignoring today the effects of general relativity, evaluate the mass defect due to the gravitational potential of the Earth.

Solution

The force on a mass $m$ located at Earth’s surface is pointing towards the center of the Earth and its magnitude is

$$ F_g = mg = \frac{GMm}{R^2} , $$

where $g = 9.80\text{m/s}^2$, $G$ is gravitational coupling constant, $M$ is the Earth mass and $R$ is the Earth radius. The introduction of $g$ allows us to sidestep the computation using large numbers.

The gravitational potential energy for a body at surface can be written in the form

$$ V_g = -\frac{GMm}{R} = -mgR , $$

and this is by how much the mass is decreased

$$ \delta mc^2 = V $$
since \( V \) is negative. The fractional gravity mass defect thus is

\[
4 \frac{\delta m}{m} = \frac{V/c^2}{m} = -\frac{gR}{c^2} = -6.94 \times 10^{-10}.
\]

The numerical value is given for the average radius of the Earth \( R = 6371 \text{ km} \). This gravitational mass defect is, as we see, small but in fact comparable to atomic and chemical mass defects, yet usually this effect is rarely discussed. That the effect is small but not negligible we notice in computing the magnitude for a human body, which turns out to be about 50 \( \mu \text{g} \).

We note that the above considerations are qualitative as we did not introduce a proper theory of gravity, general relativity. However, since the consideration is carried out in Newtonian limit and does not involve relativistic motion, it should be valid.

**End VI–3: Potential energy mass defect**

### 16.4 Atomic mass defect

Consider an electron in a high ‘excited’ orbit around an atomic nucleus. When the electron moves into an ‘orbit’ with lower energy, the difference in energy is carried away in the form of an emitted photon. As we have already seen, a photon can carry mass away from a body, so the mass of the atom decreases with the change in binding energy of the electron. In the simplest case of an electron \( e^- \) binding to a proton \( p \) we have

\[
e^- + p \rightarrow \text{H} + \overline{Q}.
\]  

(16.10)

Here the energy released per each reaction in form of radiation is marked \( \overline{Q} = 13.6 \text{ eV} \). The energy-mass balance of the reaction is

\[
m_ec^2 + m_pc^2 = m_Hc^2 + \overline{Q}.
\]  

(16.11)

The mass of the product of the reaction, here atomic hydrogen, is thus smaller than the masses of the components by the amount

\[
\delta m = \frac{\overline{Q}}{c^2} = m_e + m_p - m_H.
\]  

(16.12)

The fractional mass defect of the most bound electron in hydrogen is

\[
\frac{\delta m}{m_H} = \frac{\overline{Q}}{m_Hc^2} = 1.45 \times 10^{-8},
\]  

(16.13)

where we have used electron binding \( \overline{Q} = 13.6 \text{ eV} \) and \( m_Hc^2 = 0.938 \times 10^9 \text{ eV} \). This atomic mass defect is at the same time small, and yet large e.g. when compared to the effect of the gravitational potential.
16.5 Rotational energy mass equivalence

We have seen that kinetic energy internal to a body contributes to the body mass. Rotation is a special form of kinetic motion and thus a body that rotates acquires additional energy equivalent mass content. The rotational energy must contribute since locally it arises from the speed \( v_i \) of each of the constituents of the rotating body of mass \( m \). In the case \( (v/c) \ll 1 \) we have for the rotating body (assuming the center of mass is stationary),

\[
\frac{E(v)}{c^2} = m + \frac{E_{rot}^{NR}}{c^2} > m ,
\]

in which \( E_{rot}^{NR} \) is the non-relativistic rotational kinetic energy obtained using a classic form of moments of inertia.

The fractional increase in flywheel mass we achieve by speeding it up remains rather small. The best flywheels will not exceed the \( 10^{-11} - 10^{-13} \) fraction we discussed in exercise [VI–2]. This is so since the speed of the flywheel rotation must be below the speed scale of the effect of thermal motion, to assure that the flywheel centripetal forces do not break the device.

This effect of rotational energy mass equivalence can, in principle, be measured as is depicted in figure [16-2]. The spring-scale measures the weight of a body \( m \) at rest relative to the scale, and another identical scale on the right measures the weight of an identical body in uniform rotational motion relative to the scale.
The rotating body weighs in with $E(v)/c^2$ more mass. In ‘constructing’ the scale in figure [16-2] we tacitly used the equivalence of inertial and gravitational mass. This is known as the principle of equivalence, the consideration which leads ultimately to the general theory of relativity.

There are ongoing efforts to build a car kinetic energy recovery system (KERS) based on an ultra-rapid flywheel. Before taking a relativity class you would think of rotation energy, originating in rotational motion, as the energy source. Now, another way to think about this situation is to realize that the rotating flywheel is somewhat more massive when it stores energy. As the energy of the flywheel is consumed, it slows down and therefore a lower mass value of the rotating flywheel is reached. We can say that the car with flywheel power-assist drive derives its energy from the variability of the mass of the flywheel, which depends on the wheel rotation. Consideration of the mass and how it changes provides a more general view about the origin of the usable energy.

16.6 Chemical energy mass defect

When we burn hydro(carbon) fuels, we think that the energy comes from chemical reactions between the hydrocarbons and the oxygen in air. However, the source of the energy gain, as we now realize, is in reducing the mass of the reaction products as compared to the initial components. Let us check how big this effect is for, arguably, the most efficient of these ‘chemical’ reactions

$$H_2 + \frac{1}{2}O_2 \rightarrow H_2O + Q_h, \quad (16.15)$$

where $Q$ is the heat released, in this example, $Q = Q_h$ of hydrogen combustion. $Q$ is usually given in Joule per mol ($Q_h = 286$ kJ/mol) but when considering elementary reactions we prefer to use units of eV per single molecule created in the reaction. The conversion goes as follows: there are $N_A = 6.022 \times 10^{23}$ (≈ Avogadro constant) molecules in each mole and one $J \equiv 0.624 \times 10^{19}$eV (coefficient is the number of electron charges in a ‘Coulomb’ charge, see Insight on page [204]). The unit conversion factor thus is

$$R_Q = \frac{6.24 \times 10^{18}}{6.022 \times 10^{23}} = 1.03643 \times 10^{-5} \text{eV mol} \frac{\text{J}}{J}. \quad (16.16)$$

This leads to $\overline{Q}_h \equiv 2.86 \times 10^{5} JR_Q = 2.96$ eV. In each reaction Eq. (16.15) many particles participate. To be specific, since each hydrogen comprises one proton and one electron, and oxygen has 8 electrons, 8 protons and 8 neutrons, we in total have 10 electrons, 10 protons and 8 neutrons, and of course the nucleons carry practically all the mass. We recall the energy equivalents of each of these
particles based on eV units, see Insight on page 204:

\[ m_e c^2 = 0.5110 \text{ MeV}, \quad (16.17) \]

\[ m_p c^2 = 938.3 \text{ MeV}, \quad m_n c^2 = 939.6 \text{ MeV}. \]

We estimate the total mass of participating particles at 16,900 MeV. Each reaction Eq. (16.15) gains approx. 3 eV out of \( 1.7 \times 10^{10} \text{ eV} \), which amounts to a fractional reduction of mass at the level of \( 1.8 \times 10^{-10} \).

\[ \frac{\delta m}{m} \bigg|_{\text{chemical}} \leq 1.8 \times 10^{-10}. \quad (16.18) \]

This is, arguably, the largest mass defect we can find among chemical reactions. This chemical mass defect is also, on the order of 100 times greater than the flywheel mass defect we considered earlier. And, it is a factor 4 smaller than the gravity potential mass defect, Eq. (4) in exercise VI–3.

### 16.7 Nuclear mass defect

One of likely near term paths to fusion involves heavy hydrogen, called deuteron. Each atomic deuteron D has a nucleus \( d = pn \) made of one proton and one neutron. The deuteron abundance around us is the remainder left over from the primordial nuclear burning period of the early Universe. In water on Earth one out 6,420 of each hydrogen atoms is actually a deuterium atom. The relatively small amount remaining testifies to the relative ease with which deuterium can serve as nuclear fuel in primordial burn processes.

To make a deuteron, one proton binds with a neutron

\[ p + n \rightarrow d + \gamma + Q_d. \quad (16.19) \]

The energy released is \( Q_d = 2.2245 \text{ MeV} \). The difference in energy between the fusing nuclei and the fused nucleus formed is removed by gamma-radiation; that is, a very energetic photon of nuclear origin. The fractional nuclear mass defect for the reaction Eq. (16.19) in view of the masses Eq. (16.17) is

\[ \frac{\delta m}{m} \simeq \frac{2.2245}{938.3 + 939.6} = 1.2 \times 10^{-3}. \quad (16.20) \]

This is very large compared to atomic, chemical or gravitational fractional mass defects. The key advantage of nuclear fusion over chemical burning is evident in the scale of energies we are considering, which are about a million times greater concerning the mass defect. A simple way to see this is to remember that we count
nuclear reaction energy on the scale of MeV as compared to eV per reaction for similar atomic components.

Generally, the energy released in nuclear reactions such as the fusion of light elements or the fission of heavy elements manifests predominantly as the kinetic energy of released reaction fragments. In all these reactions just as in this example, the binding mass defect is at a noticeable level of 0.1% and often even somewhat higher: for the fusion reaction likely to power the first generation fusion reactors, the mass defect is three times as large as that given in Eq. (16.20).

The effectiveness of nuclear reactions like the one seen in Eq. (16.19) shows why the search for an acceptable form of nuclear energy continues despite considerable technological hurdles. A crucial requirement is the minimization of naturally radioactive components produced. One path is aneutronic fusion, a controlled fusion burn without a significant production of neutrons. Neutrons, being neutral, are very penetrating and can travel far in their lifespan of about $\tau_n = 15$ min.

### Exercise VI–4: Energetics of deuterium in water

Evaluate, in principle, the accessible fusion energy content of one gallon of distilled water and compare this to the chemical energy content of an equal volume of gasoline, $Q_{\text{gas}} \simeq 125$ MJ.

**Solution**

One mole of H$_2$O will weigh in at approximately 2+16=18 grams. One gallon of water, that is 3.6kg, comprises therefore 200 moles. This round number explains why we choose in this example the non-metric unit. The number of deuterons in one gallon of water is

\[ N_d = \frac{200}{6420} \times 2N_A = 3.76 \times 10^{22}, \quad N_A = 6.022 \times 10^{23}, \]

where we used the Avogadro number $N_A$ and we recall that only 1/6240 hydrogen fraction is a deuteron. The additional factor ‘2’ reminds us that there are two hydrogen atoms in each molecule of water.

We assume that on average one d can help release an energy of $\bar{Q}_d = 5$ MeV. We recall from Insight on page [204]

\[ 1\text{ eV} \equiv 1.6 \times 10^{-19} \text{J}, \quad 1\text{ MeV} \equiv 1.6 \times 10^{-13} \text{J}. \]

The estimate of the total energy content of deuterium content in water is

\[ Q_d = N_d \bar{Q}_d \frac{1.6 \times 10^{-13} \text{J}}{\text{MeV}} = N_d 8 \times 10^{-13} \text{J}. \]
In one gallon of water the energy content is thus

\[ Q_{\text{d gallon}} = 3 \times 10^{10} \text{J} = 240 Q_{\text{gas}}. \]

The energy content available with future technology originating in heavy hydrogen in water is greatly above the energy available in an equal volume of gasoline.

---

16.8 Origin of energy

To gain usable energy we arrange reactions that reduce, generally by a very small amount, the mass. This leads us to the fundamental question: where does the mass we consume in the energy production process originate? The laboratory study employing relativistic collisions of heavy nuclear ions helped us establish that matter as we know it emerged from the primordial phase of matter, the quark-gluon plasma when the Universe was about 20 µs old, see Ref.4 on page xi. Since this ‘hadronization’ epoch, the material mass of the Universe has been decreasing, providing in part the energy that feeds into Universe expansion. When we consume usable energy today, ultimately it winds up in radiation contributing to this expansion process.

However, we do not understand the origin of the small, nano-excess of nucleons over anti-nucleons present in the primordial Universe, an excess that determines the present day visible matter content. Finding an answer to the origin of matter-anti-matter asymmetry is among the great foundational science challenges of the present age.

In the course of the ensuing Universe’s lifespan, many nuclear processes contributed to the formation of the element abundance we find on Earth and observe in the Sun. The Sun is powered by nuclear fusion processes that burn the lightest element, hydrogen, into heavier nuclei. The power emitted by the Sun is transmitted in form of radiation, to Earth. Only a very tiny fraction, that is 173,000 TW, corresponds to the fraction of the surface of the Earth compared to the entire spherical surface irradiated by solar radiation. In addition, heat flows from Earth’s interior at a rate of about\(^6\) 44.2 ± 1 TW. Only about half of this heat is attributed to natural radioactive decay.

Let us establish how the power unit, TW = \(10^{12} \text{J/s}\), relates to mass consumption. We first employ the fact that the charge of the electron entering the unit ‘eV’ is \(e = 1.602177 \times 10^{-19} \text{C}\)

\[
1 \text{TW} = 10^{12} \text{CV/s} = \frac{10^{12} \text{eV}}{1.602177 \times 10^{-19} \text{s}} \quad (16.21)
\]

and that 1g of hydrogen contains Avogadro’s number $N_A = 6.022141 \times 10^{23}$ of individual atoms, each weighing in at $M_H = (938.272 + 0.511)10^6 \text{eV/c}^2$

\[ 1 \text{g} c^2 = N_A M_H c^2 = 6.022141 \times 10^{23} (938.272 + 0.511)10^6 \text{eV}. \]  

(16.22)

Combining Eq. (16.21) with Eq. (16.22) we obtain

\[ 1 \text{TW} = 10^{12} \frac{1.782662 \times 10^{-36} \text{kg} c^2}{1.602177 \times 10^{-19} \text{s}} = 1.11265 \times 10^{-5} \frac{\text{kg} c^2}{\text{s}}. \]  

(16.23)

where the numerator is the eV/kg-ratio Eq. (16.22). This awkward conversion number shows the arbitrariness of SI units originating in pre-Einstein era. We take note that the Sun delivers to Earth every second 1.9kg in radiation energy equivalent, that is in a year the equivalent of 60,000,000 kg in radiation energy equivalent. Even if this is a small amount compared to the mass of the Earth, we need to also remember that this mass-energy equivalent is re-radiated away in thermal balance.

In comparison, the total annual World energy consumption is about $5 \times 10^8 \text{TJ}$, corresponding (upon division with $3.15 \times 10^7 \text{s/year}$) to 16 TW power. Thus human activity perturbs in an unmeasurably small way the 217,000 TW natural energy balance of the Earth. However, the production of pollutants capable of catalyzing the capture of even a small fraction of the natural energy in-out-flow of 217,000 TW has the potential of creating a significant disturbance of the Earth energy balance. To escape the pollutant-caused global warming, the use of solar power is proposed, and harvesting a small 1 in 10,000 part of available energy suffices to satisfy all the power needs.

In the context of balancing the energy needs, ‘renewable’ means the exploitation on Earth of the solar fusion energy, and ‘fossil’ describes the solar energy of the past stocked in carbon-hydrates. For comparison, nuclear power, be it fusion, or fission, exploits the primordial elemental abundance and their mass, created by processes operating prior to the formation of the solar system.

17 Particle Momentum

17.1 Relation between energy and momentum

We now obtain the form of the relativistic momentum which is consistent with the relativistic energy form Eq. (15.21). Our arguments will be justified by the result. We recall that kinetic energy is imparted to a body through the action of

---

7 http://www.nist.gov/pml/wmd/metric/mass.cfm “The SI base unit of mass, the kilogram, is the last remaining physical artifact.” See also the comment in Ref.6 in Part I
a force which does work on the body. We consider an infinitesimal change in the
energy of a body according to the work-energy theorem
\[
dE = \vec{F} \cdot dx = \frac{d\vec{p}}{dt} \cdot d\vec{x} = \frac{d\vec{v}}{dt} \cdot d\vec{p} = \vec{v} \cdot d\vec{p},
\]
(17.1)
that is, the infinitesimal change in kinetic energy of a body is described by the
product of velocity \(\vec{v}\) with infinitesimal change in momentum \(d\vec{p}\). This equation
is exactly the same as the equation one considers for nonrelativistic dynamics;
one can regard Eq. (17.1) as guiding to the definition of the momentum vector.

Relativity enters this momentum definition in the specific form Eq. (15.21) for
\(E(v)\) which implies
\[
dE = \frac{m\vec{v} \cdot d\vec{v}}{(1 - (v/c)^2)^{3/2}}.
\]
(17.2)
We combine Eq. (17.1) with Eq. (17.2) eliminating \(dE\)
\[
\vec{v} \cdot d\vec{p} = \frac{m\vec{v} \cdot d\vec{v}}{(1 - (v/c)^2)^{3/2}}.
\]
(17.3)
The momentum component parallel to velocity is
\[
\vec{p} = \frac{m\vec{v}}{\sqrt{1 - (v/c)^2}} = m\vec{v}\gamma.
\]
(17.4)
To prove Eq. (17.4) we compute the left hand side of Eq. (17.3) using Eq. (17.4)
\[
\vec{v} \cdot d\vec{p} = m\gamma\vec{v} \cdot d\vec{v} + m\gamma^2 d\gamma = m\gamma^3(1 - v^2/c^2)\frac{dv^2}{2} + m\gamma^3(v^2/c^2)\frac{dv^2}{2} = m\gamma^3\frac{dv^2}{2},
\]
which agrees with Eq. (17.3).

Note that Eq. (17.3) does not allow us to find the component of momentum,
if any, which is normal to velocity. However, it may be argued that the mean-
ing of momentum implies that component orthogonal to \(\vec{v}\) does not exist. To
recapitulate, the relativistic momentum of a particle given by Eq. (17.4) has been
obtained from the relativistic energy of the moving body, Eq. (15.21) and the
work-energy theorem Eq. (17.1).

For a given momentum and energy we find the velocity of a particle from
\[
\vec{\beta} = \frac{\vec{v}}{c} = \frac{\vec{p}}{E}.
\]
(17.5)
Expanding in \(v/c\) the relativistic form of three-momentum Eq. (17.4) we obtain
\[
\vec{p} = m\vec{v} + m\vec{v} \left[ \frac{1}{2} \left( \frac{v}{c} \right)^2 + \frac{3}{8} \left( \frac{v}{c} \right)^4 + \frac{5}{16} \left( \frac{v}{c} \right)^6 + \frac{35}{128} \left( \frac{v}{c} \right)^8 + \ldots \right] .
\]
(17.6)
Particle Momentum

Figure 17-1: The relationship between energy, momentum, and rest mass, as given in Eq. (17.8).

The first term is the non-relativistic three-momentum. This reaffirms that the vector $\vec{p} = \gamma m \vec{v}$ is an appropriate relativistic generalization of the three-momentum vector.

Using Eq. (17.4) and Eq. (15.21) we evaluate

$$E^2 - \vec{p}^2 c^2 = m^2 c^4 \gamma^2 (1 - \vec{v}^2 / c^2) = m^2 c^4 ,$$

which may be conveniently written in the form

$$E / c = \sqrt{m^2 c^2 + \vec{p}^2} .$$

(17.8)

It can be helpful to relate this relationship to the sides of a right triangle, as shown in figure 17-1. Note that the base of the triangle is the same in any reference frame, as the rest mass is an invariant quantity. However, a moving observer $S'$ will always observe a different magnitude of momentum, and thus energy allowing to close the triangle.

Collecting the different relativistic energy and momentum relations for a particle moving with velocity $\vec{v}$

$$E = \gamma mc^2$$

$$= \sqrt{(mc^2)^2 + (\vec{p}c)^2} ,$$

$$c \vec{p} = \gamma mc \vec{v}$$

$$= \gamma \beta mc^2 ,$$

$$c |\vec{p}| = \sqrt{E^2 - (mc^2)^2} .$$

(17.9)

This form of energy and momentum as a function of velocity suggests an important analogy between the space-time coordinates and momentum and energy

$$(ct, \vec{x}) \leftrightarrow (E, c\vec{p}) ,$$

(17.10)
where the analogy in particular extends to the behavior under Lorentz transformation:

\[
\begin{align*}
E' &= \gamma(E + cp_x \beta), \\
cp'_x &= \gamma(cp_x + E \beta), \\
p'_y &= p_y, \quad p'_z = p_z.
\end{align*}
\] (17.11)

Note that unlike transformation of coordinates (passive transformation) we study here the transformation of particle properties, moving the particle in an active transformation. Thus the + sign is required to connect \( E \) and \( \vec{p} \) under LT.

The interesting special case is that of a particle at rest so \( E = mc^2 \) and \( \vec{p} = 0 \). In order to determine the corresponding finite velocity \( v_x \) properties we perform an active Lorentz transform, that is we boost this particle employing Eq. (17.11) and find

\[
E' = \gamma(E + (cp_x = 0) \beta), \quad cp'_x = \gamma((cp_x = 0) + E \beta), \quad p'_y = p_y = 0, \quad p'_z = p_z = 0
\]

which agrees with Eq. (17.9). By explicit computation we further confirm that

\[
E'^2 - \vec{p}'^2 c^2 = \gamma^2[(E + cp_x \beta)^2 - (cp_x + E \beta)^2] - p_y^2 - p_z^2 = \gamma^2(1 - \beta^2)(E^2 - (cp_x)^2) - p_y^2 - p_z^2.
\]

Thus we have

\[
E'^2 - \vec{p}'^2 c^2 = E^2 - \vec{p}^2 c^2 = (mc^2)^2.
\] (17.12)

This result means that the mass value of a body reported by two observers, the one at rest with respect to the body, and the other moving with a speed \( v \), are the same. This property, called Lorentz invariance of mass, means that the energy locked in the mass is the same for all observers, including in particular the comoving observer. The realization that the intrinsic energy content of a body is the same for all observers is an insight of monumental practical importance allowing the generalization of the energy conservation law to include the energy content of all bodies involved, see section 16.

Exercise VI–5: Nonrelativistic limit of energy-momentum relation

Find an approximation for the energy of an nonrelativistic particle in terms of its momentum and compare to Eq. (15.19), explain the difference.

Solution

We begin with the form seen in Eq. (17.9)

\[
E = \sqrt{(\vec{p}c)^2 + (mc^2)^2} = mc^2 \sqrt{1 + x}, \quad x = \frac{p^2}{(mc)^2}.
\]
We use the Taylor expansion
\[ \sqrt{1 + x} = 1 + \frac{1}{2} x - \frac{1}{8} x^2 + \frac{1}{16} x^3 - \frac{5}{128} x^4 + \ldots. \]
This leads to the expression for energy
\[ E = mc^2 + \frac{p^2}{2m} - mc^2 \left[ \frac{1}{8 (mc)^4} - \frac{1}{16 (mc)^6} + \frac{5}{128 (mc)^8} - \ldots \right]. \]
The coefficients of the power series we see in the velocity expansion Eq. (15.19) of energy are different from those we have obtained in Eq. (3). The reason is the relation between the momentum and velocity which also is a power series, see Eq. (17.6) and which when used in Eq. (3) will produce the earlier velocity series Eq. (15.19). The momentum expansion of energy Eq. (3) converges faster compared to velocity expansion Eq. (15.19).

End VI–5: Nonrelativistic limit of energy-momentum relation

Exercise VI–6: Ultra-relativistic limit

Find an approximation for the energy of an ultra-relativistic particle, that is a particle with kinetic energy much greater than its rest energy. What is the velocity of such a particle?

Solution

Large kinetic energy implies \( p^2 c^2 \gg m^2 c^4 \) so that the expansion
\[ E = \sqrt{p^2 c^2 + m^2 c^4} = pc \sqrt{1 + \left( \frac{mc^2}{pc} \right)^2} = pc + \frac{m^2 c^4}{2pc} + \ldots \]
is justified. To the first order \( E = pc \), which is just the relationship for massless particles e.g. photons. Seeing this result one also easily realizes an exact form
\[ E - pc = \frac{(\sqrt{p^2 c^2 + m^2 c^4})^2 - (pc)^2}{\sqrt{p^2 c^2 + m^2 c^4} + pc} = \frac{m^2 c^4}{E + pc}. \]

To determine the speed using Eq. (17.5)
\[ v = \frac{pc^2}{E} = c \frac{1}{\sqrt{1 + (mc^2/pc)^2}} = c \left( 1 - \frac{1}{2} \left( \frac{mc^2}{pc} \right)^2 + \ldots \right). \]
In most cases for ultra-relativistic particles we can simply use \( v \approx c \). Seeing this result one can also, just as before, obtain an exact form
\[ c^2 - v^2 = \frac{(Ec)^2 - (pc)^2}{E^2} = c^2 \left( \frac{mc^2}{E} \right)^2. \]
Given the energy of a relativistic particle we can use this to evaluate the small deviation from the speed of light.

We note that the deviations are all $O(m^2)$.

**17.2 Particle rapidity**

The Lorentz transformation Eq. (17.11) of energy and momentum can also be written using the rapidity $y_r$ of the Lorentz transformation introduced in section 7.5. Analogous to Eq. (7.49) for a change in coordinates, we have, using rapidity, the transformed energy and momentum for a boost in the $x$-direction:

\[
\begin{align*}
E' &= \cosh y_r E + \sinh y_r cp_x, \\
cp'_x &= \cosh y_r cp_x + \sinh y_r E, \\
p'_y &= p_y, \quad p'_z = p_z.
\end{align*}
\]  \hspace{1cm} (17.13)

We now address an interesting special case: imagine that we start with $p_x = 0$. In this case we have

\[
E(p_x = 0) = \sqrt{m^2 c^4 + (p_y^2 + p_z^2)c^2} \equiv \sqrt{m^2 c^4 + \vec{p}_\perp^2 c^2} = E_\perp. \hspace{1cm} (17.14)
\]

After we carry out the transformation Eq. (17.13) the particle is observed to have a ‘longitudinal’ momentum $p'_x \equiv p_\parallel$ and an energy $E' \rightarrow E$, where we rename the primed values and which comprises all three momentum components. Since we are starting from the case $p_x = 0$, let us call the associated rapidity of the transformation that generates the $x$-directed momentum the particle rapidity $y_p$. Then Eq. (17.13) reads

\[
\begin{align*}
E &= \cosh y_p E_\perp, \\
cp_\parallel &= \sinh y_p E_\perp, \\
p'_y &= p_y, \quad p'_z = p_z.
\end{align*}
\]  \hspace{1cm} (17.15)

Equation (17.15) is a clever way to express the energy and momentum of a particle using as reference the transverse energy $E_\perp$, and a new variable describing the longitudinal motion. At this point the choice of a $\parallel$-directed particle is discretionary. When we introduce a boost clearly we want the boost to be in
this direction. The special case that we choose as the reference direction is the
direction of motion of the particle $\vec{p}_\perp = 0, p_\parallel = p$, worth presenting explicitly:

\begin{equation}
\begin{align*}
E &= \cosh y_p mc^2, \\
cp &= \sinh y_p mc^2, \\
\vec{p}'_\perp &= p_\perp = 0.
\end{align*}
\end{equation}

As a first step we develop further the relation of particle energy and momen-
tum with particle rapidity that follows from Eq. (17.15): we see that the relation
is consistent by evaluating

$$E^2 - (c p_\parallel)^2 = (\cosh^2 y_p - \sinh^2 y_p) E_\perp^2 = E_\perp^2 \rightarrow E^2 = m^2 c^4 + c^2 p_\perp^2 + (c p_\parallel)^2.$$ 

Next we divide the two non-trivial equations Eq. (17.15) one by the other

$$\beta_\parallel \equiv \frac{c p_\parallel}{E} = \tanh y_p \rightarrow y_p = \text{arctanh}(c p_\parallel / E).$$

The first form in Eq. (17.17) reads

$$\beta_\parallel = \tanh y_p .$$

which implies

$$y_p = \frac{1}{2} \ln \left( \frac{1 + \beta_\parallel}{1 - \beta_\parallel} \right),$$

and the last form in Eq. (17.17) leads to

\begin{equation}
\begin{align*}
y_p &= \frac{1}{2} \ln \left( \frac{E + c p_\parallel}{E - c p_\parallel} \right), \\
y_p &= \ln \left( \frac{E + c p_\parallel}{E_\perp} \right), \\
y_p &= \ln \left( \frac{E_\perp}{E - c p_\parallel} \right).
\end{align*}
\end{equation}

where the last two expression arise by multiplying nominator and denominator
with $E \pm c p_\parallel$ respectively.

---

**Exercise VI–7: Properties of particle rapidity**

Show that for a particle with $\vec{p}_\perp \neq 0, p_\parallel = 0$ we have $y_p = 0$. Show that in the
nonrelativistic limit the particle rapidity is particle speed in $\parallel$-direction divided by $c$. 
Perform an explicit Lorentz boost in $\parallel$-direction and find the boosted particle rapidity.

Solution

For $p_\parallel = 0$ we see from the first form of Eq. (17.20) that $y_p \propto \ln 1 = 0$ resolving the first assertion. In order to see the behavior in the nonrelativistic limit we rewrite Eq. (17.20) to read

$$ y_p = \frac{1}{2} \ln \left( 1 + \frac{2cp_\parallel}{E - cp_\parallel} \right). $$

When the momentum is small i.e. $|\vec{p}| \ll mc$ we have the non-relativistic limit. We can replace $E - cp_\parallel \to mc^2$, which shows that the 2nd term in the logarithm is small. Expanding the logarithm we find

$$ y_p \simeq \frac{cp_\parallel}{mc^2} = \beta_\parallel. $$

Another way to reach the same conclusion employs Eq. (17.18)

$$ \beta_\parallel \equiv \frac{cp_\parallel}{E} = \tanh y_p \simeq y_p + O(y_p^3). $$

We perform an active boost of a particle in the direction of $\vec{v}$ and hence will be referring with $\perp$, $\parallel$ to the direction of $\vec{v}$. We use Eq. (17.13) where $\cosh y_r = \gamma$, $\sinh y_r = \beta \gamma$

$$ E' = \cosh y_r E + \sinh y_r cp_\parallel, $$

$$ cp'_\parallel = \cosh y_r cp_\parallel + \sinh y_r E, $$

$$ p'_\perp = p_\perp. $$

We compute the active boost using Eq. (17.20)

$$ y'_p = \frac{1}{2} \ln \left( \frac{E' + cp'_\parallel}{E' - cp'_\parallel} \right), $$

$$ = \frac{1}{2} \ln \left( \frac{\cosh y_r E + \sinh y_r cp_\parallel + \cosh y_r cp_\parallel + \sinh y_r E}{\cosh y_r E + \sinh y_r cp_\parallel + \sinh y_r cp_\parallel + \sinh y_r E} \right). $$

We recognize that the coefficients comprise $\cosh y \pm \sinh y = e^{\pm y}$ and obtain

$$ y'_p = \frac{1}{2} \ln \left( \frac{Ee^{y_r} + cp_\parallel e^{y_r}}{Ee^{-y_r} - cp_\parallel e^{-y_r}} \right) = \frac{1}{2} \ln \left[ e^{2y_r} \left( \frac{E + cp_\parallel}{E - cp_\parallel} \right) \right] = y_r + y_p. $$
This is a very general result; one can call this the theorem of addition of particle rapidities.

One can use multiple boosts to build-up a final rapidity. Starting with $p_\parallel = 0$ we have $y_p = 0$ and thus we find that the first boost rapidity $y_r$ is the particle rapidity $y'_p$. We can perform multiple boosts in sequence: if in the first boost we have moved the particle from $y_p = 0$ to $y'_p = y_{r1}$ the second boost adds on and we obtain

$$y''_p = y'_p + y_{r2} = y_{r1} + y_{r2}.$$  

We have learned that particle rapidities behave additively under LT carried out in the same direction, just like LT boost rapidities. This is so even though on first sight the expressions in Eq. (17.15) introducing particle rapidity $y_p$ are a bit more complex and not immediately apparent as comprising this property!

End VI–7: Properties of particle rapidity

Exercise VI–8: Emission polar angle transformation

This problem generalizes and expands the exercise [III–9] on page 92. We put the $x$-axis in the direction of motion of an observer moving with speed $\beta c$, or equivalently rapidity $y_r$ – this can be an observer co-moving with the source of these particles, a ‘fireball’ created in collisions of elementary particles. Within this comoving frame of reference particles are typically emitted in an isotropic distribution - there is no preferred axis. $\theta_p$ is the angle measured from the direction transverse to source motion, under which a particle is observed in the laboratory frame of reference. Let the primed angle $\theta_p'$ be the angle of emission measured by an observer at rest with the source. What is the relation of $\theta_p'$ with $\theta_p$, the angle of emission measured in the laboratory frame where the source is moving with rapidity $y_r$?

Solution

We will use the Lorentz transformation in rapidity form Eq. (17.13), where we apply the definitions of emission angles with respect to the axis of motion as follows (valid for both primed, and unprimed reference frame):

$$1 \quad p_\parallel = p \sin \theta_p, \quad p_\perp = p \cos \theta_p; \quad \parallel \equiv \hat{x}, \ \perp \equiv \hat{z}.$$  

The $z$-axis points along the transverse momentum component. Since $p = \sqrt{p_\parallel^2 + p_\perp^2}$ has a more complicated Lorentz transformation we choose to study the ratio $p_\parallel / p_\perp = \tan \theta_p$.

We thus obtain according to Eq. (17.13)

$$2 \quad \tan \theta'_p = \frac{p'_\parallel}{p'_\perp} = \frac{\cosh y_r p_\parallel - \sinh y_r (E/c)}{p_\perp} = \frac{\cosh y_r \sin \theta_p - \sinh y_r (E/cp)}{\cos \theta_p}.$$
Here the direction of motion expressed in the sign of $y_r$ is chosen such that the particle is emitted in the direction of the observer (typically comoving with the source) velocity vector.

The inverse transform is obtained by exchanging prime and unprimed quantities, and changing $y_r \rightarrow -y_r$:

\[ \tan \theta_p = \frac{\gamma y_r' p_{||} + \sinh y_r' (E'/c)}{\gamma p_{\perp}} = \frac{\cosh y_r' \sin \theta_p' + \sinh y_r' (E'/cp')}{\cos \theta_p'} . \]

We recall that $\cosh y_r = \gamma, \sinh y_r = \beta \gamma, E' = \gamma_p' mc^2, cp' = \gamma p' \beta' p mc^2$ which allows us to write

\[ \tan \theta_p = \gamma \frac{\sin \theta_p' + \beta/\beta_p'}{\cos \theta_p'} . \]

There is another way to express this result by introducing the polar angle $\tilde{\theta}$, that is an angle measured with respect to the direction of source motion. In this case $\sin \theta_p' \rightarrow \cos \tilde{\theta}_p$, $\cos \theta_p' \rightarrow \sin \tilde{\theta}_p$ and $\tan \theta_p \rightarrow \cotan \tilde{\theta}_p$ resulting in

\[ \tan \tilde{\theta}_p = \frac{1}{\gamma \cos \theta_p' + \beta/\beta_p'} . \]

For $\beta/\beta_p' \rightarrow 1$ which applies to relativistic particles and relativistic fireball motion, i.e., $\beta, \beta_p' \rightarrow 1$, we find

\[ \tan \tilde{\theta}_p \simeq \frac{1}{\gamma} \tan \frac{\theta_p'}{2} , \]

where we used the half-angle formula

\[ \tan \frac{\theta}{2} = \frac{\sin \theta}{\cos \theta + 1} , \]

easily verified by using the more familiar double angle formulas $\sin 2\alpha = 2 \sin \alpha \cos \alpha$ and $\cos 2\alpha = \cos^2 \alpha - \sin^2 \alpha$.

This result demonstrates that the CM frame emission angles under a relativistic LT are imaged into laboratory angles which are very much smaller. The effect of the relativistic (source) motion is to focus particles emitted isotropically in the CM system into a relatively small forward cone in the laboratory reference system.

**Further Reading:** When multiple source fireballs with different values of $\gamma_i$ are present, there is multiple azimuthal bunching of particles into a cone defined by the value of $\gamma$. This means that one observes in such a situation several azimuthal emission rings. Such considerations played some time ago an important role in the understanding of reaction mechanisms in reactions of strongly interacting particles\textsuperscript{8} – the meaning of

the term ‘emission angle’ in this publication is later clarified by R. Hagedorn to be the polar angle $\theta$.

End VI–8: Emission polar angle transformation

Exercise VI–9: Electron and proton beams

What is the speed and rapidity of an electron in a beam with kinetic energy 1 MeV? 1 GeV? 1 TeV? What about a proton beam? The electron and proton rest masses are respectively $m_e \approx 0.511 \text{ MeV}/c^2$ and $m_p \approx 938 \text{ MeV}/c^2$.

Solution

We recall the relationship Eq. (17.5)

$$\frac{v}{c} = \frac{p}{E},$$

which with Eq. (17.8) leads to

$$\frac{v}{c} = \frac{\sqrt{E^2 - m^2c^4}}{E} = \sqrt{1 - \frac{m^2c^4}{E^2}}.$$  

The kinetic energy $T$ is defined in Eq. (15.19) as the difference between the total energy and the rest energy, $T = E - mc^2$, so we obtain with Eq. (2) a form suitable for small $mc^2/T$,

$$\frac{v}{c} = \sqrt{1 - \frac{m^2c^4}{(T + mc^2)^2}} = \sqrt{1 + \frac{2mc^2/T}{1 + mc^2/T}}.$$

To evaluate rapidity for a given kinetic energy we employ the second form of Eq. (17.20) where we insert $E = T + mc^2$ and $cp_\parallel = cp = \sqrt{(T + mc^2)^2 - (mc^2)^2} = \sqrt{T^2 + 2Tmc^2}$, thus

$$\frac{v}{c} = \ln \left( 1 + \frac{T}{mc^2} + \sqrt{\frac{T^2}{(mc^2)^2} + 2 \frac{T}{mc^2}} \right).$$

This specific form more easily provides a more precise answer since in all cases under consideration $T/mc^2 > 1$, except for the one case of proton at 1 MeV where $T/mc^2 << 1$. However, since all terms in the argument of the logarithm are positive there is no cancellation and this expression also works in this nonrelativistic case.

---

It is important to remember that the theoretically beautiful form Eq. (17.19) is singular in the limit $\beta_\parallel \to 1$. Therefore speed should not be used in this limit to compute the rapidity since the numerical rounding errors could be important. However, we can use Eq. (17.18) to confirm the result we obtain by proceeding as follows:

$$5 \quad y_p = \ln[\gamma(1 + \beta)].$$

In the relativistic limit we can take $\beta \to 1$ and set $\gamma = 1 + \frac{T}{mc^2}$. We obtain the following speeds and rapidities: GeV-energy electrons and TeV-energy protons have velocities that differ from the speed of light by only a few parts in ten million.

<table>
<thead>
<tr>
<th>$T$</th>
<th>$T/m_e c^2$</th>
<th>$v_e$</th>
<th>$y_e$</th>
<th>$T/m_p c^2$</th>
<th>$v_p$</th>
<th>$y_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 MeV</td>
<td>1.957</td>
<td>0.941c</td>
<td>1.747</td>
<td>0.001</td>
<td>0.046c</td>
<td>0.046</td>
</tr>
<tr>
<td>1 GeV</td>
<td>1957</td>
<td>0.99999987c</td>
<td>8.27</td>
<td>1.066</td>
<td>0.875c</td>
<td>1.36</td>
</tr>
<tr>
<td>1 TeV</td>
<td>1.96 $10^6$</td>
<td>$\sim c$</td>
<td>15.2</td>
<td>1.066</td>
<td>0.99999956c</td>
<td>7.665</td>
</tr>
</tbody>
</table>

This result shows that when (kinetic) energy is much larger than the rest mass of the particle it is hard to distinguish particle speed from the speed of light. On the other hand the rapidity can characterize the relativistic particle motion quite adequately.

17.3 Relativistic rocket equation

Even though this takes us off the main track of the book briefly, we will look at a particularly special case of particle ‘decay’; that is the case of a relativistic rocket where a mass increment $\Delta m$ separates at a prescribed speed, or better, rapidity, from the main rocket body $m \gg \Delta m$. This complements section 12.4 considerations which showed how SR constrains a rocket that could accelerate with constant $a' = g$, considering now the constraints of both energy and momentum conservation.

NONRELATIVISTIC ROCKET EQUATION

First let us recapitulate the momentum conservation argument that leads to the rocket equation in nonrelativistic physics. This insight is attributed and named after Konstantin Tsiolkovsky, though the work of Tsiolkovsky on the rocket equation occurred in parallel to a mathematical study by Ivan V.
opposite to the direction desired for the change of body momentum. We begin by balancing momentum.

Immediately prior to ejection of mass increment

\[ \Delta m = -dm = |dm|, \]

(17.21)

the magnitude of momentum of the non-relativistic rocket is

\[ p_1 = (m + \Delta m)v. \]

(17.22)

The introduction of \(-dm\) in Eq. (17.21) reminds us that while \(\Delta m\) is a positive quantity, it is actually a negative of a negative change, \(i.e. -(\Delta m)\) in the mass of the rocket \(m\). We combine two contributions, the rocket and its exhaust to obtain the new momentum of the system, \(p_2\) established immediately after mass ejection

\[ p_2 = m(v + \Delta v) + \Delta m v_{\text{ex}}, \]

(17.23)

where the speed of exhausted mass is

\[ v_{\text{ex}} = v - V. \]

(17.24)

Here the propellant speed \(V\) is in the rest frame of the rocket and derives from characteristics of the engine, and we assumed that it is pointing opposite to the speed of the rocket \(v\).

Requiring momentum conservation \(p_1 = p_2\) we find after a few simple steps

\[ m \Delta v - \Delta m V = 0. \]

(17.25)

Note that all dependence on the speed of the rocket \(v\) with reference to an arbitrary inertial observer has canceled. This is so since we could have started by writing that the change of the momentum of the rocket in the rocket rest frame is the momentum that is in the ejected material as measured in rocket rest frame.

The integral form of Eq. (17.25)

\[ \int dv + V \int d\ln m = 0 \]

(17.26)

yields the ‘Tsiolkovsky’ speed-mass relationship,

\[ v - v_0 = V \ln \left( \frac{m_0}{m(v)} \right), \quad m(v) = m_0 \exp \left( -\frac{v - v_0}{V} \right). \]

(17.27)
We see that for a required speed, e.g. escape speed to orbit, the payload \( m \) decreases exponentially from the initial mass \( m_0 \) for a given \( V \ll v \). Any increase in \( V \) considering the required speed \( v \) leads to a major change in payload fraction that can be retained.

**RELATIVISTIC ROCKET EQUATION**

To implement a correctly relativistic generalization of the rocket equation we must use both momentum and energy conservation and allow that use of energy consumes mass. This is so since we learned, see section [16.1], that any energy required to drive an engine must come from conversion of rest mass into energy and thus we burn rocket fuel, thereby driving the engine of the rocket. There are two mass losses – the first is due to mass ejection just described, and the second is due to the relativistic effect: chemical mass defect accumulated throughout the burning of rocket fuel. To account for momentum conservation we substitute in Eq. (17.22) \( v \to \sinh y_r = \gamma \beta \), where \( y_r \) is the rapidity of the rocket seen by some observer.

Just before mass ejection we have

\[
p_1/c = (m + \Delta m) \sinh y_r .
\]  
\( (17.28) \)

After mass ejection, we combine the two contributions just like in Eq. (17.23)

\[
p_2/c = \tilde{m} \sinh(y_r + \Delta y_r) + \Delta m \sinh y_{r \text{ex}} .
\]  
\( (17.29) \)

Here \( \tilde{m} < m \) since we need to ‘combust’ an additional mass of the spaceship to run the engine that ejected the mass \( \Delta m \). The additional mass change, i.e., combustion mass defect

\[
\delta m = m - \tilde{m}
\]  
\( (17.30) \)

appears immediately after the acceleration step is complete.

We introduced the rapidity of exhausted mass \( \sinh y_{r \text{ex}} \). For the special case that the exhaust material is ejected against the direction of motion we can use additivity of rapidity,

\[
y_{r \text{ex}} = y_r - y_r V .
\]  
\( (17.31) \)

Here \( y_{r \text{ex}} \) is the propellant rapidity in the rocket rest frame

\[
\sinh y_{r \text{ex}} \equiv \gamma_r V/c = \frac{V/c}{\sqrt{1 - (V/c)^2}} .
\]  
\( (17.32) \)

We proceed similarly with energy conservation. Just before mass ejection we have

\[
E_1/c^2 = (m + \Delta m) \cosh y_r .
\]  
\( (17.33) \)

After mass ejection, we combine two contributions to the energy, the rocket with increased rapidity and the residual exhaust matter

\[
E_2/c^2 = \tilde{m} \cosh(y_r + \Delta y_r) + \Delta m \cosh y_{r \text{ex}} .
\]  
\( (17.34) \)
The conservation of momentum requires \( p_2 = p_1 \) and the conservation of energy \( E_2 = E_1 \), thus we find

\[
p_2 = p_1 \rightarrow \tilde{m} \sinh(y_r + \Delta y_r) + \Delta m \sinh y_{r_{\text{ex}}} = (m + \Delta m) \sinh y_r , \tag{17.35}
\]

\[
E_2 = E_1 \rightarrow \tilde{m} \cosh(y_r + \Delta y_r) + \Delta m \cosh y_{r_{\text{ex}}} = (m + \Delta m) \cosh y_r . \tag{17.36}
\]

In order to isolate \( \tilde{m} \) we square both equations and evaluate the difference motivated by \( M^2 = E^2/c^4 - p^2/c^2 \) which results in the exact expression

\[
\tilde{m}^2 + \Delta m^2 + 2 \tilde{m} \Delta m \cosh(y_{r_{\mathcal{V}}} + \Delta y_r) = (m + \Delta m)^2 . \tag{17.37}
\]

This condition is, as it must be, exactly independent of the rapidity of the rocket, that is independent of the choice of the reference frame of the observer.

Expanding Eq. (17.37) and keeping terms up to linear order in small deviations we find using Eq. (17.30)

\[
\delta m + \Delta m = \Delta m \cosh y_{r_{\mathcal{V}}} . \tag{17.38}
\]

The non-relativistic limit of this condition is obtained by remembering that

\[
\cosh y_{r_{\mathcal{V}}} - 1 \equiv \gamma_{\mathcal{V}} - 1 = \frac{\gamma^2}{2c^2} , \tag{17.39}
\]

which leads to

\[
\delta m c^2 = \Delta m \frac{\gamma^2}{2} . \tag{17.40}
\]

showing that the mass defect energy is the kinetic energy of the expelled material.

We now recognize this is the meaning of the relativistic relation Eq. (17.38) as well.

We note that a ‘star drive’ that uses antimatter-matter pair annihilation emitting material products of annihilation operates in the condition \( \delta m > \Delta m \). Annihilation at rest in the center of momentum frame of the rocket is accompanied by the production, on average, of little more than 5 pions: \( \delta m/\Delta m \approx (2m_N - 5m_\pi)/5m_\pi \approx 1.7 \).

Earlier we tracked the mass of the nonrelativistic rocket in time. Let us now look at the relativistic analogue, the energy of the rocket as observed from the base

\[
E = m(t)c^2 \cosh y_r(t) . \tag{17.41}
\]

\( E \) comprises the remaining payload and its relativistic kinetic energy. Clearly this must substantially decrease as a function of time in consideration of the relativistic properties of the emitted propellant. Equation (17.38) provides us the total change in mass, thus we obtain

\[
dE = dmc^2 \cosh y_{r_{\mathcal{V}}} \cosh y_r + mc^2 dy_r \sinh y_r . \tag{17.42}
\]
Reintroducing the energy $E$ we thus have

$$\frac{dE}{E} = \frac{dm}{m} \cosh y_r \nu + dy_r \tanh y_r .$$  \hfill (17.43)

Here $dm$ is the nonrelativistic analogue to the negative mass change while $dy_r$ is the positive increase in rapidity.

We can combine Eq. (17.35) with Eq. (17.36) to eliminate $\tilde{m}$ by multiplying with $\cosh(y_r + \Delta y_r)$ and $\sinh(y_r + \Delta y_r)$, respectively and by taking the difference. This leads to the exact relation

$$\Delta m \sinh(\Delta y_r + y_r \nu) = (m + \Delta m) \sinh \Delta y_r ,$$  \hfill (17.44)

which provides a 2nd independent equation for the rocket constraint, $y_r$. For small values of $\Delta m$, $\Delta y_r$, Eq. (17.44) yields the relativistic analogue of the non-relativistic rocket equation Eq. (17.25)

$$m \Delta y_r - \Delta m \sinh y_r \nu = 0 ,$$  \hfill (17.45)

where again Eq. (17.21) applies. However, this equation only accounts for the direct loss of mass. The first rocket equation Eq. (17.38) describes the total mass loss, including the combusted mass.

Comparing Eq. (17.45) to the nonrelativistic Eq. (17.25) we recognize that the speed increment (in units of c) is replaced by rapidity increment while $\nu \rightarrow \sinh y_r \nu$, see Eq. (17.32). This replacement introduces an additional factor $\gamma \nu = 1/\sqrt{1 - \nu^2/c^2}$ which accounts for the momentum carried by the incremental mass ejection having relativistic content. The greater value $\sinh y_r \nu$ the engine achieves, the less mass needs to be ejected to achieve an increment in rapidity $\Delta y_r$.

We now use Eq. (17.45) to eliminate in Eq. (17.43) the quantity $dm = -\Delta m$.

$$\frac{dE}{E} = dy_r(- \coth y_r \nu + \tanh y_r) = -\frac{dy_r}{\nu/c} + \tanh y_r dy_r .$$  \hfill (17.46)

We recall that $dy_r \tanh y_r = d \ln \cosh y_r$. Thus Eq. (17.46) can be integrated to yield

$$\ln \left( \frac{E}{\cosh y_r} \right) = -\frac{y_r - y_r^0}{\nu/c} .$$  \hfill (17.47)

Note that after division by $\cosh y_r$ we are back to the study of the residual rest energy (or rest mass) content of the rocket in its proper frame of reference

$$\frac{E}{\cosh y_r} \equiv m(y_r) = m_0 \exp \left( -\frac{y_r - y_r^0}{\nu/c} \right) ,$$  \hfill (17.48)

which in fact is a generalization of the nonrelativistic result Eq. (17.27).
It is remarkable how a straightforward generalization of the rocket equation emerges. While we replace \( \frac{v}{c} \rightarrow y_r \) moving from non-relativistic to relativistic case, the exhaust speed remains as it was in the nonrelativistic case Eq. (17.27). Result Eq. (17.48) could be guessed as follows: the energy that a rocket retains starting in the frame of an observer with \( y_{r0} = 0 \) and with the initial energy content \( m_0c^2 \) is found by multiplying with the Lorentz factor \( \gamma = \cosh y_r \). Allowing for \( c^2 \) to change from mass to energy unit Eq. (17.48) becomes

\[
E = m_0c^2 \cosh y_r \exp \left( -\frac{y_r}{\sqrt{V/c}} \right) \rightarrow m_0c^2 \left( \frac{1}{2} + e^{-2y_r} \right).
\] (17.49)

In the limit of \( \frac{V}{c} \rightarrow 1 \) the original observer sees the rocket split into two halves – she sees how in very small steps, ‘half’ of rest energy is given to the expelled and combusted material. According to action=reaction principle the recoiling rocket retains half of original energy content for this observer, but with a vastly smaller fraction in its rest frame, as is seen now dividing Eq. (17.49) by \( \gamma = \cosh y_r \), which returns us to the main result Eq. (17.48).
Part VII

Collisions, Decays
Introductory remarks to Part VII

The kinematic tools needed to study the conversion of massive particles from one form into another, and the conversion of energy into matter and matter into energy are the main topic of this Part VII of the book.

The concept of the center of momentum (CM) reference frame is introduced by replacing the center of mass as the primary frame of reference of a multi-particle system, and by characterizing the Lorentz transformations into the CM-frame.

We consider in detail the decay of a quasi-stable particle into several lighter particles, and look at the situation in several relevant frames of reference. The energy balance of a particle decay is then considered in the CM-frame. Among several applications the kinematic variables, energy and momentum, products of the decaying body are characterized.

Another case of interest is the decay of a body in flight, with special attention given to the decay that produces one of the reaction products at rest in the frame of reference considered. Numerous exercises sharpen a skill in the use of relativistic variables in this context. One of the exercises shows that photons do not decay into two massive particles.

We then turn to the study of two body collisions, addressing both elastic scattering and inelastic scattering in which the conversion of particle kinetic energy into matter occurs. The latter topic is vast and thus has books dedicated to it. I will present here a few select key results as an introduction to this domain of relativistic physics. This discussion will proceed based on application of energy and momentum conservation, in following Part VIII we complete the discussion introducing the Mandelstam variables, see section 21.2.

In an inelastic reaction, masses of particles participating change and new particles can be produced; in an elastic collision, the masses of particles remain unchanged and no new particles are produced. Both production and annihilation of antimatter illustrate the technical details of inelastic collisions, while the Compton process is used to illustrate the kinematic conditions present in elastic collisions.

This section ends with a discussion of the particle scattering from a moving wall, where the relativistic wall motion with the Lorentz factor $\gamma$ generates a bounced particle with up to $4\gamma^2$ enhanced energy-momentum. Our presentation generalizes the remarks found in the Einstein 1905 relativity publication.

---

1Perhaps the most clear introduction to the topic is: R. Hagedorn, *Relativistic Kinematics* W. A. Benjamin, New York (1963) and(1973).
18 The Center of Momentum Frame (CM-Frame)

18.1 A preferred frame of reference

A good choice of a frame of reference is of considerable importance in the study of the properties of any physical system. This is especially true when one or more particles are relativistic. The idea of the center of mass system, which is so useful in non-relativistic mechanics, must be extended in the relativistic domain: considering that particles can be created (and destroyed), and their masses can experience mass defect due to binding, the ‘center of mass’ is not available to us without a further elaboration\(^2\) \(^3\). In the study of an assembly of particles we are led to consider in a first step the center of momentum frame (CM-frame) of reference.

The existence of such a reference system is assured by the law of the conservation of momentum, which has very deep experimental roots, and in the theoretical formulation was understood as result of translational spatial symmetry: said simply, we expect that the laws of physics are the same everywhere. The relation of this natural postulate to the momentum conservation law was presented by E. Noether\(^3\) and is part of the renown Noether theorem. Another aspect of this theorem is that if laws of physics are unmutable as function of time, there is also energy conservation. We rely on energy and momentum conservation as the foundational principles and recognize through the Noether theorem that the Laws of Physics are the same everywhere in time and space.

We want first to consider more systematically what is the mass of a body of which some internal parts are in motion. In fact almost any material body is when looked at in greater depth composed of smaller constituents that are in motion with respect to each other. When considering a system of many particles we have the total energy \(E\) of the system

\[
E = \sum_i E_i , \tag{18.1}
\]

and the total momentum \(\vec{P}\) of the system,

\[
\vec{P} = \sum_i \vec{p}_i . \tag{18.2}
\]

\(^2\)The concept of center of mass may arise in the context of covariant Hamiltonian formulation of a many body system, see section 25.3.

\(^3\)Amalie Emmy Noether (1882 – 1935) is to this day considered the most important woman in the history of theoretical physics and mathematics.
The rest frame for a system comprised of many moving bodies is the CM-frame. This is the reference frame where all the momentum vectors of individual particles \(i\) which make up the object of interest sum to zero,

\[
\bar{P} = 0 = \sum_{i} \bar{p}_{i},
\]

where the over-bar denotes in this book that this quantity is evaluated in the CM-frame.

Each of the contributing particles is in motion, so the total energy is obtained by summing over the energy of all particles,

\[
E = \sum_{i} E_{i}.
\]

This formula is strictly true only for a gas of non-interacting particles one usually calls dust, since there is no allowance made for their mutual interactions. We have already seen that the presence of a potential describing such mutual interactions contributes to the energy and mass content of the body. For many physics considerations the motion of a body within the CM-frame is the dominant contribution to the energy.

While each of the contributing components in Eq. (18.4) is due to the motion of an individual particle, the sum of all is at rest. Therefore we also know the mass of the body that comprises all these particles;

\[
M = \frac{E}{c^{2}} > 0.
\]

We did not highlight by an over-bar the mass \(M\) of the composite system as being measured in CM-frame. The mass of a body does not depend on the frame of reference in which it is observed and only in the frame in which the body is at rest is Eq. (18.5) valid. That is by definition the CM-frame.

We can evaluate the mass \(M\) in any frame of reference. This is done by introducing an expression analogous to Eq. (17.12) which is a Lorentz invariant\(^{4}\)

\[
M^{2}c^{4} = E^{2} - \bar{P}^{2}c^{2} \equiv s.
\]

In Eq. (17.12) we considered \(m\) to be the intrinsic (rest) mass of a particle. \(M\) in Eq. (18.6) is the definition of the mass of a system of particles, and these need not be bound particles. It is common in the latter case to denote the resultant CM-frame energy \(Mc^{2}\) by the symbol \(\sqrt{s}\) as indicated in Eq. (18.6). We will

---

\(^{4}\)In Eq. (18.6) we introduce a ‘Mandelstam’ variable denoted by \(s\) which is the same letter symbol as used elsewhere in this book for the invariant event separation. For historical reasons we adhere to this usually unambiguous in the context nomenclature.
characterize this variable further in section [19.3] and place it in a more general context in section [21.2].

We note that Eq. (18.4) makes good sense for the ‘trivial’ system of a single particle. In this case there is no sum and we find that the CM-frame is the particle rest frame where \( \vec{p}_i = 0 \). The mass is given by \( M = E_{i=1}/c^2 = m_{i=1} \), where \( E_{i=1} \) is the energy of the particle in its rest frame. For the case of two particles, for the momentum in the CM-frame to vanish, particles must collide head-on; see the following exercise [VII–8] for the case of equal mass colliding objects.

The more general case of two colliding objects of different mass will be explored in section [19.3]. However, in the following we consider the available two-particle system energy in collisions on a target at rest in the laboratory and compare to the collider mode where the two particles approach from opposite directions, and the CM-frame is the laboratory-frame. We will also explore the Lorentz transformation between the two systems, CM and laboratory-frame when both differ.

### 18.2 The Lorentz transformation to the CM-frame

We select a coordinate system such that the three-momentum \( \vec{P} \) of the system in the lab frame is parallel to the \( x \)-axis, \( \vec{P} = P_x\hat{x} \). With a boost to a frame moving at dimensionless velocity \( \beta \) also along the \( x \)-axis, the transformed energy and momentum are:

\[
\begin{align*}
\bar{E} &= \gamma (E - \beta P_x c), \\
\bar{P}_x &= \gamma (P_x - \beta E), \\
\bar{P}_y &= P_y = 0, \\
\bar{P}_z &= P_z = 0.
\end{align*}
\]  

(18.7)

Since we are transforming to the CM-frame, we set \( \bar{P}_x = 0 \) in Eq. (18.7). We can now determine the dimensionless velocity \( \beta \) of the CM-frame system relative to the laboratory system, which we label with the subscript ‘s’ to indicate that it refers to a transformation to the CM-frame.

\[
\beta_{CM} = \frac{P_x c}{E}.
\]

(18.8)

We obtain also \( \gamma_s \):

\[
\gamma_{CM} = \frac{1}{\sqrt{1 - \beta_{CM}^2}} = \frac{E}{\sqrt{E^2 - P^2 c^2}},
\]

which simplifies with Eq. (18.6) to

\[
\gamma_{CM} = \frac{E}{Mc^2}.
\]

(18.9)
Note that quantities with subscript ‘CM’ describe the properties of the CM-frame of reference, in particular its motion, as observed in the reference frame in which all the other quantities are observed. Therefore we cannot characterize these with a ‘bar’. By definition $\beta_{\text{CM}} = 0$ and $\gamma_{\text{CM}} = 1$.

Exercise VII–1: Transformation to CM-frame in two-particle collisions

Obtain the Lorentz transformation parameters from the lab system $S$ to CM-frame $S'$ for the LHC beam dump event, and consider the physical relevance of such collisions.

Solution

The $E_1 = 7000$ GeV LHC proton beam collides with a nucleon in the wall, thus $E_2 = mc^2 = 0.94$ GeV, and $m_1 = m_2 = m$.

Note that to reasonable precision the energy and momentum of the beam have the same value,

1. $p_1c = \sqrt{E_1^2 - m_1^2c^4} = \sqrt{7000^2 - 0.94^2}$ GeV = 6999.999937 GeV,

yet as will be seen, it is often better to keep all terms to the end when deriving equations as significant cancellation occurs.

The (high) CM-frame system velocity in the lab frame is:

2. $\beta_{\text{CM}} = \frac{p_1c + p_2c}{E_1 + E_2} = \sqrt{\frac{E_1^2 - m_1^2c^4}{E_1^2 + m_2^2c^4}} = \sqrt{\frac{E_1 - mc^2}{E_1 + mc^2}}$,

where we used in the last equality the fact that projectile and target have the same mass $m$. In this case $\sqrt{E_1 + mc^2}$ between numerator and denominator cancel. We expand in powers of $m/E_1$:

3. $\beta_{\text{CM}} = \sqrt{\frac{E_1 - mc^2}{E_1 + mc^2}} \approx 1 - \frac{mc^2}{E_1} \approx \frac{7000 - 0.94}{7000} = 0.9998657$.

Given the simplicity of the analytic result Eq. (2) we compute the Lorentz factor $\gamma_{\text{CM}}$ directly

4. $\gamma_{\text{CM}} = \frac{1}{\sqrt{1 - \beta_{\text{CM}}^2}} = \sqrt{\frac{E_1 + mc^2}{2mc^2}} = \sqrt{\frac{7000.94}{2 \cdot 0.94}} \simeq 61.02$.

To test for consistency, we can cross-check using Eq. (18.9) and we find for the case of collision with a particle at rest in laboratory

5. $\gamma = \frac{E}{\sqrt{s}} = \frac{E_1 + m_2c^2}{\sqrt{(m_1c^2)^2 + (m_2c^2)^2 + 2m_2c^2 E_2}} = \frac{7000.94}{\sqrt{2 \cdot 0.94 \cdot 7000.94}} = 61.02$. 
When we dump protons from the LHC into a specially designated wall at rest in the laboratory, we find according to Eq. (4) that in the CM system we have 57 GeV/proton head-on collisions, thus for the pair of projectile-target nucleon \( \sqrt{s} = 114 \) GeV. This is only 0.8% of the energy available in LHC collider mode for primary physical processes such as particle production. Though this is a small energy compared to the LHC collider mode, and is in fact in the middle of the RHIC energy range, the high velocity of the CM-frame offers certain experimental advantages in terms of potential for detection of unstable particles. A fixed target LHC experiment involving a gas jet target has been realized in the context of the LHC-B experimental program.

---

Exercise VII–2: CM rapidity of colliding particles

Obtain CM-rapidity in terms of kinematic variables of two colliding particles and for collinear motion describe how CM-rapidity depends on colliding particle rapidities and masses.

Solution

The value of CM-rapidity can be obtained using Eq. (18.8)

1. \( \tanh y_{\text{CM}} \equiv \beta_{\text{CM}} = \frac{P_{||}}{E} = \frac{cp_{||1} + cp_{||2}}{E_{1} + E_{2}} \),

which leads to

2. \( y_{\text{CM}} = \frac{1}{2} \ln \left( \frac{E_{1} + E_{2} + cp_{||1} + cp_{||2}}{E_{1} + E_{2} - cp_{||1} - cp_{||2}} \right) \).

We can use the particle rapidity relations Eq. (17.15) to write

3. \( E_{1,2} \pm cp_{||1,2} = E_{1,2} e^{\pm y_{1,2}} \)

which allows to write for Eq. (2)

4. \( y_{\text{CM}} = \frac{1}{2} \ln \left( \frac{E_{1} e^{y_{1}} + E_{2} e^{y_{2}}}{E_{1} e^{-y_{1}} + E_{2} e^{-y_{2}}} \right) \).

We take in the numerator the factor \( e^{y_{1}} \) and in the denominator the factor \( e^{-y_{2}} \) out of the logarithm and obtain:

5. \( y_{\text{CM}} = \frac{y_{1} + y_{2}}{2} + \frac{1}{2} \ln \left( \frac{E_{1} + E_{2} e^{-(y_{1} - y_{2})}}{E_{1} + E_{2} e^{-(y_{1} - y_{2})}} \right) \).
We note that this expression is symmetric under renaming $1 \leftrightarrow 2$. In most cases of interest, including the collider mode we have that the rapidity gap between two relativistic colliding particles is sufficiently large $y_1 - y_2 \gg 0$ so that the exponential function in Eq. 5 is very small allowing the expansion

$$
y_{\text{CM}} \simeq \frac{y_1 + y_2}{2} + \frac{1}{2} \ln \left( \frac{m_1}{m_2} \right) + \frac{m_2^2 - m_1^2}{2m_1m_2} e^{-(y_1 - y_2)} + \cdots ,
$$

For collinear head-on particle collision we have already reset $E_{\perp 1,2} \to m_{1,2}$ in Eq. 6. When $m_1 = m_2$ mass-dependent terms vanish, and we recognize as CM-rapidity the average of both colliding particle rapidities.

We can use the additivity of rapidity, see section 7.5, to relate the laboratory and CM-frame particle rapidities,

$$y_1 = \bar{y}_1 + y_{\text{CM}} , \quad y_2 = \bar{y}_2 + y_{\text{CM}} ,$$

Combining with the result Eq. 5 we find

$$\bar{y}_1 = \frac{y_1 - y_2}{2} - \frac{1}{2} \ln \left( \frac{m_1 + m_2 e^{-(y_1 - y_2)}}{m_2 + m_1 e^{-(y_1 - y_2)}} \right) ,$$

and using $1 \leftrightarrow 2$ symmetry a similar result follows for $\bar{y}_2$.

End VII-2: CM rapidity of colliding particles

18.3 Decay of a body in CM-frame

When a heavy nucleus undergoes spontaneous or induced fission, the resulting fragments carry usable kinetic energy. This is a special case of a family of reactions called particle, and more generally, body decays. We address here the two-body decay processes, in which a mother particle of mass $M$ decays into two daughter particles with mass $m_1$ and $m_2$, illustrated in figure 18-1.
We study here the decay in the rest frame of the mother particle, which, due to conservation of momentum, must also be the CM-frame of all (two in present example) daughter particles. Conservation of energy and momentum in this reference system can be expressed as

\[ M c^2 = E_1 + E_2, \]  
\[ \vec{0} = \vec{p}_1 + \vec{p}_2. \]

Momentum conservation thus requires that the daughter particles move in opposite directions with the same momentum, i.e. that the rest frame of the mother particle is also the CM-frame of the daughter particles,

\[ \vec{p}_2 = -\vec{p}_1 \]  
and therefore naturally also

\[ |\vec{p}_2| = |\vec{p}_1|. \]

The conservation of energy as stated in Eq. (18.10) implies

\[ M = \sqrt{m_1^2 + \left(\frac{p_1}{c}\right)^2} + \sqrt{m_2^2 + \left(\frac{p_1}{c}\right)^2} > m_1 + m_2. \]  

The inequality reminds us that for the mother particle to decay, the daughter particles must be less massive; the balance of energy is found in their motion. We now solve Eq. (18.14) for the only unknown \( |\vec{p}_1| \). These manipulations are straightforward but nontrivial. We use the trick of placing one of the roots alone on one side of the expression. In this way Eq. (18.14) yields

\[ \left( M - \sqrt{m_1^2 + \left(\frac{p_1}{c}\right)^2} \right)^2 = m_2^2 + \left(\frac{p_1}{c}\right)^2, \]

and simplifying:

\[ M^2 - 2M \sqrt{m_1^2 + \left(\frac{p_1}{c}\right)^2} + m_1^2 = m_2^2. \]

We obtain

\[ \vec{p}_1^2 = c^2 \left[ \left( \frac{M^2 + m_1^2 - m_2^2}{4M} \right)^2 - m_1^2 \right], \]

and so

\[ \vec{p}_1^2 = c^2 \left( \frac{M^4 + m_1^4 + m_2^4 - 2M^2m_1^2 - 2M^2m_2^2 - 2m_1^2m_2^2}{4M^2} \right). \]

The energy of each of the daughter particles follows from Eq. (18.15)

\[ \frac{E_1}{c^2} = \sqrt{m_1^2 + \left(\frac{p_1}{c}\right)^2} = \frac{M^2 + m_1^2 - m_2^2}{2M}, \]
and
\[
\frac{E_2}{c^2} = \sqrt{m_2^2 + \left(\frac{p_1}{c}\right)^2} = \frac{M^2 + m_2^2 - m_1^2}{2M}.
\]
(18.18)

Note that the solution checks; that is \(E_1 + E_2 = Mc^2\), and the expressions are symmetric under the exchange of the names, 1 \(↔\) 2.

Exercise VII–3: Momentum of \(\pi^-\) in \(\Lambda\)-decay

\(\Lambda\) is a neutral elementary particle of mass \(m_\Lambda = 1,115.7\) MeV/\(c^2\), which decays (among other possibilities, and with about 2/3 probability) into a proton \(p\) (mass \(m_p = 938.3\) MeV/\(c^2\)) and a negatively charged pion, \(\pi^-\) (mass \(m_{\pi^-} = 139.6\) MeV/\(c^2\)). Show that the momentum of \(\pi^-\) produced in the decay \(\Lambda \rightarrow \pi^- + p\) is exactly 101 MeV/\(c\) when measured in the rest frame of \(\Lambda\), and compare this value to the momentum of the produced proton.

Solution

In the rest frame of \(\Lambda\) the decay products must emerge with opposite momenta so that momentum is conserved

1 \(\vec{p}_{\pi^-} + \vec{p}_p = \vec{p}_\Lambda = 0\).

Thus the momentum of decay products will be equal in magnitude and opposite in direction. To obtain the momentum magnitude we can solve either Eq. (18.17) or Eq. (18.18) for \(\vec{p}_{\pi^-} = |\vec{p}_{\pi^-}| = |\vec{p}_p|\). Both approaches result after elementary algebra in

2 \(p^2 = c^2\left[m_\Lambda^4 + m_p^4 + m_{\pi^-}^4 - 2m_\Lambda^2(m_p^2 + m_{\pi^-}^2) - 2m_p^2m_{\pi^-}^2\right] \frac{4m_\Lambda^2}{4m_\Lambda^2}\).

When evaluating this quantity there is a difficulty: one easily sees that

3 \(m_p^2 + m_{\pi^-}^2 = 938.3 + 139.6 = 1077.9 = m_\Lambda c^2 - 37.8\) MeV.

This near balance of masses signals considerable cancellation of large numbers in brute force computation. A more palatable expression of Eq. 2 appropriate in such a situation is

4 \(p^2 = c^2\left[\frac{m_\Lambda^2 - (m_p + m_{\pi^-})^2}{m_\Lambda^2}\right] \frac{4m_\Lambda^2}{4m_\Lambda^2}\).

Like Eq. 2, Eq. 4 is also symmetric under exchange of subscripts \(p \leftrightarrow \pi\).

We note that the first factor in the numerator of Eq. 4 can also be written

5 \(m_\Lambda^2 - (m_p + m_{\pi^-})^2 = [m_\Lambda - (m_p + m_{\pi^-})][m_\Lambda + (m_p + m_{\pi^-})]\),
allowing factorization of the small number, \( m_{\Lambda} - (m_p + m_\pi) = 37.8 \text{ MeV} \) and a straight-forward computation now succeeds.

---

**End VII–3: Momentum of \( \pi^- \) in \( \Lambda \)-decay**

### 18.4 Decay energy balance in CM-frame

To see how much energy is freed in the decay we evaluate the kinetic energy \( T_i = E_i - m_i c^2 \) for both particles \( i = 1, 2 \).

\[
T_1 = E_1 - m_1 c^2 = \frac{(M - m_1)^2 - m_2^2 c^2}{2M}
\]

and

\[
T_2 = E_2 - m_2 c^2 = \frac{(M - m_2)^2 - m_1^2 c^2}{2M}.
\]

Thus the total freed kinetic energy is:

\[
\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2 = (M - m_1 - m_2)c^2,
\]

which is what we expect: the available ‘motion’ energy is the energy of the mother particle, minus the energy locked in the mass of daughter particles. This is also a correct statement in case the final state has more than 2 particles, and Eq. (18.21) can be generalized accordingly for decays observed in the rest frame of the mother particle.

Many satellites, in particular those going far and away from the Sun use radioactive decay “batteries” which use this principle. In such an energy source, a radioactive slowly decaying isotope is employed, and the decay products are stopped e.g. by a lead mantle. The heat released in the decay process is used to both warm the electronics and to power a generator. A bit more sophisticated is the decay reaction induced in a nuclear fission reactor. Here the impact of a thermal neutron induces nuclear fission. The energy freed in the reaction is also described by Eq. (18.21).

Clearly a spontaneous (i.e., on its own) decay can only occur if the energy released is positive. Conversely, energy is gained from the binding of several components into fewer and less massive particles; the energy gain is the mass defect; that is the reduction of the total mass \( M \) due to binding of e.g. two components, \( m_1, m_2 \):

\[
(m_1 + m_2 - M)c^2 = \delta E.
\]

The reaction proceeds if the binding energy can be released, e.g. by radiation of a photon, or as is the case with many chemical reactions, recoil on a catalyst surface.

We refer for further details to our prior discussion in section 16 on page 203.
18.5 Decay of a body in flight

The above considerations yielded properties of decay particles in the rest frame of the mother particle. In the more general case the mother particle is in motion in the laboratory-frame where it has the energy

\[ E = \sqrt{M^2 c^4 + P^2 c^2} \]  \tag{18.23}

and momentum \( P \). The transformation from the lab frame to the CM-frame is given by Eq. (18.8) and Eq. (18.9). The inverse transformation from the CM-frame to the lab frame is then given here with parameters

\[ \gamma = \frac{E}{Mc^2}, \quad \beta = \frac{-\vec{P} c}{E}, \quad \gamma \beta = \frac{-\vec{P}}{Mc}. \]  \tag{18.24}

Decomposing the momenta of the daughter particles into components parallel (\( \parallel \)) and perpendicular (\( \perp \)) to \( \vec{P} \), we apply the Lorentz transformation of Eq. (18.7) with Eq. (18.24) to find the lab frame properties of both the daughter particles \( (i = 1, 2) \),

\[ E_i = \frac{1}{M} \left( \frac{E}{c^2} E_i + P \vec{p}_i \parallel \right), \]  \tag{18.25}

\[ p_i \parallel = \frac{1}{M} \left( \frac{E}{c^2} \vec{p}_{i \parallel} + \frac{P}{c^2} E_i \right), \]  \tag{18.26}

\[ \vec{p}_{1 \perp} = -\vec{p}_{2 \perp} = \vec{p}_\perp. \]  \tag{18.27}

In an experiment we usually measure the laboratory momenta and energies of particles produced and would like to obtain the CM values for better understanding of the reaction process. We find this result with help of inverse transformation or by solution of Eq. (18.25), Eq. (18.26); both procedures lead to the same result.

We multiply Eq. (18.25) with \( E \) and Eq. (18.26) with \( P \) and the difference of the two reads:

\[ \vec{E}_i = \frac{1}{M} \left( \frac{E}{c^2} \vec{E}_i - P \vec{p}_{i \parallel} \right), \]  \tag{18.28}

Note that if we use Eq. (18.8) and Eq. (18.9) to replace \( E/M \) and \( P/M \) in Eq. (18.28), this equation is confirmed as being the Lorentz transform from lab to CM-frame replacing \( \beta \) by \( -\beta \) in the inverse transformation.

Similarly, we can use Eq. (18.28) in Eq. (18.26) along with Eq. (18.23) and obtain

\[ \vec{p}_i \parallel = \frac{1}{M} \left( \frac{E}{c^2} p_{i \parallel} - \frac{P}{c^2} E_i \right). \]  \tag{18.29}

This is also easily recognized as a transformation to CM-frame replacing \( \beta \) by \( -\beta \) in the inverse transformation from lab to CM-frame.
Because of energy-momentum conservation in the lab frame we also have
\[ P = p_1 + p_2, \quad E = E_1 + E_2, \quad M c^2 = \sqrt{E^2 - P^2 c^2} \]
which we can insert to obtain the CM-frame daughter particle properties in terms of laboratory measured daughter particles momenta \( p_1, p_2 \).

**Exercise VII–4: Decay of the \( \eta \)-particle**

About 40\% of \( \eta \)-particle decays result in the production of two \( \gamma \) particles. Obtain the energy and momenta of the \( \gamma \) particles both in the CM-frame, and for an \( \eta \) having a momentum of 1 GeV/c in the direction of the motion of one of the decay photons. Note \( M\eta c^2 = 0.5475 \) GeV.

**Solution**

The decay process in the CM-frame divides the \( \eta \)-mass into two massless photons and thus we have for each

1. \( \bar{p}_{1,2} = E_{1,2} = 0.5M\eta c^2 = 0.274 \) GeV ,

with the photons moving in opposite directions. In the following we will use lower index \( \gamma \) to denote the decay particles, not to be mixed up with the Lorentz-factor.

We now consider a boost to a momentum of 1 GeV/c for the \( \eta \). The energy of the \( \eta \) is now

2. \( E_\eta = \sqrt{M^2 \eta c^4 + P^2 \eta c^2} = \sqrt{0.5475^2 + 1^2} \) GeV = 1.1401 GeV .

We also have

3. \( \frac{P_\eta}{M_\eta c^2} = \gamma \beta = 1.8265, \quad \frac{E_\eta}{M_\eta c^2} = \gamma = 2.0823, \quad \frac{P_\eta c}{E_\eta} = \beta = 0.8771 \).

We use this input in Eq. (18.25) which we write as:

4. \( E_{\gamma \pm} = (\gamma E_\gamma \pm \gamma \beta \bar{p}_\gamma c) = 2.0823E_\gamma \pm 1.8265\bar{p}_\gamma c \),

where the \( \pm \) refers to the photon being parallel and, respectively anti-parallel with reference to the motion of the \( \eta \) particle.

Remembering that each photon carries in CM-frame half of the \( \eta \) energy we obtain inserting the numerical values

5. \( E_{\gamma +} = (2.0823 + 1.8265)0.274 = 1.0710 \) GeV ,

6. \( E_{\gamma -} = (2.0823 - 1.8265)0.274 = 0.0701 \) GeV .
The energy conservation checks, \( E_{\gamma^+} + E_{\gamma^-} = 1.0710 + 0.0701 = 1.1401 \text{ GeV} \), which is the energy of the \( \eta \) in the lab frame. Since the decay particles are massless, \( E_{\gamma^\pm} = p_{\gamma^\pm} c \), thus up to the factor \( c \) the result is valid also for the momenta of the photons.

As we see the two decay photon energies are now very different; in fact one is nearly zero as measured on the scale of the total energy. This shows that a suitable frame of reference of the mother particle results in one of the daughter photons being produced with very little energy in the lab frame.

--- End VII–4: Decay of the \( \eta \)-particle ---

--- Exercise VII–5: The magic momentum ---

Sometimes we would like to take advantage of the decay process to study a daughter particle in the laboratory, in which case we would like to form it at rest in the lab. The question arises how to choose the energy-momentum of the mother particle so that this happens. It turns out that it is necessary for the mother particle to have a special momentum \( P_m \), which we call the magic momentum.

Determine \( P_m \) in general for a two-body decay process, where we want the daughter particle with mass \( m_1 \) to be at rest in the laboratory-frame. Then calculate the exact value of \( P_m \) for the example decay of

\[ \pi^- \rightarrow \mu^- + \nu, \quad M = 139.57 \text{ MeV}, \quad m_{1=\mu} = 105.66 \text{ MeV}, \quad m_{2=\nu} \simeq 0. \]

**Solution**

When the mother particle has the required magic momentum \( P_m \), we arrange to have \( p_{1||} = 0 \). However, we cannot constrain the value of \( \vec{p}_\perp \) as it has no dependence on momentum \( P \) of the mother particle, see Eq. (18.27), and thus we present the case solution for \( \vec{p}_\perp = \vec{0} \), a condition that can happen by chance.

We then need only set \( p_{1||} = 0 \) in Eq. (18.26) to obtain for the magnitude of the magic momentum:

\[ \frac{E_m}{P_m} = \frac{E_1}{p_1}. \]

Squaring both sides and simplifying by using the relativistic energy relationship Eq. (17.8) and calling the particle momentum (the CM-frame momentum) the magic momentum \( P_m \) we find

\[ P_m^2 = \frac{M^2}{m_1} \vec{p}_1^2. \]

As we have already calculated the CM-frame quantities, we insert \( \vec{p}_1^2 \) from Eq. (18.16), yielding

\[ P_m^2 = c^2 \left( \frac{M^4 + m_1^4 + m_2^4 - 2M^2m_1^2 - 2M^2m_2^2 - 2m_1^2m_2^2}{4m_1^2} \right). \]
Now we can calculate the magic momentum for specific reaction mentioned above,
\[ \pi^- \rightarrow \mu^- + \nu \ . \]

In this case \( M = m_\pi \), \( m_1 = m_\mu \), and \( m_2 = m_\nu = 0 \). With Eq. 3 we get for the magic momentum of the pion
\[
P_m^2 = c^2 \left( \frac{m_\pi^4 + m_\mu^4 - 2m_\pi^2 m_\mu^2}{4m_\mu^2} \right) ,
\]
therefore
\[
P_m = c \left( \frac{m_\pi^2 - m_\mu^2}{2m_\mu} \right) = 39.4 \text{ MeV/c} .
\]

We can also calculate explicitly the dimensionless velocity \( \beta \) of the pion and the Lorentz factor
\[
\beta = \frac{P_m c}{\sqrt{P_m^2 c^2 + m_\pi^2 c^4}} = 0.27 , \quad \gamma = \frac{\sqrt{P_m^2 c^2 + m_\pi^2 c^4}}{M c^2} = 1.04 .
\]

---

**Exercise VII–6: Are photon decay reactions possible?**

Show that a single photon cannot decay into an electron-positron pair and explain how photon pair-‘conversion’ detectors work despite this.

**Solution**

Assume that the following reaction
\[ \gamma \rightarrow e^+ + e^- \]
can occur, see figure [18-2] on left. Then the sum of the energy and momentum of the electron-positron pair must equal the energy and momentum of the photon:
\[ E_\gamma = E_{e^+} + E_{e^-} , \]
\[ \vec{p}_\gamma = \vec{p}_{e^+} + \vec{p}_{e^-} . \]

For the invariant of the photon’s energy-momentum we have
\[ E_\gamma^2 - \vec{p}_\gamma^2 c^2 = m_\gamma^2 c^4 . \]
Since the photon is massless we obtain

$$E_{\gamma}^2 - \vec{p}_{\gamma}^2c^2 = 0 .$$

The energy and momentum of the photon are given by

$$E_{\gamma} = E_{e^+} + E_{e^-}, \quad \vec{p}_{\gamma} = \vec{p}_{e^+} + \vec{p}_{e^-} .$$

Eq. 4 becomes, up to overall factor 2

$$m^2c^2 + \frac{E_{e^+}E_{e^-}}{c^2} - \vec{p}_{e^+} \cdot \vec{p}_{e^-} = 0 ,$$

where $m = m_{e^+} = m_{e^-}$ is the mass of an electron or positron. Using the relativistic energy equation Eq. (17.8) we have

$$m^2c^2 + \sqrt{\vec{p}_{e^+}^2 + m^2c^2} \sqrt{\vec{p}_{e^-}^2 + m^2c^2} - \vec{p}_{e^+} \cdot \vec{p}_{e^-} = 0 .$$

Since $m > 0$, with the Cauchy-Schwartz inequality the left side of Eq. 7 cannot equal zero. Thus our assumption must be false; the decay of a free photon in the vacuum into an electron-positron pair is not possible.

This result is easily generalized to a theorem: a particle of mass $M$ can decay into several (at least two) particles of mass $m_i$ if and only if $M \geq \sum_i m_i$. This we can recognize without any computation by going to the rest frame of the particle $M$. There, the decaying particle has energy $Mc^2$ and the decay products must also have this energy considering energy conservation. The condition we stated corresponds to the ‘threshold’; that is, all decay products are at rest, and have no kinetic energy. The effort we made above to resolve for the special case $M \to 0$ of a photon derives from the circumstance that there is no rest frame for a massless particle.

Reality is more complex – in practice, the photon ‘converts’ into a $e^+e^-$-pair. In the conversion process illustrated on the right in figure 18-2 there is another body participating in the interaction. Specifically, when $\gamma$ is in the proximity of a nucleus $N$.
and this nucleus participates in the reaction, the presence of the nucleus fixes a special CM-frame of reference, which is nearly the rest frame of the nucleus. Thus the reaction we consider is the collision of $\gamma$ with the nucleus

$$\gamma + \text{nucleus at rest} \rightarrow e^+ + e^- + (\text{moving Nucleus, momentum}=P_N).$$

Our condition Eq. 6 is now supplemented by two additional terms on the right, related to the nucleus:

$$8\ 2(\ldots) = 2 \vec{P}_N \cdot (\vec{p}_{e^-} + \vec{p}_{e^+}) - \frac{P_N^2}{2M_N} (E_{e^+} + E_{e^-}).$$

Normally the last term, the nucleus recoil energy $E_N = P_N^2/2M_N$ is negligible in consideration of the large mass of the nucleus. Still, there is the other term which must be considered, and when we look at the mathematical relations we will see that the reaction $\gamma \rightarrow e^+ + e^-$ is not anymore forbidden. The energy of the photon now measured in the rest frame of the nucleus is found in comparison to the light $e^+e^-$-pair; see figure 18-2 on the right.

The reaction Eq. 8 is of practical importance, being employed in detectors.

---

19 Particle Collisions

19.1 Two body reactions

We begin with inelastic reactions in which energy can be converted into particle rest mass and vice versa, as we have already seen considering decay reactions. In these cases, some of the physical characteristics of the participating particles change in the interaction. This change can be relatively minor, for example, the excitation of internal dynamics of a nucleus or an atom. However, more often in relativistic collisions we see a much more radical modification, such as a breakup of particles into more elementary constituents or formation of new particles, such as antiprotons in proton-proton collisions, an example we will present.

After we turn to elastic reactions in which all physical characteristics of the interacting particles are unchanged. We establish what this means for the process of scattering. Examples we consider in greater depth are a) Compton scattering (photon electron scattering) and b) scattering from a moving brick wall.

We indicate values of particle properties after the collision with a prime; for example $E_1$ refers to the energy of particle ‘1’ before the collision, and $E_1'$ refers to the energy of this particle after the collision. As in the previous section, we will identify values in the center of momentum system (CM-frame) through
the use of an over-bar; for example \( \vec{p}_1 \) refers to the momentum of particle ‘1’ before the collision in the CM-frame, and \( \vec{p}_1' \) in the CM-frame after the collision. These variables can be seen in figure [19-1] showing the scattering plane normal to the angular momentum \( \vec{L} \) – conservation of angular momentum assures that a scattering plane exists. In scattering where ‘action at a distance’ occurs, such as scattering due to electromagnetic force, the impact parameter \( b \) is important in characterizing outcome of the collision; its definition is shown in figure [19-1].

### 19.2 Inelastic reaction threshold

A ‘practical’ example of an inelastic reaction is a nuclear fusion reaction in which two heavy hydrogen isotopes (deuteron \( d = pn \), and triton \( t = pnn \)) react into an \( \alpha \)-particle (\( ppnn \)), and a neutron \( n \) – here \( p \) denotes a proton:

\[
d + t \rightarrow \alpha + n + 17.6 \text{ MeV}.
\]

While this reaction is exothermic (it releases \( Q = 17.6 \text{MeV} \) kinetic energy as indicated by the ‘+’ sign) many two-body reactions are endothermic; that is, energy needs to be supplied in order for the reaction to proceed. In such a case one often indicates the amount of threshold energy required with a ‘-’ sign.

In the field of elementary particle physics the production of new particles, in general, consumes energy due to the conversion of kinetic collision energy into the rest energy of newly produced particles. Consider as an example the production of two particles called ‘strange’, \( \Lambda \) and \( K^0 \)

\[
m_\Lambda = 1115.68 \text{ MeV/c}^2, \quad m_{K^0} = 497.65 \text{ MeV/c}^2.
\]

The new physics particles \( \Lambda, K^0 \) were called ‘strange’ mid-20th century because of their unusual properties and the name stuck. The ‘strangeness’ producing
reaction is strongly endothermic

\[ \pi^- + p \rightarrow \Lambda + K^0 - 535.5 \text{ MeV}, \]  

as shown by the negative sign of the \( Q \)-value, the last term in Eq. (19.1). We compute the energy threshold, that is required energy from the masses of the participating particles

\[ Q = m_p + m_{\pi^-} - m_\Lambda - m_{K^0}, \]

where the incoming particle masses are:

\[ m_{\pi^-} = 139.57 \text{ MeV/c}^2, \quad m_p = 938.27 \text{ MeV/c}^2. \]

This reaction and other endothermic reactions obviously require that the reactants have a kinetic energy greater than some minimum, below which there is not enough energy available to convert to the rest mass of the reaction products. The threshold energy that is shown is the energy amount needed and not what kinetic energy a projectile needs to have to make the reaction happen, which is often a much greater value.

We now determine the minimum kinetic energy and momentum of the reactants as follows. In the CM-frame of reference, the mass \( M \) of the system is related to the energy \( E \) by \( E = Mc^2 \). The Lorentz transformation equation out of the CM-frame to any other frame reads

\[ E = \gamma E = \gamma Mc^2. \]  

We see that minimizing the required mass-energy equivalent \( Mc^2 \) of a system minimizes its energy in any other frame of reference. We therefore consider the process Eq. (19.1) in the CM-frame, where we have

\[ \vec{p}_{\pi^-} = -\vec{p}_p = \vec{p}. \]  

The total energy of participants is therefore

\[ E_T = E_{\pi^-} + E_p = \sqrt{m_{\pi^-}^2c^4 + \vec{p}_{\pi^-}^2c^2} + \sqrt{m_p^2c^4 + \vec{p}_p^2c^2}. \]  

After the collision, energy conservation requires that the final state particles must have the same total energy given by Eq. (19.4). Thus we have

\[ E_T = E_\Lambda + E_{K^0} = \sqrt{m_\Lambda^2c^4 + \vec{p}_\Lambda^2c^2} + \sqrt{m_{K^0}^2c^4 + \vec{p}_{K^0}^2c^2}. \]  

The minimum energy necessary for the endothermic reaction to proceed corresponds to the state in which the reaction products are at rest in the CM-frame system, in which case we have for Eq. (19.5)

\[ E_{T\text{min}} = m_\Lambda c^2 + m_{K^0} c^2. \]  

\[ (19.6) \]
From Eq. (19.4) and Eq. (19.6) we obtain an implicit equation for the minimum CM-frame particle momentum required for the reaction,

\[
\sqrt{m^2_{\pi^-}c^4 + p_{\min}^2c^2} + \sqrt{m^2_{p}c^4 + p_{\min}^2c^2} = (m_{\Lambda} + m_{K^0})c^2 .
\]  

(19.7)

We can now transform \( p_{\min} \) into the laboratory-frame where the unstable particle, the pion, moves towards a target proton which is at rest. The pion moves toward the proton with the minimum possible laboratory momentum, \( p_{\pi^-\min} \). We find the transformation by considering how to transform the proton at rest to the CM-frame where it has momentum \( \bar{p}_{\min} \):

\[
\bar{p}_{\min} = \gamma \beta m_pc = \frac{p_{\pi^-\min}}{M} m_p ,
\]  

(19.8)

where we have substituted for \( \gamma \) and \( \beta \) from Eq. (18.8) and Eq. (18.9). The invariant mass \( M \) of the system is given in terms of the reaction products, which are both at rest in the CM-frame:

\[
M = m_{\Lambda} + m_{K^0} \equiv M_+ .
\]  

(19.9)

With this Eq. (19.8) becomes

\[
p_{\min} = \frac{m_p}{M_+} p_{\pi^-\min} .
\]  

(19.10)

It is convenient for the following to measure the momentum of the pion \( p_{\pi^-\min} \) in units of \( m_{\pi^-}c \), i.e.

\[
p_{\pi^-\min} = f m_{\pi^-}c .
\]

Substituting Eq. (19.10) into Eq. (19.7) we obtain for \( f \) the following implicit equation:

\[
\sqrt{f^2 + \left( \frac{M_+}{m_p} \right)^2} + \sqrt{f^2 + \left( \frac{M_+}{m_{\pi^-}} \right)^2} = \frac{M_+^2}{m_pm_{\pi^-}} .
\]  

(19.11)

Note that Eq. (19.11) is symmetric when the names of particles in the initial and/or final state are exchanged.

We now solve Eq. (19.11) for \( f \). Moving one of the roots to the other side, and squaring, and later isolating the remaining root and squaring again we obtain

\[
4f^2 = \frac{m_p^2}{m_{\pi^-}^2} + \frac{m_p^2}{m_{\pi^-}^2} + \frac{M_+^4}{m_p^2m_{\pi^-}^2} - 2 \left( \frac{M_+^2}{m_{\pi^-}^2} + \frac{M_+^2}{m_p^2} + 1 \right) .
\]  

(19.12)

Again we observe that this expression is symmetric in both initial particles seen in \( M \), Eq. (19.9), and final state particles \( \Lambda \) and \( K^0 \). We then insert the numeric values for \( m_{\pi^-}, m_p, M_+ = m_{\Lambda} + m_{K^0} \) and find \( f \simeq 6.4 \) which easily checks with the defining equation

\[
\sqrt{f^2 + 1.7195^2} + \sqrt{f^2 + 11.56^2} = 19.88 .
\]
The pion must hit the stationary proton with a minimum momentum $p_{\pi^-_{\text{min}}} \approx 6m_\pi c \approx 890\text{MeV}/c$ for the reaction Eq. (19.1) to be possible.

---

**Exercise VII–7: Laboratory frame annihilation**

A positron with kinetic energy of $T_e$ collides with an electron at rest. An annihilation takes place, from which two photons emerge:

$$e^+ + e^- \rightarrow \gamma + \gamma .$$

One of the photons is observed to move in the direction of motion of the incident positron. In which direction does the other photon move? How great are the laboratory energies of the two photons?

**Solution**

As the momentum must be conserved, by stipulation of parallel motion with the incoming positron of one of the photons we know that the second photon cannot have any transverse momentum.

We therefore can focus our attention on the conservation of the momentum parallel to the direction of motion of the positron with momentum $p_e$. The conservation of momentum for massless photons means, introducing photon energy $cp_{1,2} = E_{1,2}$

$$1 \quad cp_e = E_1 + \varepsilon E_2 .$$

For $\varepsilon = +1$ the 2nd photon momentum is in direction of the first, i.e. aligned with the positron motion, and for $\varepsilon = -1$ it is opposite to the first photon and the positron. We now express $p_e$ to kinetic energy term $T_e$ using the relativistic relation between energy and momentum

$$2 \quad E_e^2 \equiv (T_e + m_e c^2)^2 = (p_e c)^2 + (m_e c^2)^2 .$$

We take the $p_e$ from this relationship and insert it into the conservation of momentum

$$3 \quad E_1 + \varepsilon E_2 = \sqrt{(T_e + m_e c^2)^2 - (m_e c^2)^2} = \sqrt{T_e^2 + 2m_e c^2 T_e} .$$

We further must also conserve energy in the annihilation process

$$4 \quad E_1 + E_2 = T_e + m_e c^2 + m_e c^2 .$$

where we are equating photon energy on left with $e^+e^-$-energy on right. Taking the difference of both conservation equations Eq. 3 and Eq. 4 we find

$$5 \quad E_2(1 - \varepsilon) = T_e + 2m_e c^2 - \sqrt{T_e^2 + 2m_e c^2 T_e} > 0 .$$
One easily checks that the right hand side in Eq. 5 is always positive. However, for \( \varepsilon = 1 \) the left hand side vanishes. Thus we must have \( \varepsilon = -1 \) which means that the second photon moves in opposite direction to the first, that is against the original motion of the incoming photon.

Adding up both conservation equations Eq. 3, Eq. 4 and restating Eq. 5 we obtain

\[
2E_1 = T_e + 2m_e c^2 + \sqrt{T_e^2 + 2m_e c^2 T_e}, \quad 2E_2 = T_e + 2m_e c^2 - \sqrt{T_e^2 + 2m_e c^2 T_e}
\]

for the energies and momenta of the two photons, where the photon of smaller energy \((E_2)\) moves against the original momentum of the positron.

We look at an example, taking \( T_e = 3 \) MeV, \( m_e c^2 = 0.511 \) MeV, and obtain for the energies

\[
p_1 = \frac{E_1}{c} = 3.748 \text{ MeV/c}, \quad p_2 = -\frac{E_2}{c} = -0.274 \text{ MeV/c}.
\]

We also can check the energy conservation:

\[
E_1 + E_2 = 3.748 + 0.274 \text{ MeV} = 4.022 \text{ MeV} = 3 \text{ MeV} + 2 \times 0.511 \text{ MeV}.
\]

These values of photon energies show that in the CM-frame the annihilation photons of equal energy emerge in opposite direction and thus the LT from CM to laboratory system enhances the energy of one of the photons while reducing the energy of the other.

---

19.3 Energy available in a collision

We determine the general expression for the mass of a system of two particles observed in the lab frame \( S \) in which particle ‘1’ collides with particle ‘2’. The energy equivalent \( Mc^2 \) of these two particles is:

\[
Mc^2 = \sqrt{(E_1 + E_2)^2 - c^2(\vec{p}_1 + \vec{p}_2)^2}.
\]  \hspace{1cm} (19.13)

The physical meaning of \( Mc^2 \) is best understood by considering the reverse process; a particle of a putative mass \( M \) and in its rest frame energy \( Mc^2 \) can decay into two particles, each with energy-momentum \( E_i, \vec{p}_i, i = 1, 2 \). We thus conclude that \( M \) is the quasi-mass that two colliding particles can create, and \( Mc^2 \) is the energy available in the frame in which \( M \) is at rest.

The invariant energy is denoted by \( \sqrt{s} \)

\[
\sqrt{s} = Mc^2.
\]  \hspace{1cm} (19.14)
This is the energy available for further processes (e.g. particle production). Note that the $\sqrt{s}$ in Eq. (19.14) should not be confused with the invariant space-time interval $s^2$ defined earlier; see Eq. (11.7). Unfortunately, in the common notation these two quantities must be distinguished by their context and made clear in this book.

We can reorder in Eq. (19.13) the terms and noting that $E_1^2 - c^2 p_1^2 = m_1^2 c^4$ and similarly for particle ‘2’, we find

$$\sqrt{s} = \sqrt{m_1^2 c^4 + m_2^2 c^4 + 2(E_1 E_2 - c^2 \vec{p}_1 \cdot \vec{p}_2)}.$$  (19.15)

Two cases are of particular importance:

1. Collision with particle ‘2’ at rest. We have $E_2 = m_2 c^2, \vec{p}_2 = 0$. We find

$$\sqrt{s} = \sqrt{m_1^2 c^4 + m_2^2 c^4 + 2E_1 m_2 c^2}.$$  (19.16)

2. Collision of two particles that collide coming head-on from opposite directions: for the special case that $m_1 = m_2$ we find calling $\vec{p}_1 = -\vec{p}_2 = \vec{p}$, $m_1 = m_2 = m$ and $E_1 = E_2 = E = \sqrt{m^2 c^4 + p^2 c^2}$

$$\sqrt{s} = 2E.$$  (19.17)

The ‘power’ of an accelerator is often presented in terms of the magnitude of $\sqrt{s}$ obtainable in two-particle collisions. Presently, the most powerful accelerator is the Large Hadron Collider (LHC) at CERN. Here the particles (protons) collide from opposite directions with equal magnitudes of energy and momentum; thus the lab frame is also the CM-frame! Hence according to Eq. (19.17) we form the sum of the two proton beam energies, $\sqrt{s} = (7 + 7) \text{ TeV} = 14 \text{ TeV}$, exceeding the proton mass equivalent more than 14,000 times. This is the maximum nominal energy of the LHC collider, a value towards which the LHC, which has been working its way up in energy, is indeed reaching this maximum design energy. A considerably greater amount of energy would be required in collisions of moving protons with protons at rest. To see this, solving Eq. (19.16) for $E_1$ we find

$$E_1 = \frac{s - m_1^2 c^4 - m_2^2 c^4}{2m_2 c^2}.$$  (19.18)

We note that $s$ and not $\sqrt{s}$ enters on the right. Consequently, we find for the LHC equivalent energy of a beam impacting a laboratory fixed target $E_{LHC} = 104,000 \text{ TeV}$! Clearly, the particle colliding in collider concept allows the realization of an enormous economy: we use two 7 TeV beams and achieve as much as
with one 104,000 TeV beams, not to mention that such a proton beam of 104 PeV (PeV=10^{15}eV) is far beyond present day technological capabilities.

**Exercise VII–8: Relativistic fender bender**

Two identical rockets of masses \( m = 50,000 \text{ kg} \) are launched with missions to test special relativity. Unfortunately, after reaching their cruising speeds of \( v = 0.6c \), they collide head-on. Assuming that no debris or radiation is released, and that all energy turns into mass, no pieces fly out, and nothing moves, what is the mass of the wreck?

**Solution**

This example addresses the case of completely inelastic collisions. The energies of the two colliding rockets are given by Eq. (15.21) to be

\[
E_1 = E_2 = \frac{mc^2}{\sqrt{1 - v^2/c^2}} = \frac{5}{4}mc^2.
\]

The conservation of energy implies that the energy of the wreck is equal to the total energy,

\[
E = E_1 + E_2 = \frac{5}{2}mc^2,
\]

and since the collision was head-on of two equal objects, the wreck is stationary; that is, the collision frame of reference is the CM-frame. In that case all energy must be contained in the compound body of crashed mass \( M \),

\[
E = Mc^2.
\]

With Eq. 2 and Eq. 3 we obtain

\[
M = \frac{E}{c^2} = \frac{5}{2}m = 125,000 \text{ kg}.
\]

The mass of the wreck is 25,000 kg greater than the combined mass of the colliding rockets: 25% of additional mass is created in the materialization of the kinetic energy of two rockets.

**End VII–8: Relativistic fender bender**

**Exercise VII–9: Production of antiprotons**
Fig. 19-2: Creation of a proton-antiproton pair: (left) in the lab frame; (right) in the CM-frame. See exercise VII–9.

During the collision of a relativistic proton with a proton at rest in laboratory an additional proton-antiproton pair can be created. How large must the minimum kinetic energy of the moving proton be? With what speed does the proton move?

**Solution**

The situation in the laboratory system is depicted on the left in 19-2; here the momentum $\vec{p}_1$ of the incoming proton is shared by all final state particles. In the CM-frame shown on right in 19-2 the two protons collide head-on coming from right and left and the new pair emerges in opposite directions. The total momentum of all particles sums to zero.

In the laboratory system we have for the total energy and momentum

$1 \quad \frac{E_t}{c} = \frac{E_1}{c} + m_0c, \quad \vec{p}_t = \vec{p}_1$.

In the CM-frame the colliding beam momentum adds up to zero, but the energy of colliding particles is

$2 \quad E_i = 2E_p, \quad \vec{p}_t = 0$.

In order to be able to create a proton-antiproton pair, the two protons colliding in the CM-frame need to provide the energy to the new pair of particles, therefore minimum

$3 \quad E_i = 4m_0c^2$.

The energy equivalent $Mc^2$ of all four particles can now be written in both the laboratory and CM-frame

$4 \quad (E_1 + m_0c^2)^2 - c^2p_1^2 = (M^2c^4) = (4m_0c^2)^2$.

---

$^5$The discovery of the antiproton was achieved at the Bevatron at Berkeley in collision of protons with nucleons bound in copper nuclei. This approach reduces the threshold of the reaction here computed due to occasional collisions of incoming protons with an ‘orbiting’ nucleon that is heading onto the incoming particle. Emilio Segré and Owen Chamberlain were awarded the Nobel Prize in Physics 1959 for the discovery of the antiproton.
With $E_1^2 - c^2 p_1^2 = m_0^2 c^4$ we get

$$2E_1m_0c^2 + 2(m_0c^2)^2 = 16(m_0c^2)^2,$$

and thus

$$E_1 = 7m_0c^2.$$

The kinetic energy of the incoming proton is

$$K_p = E_1 - m_0c^2 = 6m_0c^2.$$

The speed follows from

$$E_1 = \frac{m_0c^2}{\sqrt{1 - (v/c)^2}},$$

and we find, noting $\gamma = 7$

$$v = \sqrt{\frac{48}{49}}c = 0.9897c.$$

The momentum of the proton is

$$p_1 = m_0c\gamma v = \sqrt{48} m_0c = 6.928 m_0c.$$

We must remember that this is the so-called kinematic limit defining the onset of the production process.

---

Exercise VII–10: Equal speed reference frame

A particle with velocity $u$ corresponding to a Lorentz factor $\gamma = 3$ collides with a target particle at rest in the laboratory-frame $S$. There is a system $S'$ in which the velocities of the particle and the target are equal in magnitude and opposite in direction ($\vec{u}_{\text{particle}}' = -\vec{u}_{\text{target}}'$); i.e. the particles collide head-on with each other at equal velocity, which is also equal momentum if the particles are of the same mass. Determine the magnitude of this velocity, $u' = |\vec{u}_{\text{particle}}'| = |\vec{u}_{\text{target}}'|$ and the corresponding Lorentz factor $\gamma'$. The situation in both $S$ and $S'$ is depicted in figure [19-3].

Solution

An observer at rest in system $S'$ sees the particle moving with velocity $+\vec{u}'$ and the target moving with $-\vec{u}'$. Since the target is at rest in the lab frame $S$, the velocity
Figure 19-3: Top: A particle collides with a target in lab frame $S$. Bottom: Now in a moving frame $S'$. See exercise VII–10.

of frame $S$ relative to frame $S'$ must also be $\vec{v} = -\vec{u}'$. Using Eq. (7.9a) we transform the velocity of the particle in $S'$ to $S$ (note that here we transform from primed to ‘not-primed’ coordinates, while Eq. (7.9a) transforms from ‘not-primed’ to primed),

$$1 \quad u = \frac{u' - v}{1 - u'v/c^2} = \frac{2u'}{1 + u'^2/c^2}.$$  

Solving for $u'$ yields

$$2 \quad \frac{u'}{c} = \frac{c}{u} \left( 1 \pm \sqrt{1 - \frac{u'^2}{c^2}} \right).$$

We choose a negative sign for the root so that in the nonrelativistic limit, $u \to 0$, we obtain the nonrelativistic answer $u' = \frac{1}{2}u$. Using $1/\sqrt{1 - u'^2/c^2} = \gamma$ and

$$3 \quad \frac{u}{c} = \frac{\sqrt{\gamma^2 - 1}}{\gamma},$$

we can rewrite Eq. (2) as

$$4 \quad \frac{u'}{c} = \frac{\gamma - 1}{\sqrt{\gamma^2 - 1}} = \sqrt{\frac{\gamma - 1}{\gamma + 1}},$$

and with $1 - (u'/c)^2 = 2/(\gamma + 1)$ we obtain:

$$5 \quad \gamma' = \frac{1}{\sqrt{1 - (u'/c)^2}} = \sqrt{\frac{\gamma + 1}{2}}.$$  

For $\gamma = 3$ we find for the head-on collision velocity in $S'$,

$$6 \quad \frac{u'}{c} = \frac{1}{\sqrt{2}}, \quad \gamma' = \sqrt{2},$$

which we compare to the particle velocity in the lab frame $S$ obtained from Eq. (3)

$$7 \quad \frac{u}{c} = \sqrt{\frac{8}{9}}, \quad \gamma = 3.$$
We compare the results of this exercise with those seen in exercise VII–1 and note that, for example, we can use Eq. 5 to obtain for the transformation to the CM-frame from the beam dump frame the value \( \gamma = 61.02 \). This works since we have \( m_1 = m_2 \) and thus an equal velocity frame is also an equal momentum frame, which is the CM-frame for two particles, as required in this exercise.

The physical relevance the equal speed system is considered when the objects colliding are composite and possibly of different mass, but their parts are the same. In this case for the ‘parton’ collision the equal speed system is the collision CM reference frame as long as only the single particle reactions are relevant.

---

19.4 Inelastic collision and particle production

We consider collisions of two relativistic strongly interacting particles coming from right and left in the CM-frame as shown in figure 19-4. Typically these can be protons, or more generally, heavy ions; that is atomic nuclei. One speaks of “Relativistic Heavy Ion” (RHI) collisions. It is known that in such highly energetic collisions many secondary particles can be created. Each emitted secondary particle of momentum \( \vec{p} \) can be decomposed as shown in figure 19-4 into components that are parallel and perpendicular with respect to the original collision axis

\[
|\vec{p}_\parallel| = |\vec{p}| \cos \theta, \quad |\vec{p}_\perp| = |\vec{p}| \sin \theta.
\]  

(19.19)

Similarly, in the laboratory-frame

\[
|\vec{p}_\parallel' = |\vec{p}| \cos \theta, \quad |\vec{p}_\perp' = |\vec{p}| \sin \theta.
\]  

(19.20)

The distribution of particles in the azimuthal angle \( \varphi \) is often symmetric around the collision axis. However, this is not always the case and the distribution asymmetry in \( \varphi \) contains important information about the inelastic processes that lead to abundant particle production.

The particle production is measured in detectors located in the laboratory-frame of reference: once we have measured the momentum \( \vec{p} \) and angle of production \( \theta \), and we have identified the particle produced *i.e.* we know its mass, we also know its energy \( E_p = \sqrt{m^2c^4 + \vec{p}^2c^2} \). We can then obtain the particle rapidity by evaluating

\[
y_p = \frac{1}{2} \ln \left( \frac{E_p + c\vec{p}_\parallel}{E_p - c\vec{p}_\parallel} \right) = \frac{1}{2} \ln \left( \frac{E_p + c|\vec{p}| \cos \theta}{E_p - c|\vec{p}| \cos \theta} \right).
\]  

(19.21)

In an experiment it is common to evaluate how many particles are produced per rapidity interval \( dy \); that is to form \( dN/dy \), where for example \( N \rightarrow \pi^- \) when
Figure 19-4: The CM-frame particle momentum $\vec{p}$ decomposition into the parallel $\vec{p}_\parallel$ and perpendicular $\vec{p}_\perp$ components with respect to the collision axis. Note the inclination angle $\theta$ of $\vec{p}$ and the azimuthal angle $\varphi$ of $\vec{p}_\perp$.

Table 19-1: Collisions of p+p (NA61/SHINE CERN-SPS collaboration, Ref. [6]), at the beam momentum shown in first column, the corresponding CM-rapidity in 2nd column, the available p+p energy in 3rd column, the total multiplicity of $\pi^-$ produced in 4th column, and the energy cost of producing a $\pi^-$ in last column.

<table>
<thead>
<tr>
<th>$p_{\text{beam}}$ [GeV/c]</th>
<th>$y_{\text{CM}}$</th>
<th>$\sqrt{s_{\text{pp}}}$ [GeV]</th>
<th>$\pi^-$</th>
<th>$\frac{\sqrt{s_{\text{pp}}}}{\pi^-}$ [GeV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>158</td>
<td>2.91</td>
<td>17.27</td>
<td>2.444 ± 0.130</td>
<td>7.1 ± 0.3</td>
</tr>
<tr>
<td>80</td>
<td>2.57</td>
<td>12.32</td>
<td>1.938 ± 0.080</td>
<td>6.4 ± 0.3</td>
</tr>
<tr>
<td>40</td>
<td>2.22</td>
<td>8.76</td>
<td>1.478 ± 0.051</td>
<td>5.9 ± 0.2</td>
</tr>
<tr>
<td>31</td>
<td>2.10</td>
<td>7.74</td>
<td>1.312 ± 0.069</td>
<td>5.9 ± 0.3</td>
</tr>
<tr>
<td>20</td>
<td>1.88</td>
<td>6.27</td>
<td>1.047 ± 0.051</td>
<td>6.0 ± 0.3</td>
</tr>
</tbody>
</table>

we measure the production of negatively charged pions $\pi^-$. In the following we address $\pi^-$ rapidity spectra obtained by NA61/SHINE CERN-SPS collaboration (SPS: super proton synchrotron)\(^6\). These results were obtained for a laboratory stationary proton target varying the momentum of the incoming proton, with momentum values shown in the first column in table [19-1].

In order to compare inelastic production of particles for processes where the collisions in laboratory occur at different energies we shift the laboratory mea-

sured yields $d\pi^-/dy$ to CM-frame by remembering that the CM-rapidity $\bar{y}$ arises from the laboratory measured particle rapidity $y_p$ according to

$$\bar{y} = y_p - y_{CM} .$$

(19.22)

The value $y_{CM}$ is obtained as described in exercise $\text{VII-2}$: when both colliding particles have the same mass it is the average of the rapidities of the colliding particles, see Eq. 6 on page 238.

For the target particle, here the proton $p$, being at rest in laboratory this means that CM-frame rapidity $y_{CM}$ is half of the projectile rapidity $y_{CM} = y_p/2$. To obtain $y_p$ for a given high value of incoming proton momentum $|\vec{p}_p| \gg m_p c$ we can use the 2nd form in Eq. (17.20) simplified for $\vec{p}_p \perp = 0$ to read

$$y_p = \ln \frac{E_p + c p_p \parallel}{E_p \perp} = \ln \left( \sqrt{1 + \frac{\vec{p}_p^2}{m_p^2 c^2}} + \frac{|\vec{p}_p|}{m_p c} \right) = \ln \left( \frac{2|\vec{p}_p|}{m_p c} + \frac{1}{4} \frac{m_p^2 c^2}{\vec{p}_p^2} + \ldots . \right)$$

(19.23)

The resulting values of $y_{CM}$ are shown in the second column in table 19-1.

The third column in table 19-1 shows the total available $p+p$ energy $\sqrt{s_{pp}}$, see section 19.3. The fourth column shows the result of Ref. [6], the integrated multiplicity of produced $\pi^-$ with the systematic error originating in the required extrapolation to domains of $\vec{p}$ where the experiment NA61 was not able to take data. The rapidity distributions of inelastically produced $\pi^-$ obtained in Ref. [6] are shown in figure 19-5. Each of the distributions has been shifted from the laboratory-frame to CM-frame. The localization in rapidity centered at CM-frame $\bar{y} = 0$ of observed particle yields suggests a significant non-transparency of colliding matter, a surprise finding for this light $p+p$ collision system. We show in the last fifth column in table 19-1 the energy cost to produce the $\pi^-$ which within error is a constant for the three lower momenta of the projectile proton.

### 19.5 Elastic collisions

A very important special case of two body interactions are elastic collisions, which are characterized by the following:

a) the internal structure of the interacting particles remains unchanged, and

b) no new particles are produced.

We consider this after the more general case of inelastic collisions, since at relativistic energies elastic collisions are an exceptional case. On the other hand, in the more common in daily life nonrelativistic limit elastic collisions are more frequent.

In particular, this means that the masses of the particles taking part in these collisions remain unchanged:

$$m_1^2 c^4 = E_1^2 - (cp_1)^2 = E_1'^2 - (cp'_1)^2 ,$$

(19.24)
Figure 19-5: Rapidity spectra of $\pi^-$ produced in any (the other reaction products indicated by ‘X’) inelastic collisions of protons with momentum $p_{lab} = 20, 31, 40, 80$ and $158\text{GeV}/c$, where target protons at rest in laboratory. The statistical errors are smaller than the symbol size. The systematic uncertainties are indicated by the shaded bands. Results of NA61/SHINE CERN-SPS collaboration, adapted from figure 20 in Ref. [6].

and similarly

$$m_2^2c^4 = E_2^2 - (cp_2)^2 = E_2'^2 - (cp_2')^2.$$  \hfill (19.25)

As in the other reactions we have considered, in elastic collisions energy and momentum are conserved. We consider the situation before and after collision:

$$E = E_1 + E_2, \quad \vec{p} = \vec{p}_1 + \vec{p}_2, \quad E' = E'_1 + E'_2, \quad \vec{p}' = \vec{p}'_1 + \vec{p}'_2,$$  \hfill (19.26)

which does not change:

$$E = E', \quad \vec{p} = \vec{p}'.$$  \hfill (19.27)

First we note that just as in the nonrelativistic case, in the CM-frame the elastic reaction ‘rotates’ the momentum of the participating particles, the magnitude remains unchanged as is illustrated in figure 19-6:

$$\vec{p}' = |\vec{p}'_1| = |\vec{p}'_2|.$$  \hfill (19.28)
Show that the magnitude of particle momentum in an elastic collision considered in the CM-frame does not change, discuss graphically the meaning of this result.

Solution
We consider the conservation of energy relation:

1. \( E_1 + E_2 = E'_1 + E'_2 \),

and insert the unchanged masses before and after (Eq. (19.28)) collision:

2. \( \sqrt{m_1^2 + \vec{p}^2} = \sqrt{m_1'^2 + \vec{p}'^2} \).

This equation can be satisfied if and only if the magnitude of the momenta of particles remains equal,

3. \( \vec{p} = \vec{p}' \).

This means that in the CM-frame, the elastic scattering process rotates the particle momentum vectors without changing their magnitude, as is shown in figure 19-6.
an electron. This specific process is called Compton scattering\(^7\). In the lab frame, the target electron has no momentum; \(\vec{p}_e = 0\). The photon, though massless, has momentum \(|\vec{p}_\gamma| = E_\gamma/c\). Conservation of momentum requires that the initial momentum of the photon must be equal to the sum of the final momenta of the electron and photon:
\[
\vec{p}_\gamma - \vec{p}_\gamma' = \vec{p}_e'.
\] (19.29)

Squaring and expanding this equation gives us
\[
p_{e}'^2 = p_\gamma^2 + p_\gamma'^2 - 2p_\gamma p_\gamma' \cos \theta,
\] (19.30)

where the angle of the scattered photon is measured relative to the original axis of motion, see figure [19-7]. Rewriting Eq. (19.30) using again \(p_\gamma = E_\gamma/c\), we have
\[
p_{e}'^2 = \frac{E_\gamma^2}{c^2} + \frac{E_\gamma'^2}{c^2} - \frac{2E_\gamma E_\gamma' \cos \theta}{c^2}.
\] (19.31)

Considering conservation of energy we have
\[
E_\gamma + E_e = E_\gamma' + E_e'.
\] (19.32)

The electron has no momentum before the collision, thus
\[
E_e = m_e c^2.
\] (19.33)

After the collision, we remain in the electron’s initial rest frame, and so we must consider the momentum imparted to the electron by the collision
\[
E_e' = \sqrt{p_{e}'^2 c^2 + m_e^2 c^4}.
\] (19.34)

Substituting Eq. (19.33) and Eq. (19.34) into the conservation of energy equation Eq. (19.32):
\[
E_\gamma + m_e c^2 = E_\gamma' + \sqrt{p_{e}'^2 c^2 + m_e^2 c^4}.
\] (19.35)

Solving for the final momentum of the electron we find
\[ p'_e^2 = \frac{(E_\gamma - E'_\gamma + m_e c^2)^2 - m_e^2 c^4}{c^2}. \]  
(19.36)

Setting equal the two expressions Eq. (19.31) and Eq. (19.36) gives us
\[ \frac{(E_\gamma - E'_\gamma + m_e c^2)^2 - m_e^2 c^4}{c^2} = \frac{E_\gamma^2}{c^2} + \frac{E'_\gamma^2}{c^2} - \frac{2E_\gamma E'_\gamma \cos \theta}{c^2}. \]  
(19.37)

Simplifying, we have the Compton energy formula:
\[ E_\gamma E'_\gamma (1 - \cos \theta) = (E_\gamma - E'_\gamma) m_e c^2, \]  
(19.38)

which expresses how the energy of the photon changes in terms of the scattering angle. This Compton formula is usually presented in terms of the shift in wavelength of the photon. We recall that the energy of a photon \( E \) is proportional to the photon frequency \( \nu \), \( E = h \nu \), with Planck constant \( h \). This is the corpuscular property of light proposed by Einstein in 1905. Since light propagates with velocity \( c \) we further have the wavelength-frequency relation \( \lambda \nu = c \). Thus, we find a relation between photon energy and wavelength:
\[ E = \frac{hc}{\lambda}. \]  
(19.39)

Substituting this into the Compton energy formula Eq. (19.38), we obtain the wavelength shift Compton condition:
\[ \lambda' - \lambda = \lambda_C (1 - \cos \theta), \quad \lambda_C = \frac{h}{m_e c} = 2.42631 \times 10^{-12} \text{m}. \]  
(19.40)

The Compton wavelength shift result Eq. (19.40) agrees with our prior experience concerning the energy and momentum conservation. As the angle of deflection increases, \( \lambda' \) also increases, meaning that the final energy of the photon decreases. This is what we would expect from dealing with nonrelativistic collisions. The maximum possible increase in wavelength is \( 2\lambda_C \), corresponding to the back-scattering of the photon, \( \cos \theta = -1 \). When \( \lambda' - \lambda = 0 \) we know that \( \cos \theta = 1 \). This corresponds to the photon passing straight through the atom without colliding with an electron. This is, in fact, the most common occurrence.

As the above remark highlights, we have merely evaluated a straightforward consequence of energy-momentum conservation, that is, we found the relationship between the scattering angle and the wavelength shift, and have not yet touched upon the question of how often the photon-electron scattering will occur, and if it does, how often a particular scattering angle \( \theta \) is to be expected.

The study of photon-electron quantum reactions is part of the field of quantum electrodynamics (QED). The result of the study of photon-electron scattering is
the reaction cross-section describing the effective size of an electron in an interaction with a photon. The process of Compton scattering is then described by the “Klein-Nishina” formula:

\[
\frac{d\sigma}{d\Omega} = \frac{r_e^2 \lambda^2}{2 \lambda'^2} \left( \frac{\lambda}{\lambda'} + \frac{\lambda'}{\lambda} - 2 \sin^2 \theta \cos^2 \phi \right),
\]

(19.41)

where \(\phi\) is the azimuthal angle measured from the polarization direction of the incoming photon, and

\[
r_e \equiv \frac{e^2}{4\pi\varepsilon_0 m_e c^2} = 2.817940 \text{ fm},
\]

(19.42)
is the ‘classical’ electron radius.

We see that the probability that the scattering reaction occurs is governed by the ‘classical’ radius of the electron which is of a similar magnitude as the size of the atomic nucleus – along with Thompson scattering limit, see exercise VII–12 the value of \(r_e\) governing electron-photon scattering is well understood. It is of some practical importance that the effective size of an electron as ‘seen’ by a photon is comparable to the size of a small atomic nucleus ‘seen’ by an \(\alpha\)-particle in the pivotal Rutherford scattering experiment\(^9\). There is another way to characterize this result. We rewrite Eq. (19.42) in the form (note that here \(\alpha\) is the fine-structure constant and not the \(\alpha\)-particle)

\[
r_e \equiv \frac{e^2}{4\pi\varepsilon_0 \hbar c m_e c^2} \equiv \alpha \lambda_C = \frac{\lambda_C}{137.036},
\]

(19.43)
The (reduced) Compton wavelength \(\lambda_C\) is the quantum ‘size’ of the electron, compare discussion in section \(28.2\) on page \(429\). Seen from this quantum size perspective, the smallness of \(r_e\) is recognized as being due to the relative smallness of the electromagnetic charge \(e\) described by the quantum-coupling constant of electromagnetism, the fine-structure constant \(\alpha\).

Exercise VII–12: Thompson scattering

Obtain the low energy limit of the Klein-Nishina cross section, Eq. (19.41) which is the classical Thompson limit. Does \(\hbar\) appear in this result?

Solution


The magnitude of Klein-Nishina cross section Eq. (19.41) is controlled by the classical electron radius Eq. (19.42), a quantity which is manifestly independent of \( \hbar \); \( \hbar \) appears in the Compton wavelength shift only Eq. (19.40). We use the Compton wavelength shift formula Eq. (19.40) to eliminate from the Klein-Nishina cross-section Eq. (19.41) the ratio \( \lambda'/\lambda \). We find

\[
\frac{1}{2} \frac{d\sigma}{d\Omega} = \frac{r_e^2}{2} \left( \frac{1}{(1+\epsilon)^3} + \frac{1}{1+\epsilon} - \frac{2\sin^2 \theta \cos^2 \phi}{(1+\epsilon)^2} \right), \quad \epsilon = \frac{\lambda_C(1 - \cos \theta)}{\lambda}.
\]

Here, by assumption, we are in the low energy, high wavelength limit, so \( \epsilon \) can be made as small as needed.

Upon integration over angular variables, we obtain the Thompson limit

\[
\sigma_{\text{Thompson}} = 4\pi r_e^2 \left( 1 - \sin^2 \theta \cos^2 \phi \right) = 4\pi r_e^2 \left( 1 - \frac{2}{3} \right) = \frac{8\pi}{3} r_e^2 = 665 \text{ mb}.
\]

This result is manifestly independent of \( \hbar \). Since \( r_0 \propto 1/m^2 \), the lightest free particle carrying an electrical charge interacts strongest with photons, with a remarkable strength, an order of magnitude above strong interaction scale.

---

**19.7 Elastic bounce from a moving wall**

The reflection of a photon from a mirror is a well known phenomenon: In the rest frame of the mirror the photon momentum normal to the mirror is reflected. In his landmark paper on relativity, Einstein considered the case of photon reflection from a moving mirror. He obtained the fascinating result that the reflected energy of the photon scales with \( \gamma_{\text{mirror}}^2 \) (Lorentz factor). This is so since in the rest frame of the mirror the incoming photon is Doppler-shifted to higher energy, and after reflection in the mirror rest frame, transformation back into the lab frame produces a second, energy-enhancing Doppler effect, since the longitudinal momentum has been reversed.

For a laboratory observer, the reflection is a ‘squared’ Doppler effect. For the case that the mirror velocity is relativistic this leads to the expectation of a very large boost in photon energy, the source of energy being the kinetic energy of the mirror, which is assumed to be so large that we can ignore the back reaction. We show in exercise VIII–13 that in the nonrelativistic limit the ball bounces off much harder from such a moving wall, as the normal component picks up twice the speed of the wall.

The scattering process is illustrated in figure 19-8 we solve this elastic scattering problem of a particle of mass \( m \) (photons \( m \rightarrow 0 \)) impacting with velocity \( \vec{v} \) a much heavier mirror moving with velocity \( \vec{u} \). Elastic here means that the
Figure 19-8: Scattering of massive particle $m$, impacting a brick wall moving with velocity $\vec{u}$. Directions $\perp$ and $\parallel$ denote the component of particle velocity $\vec{v}$ or momentum $\vec{p}$ perpendicular and parallel to $\vec{u}$ and the coordinate system is selected such that the wall moves against the $x$-axis, while the $y$-axis points in the direction parallel to transverse motion of the particle.

momentum normal to the wall in the rest frame of the wall is reversed, while the transverse momentum is unchanged.

Let $\perp$ and $\parallel$ denote the component of particle momentum perpendicular and parallel to the direction of motion of the brick wall, see figure 19-8. These will relate to the Lorentz transformation, which we will introduce to solve the problem. The total energy of the incoming particle ‘$i$’ is given by

$$E = \sqrt{m^2c^4 + p_\perp^2c^2} = \sqrt{m^2c^4 + p_\perp^2c^2 + p_\parallel^2c^2} \equiv \sqrt{E_\perp^2 + p_\parallel^2c^2},$$

where $E_\perp = \sqrt{m^2c^4 + p_\perp^2c^2}$. With the $x$-axis being the direction (opposite) of wall motion defining the $\parallel$ component, and the $y$-axis parallel to $\vec{p}_\perp$, we further have for the $x$-component $p_\parallel = m \gamma v_\parallel$, and the $y$-component $p_\perp = m \gamma v_\perp$, while the $z$-component of both the momentum and velocity vectors vanishes. Note that we can also write

$$E = E_\perp \cosh y_p, \quad p_\parallel = E_\perp \sinh y_p,$$

where $y_p$ is particle rapidity, see Eq. (17.15). We will also use $y_{rb}$, the brick wall rapidity related as usual to the speed $u$ of the brick wall, see Eq. (7.41)

$$y_{rb} = \frac{1}{2} \ln \left( \frac{1 + \beta_b}{1 - \beta_b} \right), \quad \beta_b = \frac{u}{c}, \quad \gamma_b = \frac{1}{\sqrt{1 - \beta_b^2}} = \cosh y_{rb}, \quad \beta_b \gamma_b = \sinh y_{rb}.$$
We will present the energy and momentum components as columns: the co-
ordinate system is as shown in figure 19-8:

\[
\begin{bmatrix}
E \\
\beta \gamma mc^2 \\
\beta_\perp \gamma mc^2 \\
0
\end{bmatrix}
= 
\begin{bmatrix}
E_\perp \cosh y_p \\
E_\perp \sinh y_p \\
0
\end{bmatrix}
\rightarrow [LT]_{\text{brick wall}} 
\rightarrow 
\begin{bmatrix}
E_\perp \cosh(y_p + y_{rb}) \\
E_\perp \sinh(y_p + y_{rb}) \\
cp_\perp
\end{bmatrix}
\]

In the last step we applied an active Lorentz transformation taking us to the rest
frame of the brick wall. Since we have decomposed the particle momentum into
the appropriate components, and the brick wall motion is opposite the motion
of the particle, the rapidities add as we saw in exercise VI-7. Seen in the rest
frame of the brick wall the incoming particle is always more energetic as we see
explicitly in Eq. (19.44). From now on we omit the last two entries in columns as
the transverse momentum does not change.

The next step is for the particle to be reflected off the wall. This of course
happens that way in the frame of reference of the wall. This means that the mo-
mentum parallel to the direction of the wall’s motion is reversed; the superscript
‘bb’ indicates ‘bounced from the brick wall as seen in the brick wall rest-frame’:

\[
\begin{bmatrix}
E^{bb} \\
\beta \gamma mc^2 \\
\beta_\perp \gamma mc^2 \\
0
\end{bmatrix}
= 
\begin{bmatrix}
E_\perp \cosh(y_p + y_{rb}) \\
E_\perp \sinh(y_p + y_{rb}) \\
0
\end{bmatrix}
\]

where we introduced \( y_b \) as the rapidity of the particle after bounce

\[ y_b = -y_p - y_{rb} \]  \hspace{1cm} (19.46)

Note the negative sign implying that the particle is now moving in the opposite
direction.

We now transform back to the laboratory-frame, where the prime indicates
‘bounced from brick wall as seen in the lab’:

\[
\begin{bmatrix}
E' \\
\beta \gamma mc^2 \\
\beta_\perp \gamma mc^2 \\
0
\end{bmatrix}
= 
\begin{bmatrix}
E_\perp \cosh(y_b - y_{rb}) \\
E_\perp \sinh(y_b - y_{rb}) \\
0
\end{bmatrix}
\]

This result shows that the reflected particle acquires in the direction of the brick
wall motion twice the rapidity of the brick wall, and a changed parallel momentum
sign shows it is moving in a direction opposite from the original.
This is not quite the final result; we need to revert to the original quantity \( p_\parallel \) for the result to become more transparent. This is accomplished by employing the addition theorems for hyperbolic functions

\[
\cosh(a + b) = \cosh a \cosh b + \sinh a \sinh b ,
\]

\[
\sinh(a + b) = \cosh a \sinh b + \sinh a \cosh b .
\]

This leads to

\[
\begin{bmatrix}
E' \\
\cp' \\
\cp'_\perp \\
0
\end{bmatrix}
= 
\begin{bmatrix}
E \cosh 2\yr - E \sinh 2\yr \\
\frac{\cosh 2\yr}{\gamma_b} - \frac{E}{\gamma_b} \sinh 2\yr \\
\frac{2\gamma_b \cp}{\gamma_b} \\
0
\end{bmatrix}.
\]

(19.49)

Remembering Eq. (19.44) we can revert to the original laboratory energy and momentum of the incoming particle

\[
\begin{bmatrix}
E' \\
\cp' \\
\cp'_\perp \\
0
\end{bmatrix}
= 
\begin{bmatrix}
2\gamma_b \gamma_b^2 \\
-\gamma_b^2 \gamma_b^2 \\
2\gamma_b^2 \\
0
\end{bmatrix}.
\]

(19.50)

which is the final result wherein we retain \( \yr \) to characterize the effect of the moving mirror attached to a brick wall.

For applications in non-relativistic and ultra-relativistic brick wall motion we seek a result which employs velocity vectors. Using Eq. (19.48) we obtain

\[
\cosh 2\yr = \cosh^2 \yr + \sinh^2 \yr = (1 + \beta_b^2)\gamma_b^2 \rightarrow 2\gamma_b^2 ,
\]

\[
\sinh 2\yr = 2\cosh \yr \sinh \yr = 2\beta_b \gamma_b^2 \rightarrow 2\gamma_b^2 ,
\]

where we show the ultra-relativistic limit explicitly. We now can write Eq. (19.50) using the brick wall velocity and Lorentz factor \( \gamma_b \), restoring for completeness the transverse momentum in explicit form

\[
\begin{bmatrix}
E' \\
\cp' \\
\cp'_\perp \\
0
\end{bmatrix}
= 
\begin{bmatrix}
(1 + \beta_b^2)E + 2\beta_b \cp_\parallel \\
-(1 + \beta_b^2)\cp_\parallel - 2\beta_b E \\
\cp_\perp \\
0
\end{bmatrix}.
\]

(19.52)
which makes explicit the coefficient $\gamma_b^2$. The appearance of $\gamma_b^2$ factor boosting the incoming energy and momentum suggests that ultra-relativistic particle beams could be realized employing brick wall bounce.

Both Eq. (19.50) and Eq. (19.52) remain valid considering the special case of massless particles (i.e., photons), since in the presented derivation we used the particle energy and momentum, and did not specify the value of particle mass. We recover from Eq. (19.52) the well known Einstein limit $E' \simeq c p' \simeq 4\gamma_b^2 E$ for head-on collision.

This remarkable result is today no longer just an academic topic. For photons, a ‘relativistic mirror’ can be realized by considering a speeding electron cloud generated by either a high contrast ultra short laser pulse interacting, for example, with thin $O(10\mu)$ Aluminum targets\footnote{S. V. Bulanov et al. “Relativistic mirrors in plasmas. Novel results and perspectives,” Phys.-Usp. 56 429 (2013).} or using conventional linear accelerator technology to create intense electron pulses. In order to achieve, in back scattering MeV range, a coherent gamma ray wave, a laser beam with 1 eV photons would need to be reflected from an electron cloud that has $\gamma = 1000$, i.e. 500 MeV. The difficulty in realizing this in practice is that in the rest frame of the mirror, where the reflection must occur, the incoming laser beam has a $\gamma$-shortened wavelength, thus requiring that photons at 1 keV energy corresponding to wavelength of $\lambda = h c/keV = 1.24 \AA$ are coherently reflected. If it’s possible to build such a mirror, this is certainly not easy!

Exercise VII–13: Nonrelativistic Bounce

Obtain the velocity vector of a particle scattered from a moving wall from consideration of the nonrelativistic limit of the general result presented in Eq. (19.52).

Solution

In the nonrelativistic limit we employ the leading terms

\[
E = \left(1 + \frac{\bar{v}^2}{c^2} + \ldots\right) m c^2, \quad p_\parallel = m v_\parallel + \ldots, \quad p_\perp = m v_\perp + \ldots, \quad \gamma_b^2 = 1 + \frac{\bar{u}^2}{c^2} + \ldots,
\]

in Eq. (19.52) to obtain

\[
\begin{bmatrix}
E' \\
\bar{c} p_\parallel' \\
\bar{c} p_\perp' \\
0
\end{bmatrix}
= \begin{bmatrix}
mc^2 \left(1 + \frac{2u^2}{c^2}\right) \left(1 + \frac{\bar{v}^2}{2c^2}\right) + 2mu v_\parallel \\
- mc \left(1 + \frac{2u^2}{c^2}\right) v_\parallel - mc^2 \left(1 + \frac{\bar{v}^2}{2c^2}\right) \frac{2u}{c} \\
c p_\perp \\
0
\end{bmatrix}
= \begin{bmatrix}
\frac{(2\bar{u} + \bar{v})^2 m}{2} + mc^2 \\
-(2u + v_\parallel) mc \\
v_\perp mc \\
0
\end{bmatrix}.
\]
This shows that the bounce off a moving wall adds twice the velocity of the wall to the
reversed (bounced) velocity vector component, \( v_\parallel \), that is parallel to the motion of the
wall. The total velocity vector of a body after a bounce from a moving wall is thus

\[
\vec{v}' = -2\vec{u} \left( 1 + \frac{\vec{u} \cdot \vec{v}}{\vec{u}^2} \right) + \vec{v}.
\]

As a cross check we compute the square of the velocity vector which provides us with
the kinetic energy in Eq. 2.

\[
4 (\vec{v}')^2 = 4(\vec{u})^2 + 8\vec{u} \cdot \vec{v} + 4\left( \frac{\vec{u} \cdot \vec{v}}{\vec{u}^2} \right)^2 - 4\vec{u} \cdot \vec{v} - 4\left( \frac{\vec{u} \cdot \vec{v}}{\vec{u}^2} \right)^2 + (\vec{v})^2.
\]

Upon effecting two trivial cancellations we obtain the result agreeing with the kinetic
energy in Eq. 2.

\[
5 (\vec{v}')^2 = (2\vec{u} + \vec{v})^2.
\]

The bounce from a wall provides a first impression of the ‘flyby’ impulse-inducing
method used to boost (or reduce) the velocity of a satellite ‘bouncing’ due to gravity
potential, that is, not by actual impact, off a moving stellar body.

---

End VII–13: Nonrelativistic Bounce ---
Part VIII

Four-Vectors and Four-Force
Introductory remarks to Part VIII

Up to this point we gave little attention to time being a new, fourth coordinate in Part I: In SR time is united with space to form Minkowski space-time. We now return to take advantage of the space-time symmetries, implementing explicitly the unity of space and time. Thus we approach the theory of relativity in a more symmetric and appealing format. In doing this we not only rewrite what we know but also uncover new physics that is otherwise hidden from view.

We show that with time seen as a 4th coordinate there are generalized four-vectors and associated rotations, the Lorentz transformations. The ‘Lorentz-covariant’, or simply covariant, vector notation in four dimensions naturally shows the unity of space-time. Just like vector notation in normal space shows equivalence of all directions, the covariant form implies equivalence of all four space-time directions. Vector length does not change under a rotation and this also applies to the new situations. There are ‘Lorentz invariant’, or simply invariant, quantities such as magnitudes of 4-vectors, and generalized scalar products of 4-vectors.

Results of experiments are best studied using Lorentz-invariant quantities since they preserve their values for all observers such that the conditions of experiment and even in special cases the outcome is reported to be the same when different inertial observers consider the observable. As an example of the use of this idea we discuss the Mandelstam variables which replace the scattering angle and collision energy in the usual formulation of the scattering processes.

An important recurring issue in this book is the distance traveled by an unstable particle. We solve this problem one more time by considering the 4-dimensional momentum definition and its conservation in the particle rest-frame and, a second time, in the laboratory frame. The objective is to show that no matter in what way one looks at the motion of the muon or, for that matter, the Earth, the outcome is the same and properties of space-time are not part of the solution at all.

The classic 4-vectors described in this section – the four coordinates of a particle, its four-velocity, and arising from it, the four-acceleration. We show that our usual physical reality is contained in the domain of truly negligible acceleration – we return in Part XI to the more general case of strong acceleration. We propose in 4-dimensional vector notation a straightforward generalization of Newton’s force to a covariant format. A deeper look at these equations allows us to identify a new physics content that the effective inertial resistance of a particle depends on its state of motion parallel or perpendicular to the applied 3-force.
20 Four-Vectors in Minkowski Space

20.1 Lorentz invariants and covariant equations

Consider a familiar three-dimensional vector, which we will hence call a three-vector, constrained for simplicity to the $xy$ plane. Its length is conserved under a rotation about the orthogonal $z$-axis,

$$\vec{r}^2 = x^2 + y^2 = x'^2 + y'^2 = \vec{r}'^2.$$

As shown in figure 20-1, the endpoints of a position vector, and its rotated image lie on a circle. We say that the length of a vector is the invariant under spatial rotations, which means that all observers agree to the length of this vector.

By analogy, we wish to identify new ‘vectors’ that under Lorentz transformation will transform in a similar fashion to a spatial rotation, and which have a magnitude invariant under LT, retaining the same value for different inertial observers. We call a quantity that does not change under a Lorentz transformation a Lorentz-invariant (LI) or just ‘invariant’. Our search for such invariant quantities is justified also by our earlier observation that the Lorentz transformation written in terms of rapidity appear in a form rather similar to a rotation, see Eq. (7.49).

Such new Lorentz vectors define the events in Minkowski space, and therefore must have four components, and thus one calls these four-vectors also often written as 4-vectors; when there is a chance that meaning is not clear we will also call regular vectors as three-vectors, or 3-vectors. As observed below Eq. (7.50) we expect a modified, LI definition of ‘length’ as compared to a regular three-vectors. We will in the following find 4-vector analogues to most 3-vectors along with their corresponding LI lengths.

Figure 20-1: The rotation of a position vector preserves its length: $|\vec{r}| = |\vec{r}'|$ is an invariant under rotations.
Here are a few examples of LIIs we already explored: mass of a body must be a Lorentz-invariant, as all observers must agree to what the mass of a body is. Another natural LI is the distance \( s \) between two events in Minkowski space. We choose the location of one of the events, just like in Eq. (20.1), at the ‘origin’ \( ct = x_1 = x_2 = x_3 = 0 \), and the other to be at any \((ct, x_1, x_2, x_3)\), and form:

\[
s^2 = c^2t^2 - (x_1^2 + x_2^2 + x_3^2) = c^2t^2 - \vec{r}^2.
\]  

(20.2)

We chose the minus sign between timelike and spacelike coordinates because we already know that proper time \( \tau \) is not modified by a Lorentz transformation, see section 7.3. From this point forward Eq. (20.2) is the prototype relation that we explore as the definition of the length of a 4-vector.

20.2 The ‘position’ 4-vector

We are motivated by the form of Eq. (20.2) to introduce a 4-vector such that its length can be interpreted as \( c^2\tau^2 \). It is usual to write this 4-vector as a set of four coordinates in the following manner:

\[
x^\mu = \{ct, x, y, z\}
\]

(20.3a)

\[
= \{ct, \vec{r}\}, \quad \mu = \{0, 1, 2, 3\},
\]

and thus

\[
x^0 = ct, \quad x^1 = x, \quad x^2 = y, \quad x^3 = z,
\]

(20.3b)

\[
x^\mu = \{x^0, x^1, x^2, x^3\}.
\]

If we were to define the length in identical format as in usual 3-dimensional space

\[
\sum_{\mu=0}^{3} x^\mu x^\mu = c^2t^2 + x_1^2 + x_2^2 + x_3^2 = c^2t^2 + \vec{r}^2,
\]

(20.4)

we obtain the non-invariant under LT expression: it is easy to see repeating the steps seen section 7.3 on page 104 that Eq. (20.4) is not an invariant under LT.

We need to find a relative minus sign visible in Eq. (20.2) and missing in Eq. (20.4). There are even today special relativity texts which introduce time as a coordinate which is imaginary. This was also what Einstein used for a few years until he realized that this was at odds with the precise mathematical treatment of SR and with all of GR he was developing. To a large extent this idea is abandoned today in favor of Minkowski space-time.
A better way to ‘find’ the missing minus sign is to introduce the following new vector (distinguished from the former by replacing the superscript with a subscript)

\[ x_{\mu} = \{ ct, -x, -y, -z \} \]

\[ x_{\mu} = \{ ct, -\vec{r} \} , \quad \mu = \{ 0, 1, 2, 3 \} , \]

and thus

\[ x_0 = ct , \quad x_1 = -x , \quad x_2 = -y , \quad x_3 = -z , \]

\[ x_{\mu} = \{ x_0, x_1, x_2, x_3 \} , \]

One calls the above \( x_{\mu} \) the covariant components of the position vector, and \( x^{\mu} \) in Eq. (20.3b) the contravariant components of the position vector.

The reason we will write vectors in the format one introduces in general relativity, with distinct indices ‘up’ or ‘down’, is that we want to establish good habits. In this book, the space-time manifold is flat, known as the ‘Minkowski’ space. The reader should thus take what follows as definitions and not seek at this point a deeper understanding or believe the book lacks an explanation.

The desired LI is now easy to form, and we obtain:

\[ s^2 = \sum_{\mu=0}^{3} x_{\mu} x^{\mu} = c^2 t^2 - (x_1^2 + x_2^2 + x_3^2) = c^2 t^2 - \vec{r}^2 . \]

(20.6)

### 20.3 Metric in Minkowski space

However, Eq. (20.6) is a trivial restatement of what we already know. The new insight about Minkowski space comes with an effort to relate \( x^{\mu} \) to \( x_{\mu} \). We introduce a matrix \( g \) called ‘metric’, with components \( g_{\mu\nu} \),

\[ g = (g^{\mu\nu}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \]

(20.7)

From the definition of the covariant components of the 4-vector (Eq. (20.5b)) and the contravariant components (Eq. (20.3b)), it follows that

\[ x_{\mu} = \sum_{\nu=0}^{3} g_{\mu\nu} x^{\nu} . \]

(20.8a)
Thus the metric $g$ ‘lowers’ the contravariant (upper) index to the covariant (lower) index. This can be also written in terms of matrix multiplication

\[
\begin{pmatrix}
x_0 \\
x_1 \\
x_2 \\
x_3
\end{pmatrix} =
\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix}
\begin{pmatrix}
x^0 \\
x^1 \\
x^2 \\
x^3
\end{pmatrix} =
\begin{pmatrix}
x^0 \\
x^1 \\
x^2 \\
x^3
\end{pmatrix},
\tag{20.8b}
\]

or using matrix notation

\[
(x_\mu) = (g_{\mu\nu})(x^\nu).
\tag{20.8c}
\]

From now on we write for matrices with components $M_{ij}$ the form $(M_{ij})$, or in absence of an ambiguity simply $M$ and the matrix multiplication summation is implied. Note that in Eq. (20.8c) this simplification cannot be done since in the equation $x = gx$ we would need to make sure that $x$ is distinct from other related meanings, and that we distinguish upper and lower index $x$.

Another short-hand notation is provided by the Einstein summation convention: as in the matrix multiplication implied in Eq. (20.8c) the summation of indices is carried out even if and when we omit matrix parenthesis shown in Eq. (20.8c)

\[
x_\mu = g_{\mu\nu}x^\nu.
\tag{20.8d}
\]

The general rule of Einstein summation convention in this book is that if one sees the same Greek letter index once down and once up in a single mathematical expression, in Eq. (20.8d) it is $\nu$, the sum over the four terms, $\nu = 0, 1, 2, 3$ is always implied. From now on we will begin to use this convention, however not exclusively as sometimes it is helpful to be reminded that of implicit sums. The use will be increasing as we are getting used to the implicit presence of this summation convention.

The inverse metric matrix $g^{-1}$ has components we call $g^{\nu\mu}$. Using the identity matrix, called the Kronecker-delta

\[
I = (\delta^\nu_\nu),
\tag{20.9}
\]

we obtain

\[
I = g g^{-1} \rightarrow (\delta^\nu_\nu) = (g_{\nu\alpha})(g^{\alpha\mu}) = \left(\sum_{\alpha=0}^{3} g_{\nu\alpha} g^{\alpha\mu}\right).
\tag{20.10}
\]

As introduced and by convention, $g_{\mu\nu}$ always denotes the components of the metric matrix $g$, and $g^{\mu\nu}$ always denote the components of the inverse metric matrix $g^{-1}$. However, in Minkowski space the inverse metric is the same as the metric

\[
g^{-1} = g.
\tag{20.11}
\]
We show this by computing explicitly Eq. (20.10) using matrix notation

\[ I \equiv \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & -1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix}. \tag{20.12} \]

Just as we have lowered an index using \( g = (g_{\mu\nu}) \), see Eq. (20.8a), we can also raise an index using \( g^{-1} = (g^{\mu\nu}) \), for example inverting Eq. (20.8a) to read

\[ x^\mu = \sum_{\nu=0}^{3} g^{\mu\nu} x_\nu = g^{\mu\nu} x_\nu. \tag{20.13} \]

The procedure of raising and lowering of indices is consistent as we can test computing

\[ x^\mu = g^{\mu\nu} x_\nu = g^{\mu\nu} g_{\nu\kappa} x_\kappa = \delta^\mu_\kappa x_\kappa = x^\mu, \tag{20.14} \]

where we used both Eq. (20.8a) and Eq. (20.9). Nothing in the above relies on the fact that we used in this introduction to lowering and raising the 4-vector designating an event in space-time. Our finding in Eq. (20.14) depends only on the lowering and raising of indices being done using the metric and its inverse, respectively. Moreover, we already have raised and lowered the indices on the metric itself, this is Eq. (20.10) presented in the format

\[ g^{\mu\nu} \equiv \sum_{\alpha=0}^{3} g_{\nu\alpha} g^{\alpha\mu} = \delta^\mu_\nu. \tag{20.15} \]

Clearly, the procedure to lower and raise indices is a more general feature. The notion of covariants, contravariants, and metrics becomes considerably more useful in General Relativity. We will not address this further, only mention that in more advanced texts one refers to the metric matrix as 'second rank tensor'. The name is in analogy to naming an object a 4-vector where Lorentz transformation behavior of 4-tuplets define these as a vector similarly a second rank tensor is a matrix with two indices and the transformation rules between matrix elements are according to two Lorentz transformations, one for each index.

The invariant \( c^2 \tau^2 \), Eq. (20.6), can now be written as

\[ s^2 = \sum_{\nu=0}^{3} \sum_{\mu=0}^{3} g_{\mu\nu} x^\mu x_\nu = g_{\mu\nu} x^\mu x_\nu, \tag{20.16a} \]

or as

\[ s^2 = \sum_{\nu=0}^{3} \sum_{\mu=0}^{3} g^{\mu\nu} x^\mu x_\nu = g^{\mu\nu} x^\mu x_\nu. \tag{20.16b} \]
It is common to replace Eq. (20.16a) and Eq. (20.16b) by the shorthand notation analog to usual scalar dot-product

\[ s^2 = x \cdot x, \]

(20.16c)

and to speak of ‘squaring a 4-vector \( x \)’. Note that in this book as in many others you recognize the squaring of a 4-vector to be different from the squaring of a 3-vector by the absence of the vector arrow. We will also use in a context that will avoid this misunderstanding the shorthand

\[ x \cdot x \equiv x^2, \]

(20.16d)

which omits the center dot.

All these notational comments of course apply to any 4-vector we study, not only to \( x^\mu \) in Eq. (20.3b). When we refer to a 4-vector we will use the index-up contravariant form like Eq. (20.3b) as it does not contain the ‘minus’ sign one sees in the covariant vectors, Eq. (20.5b). One loosely says that the contravariant form is the ‘physically measurable’ form; further discussion is left to a more advanced text that addresses General Relativity.

### 20.4 Lorentz boosts as generalized rotation

We now look for the explicit form of the ‘rotation’ in Minkowski space which involves time such that the length of a 4-vector is preserved. Let us consider the LT for motion in the \( z \) direction (also called ‘\( z \)-boost’) which is given by

\[
\begin{align*}
ct' &= \frac{ct - (v/c)z}{\sqrt{1 - \vec{v}^2/c^2}} = \gamma (ct - \beta z), \\
x' &= x, \quad y' = y, \\
z' &= \frac{z - (v/c)(ct)}{\sqrt{1 - \vec{v}^2/c^2}} = \gamma (z - \beta ct),
\end{align*}
\]

(20.17)

We seek a ‘Lorentz-rotation’ matrix \( \Lambda \) such that the Lorentz-transformation can be represented by

\[ x'^\mu = \sum_{\nu=0}^{3} \Lambda^\mu_{\nu} x^\nu. \]

(20.18)

We see that the components \( \Lambda^\mu_{\nu} \) of matrix \( \Lambda_z = (\Lambda^\mu_{\nu}) \), where both \( \mu, \nu = 0, 1, 2, 3 \). \( \Lambda_z \) is describing a ‘boost’ along the \( z \)-direction as indicated by the subscript \( z \).
\[ \Lambda_z = (\Lambda^\mu_\nu) = \begin{pmatrix} \nu \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma & 0 & 0 & -\beta \gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\beta \gamma & 0 & 0 & \gamma \end{pmatrix} . \] (20.19)

This can easily be verified by writing out the four equations explicitly. Then we find that
\[ x'^0 = ct' = \Lambda^0_0 x^0 + \Lambda^0_3 x^3 = \Lambda^0_0 ct + \Lambda^0_3 z \]
\[ x'^1 = x' = \Lambda^1_1 x^1 = x, \quad x'^2 = y' = \Lambda^2_2 x^2 = y, \] (20.20)
\[ x'^3 = z' = \Lambda^3_0 x^0 + \Lambda^3_3 x^3 = \Lambda^3_0 ct + \Lambda^3_3 z. \]

If one inserts the components of \( \Lambda^\mu_\nu \), the equations of the Lorentz transformation (Eq. (20.17)) are found immediately.

Even more “rotation-like” is the form that we obtain using rapidity ‘angle’ \( \eta_z \)
\[ \Lambda_z = \begin{pmatrix} \gamma & 0 & 0 & -\beta \gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\beta \gamma & 0 & 0 & \gamma \end{pmatrix} = \begin{pmatrix} \cosh \eta_z & 0 & 0 & -\sinh \eta_z \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh \eta_z & 0 & 0 & \cosh \eta_z \end{pmatrix} . \] (20.21)

There are two more boosts describing Lorentz transformation along the other \( x \) and \( y \) axis denoted \( \Lambda_x, \Lambda_y \). Explicitly, we have,
\[ \Lambda_x = \begin{pmatrix} \cosh \eta_x & -\sinh \eta_x & 0 & 0 \\ -\sinh \eta_x & \cosh \eta_x & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \]
\[ \Lambda_y = \begin{pmatrix} \cosh \eta_y & 0 & -\sinh \eta_y & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh \eta_y & 0 & \cosh \eta_y & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \] (20.22)

The three boosts \( \Lambda_x, \Lambda_y, \Lambda_z \) remind of a rotation \( R_x, R_y, R_z \) around an axis in three dimensions, for example, around the \( z \)-axis, showing only the transformation of \( x, y, z \) coordinates and omitting time,
\[ R_z = \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} . \] (20.23)
This is the rotation shown in figure 20-1 on page 275. The main difference between boosts $\Lambda$ with rotations $R$ is their non-compact character, with values of $\eta_x, \eta_y, \eta_z$ unbound.

The inverse transformation $\Lambda^{-1}$ matrix is obtained by changing $\beta \to -\beta$ or in rapidity representation, changing $\eta \to -\eta$, for any of the three boost directions

$$\Lambda_x^{-1} = \Lambda_x(-\eta_x), \quad \Lambda_y^{-1} = \Lambda_y(-\eta_y), \quad \Lambda_z^{-1} = \Lambda_z(-\eta_z),$$  \hspace{1cm} (20.24)

We verify Eq. (20.24) evaluating

$$\Lambda_x \Lambda_x^{-1} = \begin{pmatrix} \cosh \eta_x & -\sinh \eta_x \\ -\sinh \eta_x & \cosh \eta_x \end{pmatrix} \begin{pmatrix} \cosh \eta_x & \sinh \eta_x \\ \sinh \eta_x & \cosh \eta_x \end{pmatrix}$$

$$= \begin{pmatrix} \cosh^2 \eta_x - \sinh^2 \eta_x & \cosh \eta_x \sinh \eta_x - \sinh \eta_x \cosh \eta_x \\ -\sinh \eta_x \cosh \eta_x + \cosh \eta_x \sinh \eta_x & -\sinh^2 \eta_x + \cosh^2 \eta_x \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \Lambda_x(-\eta_x) \Lambda_x(\eta_x),$$  \hspace{1cm} (20.25)

where writing the matrices we have suppressed the trivial $y$ and $z$ components for compactness. In the last equality we have noted that the result is the same for right- and left-inverse.

Finally, let us remark that the way we introduced the three boosts $\Lambda_i, i = x, y, z$ as is evident in Eq. (20.21) and Eq. (20.22) we have

$$\Lambda_i^\top = \Lambda_i, \quad i = x, y, z.$$  \hspace{1cm} (20.26)

However, a general Lorentz transformation that includes a space rotation or, equivalently, is a product of two different boosts does not satisfy condition Eq. (20.26) as is evident inspecting any of the rotation matrices or simply looking closer at exercise VIII–2 on page 283.

---

**Exercise VIII–1: Consecutive collinear LT**

Using the matrix representation of the Lorentz transformation show that two sequential Lorentz boosts in same direction corresponding to velocities $v_1$ and $v_2$ can be replaced by one new Lorentz transformation with a new velocity $v_3$ to be determined.

**Solution**

We have already described the solution to this problem in section 7.4 on page 107. Here we use the matrix representation of the Lorentz transformation for $v_1$ and $v_2$ in
the x direction, where writing the matrices we have suppressed the trivial y and z components for compactness. We have three different matrices

1 \[ \Lambda_i = \begin{pmatrix} \gamma_i & -\beta_i \gamma_i \\ -\beta_i \gamma_i & \gamma_i \end{pmatrix}, \quad i = 1, 2, 3 \]

where ‘1,2’ are the two boosts that lead to the third ‘3’. We calculate the matrix product

2 \[ \Lambda_3 = \Lambda_1 \Lambda_2 = \begin{pmatrix} \gamma_1 & -\beta_1 \gamma_1 \\ -\beta_1 \gamma_1 & \gamma_1 \end{pmatrix} \begin{pmatrix} \gamma_2 & -\beta_2 \gamma_2 \\ -\beta_2 \gamma_2 & \gamma_2 \end{pmatrix}, \]

and obtain

3 \[ \Lambda_3 = \begin{pmatrix} \gamma_1 \gamma_2 + \beta_1 \beta_2 \gamma_1 \gamma_2 & -\gamma_1 \gamma_2 \beta_2 - \gamma_1 \gamma_2 \beta_1 \\ -\gamma_1 \gamma_2 \beta_2 - \gamma_1 \gamma_2 \beta_1 & \gamma_1 \gamma_2 + \beta_1 \beta_2 \gamma_1 \gamma_2 \end{pmatrix} = \begin{pmatrix} \gamma_3 & -\beta_3 \gamma_3 \\ -\beta_3 \gamma_3 & \gamma_3 \end{pmatrix}. \]

We thus have, compare Eq. 4 in exercise III–19 on page 114,

4 \[ \gamma_3 = \gamma_1 \gamma_2 (1 + \beta_1 \beta_2), \]

and, compare Eq. (7.33) on page 108

5 \[ -\beta_3 \gamma_3 = -\gamma_1 \gamma_2 (\beta_2 + \beta_1). \]

Inserting Eq. 4 into Eq. 5 we obtain, compare Eq. (7.34) on page 108

6 \[ \beta_3 = \frac{\beta_1 + \beta_2}{1 + \beta_1 \beta_2}, \]

which is the same as the relativistic velocity addition Eq. (7.9a) on page 92.

--- End VIII–1: Consecutive collinear LT ---

--- Exercise VIII–2: Non-commuting LT ---

Show by example and using matrix representation of LT that two Lorentz boosts carried out in two different directions are not commutative.

Solution
We consider one Lorentz transformation in the \(x\)-direction and a second in the \(y\)-direction just like in exercise III–16 on page 108. They are represented by the matrices

\[ \Lambda_x = \begin{pmatrix} \gamma_x & -\beta_x \gamma_x & 0 & 0 \\ -\beta_x \gamma_x & \gamma_x & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \Lambda_y = \begin{pmatrix} \gamma_y & 0 & -\beta_y \gamma_y & 0 \\ 0 & 1 & 0 & 0 \\ -\beta_y \gamma_y & 0 & \gamma_y & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \]

We calculate the two products \(\Lambda_x \Lambda_y\) and \(\Lambda_y \Lambda_x\)

\[ \Lambda_x \Lambda_y = \begin{pmatrix} \gamma_x \gamma_y & -\gamma_x \beta_x & -\gamma_x \gamma_y \beta_y & 0 \\ -\gamma_x \gamma_y \beta_x & \gamma_x & \gamma_x \gamma_y \beta_x \beta_y & 0 \\ -\gamma_y \beta_y & 0 & \gamma_y & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \]

and

\[ \Lambda_y \Lambda_x = \begin{pmatrix} \gamma_x \gamma_y & -\gamma_x \gamma_y \beta_x & -\gamma_y \beta_y & 0 \\ -\gamma_x \beta_x & \gamma_x & 0 & 0 \\ -\gamma_x \gamma_y \beta_y & \gamma_x \gamma_y \beta_x \beta_y & \gamma_y & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \]

We see that both results, Eq. 2 and Eq. 3 are different which indicates that two arbitrary Lorentz transformations in general do not commute; the exception are two LT in same directions.

**End VIII–2: Non-commuting LT**

### 20.5 Metric invariance

Our objective is to learn how the metric changes under Lorentz transformations. We consider the invariance of proper time seen by two different observers with two coordinate systems connected by the Lorentz boost

\[ s^2 = c^2 \tau^2 = \sum_{\mu=0}^{3} x'_\mu x'^{\mu} = \sum_{\kappa,\mu=0}^{3} x'^{\kappa} g'_{\kappa \mu} x'^{\mu}. \] (20.27)
By substituting in Eq. (20.18) and rearranging the terms we obtain
\[ s^2 = \sum_{\nu,\lambda=0}^{3} x^\nu \left( \sum_{\kappa,\mu=0}^{3} \Lambda^\kappa_{\nu} g'_{\kappa\mu} \Lambda^\mu_{\lambda} \right) x^\lambda . \]  

(20.28)

We thus have found the transformed metric, using Einstein summation convention
\[ g_{\nu\lambda} = \Lambda^\kappa_{\nu} g'_{\kappa\mu} \Lambda^\mu_{\lambda} . \]

(20.29)

This equation can also be written using matrix notation,
\[ g = \Lambda^\top g' \Lambda , \quad g' = (\Lambda^\top)^{-1} g \Lambda^{-1} , \]

(20.30)
the last equality obtained applying the inverse matrix, Eq. (20.24), from right and left. For the Lorentz-boost we have \( \Lambda^\top = \Lambda \), Eq. (20.26), and hence we obtain considering as first example a boost, here in \( z \)-direction
\[
\begin{bmatrix}
\gamma^2 (1 - \beta^2) & 0 & \beta \gamma^2 & -\gamma^2 (1 - \beta^2) \\
0 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & -1 & 0
\end{bmatrix}
= \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & -1
\end{bmatrix} = g . \]

(20.31)

The above argument applies to any of the three boosts. Consider now without loss of generality a Lorentz transformation that can be written as product of two boosts, \( \Lambda = \Lambda_i (v_i) \Lambda_j (v_j) \). Using
\[
(\Lambda_i \Lambda_j)^\top = \Lambda_j^\top \Lambda_i^\top = \Lambda_j \Lambda_i , \quad ((\Lambda_i \Lambda_j)^\top)^{-1} = (\Lambda_j^\top \Lambda_i^\top)^{-1} = (\Lambda_i^\top)^{-1} (\Lambda_j^\top)^{-1} , \]

(20.32)
we obtain for Eq. (20.30)
\[
g' = (\Lambda^\top)^{-1} g \Lambda^{-1} = (\Lambda_i^\top)^{-1} \left\{ (\Lambda_j^\top)^{-1} g \Lambda_j^{-1} \right\} \Lambda_i^{-1} = (\Lambda_i^\top)^{-1} g \Lambda_i^{-1} = g . \]

(20.33)

This procedure can be of course be continued to include a LT that is product of any number of different boosts.
We have therefore obtained the interesting result
\[ g' = g . \]  
(20.34)
showing that LT is a generalized rotation since the metric remains invariant. This means the metric is the same independent of the observer frame of reference, confirming \( g \) as a property of space-time. Therefore name ‘metric’ for \( g \) is justified.

***Exercise VIII–3: Invariance of 4-volume***

Show Lorentz invariance of the volume element \( d^4x = c \, dt \, d^3x \).

**Solution**

Any transformation of coordinates

1. \( x' \mu = \Lambda^{\mu}_{\alpha} x^\alpha \),

requires for purpose of evaluation of a 4-volume integral a transformation of the 4-volume element according to

2. \( d^4 x' = \left| \frac{\partial x' \alpha}{\partial x^\mu} \right| d^4 x \)

where the coefficient is the Jacobian determinant. We recognize looking back at section 20.4.

3. \( \frac{\partial x' \alpha}{\partial x^\mu} = \Lambda^{\alpha}_{\mu} \).

\( \partial x^\mu \) transforms just as \( x^\mu \) under LT and only one out of three spatial coordinates transforming under Lorentz boost. Suppressing for purpose of computation of the determinant the two non-transforming coordinates we have for a boost in direction ‘i’

4. \( \left| \frac{\partial x' \alpha}{\partial x^\mu} \right| = |(\Lambda^\alpha_{\mu})|_i = \left| \begin{pmatrix} \gamma_i & -\beta_i \gamma_i \\ -\beta_i \gamma_i & \gamma_i \end{pmatrix} \right| = \gamma_i^2 (1 - \beta_i^2) = 1 \quad i = 1, 2, 3 . \)

Thus Eq. 2 reads

5. \( d^4 x' = d^4 x ; \quad dt' d^3 x' = dt d^3 x . \)

***End VIII–3: Invariance of 4-volume***
20.6 Finding new 4-vectors and invariants

The Lorentz transformation property of the 4-position vector \( x^\mu \) can be used to recognize other 4-vectors. We denote the general contravariant 4-vector \( W^\mu \),
\[
W^\mu = \{W^0, W^1, W^2, W^3\},
\]
and the corresponding covariant 4-vector
\[
W_\mu = \{W_0, W_1, W_2, W_3\} = g_{\mu\nu}W^\nu = \{W^0, -W^1, -W^2, -W^3\},
\]
and thus we have
\[
\]
Let us assume that the Lorentz-transformed vector is
\[
W'^\mu = \Lambda^\mu_\nu W^\nu.
\]
If we substitute \( W^\mu \) for \( x^\mu \) in the proof of the invariance of \( c^2\tau^2 \) in equations 20.28 through 20.34, we immediately obtain the Lorentz invariance of the quantity
\[
W \cdot W = W_\mu W^\mu = W'^\mu W'^\mu,
\]
which follows from the invariance of the metric, Eq. (20.34). Thus the assumption of the transformation property Eq. (20.36) establishes the 4-magnitude \( W \cdot W \), Eq. (20.37), as a Lorentz invariant.

However, we must check assembling four quantities into the 4-vector \( W^\mu \), Eq. (20.35a), for their 4-vector nature by considering their Lorentz transformation properties. Only very special sets of four quantities \( \{W^0, W^1, W^2, W^3\} \) obtained in a reference frame \( S \) transform into corresponding quantities \( W'^0, W'^1, W'^2, W'^3 \) as measured in another frame of reference \( S' \) according to a Lorentz transformation, that is
\[
S \xrightarrow{\Lambda} S', \quad \Lambda \begin{pmatrix} a^0 \\ a^1 \\ a^2 \\ a^3 \end{pmatrix} = \begin{pmatrix} a'^0 \\ a'^1 \\ a'^2 \\ a'^3 \end{pmatrix}.
\]
This is the required condition for \( W^\mu \) to be a 4-vector, where \( \Lambda \) is the boost from \( S \) to \( S' \).

In the following exercises we learn how to construct 4-vectors and we explore some of their properties. We also will see an example of an object that is not a 4-vector.
Show that for any two 4-vector $a^\nu$ and $b^\nu$, $a \cdot b = \sum_\nu a^\nu b_\nu$ is also a LI.

**Solution**

In a frame corresponding to the boost $\Lambda^\mu_\nu$, the transformed contravariant components of the 4-vector $a^\mu$ are

1. $a'^\mu = \sum_\alpha \Lambda^{\mu}_\alpha a^\alpha$,

and the transformed covariant components of the vector $b^\mu$ are

2. $b'_\mu = \sum_\nu g_{\mu\nu} b'^\nu = \sum_\nu g_{\mu\nu} \left( \sum_\kappa \Lambda^{\nu}_\kappa b^\kappa \right)$.

The invariant in the boosted frame is then

3. $\sum_\mu a'^\mu b'_\mu = \sum_{\alpha,\kappa} a^\alpha b^\kappa \left( \sum_\nu g_{\mu\nu} \Lambda^{\mu}_\alpha \right)$.

In the parenthesis we see the transformed metric Eq. (20.34), and due to its invariance property, Eq. (20.34) we have

4. $\sum_\mu a'^\mu b'_\mu = \sum_{\alpha,\kappa} a^\alpha g_{\kappa\alpha} b^\kappa = \sum_\alpha a^\alpha b^\alpha$.

Thus $a \cdot b$ is a Lorentz-invariant.

---

**Exercise VIII–5: Constructing new 4-vectors**

Show that any linear combination of several 4-vectors $a^\mu_i$, i.e. $A^\mu = \sum_i c_i a^\mu_i$, where $c_i$ are arbitrary coefficients which do not Lorentz transform, is also a 4-vector.

**Solution**

Because the LT is linear, we know that $W^\mu$ is defined consistently between reference frames. We can see this by rearranging the sums in the transformation:

1. $A'^\mu = \sum_\nu \Lambda^{\mu}_\nu \left( \sum_i c_i a^\nu_i \right) = \sum_i c_i \left( \sum_\nu \Lambda^{\mu}_\nu a^\nu_i \right) = \sum_i c_i a'^\mu_i$.

Thus a linear superposition of 4-vectors is observed by another observer $S'$ to be the same linear superposition of transformed 4-vectors. Thus if $A^\mu$ has ‘good’ properties, so will $A'^\mu$. 
The one property that is essential is that $A \cdot A$ is a Lorentz invariant. We have

$$A \cdot A = \sum_{\nu} A^\nu A_\nu = \sum_{\nu} \left( \sum_i c_i a^\nu_i \right) \left( \sum_j c_j a_{j\nu} \right)$$

\[= \sum_{i,j} c_i c_j \left( \sum_{\nu} a_i^\nu a_{j\nu} \right) = \sum_{i,j} c_i c_j a_i \cdot a_j.\]

We have shown in exercise VIII–4 that for any two 4-vectors $a^\nu$ and $b^\nu$, $a \cdot b$ is an invariant. $A \cdot A$ consists of all possible pairs $(i, j)$, weighted by $c_i c_j a_i \cdot a_j$, and summed over. A sum over invariants is itself an invariant. We conclude that any linear combination of 4-vectors is itself a 4-vector and its square presents us with a new invariant.

---

**Exercise VIII–6: Not every ‘4-thing’ is a 4-vector**

Given two 4-vectors $a^\nu$ and $b^\nu$ show an example of an object that clearly cannot be a 4-vector.

**Solution**

Any ‘false’ combination of the components of these two 4-vectors will in general not be a 4-vector. Such an example is

$$1 \quad F = \begin{pmatrix} a^0 + |\vec{b}| \\ b^0 + a^0 \\ b^0 + a^0 \\ b^0 + a^0 \end{pmatrix},$$

where we have mixed timelike and spacelike components of 4-vectors $a^\nu$ and $b^\nu$. This is highly unusual procedure. A quick check if this makes good sense is to form

$$2 \quad F \cdot F = (a^0 + |\vec{b}|)^2 - (b_x + a^0)^2 - (b_y + a^0)^2 - (b_z + a^0)^2.$$

A short algebraic computation shows

$$3 \quad F \cdot F = 2a^0 \left( |\vec{b}| - b_x - b_y - b_z - \frac{n-1}{2} a^0 \right),$$

where $n$ is the number of spatial dimensions, here $n = 3$. The consideration of $n = 1$ generates $F \cdot F = 0$ which is invariant. However, for the more general case here considered as our example, we clearly see that $F \cdot F$ cannot be an invariant. Still, the case $n = 1$ introduces small uncertainty and we want to reassure ourselves.
Thus we proceed by looking at explicit LT: For the object \( F \), Eq. (1), to be a 4-vector we must have under \( z \)-boost

\[
\begin{pmatrix}
a^0 + |\vec{b}| \\
\gamma(a^0 + |\vec{b}|) - \gamma \beta (b^2 + a^0) \\
\gamma(b^2 + a^0) - \gamma \beta (a^0 + |\vec{b}|)
\end{pmatrix} = \begin{pmatrix}
a^0' + |\vec{b}'| \\
\gamma(a^0' + |\vec{b}'|) - \gamma \beta (b^2 + a^0) \\
\gamma(b^2 + a^0) - \gamma \beta (a^0' + |\vec{b}'|)
\end{pmatrix},
\]

where the last equality is obtained performing the \( z \)-boost on the individual components of both 4-vectors \( a^\nu \) and \( b^\nu \). For these boosted components we find applying the rules of 4-vector transformations

\[
\begin{pmatrix}
a^0 \\
a^x \\
a^y \\
a^z
\end{pmatrix} = \begin{pmatrix}
\gamma(a^0 - \gamma a^z) \\
\gamma a^z - \gamma a^0
\end{pmatrix},
\]

and similarly for \( b^\nu \).

Inserting the boosted values \( a' \) (and \( b' \)) from Eq. 5 into Eq. 4 we find the two consistency requirements

\[
\begin{pmatrix}
\gamma(a^0 + |\vec{b}|) - \gamma \beta (b^2 + a^0) \\
\gamma(b^2 + a^0) - \gamma \beta (a^0 + |\vec{b}|)
\end{pmatrix} = \begin{pmatrix}
\gamma a^0 - \gamma \beta a^z + \sqrt{b^2 + b^2 + (\gamma b^2 - \gamma \beta b^0)^2} \\
(\gamma b^2 - \gamma \beta b^0) + (\gamma a^0 - \gamma \beta a^z)
\end{pmatrix},
\]

which simplify to

\[
\begin{pmatrix}
-\gamma \beta (a^0 - a^z) \\
-\gamma \beta (a^0 - a^z)
\end{pmatrix} = \begin{pmatrix}
\sqrt{b^2 + b^2 + (\gamma b^2 - \gamma \beta b^0)^2 - \gamma |\vec{b}| + \gamma \beta b^2} \\
-\gamma \beta (b^0 - |\vec{b}|)
\end{pmatrix}.
\]

We see that both equations are inconsistent and clearly cannot be satisfied. Thus the 4-thing \( F \) is not a 4-vector.

---

End VIII–6: Not every ‘4-thing’ is a 4-vector

21 Four-Velocity and Four-Momentum

21.1 Four-velocity \( u^\mu \)

We have introduced the 4-vectors associated with an event in Minkowski space having four coordinates \( ct, x, y, \) and \( z \) in section 20. This 4-vector of position \( x^\mu \) can be used to describe a particle’s motion through space-time. To do so, we
Figure 20-2: The world line of a particle and its position 4-vector $x^\mu(t)$.

consider the spatial position as a function of time $t$. That means that we form a world line parametrized as a function of $t$:

$$x^\mu(t) = \{ct, \vec{x}(t)\}.$$  \hspace{1cm} (21.1)

A world line is a sequence of events following each other in time. Since we have shown that we cannot loop back in time and that the sequence of events remains unchanged under LT (this is the causality), this parametrization which uses this specific coordinates, the time, as an evolution parameter, is allowed.

We thus refer to the entire world line of the particle, as is illustrated in figure 20-2 and characterize it in terms of parameter $t$.

We seek now to identify a ‘natural’ 4-vector $u^\mu$ characterizing the particle’s velocity, that transforms in the usual way as a 4-vector

$$u'^\mu = \Lambda^\mu_\nu u^\nu,$$  \hspace{1cm} (21.2)

and which contains in the non-relativistic limit the usual velocity.

The lesson of exercise [VIII–7] presented below is that in order to arrive at a covariant velocity 4-vector we need to take the derivative of 4-position with respect to an invariant quantity. It seems natural to try to use the proper time $\tau$ Eq. (20.2) for this purpose. Consider two neighboring events, $x^{1\mu}, x^{2\mu},$ such that

$$(\Delta s)^2 = c^2 \Delta \tau^2_{1,2} = (x^{1\mu} - x^{2\mu})(x^{1}_{\mu} - x^{2}_{\mu}) = c^2(t^1 - t^2)^2 - (\vec{x}^1 - \vec{x}^2)^2.$$  \hspace{1cm} (21.3)

With $dx^\mu = (c \, dt, d\vec{x})$ where $t^2 - t^1 \equiv dt$ and $\vec{x}^2 - \vec{x}^1 \equiv d\vec{x}$, we also can write the above in the form

$$ds^2 = c^2(d\tau)^2 = dx^\mu dx_\mu = c^2 dt^2 - (d\vec{x})^2.$$  \hspace{1cm} (21.4)
As $dx^\mu$ is the difference of two position 4-vectors, it is also a 4-vector and hence $d\tau$ is a Lorentz invariant, like $\tau$. Thus $dx^\mu/d\tau$ is a 4-vector. We have

$$\frac{dx^\mu}{d\tau} = \left\{ \frac{c}{d\tau}, \frac{d\vec{x}}{d\tau} \right\} = \left\{ \frac{c}{d\tau}, \frac{d\vec{x}}{d\tau}, \frac{dt}{d\tau} \right\} = \frac{dt}{d\tau} \{c, \vec{v}\}. \tag{21.5}$$

In general, $\vec{x}$ is given as a function of $t$ and $d\vec{x}/dt = \vec{v}$ is the usual laboratory velocity of a body. To evaluate $dt/d\tau$, we use Eq. (21.4); we factor out $(dt)^2$ and divide by $c^2$

$$(d\tau)^2 = (dt)^2 \left( 1 - \frac{1}{c^2} \left( \frac{d\vec{x}}{dt} \right)^2 \right) = (dt)^2 \left( 1 - \left( \frac{\vec{v}}{c} \right)^2 \right).$$

This yields

$$\frac{dt}{d\tau} = \frac{1}{\sqrt{1 - \vec{v}^2/c^2}} = \gamma, \quad dt = \gamma d\tau. \tag{21.6}$$

We used this same relationship in the study of the twin paradox; we further recall that the proper time of an object is always the shortest time – any observer other than the observer placed in the particle rest-frame measures a longer time.

For $dx^\mu/d\tau$ we obtain now, combining Eq. (21.5) and Eq. (21.6)

$$\frac{dx^\mu}{d\tau} = \left\{ \frac{c}{\sqrt{1 - \vec{v}^2/c^2}}, \frac{\vec{v}}{\sqrt{1 - \vec{v}^2/c^2}} \right\} \equiv u^\mu, \tag{21.7}$$

This equation then gives the 4-velocity $u^\mu$ of the particle at any point along its world line. A more compact form is

$$u^\mu \equiv \frac{dx^\mu}{d\tau} = c\{\gamma, \gamma \vec{\beta}\} \equiv \{u^0, \vec{u}\}, \quad u^2 = c^2, \tag{21.8}$$

where we also stated the result we show in exercise VIII–8 that the magnitude of the 4-velocity is a constant.

Choosing the coordinate system such the velocity vector is aligned with a coordinate axis, for example the direction of the $z$-axis, we can use particle rapidity to write the 4-velocity vector in the format

$$u^\mu = c\{\cosh y_p, 0, 0, \sinh y_p\}. \tag{21.9}$$

By construction, the 4-velocity $u^\mu$ is the derivative of the 4-position $x^\mu$ with respect to the particle’s proper time $\tau$. The magnitude of $u^\mu$ is a constant,
\(u^2 = c^2\). This means that all bodies travel through space-time with a magnitude of 4-velocity, the 4-speed

\[ u \equiv \sqrt{u^2} = c, \]  

(21.10)

of the same magnitude, the speed of light \(c\): even if we do not move in space, \(\vec{v} = 0\), Eq. (21.10) shows that there is an invariant of motion, hence it is possible to say that we ‘move’ in time \(u^0 = c\). When we have a proper velocity \(\vec{u} \neq 0\), that is we move in space, we move differently in time, such that always \(u^2 = c^2\).

**Exercise VIII–7: Four-velocity \(u^\mu\) cannot be a laboratory time derivative**

Show that it is not possible to define a 4-velocity \(u^\mu\) by applying to a position 4-vector a differentiation with respect to the laboratory time \(t\).

**Solution**

The three-velocity vector of the particle is defined as \(\vec{v} = d\vec{x}/dt\), which corresponds to the reciprocal slope of the world line depicted in figure [20-2]. However, the three-velocity does not transform in the same way as an event; see for example Eq. (7.9a), and thus it does not lend itself to a generalization to 4-velocity which we would like to be just like \(x^\mu\) a 4-vector.

We show this by contradiction to the assumption that 4-velocity is a regular time derivative of the position 4-vector, \(x^\mu\). Thus we explore

1. \(\tilde{u}^\mu \equiv \frac{dx^\nu}{dt}\).

If this were true, Eq. (21.2) implies that

2. \(\tilde{u}^{\prime \mu} = \Lambda^{\mu \nu} \frac{dx^{\prime \nu}}{dt} = \frac{d}{dt} \Lambda^{\mu \nu} x^{\nu} = \frac{dx^{\prime \mu}}{dt}\).

However, the 4-vector must be defined consistently between different inertial frames, so we must also have

3. \(\tilde{u}^{\prime \mu} \equiv \frac{dx^{\prime \mu}}{dt'}\).

Since \(dt' \neq dt\), we have

4. \(\frac{dx^{\nu}}{dt} \neq \frac{dx^{\prime \nu}}{dt'}\),

and we see that Eq. 2 and Eq. 3 can never be consistent with one another and thus assumption Eq. 1 is false by contradiction. \(\tilde{u}^\mu\) defined in Eq. 1 can not be a 4-vector.

**End VIII–7: Four-velocity \(u^\mu\) cannot be a laboratory time derivative**
Exercise VIII–8: Speed in Minkowski space

Determine the magnitude of the 4-velocity of any moving body and compare this with the result for a body at rest.

Solution

The magnitude of the 4-velocity does not depend on the frame of reference, thus we expect that the result is the same for an observer at rest with respect to the body, and an observer moving with velocity \( \vec{v} \).

We first evaluate the magnitude of the 4-velocity in the rest-frame of the body. For \( \vec{v} = 0 \) we have \( \gamma = 1 \) and,

\[
1 \quad u^\mu|_{\text{rest-frame}} = \{c, \vec{0}\}, \quad \rightarrow \quad u^2|_{\text{rest-frame}} \equiv \sum_{\mu=0}^{3} u^\mu u_\mu|_{\text{rest-frame}} = c^2 - \vec{0}^2 = c^2.
\]

In a frame of reference in which the body has a velocity \( \vec{v} \) we compute

\[
2 \quad u^2 \equiv \sum_{\mu=0}^{3} u^\mu u_\mu = c^2 = \gamma^2(c^2 - \vec{v}^2) = \frac{1}{1 - \vec{v}^2/c^2} c^2(1 - \vec{v}^2/c^2) = c^2.
\]

Since \( u^2 \) is a Lorentz-invariant we find the same result \( u^2 = c^2 \) in all frames of reference.

End VIII–8: Speed in Minkowski space

21.2 Energy-momentum 4-vector

The 4-velocity allows us to consider a simple corollary, the energy-momentum 4-vector \( p^\mu \). We multiply the 4-velocity Eq. (21.8) by the mass \( m \)

\[
m \frac{dx^\mu}{d\tau} = mu^\mu \equiv p^\mu. \tag{21.11}
\]

In explicit terms we have

\[
p^\mu = \left\{ \frac{mc}{\sqrt{1 - \vec{v}^2/c^2}}, \frac{m\vec{v}}{\sqrt{1 - \vec{v}^2/c^2}} \right\} \equiv \left\{ \frac{E}{c}, \vec{p} \right\}. \tag{21.12}
\]

We recall the key result of section 17.1, see Eq. (17.12) where we have shown that the energy locked in the rest mass of the particle is the same for all observers,
thus the rest mass is a Lorentz invariant. Now we see that this is so since it is a square of the 4-momentum:

\[ p \cdot p \equiv p^2 = m^2 u^2 = m^2 c^2 . \] (21.13)

The four-momentum \( p^\mu \) is a convenient quantity to extend the study of two particle kinematics we begun in section 19. A particularly useful tool are the Mandelstam variables \( s, t, u \) describing the two particle scattering with momenta \( p_1^\mu, p_2^\mu \) turning into two (eventually different) particles \( p_3^\mu, p_4^\mu \), as shown in figure 21-1. The scattering process can only depend on quantities that all observer agree are the same -- since if this would not be the case, different inertial observers could observe different outcome of the scattering process. We are interested to recognize all independent invariants that can be contracted.

Naively one could expect that any of the products \( p_i \cdot p_j \) is an independent invariant, and this would lead for \( i, j = 1 \ldots 4 \) to 10 invariants, that is 10 numbers that describe a scattering process. However, the conservation of four-momentum (energy conservation, and three-momentum conservation) now written in the simple format

\[ p_1^\mu + p_2^\mu = p_3^\mu + p_4^\mu, \] (21.14)

introduces four constraints. Furthermore, we are given four masses (keep in mind here \( i \) counts particles involved in scattering process)

\[ p_i^2 = m_i^2 c^2 , \quad i = 1, 2, 3, 4 . \] (21.15)

We are thus looking for two \((10-4-4=2)\) variables only. These have to be written in a Lorentz invariant way.

### 21.3 Properties of Mandelstam variables

In classical mechanics the scattering process is characterized in terms of two variables, the energy of the incoming particle, and the laboratory scattering angle; both are not Lorentz invariant. A complicated set of transformations needs to be
developed to understand how different observers measure the outcome of the one and the same scattering process, and this is true also in non-relativistic limit.

In the context of special relativity the challenge is to find, aside of the four particle masses, these two new invariant quantities that characterize the scattering process. Mandelstam\textsuperscript{1} recognized the form of variables that are invariant and can be used instead. We show by construction how these arise.

The energy-momentum conservation Eq. (21.14) can be written in three equivalent forms

\[ p_1^\mu + p_2^\mu = p_3^\mu + p_4^\mu , \]

\[ p_1^\mu - p_3^\mu = p_4^\mu - p_2^\mu , \]

\[ p_1^\mu - p_4^\mu = p_3^\mu - p_2^\mu , \]

where the last two equations arise from simple right to left shift of terms from the first. The Lorentz-invariants that are known as the Mandelstam-variables \( s, t, u \) are obtained squaring these three different 4-vectors shown on right- and left- hand side in Eq. (21.16):

\[ s \equiv c^2 (p_1 + p_2)^2 = c^2 (p_3 + p_4)^2 , \]

\[ t \equiv c^2 (p_1 - p_3)^2 = c^2 (p_2 - p_4)^2 , \]

\[ u \equiv c^2 (p_1 - p_4)^2 = c^2 (p_2 - p_3)^2 . \]

Either of the each two formats on the right hand side for each variable can be used.\textsuperscript{2} The variable \( s \) appeared before in Eq. (18.6), and in section 19.3.

Since only two invariants should be of relevance as discussed, these three invariants are linearly dependent. To obtain the constraint we square the momentum conservation Eq. (21.16) written in the form

\[ 0 = (p_1^\mu + p_2^\mu - p_3^\mu - p_4^\mu)^2 \rightarrow \]

\[ 0 = p_1^2 + p_2^2 + p_3^2 + p_4^2 + 2p_1 \cdot p_2 - 2p_1 \cdot p_3 - 2p_1 \cdot p_4 - 2p_2 \cdot p_3 - 2p_2 \cdot p_4 + 2p_3 \cdot p_4 . \]  

This is a fancy way to write a ‘zero’. Next, we add up the three forms of Eq. (21.17) in explicit computation using once both the right, and the left hand


\textsuperscript{2}The use of \( s, t \) and \( u \) symbols for Mandelstam variables can in principle be confused with relativistic distance \( s \), time \( t \) and 4-velocity \( u \). The contextual meaning will be always clear or made clear in this book.
side, and obtain for the sum of all these terms

\[ 2(s + t + u) = 3c^2(p_1^2 + p_2^2 + p_3^2 + p_4^2) \]

\[ + 2c^2(p_1 \cdot p_2 - p_1 \cdot p_3 - p_1 \cdot p_4 - p_2 \cdot p_3 - p_2 \cdot p_4 + p_3 \cdot p_4) . \]

We thus recognize in the last line the second part of the fancy ‘zero’, Eq. (21.18).

Upon its elimination we obtain

\[ 2(s + t + u) = 2c^2(p_1^2 + p_2^2 + p_3^2 + p_4^2) . \]

(21.20)

Upon division by factor two we are left with

\[ s + t + u = \sum_{i=1}^{4} (m_i c^2)^2 . \]

(21.21)

We learn that the scattering process is described by four particle masses, and two among the three parameters \( s, t, u \) seen in Eq. (21.17), thus a total of 6 parameters.

For elastic collisions where final state masses \( m_2^2, m_4^2 \) are the same as the initial state \( m_1^2, m_2^2 \) the constraint Eq. (21.21) is

\[ s + t + u = 2(m_1 c^2)^2 + 2(m_2 c^2)^2 \quad \text{(elastic collisions)} . \]

(21.22)

Further reading: The Mandelstam variables \( s, t \) and \( u \) play an important role in the physics of elementary particles: the quantum mechanical scattering amplitude \( T \) can be written as an analytical function \( T(s, t, u) \) and permutations of variables, i.e. \( T(t, u, s), T(u, s, t) \) describes ‘crossed’ amplitudes that characterize diagrammatically a 90° ‘rotated’ process. The interested reader should consult the specialized texts for further study\(^3\).

---

**Exercise VIII–9: Meaning of Mandelstam variables**

Identify the relation of the Mandelstam variables with the familiar physical quantities.

**Solution**

We have introduced in section 21.2 for two bodies the variable \( s \). However, \( s \) defined as a square of a 4-vector is an invariant irrespective of how many particles ‘\( i \)’ are included in the total 4-momentum \( p^\mu \)

\[
\begin{align*}
s &= c^2 p_\mu p^\mu \\
&\rightarrow c^2 (p_{1\mu} + p_{2\mu} + \ldots)(p^{1\mu} + p^{2\mu} + \ldots) \\
&= c^2 P_\mu P^\mu = c^2 \left( \sum_i p_i \right) \left( \sum_i p^i \right) .
\end{align*}
\]

\(^3\)See for example: E. Byckling, and K. Kajantie *Particle Kinematics* John Wiley & Sons Ltd, New York (1973).
We see that the definition of \( s \) indeed applies to one, to two, and from several to many bodies.

Let us look at the meaning of \( s \) in the CM frame of reference.

\[ s = \left( \sum_i \sqrt{ (m_i c^2)^2 + p_i^2 c^2} \right)^2 - \left( \sum_i c p_i \right)^2. \]

The CM frame is where there is no motion of a particle system and therefore \( \vec{P} = \sum_i \vec{p}_i = 0 \). Thus the last term vanishes. The form of the first term in Eq. 2 reminds us that although the sum of momenta vanishes, the individual momentum of each particle will in general not vanish. Thus \( \sqrt{s} \) is the generalized ‘rest energy’ of an assembly of particles.

\[ \sqrt{s} = \sum_i E_i = \sum_i \sqrt{ (m_i c^2)^2 + p_i^2 c^2}. \]

When only one particle is present, \( \vec{p}_1 = 0 \) and Eq. 3 becomes \( \sqrt{s} = m c^2 \), the rest energy of the particle. For two particles in the CM frame the magnitude of the individual momentum must be the same, i.e. \( |\vec{p}_1| = |\vec{p}_2| = \vec{p} \) while their directions are opposite, thus we have (subscript ‘2’ denotes the 2-particle system).

\[ \sqrt{s_2} = \sqrt{(m_1 c^2)^2 + \vec{p}^2 c^2} + \sqrt{(m_2 c^2)^2 + \vec{p}^2 c^2}. \]

This reminds us of the 2-particle decay: we know from section 18.3 that when a particle with rest energy \( M c^2 \equiv \sqrt{s_2} \) decays into two particles, each shoots in the CM frame (rest-frame of decaying particle) in opposite direction, each carrying the energy shown in Eq. 4. Conversely the value of \( \sqrt{s_2} \) defines the energy that is available when two particles collide to transform into new particles or otherwise be of use.

These two special case confirm that Eq. 4 is a generalization of the concept of rest energy locked for one particle in the mass, and for two particles being the particle mass that can be produced, and we conclude that Eq. 3 describes the energy available in one, two and more particle systems.

We next consider the meaning of Mandelstam variable \( t \): if particle-1 continues as particle-3, that means that \( t \) describes how much momentum moved into particle-2 that continues as particle-4.

\[ t/c^2 = (p_1 - p_3)^2 = -Q^2 = - (q^2 - (E_1 - E_3)^2/c^2), \]

where we make explicit that in the invariant formulation, aside of momentum transfer \( q \), energy transfer is also included: particle-1 as it turns into particle-3 delivers also energy to the target particle-2. In relativistic collisions in which particle-2 is shattered...
into pieces, that is particle-4 is a flying cloud of pieces, the energy transfer term in $t$
can dominate. We speak in this case of deep inelastic collisions$^4$.

However, in non-relativistic elastic scattering the energy transfer (second term in
Eq. 5) compared to momentum transfer (first term in Eq. 5) is in general small, a
situation demonstrating the Mandelstam variable $t$ as being the generalized momentum
transfer, directly related to the elastic collision scattering angle.

---

End VIII–9: Meaning of Mandelstam variables

---

Exercise VIII–10: Lorentz invariance and the Compton condition

Use an invariant to derive the Compton condition obtained in section 19.6.

Solution

We consider the center of momentum (CM) frame (of the electron and the photon)
having the 4-velocity $u^\mu_{\text{cm}}$, and combine it with with the change in 4-momentum of the
scattered particle $\Delta p_1^\mu = p_1^\mu - p_1^\mu$ (here the photon) to form an invariant $I$

$$I \equiv \Delta p_1^\mu u_\mu = \Delta E_1 u^0 - \Delta \vec{p}_1 c \cdot \vec{u}.$$  

We evaluate $I$ in two different frames of reference : a) the CM frame, and b) the
laboratory frame where the scattering process is observed.

In the CM frame of reference there is no motion in space, $\vec{P} = 0$ and so

$$\vec{u}_{\text{cm}} = (c, 0), \quad I = \vec{u}_{\text{cm}} \Delta p_1^\mu = c(\Delta E_1 - \vec{E}_1') = 0$$

The last equality follows since the energy of each particle undergoing elastic scattering
is unchanged in the CM frame of reference. This is so since only the direction but not
the magnitude of the momentum changes.

When considered in the laboratory frame of reference,

$$\vec{u}_{\text{cm}}' = \gamma_{\text{cm}}c(1, \vec{\beta}_{\text{cm}}') .$$

We combine Eq. 3 with the change in 4-momentum expressed in laboratory frame to
form the invariant value

$$I = u^\mu \Delta p_\mu = \gamma_{\text{cm}} c(\Delta E_1/c - \Delta \vec{p}_1' \cdot \vec{\beta}_{\text{cm}}') .$$
\( I \) is a Lorentz invariant, with the same value observed by all observers in all frames of reference. Hence we can use Eq. 2 to set Eq. (4) to zero

\[ \Delta E_1/c - \Delta \vec{p}_1 \cdot \vec{\beta}_{cm} = 0. \]

Thus to understand the kinematic scattering condition we only need to know

\[ \vec{\beta}_{cm} = \frac{\vec{P}_c}{E}, \]

where the total momentum \( \vec{P} \) and total energy \( E \) of the colliding system is measured in the laboratory frame.

In the case of Compton scattering the scattering involves an electron target at rest in the laboratory, and the photon as the incoming particle

\[ \vec{P} = \vec{p}_1 + \vec{p}_2 = \vec{p}_1 + 0 = p_1 \hat{p}_1, \quad E = E_1 + E_2 = E_1 + m_e c^2, \]

where we used that particle ‘2’, the electron, is, in the case of Compton scattering, at rest in the laboratory frame.

Plugging Eq. 7 into Eq. 5 gives us

\[ \frac{(E_1 - E'_1)}{\Delta E_1} = \frac{c(p_1 - p'_1) \cdot \frac{cp_1 \hat{p}_1}{E_1 + m_e c^2}}{\Delta \vec{p}_1 c \vec{\beta}_{CM}}, \]

where the under braces show the origin of the terms, compare Eq. 5. We introduce the scattering angle of the photon \( \vec{p}_1 \cdot \hat{p}_1 = p'_1 \cos \theta \) and use \( cp_1 = E_1, \ cp'_1 = E'_1 \). Multiplying with \( E_1 + m_e c^2 \) we can simplify Eq. 8 to read

\[ \frac{E'_1 - E_1}{E'_1 E_1} = \frac{1}{m_e c^2} (\cos(\theta) - 1). \]

Using \( \lambda_1 = hc/E_1 \) this equation becomes the Compton condition

\[ \lambda'_1 - \lambda_1 = \frac{h}{m_e c} (1 - \cos \theta). \]
Exercise VIII–11: Distance traveled in space-time by an unstable particle

Determine the distance $D$ traveled by an unstable particle (e.g. a muon) in its proper lifespan $\tau_0$.

**Solution**

We return to the muon travel distance problem, which we have already considered in exercise II–2, exercise III–15, and discussed in several conversations in order to clarify with more advanced methods how it is possible for the muon created far-up in the atmosphere to reach the surface of the Earth.

**SOLUTION IN THE EARTH LABORATORY REFERENCE FRAME**

As reported by an Earth-bound laboratory observer, the muon is created in a cosmic particle collision in the upper atmosphere, some distance $D$ from the laboratory with four-momentum $p^\mu$ observed in the laboratory frame at laboratory time $t = 0$ corresponding to the muon birth at proper time $\tau = 0$. We recall the result of section 21.2

\[ p^\mu = m \frac{dx^\mu}{d\tau}. \]

In absence of a force the distance traveled by the particle is obtained from the first integral of Eq. 1.

\[ x^\mu = \frac{p^\mu}{m} \tau_0. \]

This is a Lorentz covariant equation describing the world line of the particle using laboratory coordinate system $x^\mu$ with the particle measured to have the 4-momentum $p^\mu$. Note that

\[ x^2 = \frac{p^2}{m^2} = \frac{c^2 \tau_0^2}{m^2} = c^2 \tau_0^2. \]

which clarifies the meaning of $\tau_0$ as the proper lifespan of the particle, a value which is the same for all observers.

Checking the ‘zero-th’ component of Eq. 2 we note that for the Earth-bound observer who measured the particle momentum $\vec{p}$, or equivalently, particle velocity $\vec{v}$, the elapsed time while the particle is traveling is

\[ ct = x^0 = \frac{E}{cm} \tau_0 = \gamma c \tau_0. \]

Checking the ‘spatial’ components of Eq. 2 we see that the range of the muon as reported by the same observer is

\[ D \equiv |\vec{x}| = |\vec{p}| \frac{\tau_0}{m} = \gamma v \tau_0. \]
This means that the muon, within its lifespan, can travel a distance $D$ as determined
by the initial momentum $|\vec{p}|$ or, equivalently, speed $v$ observed from Earth. Given a
proper lifespan $\tau_0$, $D$ grows without limit for $v \to c$. As the muon travels the observer
reports that, see Eq. 4, the time $\gamma \tau_0$ has elapsed in her reference frame; thus for this
observer the distance traveled is obtained combining Eq. 4 with Eq. 5.

$$ D = vt. $$

We recognize that a muon can penetrate the atmosphere and hit the Earth, since the
range is stretched by the Lorentz factor $\gamma$ which enters the kinetic momentum $|\vec{p}|$ of
the muon. The distance traveled by the muon was obtained here strictly as a solution
of equations of motion, and did not require any discussion of any effects due to changes
to the coordinate system.

This straightforward explanation relies on the fact that the momentum of a particle
is growing with the the Lorentz-factor $\gamma$, a fact that has been experimentally measured
and is incontestable. There is no space, or distance, or body ‘stretching’ that can be
introduced into this argument. If we need to use a catch phrase to explain the effect, it
should always be ‘time dilation’ as this effect enters into the theoretical argument that
leads us to the 4-energy-momentum growing with the Lorentz factor $\gamma$, see section [17].

**SOLUTION IN THE THE PARTICLE REFERENCE FRAME**

An interesting alternative view on this problem arises for an observer comoving with
the muon, that is we let the ‘Earth’-observer approach the particle. This alternate
collection must work if relativity is correct as long as we apply the laws of physics
correctly. For the following discussion it is important to remember that the observer
at rest in the laboratory synchronized his clock time with the birth event of the muon
in Earth laboratory frame of reference. Simultaneity cannot be present for any other
observer. Therefore, for the moving Earth observer the creation event of a particle will
be recorded at a displaced value of time coordinate compared to the earlier evaluation.

The following shows how one finds the same answer irrespective in which frame of
reference the measurement is considered as long as simultaneity of clock synchronization
is done in one and the same frame of reference. There will be an extra step as we need,
in the ‘moving Earth’ approach, to intersect two world lines (Earth and particle) in
order to determine the distance traveled by Earth-observer in the time corresponding
to the birth and death events of the particle.

The world lines of the Earth and of the muon seen by an observer comoving with
the muon are obtained applying a Lorentz transformation with

$$ \gamma = \frac{E}{mc^2}, \quad \gamma \beta = \frac{p}{mc}. $$

For the moving Earth we have

$$ t'_{(E)} = \gamma (t_{(E)} - \beta x_{(E)}/c); \quad x'_{(E)} = \gamma (x_{(E)} - \beta ct_{(E)}). $$

The world line of the muon is simpler; it is at rest, and the time is proper time,

$$ t'_{(\mu)} = \tau; \quad \vec{x}'_{(\mu)} = 0. $$
The world lines intersect when the moving Earth hits the muon at $x' = 0$, after time $\tau = \tau_0$ has elapsed

\[ t'_{(E,f)} = t'_{(\mu,f)} = \tau_0; \quad x'_{(E,f)} = x'_{(\mu,f)} = 0. \]

We have

\[ t_{(E,f)} = \tau_{(E)} = \gamma \tau_0; \quad x_{(E,f)} = -\gamma \beta c \tau_0. \]

These are the same relations as Eq. 4 and Eq. 5, but the sign of $x_{(E,f)}$ indicates that we now look in the opposite direction.

As a cross-check we also obtain the coordinates of the Earth at the world event at which the muon was created, that is for $t_{(E,0)} = t_0 = 0$, and $x_{(E,0)} = D$. In the reference frame of the observer comoving with the muon this corresponds, see Eq. 8, to the initial coordinates

\[ ct'_{(E,0)} = -\gamma \beta D; \quad x'_{(E,0)} = \gamma D, \]

and thus the final coordinates

\[ ct'_{(E,f)} = \gamma (c\tau_{(E)} - \beta D) = \tau_0; \quad x'_{(E,f)} = \gamma (D - \beta c \tau_{(E)}) = 0. \]

Between the birth event and the collision event the earth coordinates change as follows:

\[ \Delta t' = t'_{(E,f)} - t'_{(E,0)} = \tau_0 + \gamma \beta D/c; \quad \Delta x' = x'_{(E,f)} - x'_{(E,0)} = -\gamma D. \]

We find the proper time elapsed relative to an Earth’s observer by evaluating

\[ \tau_{(E)} = \sqrt{\Delta t'^2 - \Delta x'^2/c^2} = \sqrt{\tau_0^2 + 2\gamma \beta \tau_0 D/c - D^2/c^2} = \sqrt{\tau_0^2 + D^2/c^2}, \]

where to obtain the last equality we used the second term in Eq. 13, i.e. $\tau_0 = D/(\gamma \beta c)$, which is plugged into the cross term in Eq. 15. When rearranged this results in

\[ c^2 \tau_0^2 = c^2 \tau_{(E)}^2 - D^2, \]

We recognize Eq. 16 to be the same as Eq. 3.

End VIII–11: Distance traveled in space-time by an unstable particle
### 22 Acceleration and Relativistic Mechanics

#### 22.1 Small acceleration

We have acknowledged the presence of acceleration when we considered the Bell-rocket example, and more generally each time we spoke of relativistic effects that impact the property of a body. We considered this acceleration with reference to the inertial observer’s frame of reference and introduced it as means of moving a body from one to another inertial system in small steps. The magnitude of the acceleration is assumed to be as small as needed; yet crucially, we have always been able to tell which of the compared bodies has been accelerated.

This of course means that the accelerated and inertial observers are not equivalent, we will say there is no relativity between accelerated and inertial observers: unlike the case of two inertial observers where the measured velocity is relative, the concept of acceleration is not-relative — in this book we have argued that we can tell which body is inertial and which body is accelerated.

In order to tell what ‘small’ means we need to have a reference benchmark. We look for natural scales of length and time to construct a ratio that describes the acceleration. The situation with a material body is different from that of an elementary particle as in the latter case the choice of a scale is usually not in question, and is related either to the (inverse) mass or, if available, some internal structure parameter. Once we agree which is the ‘natural’ object of size $L$, the large ‘critical’ acceleration is by dimensional counting

$$ a_{cr} = \frac{c^2}{L} . $$

(22.1)

We first consider a material object where we choose as the natural length scale $L$ the macroscopic thickness in direction of acceleration. For an object with about $L_{\text{matter}} = 10^{-3}$ m, the top laboratory acceleration that can be created today is $a \approx 10^{11}$ m/s$^2$. Considering Eq. (22.1) this is $10^{-9} a_{cr}^{\text{matter}}$. It is possible to make the reference critical acceleration $a_{cr}^{\text{matter}}$ bigger by choosing smaller value of $L_{\text{matter}}$ in Eq. (22.1). This will lead to even a smaller value of achieved acceleration compared to the reference critical value. Thus we conclude that within the limits of present technology

$$ a_{\text{matter}} < 10^{-9} a_{cr}^{\text{matter}} . $$

(22.2)

Turning to elementary particles, the natural scale of the length is the (reduced) Compton wavelength $\lambda_C = \lambda_C/2\pi$ and the natural unit of time is $\tau_C$, the time

---

needed by light to travel across this distance $\lambda_C$ where

$$\lambda_C = \frac{\hbar}{mc} = 386.16 \text{ fm}, \quad \tau_C = \frac{\lambda_C}{c} = 1.288 \times 10^{-21} \text{ s}.$$  \hspace{1cm} (22.3)

Thus for a particle of mass $m$ we adopt

$$a_{cr} \equiv \frac{c^2}{\lambda_C} = mc^2 \frac{c}{\hbar}.$$  \hspace{1cm} (22.4)

For an electron,

$$a_{cr}^e = 2.331 \times 10^{29} \text{ m/s}^2.$$  \hspace{1cm} (22.5)

An electron moving with speed of light across a maccroscopic bending magnetic field of 4.414 T experiences an acceleration

$$a_{\text{particle}} = ceB < 10^{-9} a_{cr}^\text{particle}.$$ \hspace{1cm} (22.6)

compare Eq. (29.4) on page 444; we return to this question in much more depth in section \[29\]. It is of course possible to make the reference acceleration bigger by choosing shorter ‘natural’ particle size, for example the classical electron radius $r_e = \lambda_C/137$ could be such a choice, yet this does not alter the fact that in laboratory conditions elementary particles experience nano-scale acceleration.

We see that both in macroscopic and elementary domains the most extreme acceleration that can be achieved today in laboratory is at least eight orders of magnitude below acceleration that one could call ‘strong’. Furthermore extremely small acceleration is present when testing and exploring SR. For example SR is tested in Earth’s surface reference frame, where we are 28 orders of magnitude below what one could argue is ‘strong’ acceleration. Equipped with this insight we now understand why it is possible to achieve very high precision verification of SR as we discussed in section \[14\]. It is in this ‘small’ acceleration context that we now work towards the introduction of a relativistic form of Newton’s law, postponing discussion of the problems that large acceleration can produce to the final section \[29\].

### 22.2 Definition of four-acceleration

We have previously addressed the example of a rocket engine which converts fuel mass into usable energy and momentum, see section \[17.3\], leading to continuous change of rocket speed. To describe this situation in covariant fashion we need to introduce the acceleration 4-vector. We proceed in the same way as argued in section \[21.1\] introducing the 4-velocity. Here, we differentiate the velocity 4-vector with respect to particle proper time, to obtain the relativistic acceleration 4-vector

$$a^\mu \equiv \frac{du^\mu}{d\tau}.$$ \hspace{1cm} (22.7)
This definition makes the acceleration 4-vector $a^\mu$ ‘belong’ to the body as it measures the change of its 4-velocity vector $u^\mu$ per unit of body proper time.

Not all components of this new 4-vector $a^\mu$ are independent dynamically. Since the 4-velocity magnitude is a constant, Eq. (21.8), we can write

$$u^2 = c^2 \longrightarrow \frac{d}{d\tau} u^2 = \frac{d}{d\tau} u_\mu u^\mu = \frac{d}{d\tau} c^2 = 0.$$  (22.8)

Differentiating the parts individually with respect to proper time and applying the product rule we obtain

$$0 = \frac{d}{d\tau} (u_\mu u^\mu) = \frac{d u^\mu}{d\tau} u_\mu + \frac{d u_\mu}{d\tau} u^\mu = 2 \frac{d u^\mu}{d\tau} u_\mu = 2 a \cdot u.$$  (22.9)

We found a new and useful invariant relation

$$u^2 = c^2 \longrightarrow a \cdot u \equiv \frac{d u^\mu}{d\tau} u_\mu = 0.$$  (22.10)

Thus the 4-acceleration is orthogonal to the 4-velocity. Just like the 4-velocity having in view of Eq. (21.8) three independent components, the 4-acceleration in view of Eq. (22.10) also has only three independent components.

Given $u^\mu = \gamma(c, \vec{v})$, we evaluate $a^\mu$ explicitly from Eq. (22.7). Using the chain rule we obtain

$$a^\mu = \frac{dt}{d\tau} \frac{du^\mu}{dt} = \frac{dt}{d\tau} \frac{d}{dt} \left[ \gamma(c, \vec{v}) \right].$$  (22.11)

Noting $dt/d\tau = \gamma$, Eq. (21.6), and using the product rule, we obtain

$$a^\mu = \frac{1}{2} \left( \frac{d}{dt} \gamma^2 \right) \{c, \vec{v}\} + \gamma^2 \{0, \vec{a}\}, \quad \vec{a} \equiv \frac{d\vec{v}}{dt},$$  (22.12)

where just as we use the classical velocity $\vec{v}$ vector, we also introduced the classical acceleration vector $\vec{a}$, which must be distinguished from $\vec{a}_4$, the spatial components of the 4-acceleration, which is used rarely. We recognize that in general $\vec{a}_4$ is not parallel to $\vec{a}$. This differs from the cases of the 4-position and 4-velocity, where the spatial components of the 4-vector are proportional to their respective three-vectors.

Evaluating the derivatives in Eq. (22.12) yields explicitly

$$a^\mu = \left\{ \frac{(\vec{v}/c) \cdot \vec{a}}{(1 - \vec{v}^2/c^2)^2}, \frac{\vec{a}}{1 - \vec{v}^2/c^2} + \frac{(\vec{v}/c) [(\vec{v}/c) \cdot \vec{a}]}{(1 - \vec{v}^2/c^2)^2} \right\}. \quad (22.13)$$

In the instantaneous rest-frame where $\vec{v} = 0$ Eq. (22.13) shows that there can be non-vanishing acceleration $a^\mu = \{0, \vec{a} = \vec{a}_4\}$. This also demonstrates that $a^\mu$ is spacelike vector, more generally

$$a \cdot a = -\gamma^4 \vec{a}^2 - \gamma^6 \frac{(\vec{v} \cdot \vec{a})^2}{c^2} < 0, \quad a \cdot \frac{\vec{v}}{\vec{v} \parallel \vec{a}} = -\gamma^6 \vec{a}^2.$$  (22.14)
A yet more general way to argue that $a^\mu$ is spacelike is to note that Eq. (22.10) in the rest frame of the particle where $u^\mu|_0 = \{c, \vec{0}\}$ requires $a^0|_0 = 0$, since $0 = u \cdot a = u \cdot a|_0 = ca^0|_0$. If in some frame one can find a vanishing $a^0 = 0$, the vector is either null, or spacelike. We conclude $u \cdot u = c^2 \rightarrow u \cdot a = 0$ leads on to $a \cdot a < 0$, see exercise VIII–13 for further related insights.

Exercise VIII–12: Orthogonality of $u^\mu$ and $a^\mu$

Using Eq. (22.13) show $u \cdot a = u_\mu a^\mu = 0$.

Solution

We form the product of Eq. (21.8) on page 292 with Eq. (22.13) and obtain:

$$1 \quad u \cdot a = \gamma \left( \frac{\vec{v} \cdot \vec{a}}{(1 - \vec{v}^2/c^2)^2} - \frac{\vec{v} \cdot \vec{a}}{1 - \vec{v}^2/c^2} - \frac{(\vec{v}^2/c^2)\vec{v} \cdot \vec{a}}{(1 - \vec{v}^2/c^2)^2} \right).$$

The first and last term can be combined, canceling one of the powers in the denominators and we obtain

$$2 \quad u \cdot a = \gamma \left( \frac{\vec{v} \cdot \vec{a}}{1 - \vec{v}^2/c^2} - \frac{\vec{v} \cdot \vec{a}}{1 - \vec{v}^2/c^2} \right) = 0.$$

End VIII–12: Orthogonality of $u^\mu$ and $a^\mu$

Exercise VIII–13: Why is $a^\mu$ spacelike

Present a general reason why 4-acceleration $a^\mu$ must be a spacelike vector.

Solution

We aim to show that any 4-vector $A^\mu$ that is 4-orthogonal to some timelike 4-vector $U^\mu$, is spacelike. We use for our timelike 4-vector the symbol $U^\mu$ in analogy to the 4-velocity which is timelike $u^2 = (u^0)^2 - \vec{u}^2 = c^2 > 0$. Thus we have for any timelike vector

$$1 \quad U^0 > |\vec{U}|, \quad \frac{U^0}{|U|} > 1.$$

The proof follows from simple algebraic manipulation. Given

$$2 \quad A \cdot U = 0 \rightarrow A_0 U^0 = \vec{A} \cdot \vec{U} \rightarrow 1 = \frac{\vec{A} \cdot \vec{U}}{A_0 U^0} < \frac{|\vec{A}| |\vec{U}|}{A_0 U^0} < \frac{|\vec{A}| |\vec{U}|}{A_0 U^0} |U|.$$
where in the last step we used Eq. 1. We now cancel the common factor in last term in
Eq. 2 to obtain upon squaring

\[ A_0 < |\vec{A}|, \quad (A_0)^2 < (\vec{A})^2, \quad \rightarrow A \cdot A < 0. \]

We see in Eq. 3 that any 4-vector \( A^\mu \) orthogonal to a timelike vector \( U^\mu \) must be
spacelike.

We recall that the orthogonality of \( a^\mu \) with timelike \( u^\mu \) is based in the physics rule
that the speed of light is a constant of nature, see Eq. (22.9). The argument thus goes:
since the speed of light is a universal constant, the 4-acceleration \( a^\mu \) is 4-orthogonal to
the timelike 4-velocity \( u^\mu \), and hence 4-acceleration \( a^\mu \) must be a spacelike vector.

End VIII–13: Why is \( a^\mu \) spacelike

Exercise VIII–14: \( a^\mu \) and particle rapidity

For the case that \( \vec{a} \parallel \vec{v} \) simplify the presentation of 4-acceleration using the particle
rapidity \( y_p \).

**Solution**

We begin with the 4-velocity in the form Eq. (21.9), where the velocity is aligned with
the \( z \)-axis

\[ u^\mu = c \{ \cosh y_p, 0, 0, \sinh y_p \}. \]

Only in case that the force that imposes a change in 4-velocity is such that the motion
remains aligned with the \( z \)-axis we can find the 4-acceleration vector by differentiation
of Eq. 1 with respect to the proper time

\[ a^\mu = \frac{dy_p}{d\tau} c \{ \sinh y_p, 0, 0, \cosh y_p \}, \quad \vec{a} \parallel \vec{v}. \]

Combining Eq. 1 and Eq. 2 we confirm \( a \cdot u = 0 \). Comparing with Eq. (22.13) we can
identify

\[ c \frac{dy_p}{d\tau} = \frac{|\vec{a}|}{(1 - \vec{v}^2/c^2)^{3/2}} = |\vec{a}| \gamma^3, \quad \vec{a} \parallel \vec{v}. \]

Since \( dt/d\tau = \gamma = \cosh y_p \) and using chain rule we can write \( a^\mu \) as a derivative with
respect to laboratory time

\[ a^\mu = \frac{dy_p}{dt} c \cosh y_p \{ \sinh y_p, 0, 0, \cosh y_p \}. \]
Evaluating the magnitude of both forms of $a^\mu$ shown in Eq. 2 and Eq. 4.

$$5 \quad a^2 = -c^2 \left( \frac{dy_p}{d\tau} \right)^2 = -c^2 \left( \frac{dy_p}{dt} \right)^2 \cosh^2 y_p. $$

The last form Eq. 5 applies to a laboratory observer who also measures a particle rapidity $y_p \neq 0$, while the first form of Eq. 5 applies to a particle comoving observer using particle proper time as coordinate time.

We note that the sign in Eq. 5 shows that the 4-acceleration is a spacelike vector, with the magnitude

$$6 \quad a \equiv \frac{dy_p}{d\tau} c, \quad \text{or} \quad a \equiv \frac{dy_p}{dt} \cosh y_p c. $$

---

### 22.3 Relativistic form of Newton’s 2nd Law

We now search for a relativistic generalization of the key equation of Newtonian dynamics,

$$\vec{F} = \frac{d\vec{p}}{dt} = m \frac{d\vec{v}}{dt} = ma. $$

Given that we have already introduced in section 22.2 the 4-acceleration, we need to generalize the concept of the force to be a 4-vector, *i.e.* a 4-force. Thus we consider

$$K^\mu = (K_0, \vec{K}), \quad (22.15)$$

where the spatial components $\vec{K}$ in the non-relativistic limit relate to the non-relativistic force $\vec{F}$

$$\vec{K} \xrightarrow{c \to \infty} \vec{F}. \quad (22.16)$$

Using the 4-acceleration the generalized Newton equation takes the form

$$\frac{dp^\mu}{d\tau} = \frac{d(mu^\mu)}{d\tau} = ma^\mu = K^\mu, \quad (22.17)$$

where we assume and keep, unless otherwise stated, the rest mass $m$ of the body a constant. The quantity $K^\mu$ is a 4-vector characterizing the force. Since we know that $u \cdot a = 0$, Eq. (22.10), the 4-force is subject to a constraint

$$u \cdot K = mu \cdot a = 0. \quad (22.18)$$

The implications of this constraint are discussed in the next section 22.4.
We consider next the spatial portion of Newton’s force Eq. (22.17):\[
\vec{K} = \frac{d}{dt} \frac{d}{d\tau} \left( \frac{m\vec{v}}{\sqrt{1 - \vec{v}^2/c^2}} \right) = \gamma \frac{dm\gamma\vec{v}}{dt}.
\] (22.19)

The prefactor \(\gamma\) arises from the change of the proper time \(\tau\) dependence inherent in \(\vec{K}\) to the laboratory time \(t\) dependence needed to introduce the three force \(\vec{F}\)
\[
\vec{K} \equiv \gamma \vec{F}.
\] (22.20)

Thus, written explicitly, we have for the relativistic form of the three force
\[
\vec{F} = \frac{d(m\gamma\vec{v})}{dt} = \frac{d\vec{p}}{dt}.
\] (22.21)

For \(dm/dt = 0\), only \(v(t)\) and \(\gamma(t)\) contribute on the left-hand side of Eq. (22.21), and we obtain two terms just as we did in Eq. (22.13), where the second arising from differentiation of the Lorentz factor \(\gamma\) is new:
\[
m \left( \frac{\ddot{a}}{\sqrt{1 - \vec{v}^2/c^2}} + \frac{1}{c^2} \frac{\vec{v} (\vec{v} \cdot \ddot{a})}{(1 - \vec{v}^2/c^2)^{3/2}} \right) = \vec{F},
\] (22.22)

and in simplified notation
\[
\frac{d\vec{p}}{dt} = m\gamma \ddot{a} + m\gamma^3 \frac{1}{c^2} \vec{v} (\vec{v} \cdot \ddot{a}) = \vec{F}, \quad m = \text{Const.}
\] (22.23)

The velocity of a body is a preferred direction with regard to which we usually consider all dynamics. For this reason we decompose all vectors in their parallel and orthogonal components with respect to \(\vec{v}\) using the unit vectors \(\vec{e}_\parallel\) and \(\vec{e}_\perp\)
\[
\vec{v} = v\vec{e}_\parallel, \quad \ddot{a} = a_\perp \vec{e}_\perp + a_\parallel \vec{e}_\parallel.
\] (22.24)

In the next step we rewrite Eq. (22.23) using Eq. (22.24)
\[
\frac{d\vec{p}}{dt} = m\gamma a_\perp \vec{e}_\perp + m\gamma a_\parallel \vec{e}_\parallel + m\gamma^3 \frac{\vec{v}^2}{c^2} a_\parallel \vec{e}_\parallel = \vec{F}.
\] (22.25)

Introducing a common denominator for the two \(\vec{e}_\parallel\) terms we obtain the much simplified equation
\[
\frac{d\vec{p}}{dt} = m\gamma a_\perp \vec{e}_\perp + m\gamma^3 a_\parallel \vec{e}_\parallel = \vec{F}.
\] (22.26)

We see in Eq. (22.26) that a particle reaching relativistic motion \(\gamma > 1\) responds to an applied force in very different manner. The force component transverse to the velocity is \(\gamma^2\) more effective at accelerating the particle as compared to the force component parallel to the velocity. To make this explicit some authors call
Figure 22-1: A particle moving with velocity $\vec{v}$ subject to a non-parallel force $\vec{F}$ yields a resultant acceleration vector $\vec{a}$ shown with the two relativistic corrections, $\frac{1}{2} (\frac{v}{c})^2 \vec{a}$ and $(\frac{v}{c}) \vec{a}$, see Eq. (22.28).

$m\gamma$ the transverse mass and $m\gamma^3$ the parallel mass. However, there is just one body mass; so this distinction is better made using the words ‘transverse inertial resistance $m\gamma$’ and ‘parallel inertial resistance $m\gamma^3$’.

As an example recall that the force due to a magnetic field is always transverse to the velocity vector. Therefore this force goes along with the first term in Eq. (22.26). Any other force component parallel to the velocity vector acting together with the magnetic field has for ultra-relativistic particle dynamics to overcome a much greater, by factor $\gamma^2$, effective inertia, and as result is significantly less effective in producing a change in body motion. One concludes that to guide relativistic particles, magnetic fields are the natural choice.

We next consider the nonrelativistic limit of the relativistic 3-force $\vec{F}$. We rewrite Eq. (22.26)

$$\vec{F} = m\gamma \vec{a} + m(\gamma^3 - \gamma) a_t \vec{e}_t,$$  \hspace{1cm} (22.27)

which clarifies the coefficients of the post-Newtonian limit expansion

$$\frac{d\vec{p}}{dt} = m \left(1 + \frac{1}{2} \frac{\vec{a}^2}{c^2} + \ldots \right) \vec{a} + m \left(\frac{\vec{a}^2}{c^2} + \ldots \right) a_t \vec{e}_t = \vec{F}. \hspace{1cm} (22.28)$$

In Eq. (22.28) there are two $O(\vec{v}^2/c^2)$ relativistic corrections. The second of these two terms is parallel to $\vec{a}$ and thus $\vec{a}$ and $\vec{F}$ are not parallel, as illustrated in figure 22-1. When the force $\vec{F}$ is either exactly parallel or exactly perpendicular to the instantaneous direction of motion, $\vec{F}$ and $\vec{a}$ are parallel. In the general case the vector describing acceleration of the body deviates in the first post-Newtonian correction from the direction of the force, a situation familiar to sailors, given that
the sailboat subject to the force $\vec{F}$ of wind in some direction, is also subject to the drag force of water acting against the direction of boat velocity $\vec{v}$. Looking deeper into this analogy note that the medium drag is only present when the particle is accelerated (non-inertial motion); the drag vanishes also when the velocity is normal to the direction of acceleration as we see in figure 22-1.

Exercise VIII–15: Acceleration vector $\vec{a}$ in term of the force

Cast Eq. (22.23) and Eq. (22.25) into a form that presents the 3-acceleration vector $\vec{a}$ directly as a function of the 3-force $\vec{F}$ and the momentary particle velocity $\vec{v}$.

Solution

We multiply Eq. (22.23) with $\vec{v}$ to obtain

1. $m\gamma^3 \left( (1 - \vec{v}^2/c^2) + \vec{v}^2/c^2 \right) \vec{v} \cdot \vec{a} = \vec{v} \cdot \vec{F}$.

Using Eq. 1 in Eq. (22.23) we find

2. $m\gamma \vec{a} = \vec{F} - \frac{\vec{v} (\vec{v} \cdot \vec{F})}{c^2}.$

We can elaborate this result further: decomposing the force $\vec{F}$ into the parallel and perpendicular to direction of motion components, and comparing with Eq. (22.24), we obtain

3. $m\gamma \vec{a} = \vec{e}_\perp F_\perp + \vec{e}_\parallel F_\parallel - \frac{\vec{v}^2}{c^2} \vec{e}_\parallel F_\parallel.$

This can be written in a manner analogous to Eq. (22.26)

4. $m\gamma \vec{a} = \vec{e}_\perp F_\perp + \vec{e}_\parallel F_\parallel \gamma^2.$

We note that Eq. 4 has exactly the same form and content as does Eq. (22.26). In fact both vector formats look the same when written in component form:

5. $m\gamma (\vec{a} \cdot \vec{e}_\perp) = (\vec{F} \cdot \vec{e}_\perp),$

When solving a problem of relativistic motion one must remember that since the velocity vector $\vec{v}$ undergoes a continuous change in time, the unit directional vectors $\vec{e}_\perp, \vec{e}_\parallel$ with respect to $\vec{v}$ are in general time-dependent.

End VIII–15: Acceleration vector $\vec{a}$ in term of the force
22.4 The 4-force and work-energy theorem

We now consider the 0-component of the generalized Newtonian equation Eq. (22.17):

\[ K_0 = \frac{1}{c} \frac{dE}{d\tau} = \frac{1}{c} \frac{dt}{d\tau} \frac{dE}{dt} = \gamma \frac{dE}{dt}. \]  

(22.29)

The term \( dE/dt \) describes the energy absorbed or released by the body per unit of laboratory time, i.e. the power. The use of laboratory time is appropriate when the force that a particle experiences is maintained with laboratory based equipment, meaning that the source of the force is kept at rest in the laboratory. \( K_0 \) is therefore the instantaneous power that is either absorbed or released by an accelerated body, as observed from an inertial system from which the process of acceleration is observed.

We can relate the value of the power to the acting force. We recall that the 4-force is orthogonal to 4-velocity, Eq. (22.10), and exercise VIII–12:

\[ u \cdot K = u_\mu K^\mu = 0. \]  

(22.30)

Expanding the summation yields

\[ c \frac{dt}{d\tau} K_0 - \frac{d\vec{x}}{d\tau} \cdot \vec{K} = 0. \]  

(22.31)

We substitute in \( K_0 \) from Eq. (22.29) and after a simple rearrangement using Eq. (22.20), we obtain:

\[ \frac{dE}{dt} - \frac{d\vec{x}}{dt} \cdot \vec{K} = \frac{dE}{dt} - \frac{d\vec{x}}{dt} \cdot \vec{F} = 0. \]  

(22.32)

We have thus shown that the 0-th component of the force entering the invariant Eq. (22.30) describes the work-energy theorem of classical mechanics,

\[ dE = \vec{F} \cdot d\vec{x}. \]  

(22.33)

As in the nonrelativistic case, a force normal to velocity \( \vec{v} = d\vec{x}/dt \) does not do any work on the particle; it can deflect the direction of motion but not directly change particle energy.

A special case of interest concerns the conservative forces. These are forces that are time independent, and where the force can be written in terms of a gradient of a potential \( V \)

\[ \vec{F} = -\vec{\nabla} V, \quad \vec{v} \cdot \vec{F} = \frac{-d\vec{x}}{dt} \cdot \frac{\partial V}{\partial \vec{x}} = -\frac{dV}{dt}, \]  

(22.34)

where \( dV/dt \) is the total time derivative arising from time dependent motion \( \vec{x}(t) \) of a particle in a time independent potential \( V(\vec{x}) \). Equation (22.32) then provides energy conservation

\[ 0 = \frac{d}{dt}(E + V). \]  

(22.35)
Here $E$ is energy of motion of a particle and $V(x)$ is the potential that a particle experiences at position $x$. In Eq. (22.35), the sum $E + V$ appears which is conserved when evaluated in the laboratory frame of reference. This energy conservation must, however, be a meaningful statement for all observers. Since $E$ is the timelike (zero) component of the 4-vector $p^\mu$, $V$ must behave under a Lorentz transformation like timelike component of a 4-vector.

Since the position $x$ changes as function of time, the motion energy changes accordingly, and the sum of kinetic energy and potential energy defined through equation (22.34) remains constant. This is familiar from nonrelativistic dynamics but for the relation of $E$ with either the velocity, or the momentum. We integrate Eq. (22.35) and move $V$ to the other side of equation to obtain

$$E_0 - V(x) = E = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}} = \sqrt{p^2c^2 + m^2c^4} ,$$

(22.36)

where $E_0$, an integration constant includes $V(x_0)$.

**Exercise VIII–16: Acceleration and relativistic dynamics**

Obtain and/or recapitulate all constraints arising for relativistic particle dynamics and place these in context of accelerated motion.

**Solution**

We know that particles arrive from the observer’s past light cone and cannot leave the future light cone; that is, all position vectors $x^\mu(\tau)$ satisfy the timelike constraint

$$1 \quad x^2(\tau) > 0 ,$$

where the observer is in the coordinate origin. We also know that

$$2 \quad u^2(\tau) = c^2 , \quad u^\mu = \frac{dx^\mu}{d\tau} .$$

Upon differentiation of Eq. 2 with respect to $\tau$ we obtain

$$3 \quad u(\tau) \cdot a(\tau) = 0 , \quad a^\mu = \frac{du^\mu}{d\tau} .$$

In view of Eq. 3 a 4-velocity increasing linearly with the 4-acceleration

$$4 \quad u^\mu = u^\mu(0) + a^\mu \tau ,$$

makes no physical sense: it implies

$$5 \quad 0 = u(\tau) \cdot \frac{du}{d\tau} = u(\tau) \cdot a = 0 ,$$
and hence

$$
c^2 = u^2(\tau) = u^2(0) + 2u^\mu(0)a^\mu \tau + a^2 \tau^2 = c^2 + 0 + a^2 \tau^2 \rightarrow a^2 = 0.
$$

which contradicts the requirement that acceleration is a spacelike vector.

This finding implies that the 4-force cannot take the ‘natural’ form common in 3-dimensional consideration leading to the form of motion Eq. 4 when force is a constant. Specifically a 4-force cannot be a 4-gradient of a Lorentz-invariant function $f(x^2)$

$$K^\mu_\nabla = \partial^\mu f(x^2) = 2x^\mu \frac{df}{dx^2}, \quad \Rightarrow K^2_\nabla = x^2 4 f'^2 > 0.
$$

We see that the $K^\mu_\nabla$ is timelike. Since $K^\mu_\nabla \propto a^\mu$ we recognize this conflicts with the requirement that $a^\mu$ is spacelike. Furthermore, a 4-force which generates a 4-acceleration must satisfy $u \cdot K = 0$. Imposing this condition we find that the 4-force must vanish.

$$0 = u_\mu K^\mu_\nabla = u_\mu \partial^\mu f(x^2) = \frac{df(c^2 \tau^2)}{d\tau} = 2c^2 \tau f' \Rightarrow f' = 0 \Rightarrow K^\mu_\nabla = 0.
$$

End VIII–16: Acceleration and relativistic dynamics

Exercise VIII–17: Types of 4-forces

Identify the format that a velocity dependent 4-force can take that is consistent with the constraints on 4-velocity and 4-acceleration.

Solution

In exercise VIII–16 we have seen that not every 4-vector can be a 4-force in consideration on the constraints on relativistic particle dynamics. The key constraint $u \cdot K = 0$ tells us that the 4-force must involve the 4-velocity. There are three ways that allow us to accomplish this, in which the third example is a different way of presenting the second.

I: We consider

$$K^\mu = \left(g^{\mu\nu} - \frac{u^\mu u^\nu}{c^2}\right) B_\nu(x), \quad u \cdot K = 0.
$$

$B_\nu$ can be any 4-vector including $B_\nu(x) = \partial f(x^2)/\partial_\nu$. We see explicit dependence of $K^\mu$ on 4-velocity which assures $u \cdot K = 0$ constraint can be satisfied. To show that $K^\mu$ is spacelike we can choose any convenient frame of reference since the magnitude of a 4-vector is invariant under LT. We consider $K^\mu$ in the rest-frame of the particle $u^\mu = \{c, \vec{0}\}$

$$K^0_\nu(0) = \left(1 - \frac{c^2}{c^2}\right) B_0(x) = 0, \quad \vec{K}(0) = -\vec{B}.
$$
This means that $K^\mu$ is spacelike since it is spacelike in the rest-frame of the particle $K \cdot K = K \cdot K_{(0)} = 0 - (\vec{B})^2 < 0$, which assures that the constraint $a^2 < 0$ is satisfied.

II: We consider $K^\mu = q c F^\mu_{\nu} u^\nu$.

This is the Lorentz-force which we will discuss at length in section 26.4, where we show that it satisfies the constraint $u \cdot K = 0$, see also exercise X–7 on page 390, and in exercise X–8 on page 392 we show that it is a spacelike vector.

III: We consider $K^\mu = q c \left( u \cdot \frac{\partial A}{\partial x^\mu} - \frac{d A^\mu}{d \tau} \right)$, where $A^\mu$ is an arbitrary 4-vector and the last term includes effect of particle motion $x^\mu(\tau)$. We will see in section 25.1 that this is the Lorentz-force written in terms of potentials. In section 26.4 the relationship to case II is established. The constraints on $K^\mu$ are considered there.

End VIII–17: Types of 4-forces
Part IX

Motion of Charged Particles
Introductory remarks to Part IX

The objective in this Part is to extend the well known nonrelativistic Lorentz-force dynamics of charged particles to relativistic motion, addressing specific field examples. We begin with discussion of motion of a charged particle in a constant magnetic field, then follow with the more complex case of motion in a constant electric field. The difficulty that arises in the study of electrical fields is that even if we choose initial particle motion normal to the electrical field, particle velocity turns into the field direction resulting in a catenary particle path.

The Lorentz-force is most often presented in terms of electric and magnetic fields. However, the motion of charged particles can also be described in terms of the EM-potentials \( V, \vec{A} \) that generate these fields. We rewrite and present the modified Lorentz-force written in terms of these four electromagnetic potential. We introduce a relativistic, but not yet covariant formulation finding a conserved canonical momentum. Since the potentials are not defined in unique fashion we introduce the transverse gauge constraint.

In the next step we generalize the nonrelativistic variational action principle for particle dynamics by introducing the relativistic form of kinetic energy. Derivation of the relativistic dynamical equations from action principle allows for example a better understanding of canonical momentum conservation. We identify another conserved quantity, Hamiltonian (energy). We establish a constraint that bears the signature of a 4-dimensional energy-momentum invariant.

The conservation laws play an important role in the understanding of charged particle orbits in the presence of the Coulomb force of an atomic nucleus, our third example of relativistic particle motion. We point out parallels with relativistic atomic quantum orbits. This shows that the relativistic atomic quantum theory is firmly rooted in the classical Lorentz-force. We also derive the SR aphelion precession that is significantly different from the GR case.

Another application of newly gained insights is the case of an electron surfing on the electromagnetic plane wave. We address this complex relativistic dynamics problem in non-covariant transverse gauge, a formulation convenient to this problem. We first explore how conserved quantities constrain particle motion. We than solve for particle motion explicitly considering both linear and circular wave polarization. We demonstrate a significant change in particle dynamics when a threshold intensity of the plane wave is exceeded. This opens up the realm of ‘relativistic optics’.

This entire chapter, an introduction to relativistic charged particle dynamics, relies on physics concepts that do not require introduction of 4-dimensional Minkowski space. This modus of the presentation is chosen here since one can argue that in case of the EM theory covariant notation leads to a loss of connection to laboratory reality without adding to the insight about EM theory. We will see how EM in 4-dimensions works, in the following Part X and Part XI.
23 The Lorentz Force

23.1 Motion in magnetic and electric fields

The well-known Lorentz Force \( \vec{F}_L \), the force acting on an electron is

\[
\frac{d\vec{p}_i}{dt} = q\left( \vec{E} + \vec{v} \times \vec{B} \right) \equiv \vec{F}_L , \quad \vec{p}_i = m\gamma\vec{v} = \frac{mv}{\sqrt{1 - \frac{v^2}{c^2}}} .
\] (23.1)

where in SI-units the EM-fields \( \vec{E}, \vec{B} \) are measured respectively, in V(olt)/m and T(esla). We will typically look at dynamics of an electron. The electron charge is \( q \rightarrow q_e \equiv e = -|e| = -1.602177 \times 10^{-19} \text{C} \). The modification from the nonrelativistic dynamics is seen in the relativistic form of the kinetic or as we prefer to say in this book, inertial momentum \( \vec{p}_i \).

Both \( \vec{E} \) and \( \vec{B} \) fields are additive; that is, they can be superposed. However, except for the case of EM-plane-wave we explore in the following examples the dynamics described in the simplified case that either \( \vec{E} \) or \( \vec{B} \) are non vanishing:

**Exercise IX–1**: The motion of a charged particle in a homogeneous time-independent magnetic field \( \vec{B} \), with vanishing electrical field \( \vec{E} = 0 \). In this example the force is always normal to the direction of motion, irrespective of the motion along the field lines of the field. Thus we find dynamics with which we are familiar from classic nonrelativistic electromagnetism, up to the detail that it is the energy content of the particle, and not its mass, that provides the inertia that the force must overcome, see Eq. (22.26).

**Exercise IX–2**: The motion of a charged particle in a homogeneous time independent electric field \( \vec{E} \) with vanishing magnetic field \( \vec{B} = 0 \), choosing the initial momentum transverse to the direction of the field. We find that the force acts to align the particle with the force field. One can say a particle is trying to cross an ‘electric river’, and the flow will turn its motion into the direction of the field. Unlike the non-relativistic case, the inertial resistance to the force depends on the velocity of the particle, which continuously changes in magnitude and direction under the influence of the electrical field.

Further below in **exercise IX–5 on page 337** we explore the motion and orbits of a classical relativistic electron in the presence of a radial Coulomb field of a point-like atomic nucleus \( q\vec{E} = e_r Ze^2/r^2 \). The dynamical equations differ the nonrelativistic mechanics analogue, allowing the aphelion precession. Comparing with the well known GR test result we determine that SR predicts 1/6 of the residual Mercury precession.

In **section 24** we will study the motion in the electromagnetic field of a plane wave i.e. time-dependent orthogonal fields \( \vec{E} \neq 0 \) and \( \vec{B} \neq 0 \). In **Part X, section 26.5** we will study motion in the presence of homogeneous time-independent
Insight: SI and Gauss EM units

In a book written for students we adhere to the SI-units of the electrical field \(\vec{E}\): ‘volt/meter’; and of the magnetic field: \(\vec{B}\): T=tesla=volt sec/meter\(^2\). According to Eq. (23.1) a charged particle of charge \(q\) experiences the same strength E&M-force when it moves with the speed \(v = 1\) m/s. For relativistic particles \(v \to c\), the \(B = 1\) T produces a force equivalent to a force caused by \(E = 3 \times 10^8\) V/m

This factor 300 million is introduced on purpose in the SI-unit system. The Lorentz force law Eq. (23.1) is in first place written more naturally with dimensionless factor \(\vec{v} \to \vec{v}/c\). The \(c\) denominator can be traced back to the inclusion of \(c\) in \(x^0\) when introducing the position four-vector, see Eq. (20.3b). In the cgs+Gauss-unit system one keeps the \(c\) in the force definition and the asymmetry in force units in presence of relativistic motion disappears. In the SI-unit system this factor \(c\) is absorbed into the product \(eB\). We thus discover the first two of the following rules of transcription of equations between cgs+Gauss and SI, for third relation see exercise XI–2 on page 405

\[
e E|_{\text{cgs}} \to e E|_{\text{SI}}, \quad e B|_{\text{cgs}} \to c e B|_{\text{SI}}, \quad e^2|_{\text{cgs}} \to \frac{e^2}{4\pi\epsilon_0|_{\text{SI}}}.
\]

(23.2)

In the SI-system the Maxwell equations are written exactly in the same format as seen in Maxwell work. The SI-unit system introduces dimensioned vacuum properties: the vacuum permittivity \(\epsilon_0\) and vacuum permeability \(\mu_0\) (assuming here isotropic response)

\[
\vec{D} = \epsilon_0\vec{E}, \quad \vec{B} = \mu_0\vec{H}, \quad \epsilon_0\mu_0 = \frac{1}{c^2}.
\]

(23.3)

The last relation defines propagation speed of Maxwell waves. To some (including Maxwell) Eq. (23.3) implies that the vacuum is another form of matter. However, vacuum is not material as Einstein explained – see Preamble, Ref. [10] on page xv. \(\epsilon_0\) and \(\mu_0\) are not appearing in the cgs+Gauss unit system which made an effort to remove the æther from Maxwell equations.

To obtain physical results we will use electron properties as follows

\[
\frac{e^2}{4\pi\epsilon_0 m_e c^2} \equiv r_e \to \frac{e^2}{4\pi\epsilon_0} = 1.4403 \times 10^{-9}\text{ eV m}, \quad \leftrightarrow m_e c^2 = 0.510999\text{ MeV},
\]

(23.4)

see Eq. (19.42) for the electron classical radius \(r_e\), and Insight on page 204 for the energy unit eV. In this book \(\epsilon_0\) will always appear in the form shown in Eq. (23.4) thus we will neither need to know its numerical value, nor that of \(\mu_0 = 1/\epsilon_0 c^2\). This shows that \(\epsilon_0\) (and \(\mu_0\)) are not independently known natural constants, but reconcile several choices made for the elementary measurement units, a situation similar to the speed of light, \(c\), see section 1.2.
orthogonal $\vec{E} \neq 0$ and $\vec{B} \neq 0$ fields of equal magnitude, using as the parameter of evolution the particle proper time.

Beyond these exercises we aim to demonstrate that the Lorentz-force is not the entire electromagnetic force. In exercise XI–8 on page 417 we include the force acting on an elementary point magnetic dipole moment, and in section 29.4 we address radiation-friction force, the reaction to radiation emission. No universal electromagnetic force is known incorporating these well-known effects.

---

Exercise IX–1: Relativistic motion of an electron in constant $\vec{B}$

Examine the movement of an electron in a constant homogeneous magnetic field $\vec{B}$ and compare with the non-relativistic case.

**Solution**

The force on an electron in magnetic field $\vec{B}$ is known to be

$$1 \quad \frac{d\vec{p}}{dt} = \vec{F} = e\vec{v} \times \vec{B}. $$

we multiply with $\vec{p} = m\gamma\vec{v}$ and find

$$2 \quad \vec{p} \cdot \frac{d\vec{p}}{dt} = m\gamma e\vec{v} \cdot (\vec{v} \times \vec{B}) = 0. $$

Which leads us to

$$3 \quad \frac{d\vec{v}^2}{dt} = 0, \quad \rightarrow m^2 \frac{\vec{v}^2}{1 - \vec{v}^2/c^2} = \text{Const.}, \quad \rightarrow \vec{v}^2 = \text{Const.}, \quad \rightarrow \gamma = \text{Const}. $$

The magnetic field $\vec{B}$ is only capable of altering the direction of motion, but not the speed of a particle. This is clearly true for both relativistic and nonrelativistic particle motion as we see in Eq. 3. In nonrelativistic case the particle mass $m$ enters into dynamics, in relativistic case we recognize that this is the energy $E/c^2 = \gamma m$.

We decompose $\vec{v}$ into components normal ($\vec{v}_\perp$) and parallel ($\vec{v}_\parallel$) to the $\vec{B}$ field, so that

$$4 \quad \vec{v} = \vec{v}_\parallel + \vec{v}_\perp, \quad \vec{v}^2 = \vec{v}_\parallel^2 + \vec{v}_\perp^2. $$

Inserting the first form of Eq. 4 in Eq. 1 and keeping in mind that in view of Eq. 3 also $\gamma = \text{Const.}$, we find

$$5 \quad m\gamma \frac{d(\vec{v}_\parallel + \vec{v}_\perp)}{dt} = e(\vec{v}_\parallel + \vec{v}_\perp) \times \vec{B} $$

and keeping in mind that by definition $\vec{v}_\parallel \times \vec{B} = 0$, we see that

$$6 \quad m\gamma \frac{d\vec{v}_\parallel}{dt} = 0 \rightarrow \vec{v}_\parallel = \mathcal{C}\text{Const.}, \quad \rightarrow |\vec{v}_\perp| = \text{Const.}$$
The last condition follows since we know that the speed of the particle is constant, Eq. 3. The motion of the particle normal to the $\vec{B}$-field is according to Eq. 5 described by

$$m\gamma \frac{d\vec{v}_\perp}{dt} = e(\vec{v}_\perp \times \vec{B}).$$

To summarize, a charged particle entering a magnetic field domain experiences a force normal to the field, and to its instantaneous velocity vector. Therefore the motion is a superposition of linear motion along the field lines Eq. 6, not affected by the magnetic field, with the motion normal to the field subject to deflection, but preserving the transverse speed Eq. 7.

We choose a coordinate system such that the magnetic field is aligned with the $z$-axis. It follows from Eq. 7 that we must solve the equations

$$\frac{dv_x}{dt} = +\omega_c v_y, \quad \frac{dv_y}{dt} = -\omega_c v_x.$$  

The characteristic ‘cyclotron’ frequency $\omega_c$ is defined by

$$\omega_c = \frac{eB}{\gamma m},$$

Inserting one of the component equations Eq. 8 into another we obtain the 2nd order differential equations.

$$\frac{d^2v_i}{dt^2} = -\omega_c^2 v_i \quad i = x, y.$$  

We recognize the harmonic oscillator differential equation, hence the solution of Eq. 8 is

$$v_x = v_0 \cos \omega_c t, \quad v_y = -v_0 \sin \omega_c t,$$

where we chose the $x$-axis aligned with the initial transverse tangential speed $\vec{v}_\parallel(t = 0)$ pointing out of the plane of figure 23-1.

Time integration of Eq. 11 provides

$$x = \frac{v_0}{\omega_c} \sin \omega_c t, \quad y = \frac{v_0}{\omega_c} \cos \omega_c t,$$

showing that the cyclotron frequency governs the particle orbital rotation with radius

$$\rho \equiv \sqrt{x^2 + y^2} = \frac{v_0}{\omega_c}, \quad \rho \omega_c = v_0.$$  

This is the situation shown in figure 23-1 where we chose the initial condition such that the particle is at position $x_0 \equiv x(t = 0) = 0, y_0 \equiv y(t = 0) = \frac{v_0}{\omega_c}$. The electron path
Figure 23-1: The helical movement of an electron in a constant magnetic field.

Figure 23-2: The projection of the path from figure 23-1 onto a plane orthogonal to $\vec{B}$.

Projected onto a plane orthogonal to $\vec{B}$ corresponds to a motion on a circle with radius $\rho$, as shown in figure 23-2. The path of motion shown in figure 23-1 is called a helix, and the motion called 'helical'.

Of interest is the magnitude of acceleration which we obtain differentiating Eq. 11

$$|\vec{a}| \equiv a = \sqrt{a_x^2 + a_y^2} = v_0 \omega_c .$$

Combining Eq. 14 with Eq. 13 to eliminate $\omega_c$ we fix the orbital acceleration, writing it in several applicable forms using two out of the three variables $v_0, \omega_c, \rho$

$$a = \frac{v_0^2}{\rho} = \rho \omega_c^2 = v_0 \omega_c .$$

Only the cyclotron frequency $\omega_c$ and the orbital velocity $v_0$ are experimental parameters, $\rho 0$ follows from Eq. (13). There is no limit on acceleration that an electron can experience with an increasing strength of a magnetic field which drives the magnitude of cyclotron frequency. We see that when the electron moves at relativistic speed entering a magnetic field, the cyclotron frequency $\omega_c$ decreases and the radius of helical
motion Eq. 13 increases. This means that a charged relativistic particle traversing a magnetic field curves its path as if the mass of the particle is larger by a factor of $\gamma$. In fact what we found is that it is not the mass but the laboratory energy of charged particles that determines their path in a magnetic field. Particle detectors take advantage of the path curvature as a measure of particle energy.

The helical motion we obtained is practically the same as in the case of nonrelativistic dynamics. The difference is the appearance of the Lorentz factor $\gamma$ in the definition of cyclotron frequency $\omega_c$, Eq. 9, an insight with considerable influence in the domains of plasma physics, accelerator physics, and radiation emission Eq. (29.26).

End IX–1: Relativistic motion of an electron in constant $\vec{B}$

Exercise IX–2: Relativistic motion of an electron in constant $\vec{E}$

An electron moves in a space and time-independent (i.e. constant) electric field $\vec{E} = E\hat{e}_x$, where we oriented the $x$-coordinate along the field direction. Obtain the velocity $\vec{v}$ of the electron as a function of laboratory time $t$ and determine the path $x(y)$ of the electron. Describe these results in quantitative terms considering the initial values $p_{0x} = 0$, $p_{0y} = mc$. As a cross-check, obtain the non-relativistic limit for the path $x(y)$ and compare with a direct nonrelativistic solution of Newton’s force equation.

Solution

The 3-force on the electron due to electrical field $\vec{E}$ is known to be $\vec{F} = e\vec{E}$

$$1 \quad \frac{d\vec{p}}{dt} = \frac{d(m\gamma\vec{v})}{dt} = e\vec{E}, \quad \gamma = \frac{1}{\sqrt{1 - (\vec{v}/c)^2}}.$$ 

Without restriction of generality we do not need to consider motion in the third $z$-coordinate as we have oriented the coordinated system conveniently. We decompose Newton’s force equation into the $x$-component parallel to $e\vec{E}$, and the orthogonal $y$-component

$$2 \quad \frac{d(m\gamma v_x)}{dt} = eE, \quad \frac{d(m\gamma v_y)}{dt} = 0.$$ 

We integrate both equations with respect to $t$ and obtain

$$3 \quad p_x = m\gamma v_x = eEt + p_{0x}, \quad p_y = m\gamma v_y = p_{0y}.$$ 

Since we know the momentum of the particle as a function of time, we also have determined its energy

$$4 \quad E(t) = \sqrt{m^2c^4 + p_x^2c^2 + p_y^2c^2} = \sqrt{m^2c^4 + (eEt + p_{0x})^2 c^2 + p_{0y}^2c^2}.$$
Figure 23-3: The velocity as a function of normalized laboratory time $t/(mc/eE)$ for initial $p_0x = 0, p_0y = mc$: component in direction of the field $v_\parallel = v_x$ (short dashed, blue); the component normal to field $v_\perp = v_y$ (long dashed, green); and the total speed of the particle $v/c$ (solid, red).

We can now determine the velocity vector of the particle

$$v_x = \frac{p_x}{E} = c\frac{eE t + p_{0x}}{\sqrt{(mc)^2 + (eE t + p_{0x})^2 + p_{0y}^2}}, \quad v_y = \frac{p_y}{E} = c\frac{p_{0y}}{\sqrt{(mc)^2 + (eE t + p_{0x})^2 + p_{0y}^2}}.$$ 

Similarly we obtain the Lorentz-$\gamma$ factor

$$\frac{E}{mc^2} \equiv \gamma = \sqrt{1 + \frac{(eE t + p_{0x})^2 + p_{0y}^2}{(mc)^2}}.$$ 

We find for the speed $v$

$$v = \sqrt{\frac{v_x^2 + v_y^2}{\gamma}} = c\sqrt{\frac{(eE t + p_{0x})^2 + p_{0y}^2}{(mc)^2 + (eE t + p_{0x})^2 + p_{0y}^2}}.$$ 

These solutions $v_x$, $v_y$ and $v$ for the initial values $p_{0x} = 0$, $p_{0y} = mc$ are depicted in figure 23-3. The solid (red) line shows that the total speed $v$ monotonically increases, as does the component of the velocity in direction of the field $v_x$ (blue, short dashed). However, the component of the velocity $v_y$ (green, long dashed) transverse to field monotonically decreases from its initial value so that the transverse momentum can remain constant, see Eq.3 given that $\gamma(t)$ Eq.6 increases. The decrease of the transverse to the field velocity component $v_y$ when the initial value is non-zero is an important lesson of relativistic electron dynamics in constant $\vec{E}$ fields.

Setting $v_x = dx/dt$ and $v_y = dy/dt$ in Eq.5 we integrate with respect to $t$. Setting at $t = 0$, $x = 0$ and $y = 0$, the results are

$$x = \frac{\sqrt{m^2c^4 + p_{0y}^2c^2 + (ceE t + cp_{0x})^2} - E_0}{eE}, \quad E_0 = \sqrt{m^2c^4 + p_{0y}^2c^2 + p_{0x}^2c^2},$$

where $E_0$ is the initial energy of the particle.
Figure 23-4: The distance (in units of $mc^2/e\mathcal{E}$) traveled by a particle for the initial momentum $p_0x = 0$, $p_0y = mc$ as a function of time $t/(mc/e\mathcal{E})$: in direction of the field $x(t)$ (solid, blue); normal to the field $y(t)$ (dashed, green).

where $E_0 \equiv E(t = 0)$, and

$$y = \frac{p_0yc}{e\mathcal{E}} \left[ \text{arsinh} \left( \frac{e\mathcal{E}t + p_0x}{\sqrt{(mc)^2 + p_0y^2}} \right) - A \right], \quad A = \text{arsinh} \left( \frac{p_0x}{\sqrt{(mc)^2 + p_0y^2}} \right).$$

For the case $p_{0x} = 0$, $p_{0y} = mc$, the distances traveled are shown in figure 23-4. We see that the distance traveled in transverse direction dominates at first but the electron soon goes further in direction of the field lines.

We now are interested in finding the path $x(y)$, and thus we use Eq. 9 to eliminate $t$ in Eq. 8. We find in first step

$$\sinh^2 \left( \frac{ye\mathcal{E}}{p_{0y}c} + A \right) = \frac{(e\mathcal{E}t + p_{0x})^2}{(mc)^2 + p_{0y}^2}, \quad \frac{x e\mathcal{E} + E_0}{\sqrt{m^2c^4 + p_{0y}^2c^2}} = \sqrt{1 + \frac{(e\mathcal{E}t + p_{0x})^2}{(mc)^2 + p_{0y}^2}},$$

to yield

$$\frac{x e\mathcal{E} + E_0}{\sqrt{m^2c^4 + p_{0y}^2c^2}} = \cosh \left( \frac{ye\mathcal{E}}{p_{0y}c} + A \right), \quad A = \text{arcosh} \left( \frac{E_0/c}{\sqrt{(mc)^2 + p_{0y}^2}} \right).$$

which defines the catenary path. The word catenary is derived from the Latin word catena=chain. A hanging chain is described by the ‘cosh’ function with $x \Leftrightarrow y$ compared to our form.

For the special case $p_{0x} = 0$ we are interested in we have $E_0/c = \sqrt{(mc)^2 + p_{0y}^2}$, and Eq. 11 simplifies to

$$x = \frac{E_0}{e\mathcal{E}} \left( \cosh \left( \frac{ye\mathcal{E}}{p_{0y}c} \right) - 1 \right), \quad \lim_{c \to \infty} x = \lim_{c \to \infty} \frac{1}{2} \frac{E_0}{p_{0y}^2} \frac{e\mathcal{E}y^2}{c^2} = \frac{1}{2} \frac{e\mathcal{E}}{mv_{0y}^2} y^2.$$
The path is depicted in figure 23-5, and the nonrelativistic limit is shown on right in Eq. (12). We see in the non-relativistic limit that the electron follows the parabolic path. To cross-check this result we integrate the nonrelativistic equations of motion with the boundary conditions $v_{0x} = 0$, $x_0 = 0 = y_0$

$$\ddot{x} = \frac{eE}{m}, \quad \dot{x} = \frac{eE}{m} t \quad \rightarrow \quad x = \frac{eE}{2m} t^2, \quad \dot{y} = 0 \quad \rightarrow \quad \dot{y} = v_{0y} \quad \rightarrow \quad y = v_{0y} t.$$

Solving the last form in Eq. 13 for $t$ and inserting into the solution for $x$ we obtain the last form in Eq. 12.

End IX–2: Relativistic motion of an electron in constant $\vec{E}$

23.2 EM-potentials and homogeneous Maxwell equations

In exploration of particles dynamics where there is a given ‘external’ electromagnetic fields, such as that belonging to a light plane-wave, it is of advantage to work not with the EM-fields but with electromagnetic potentials. These can be introduced considering the ‘Maxwell equations’: on left the Maxwell-Faraday vector-equation and on right the Gauss’ magnetism law

$$\nabla \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = \vec{0}, \quad \nabla \cdot \vec{B} = \vec{0}. \quad (23.5)$$

---

Michael Faraday (1791 – 1867) Preeminent English scientist, discoverer of the principles underlying electromagnetic induction that allows the generation and use of electrical power.
These are called the homogeneous Maxwell equations since there are no spatially
distributed sources driving the fields. The other four Maxwell equations which
are driven by applied sources will be introduced when we address the dynamics
of EM-fields in section 27.1 on page 401.

As we now demonstrate given Eq. (23.5) the electric and magnetic fields can
be obtained from the ‘scalar’ potential\(^2\) \(V\) and the 3-vector potential \(\vec{A}\)
\[
\vec{E} = -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}V, \quad \vec{B} = \vec{\nabla} \times \vec{A}.
\]

We easily verify
\[
\vec{\nabla} \times \vec{E} = \vec{\nabla} \times \left( -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla}V \right) = -\frac{\partial (\vec{\nabla} \times \vec{A})}{\partial t} = -\frac{\partial \vec{B}}{\partial t},
\]
\[
\vec{\nabla} \cdot \vec{B} = \vec{\nabla} \cdot (\vec{\nabla} \times \vec{A}) = 0,
\]
we must be able to exchange second derivatives of the potentials everywhere for
the above to be true, or else we would for example have \(\vec{\nabla} \cdot \vec{B} \neq 0\), implying the
presence of magnetic monopoles. The question if magnetic monopoles exist or
not is answered experimentally. Since none have been seen, we assume validity
of Eq. (23.5) which in compact format is Eq. (23.5).

An astute reader will notice that we have replaced the six EM-field compo-
nents \(\vec{E}, \vec{B}\) which are constrained by four EM-field equations Eq. (23.5) by four
potentials \(V, \vec{A}\). This means that there are certainly additional constraints that
need to be imposed on the potentials. Indeed, given the EM-fields, the corre-
sponding set of potentials is not unique: a ‘gauge’ transformation,
\[
\vec{A} \rightarrow \vec{A}' = \vec{A} - \vec{\nabla}\Lambda(\vec{r}, t),
\]
\[
V \rightarrow V' = V + \frac{\partial \Lambda(\vec{r}, t)}{\partial t},
\]
leaves in Eq. (23.6) the fields unchanged for any regular differentiable field \(\Lambda(\vec{r}, t)\).

For \(\vec{E}\) to remain invariant we need to exchange time and space differentiation, and
for \(\vec{B}\) to remain invariant we must have \(\vec{\nabla} \times \vec{\nabla}\Lambda(\vec{r}, t) = 0\), true if \(\partial^2 \Lambda/\partial x_i \partial x_j = \partial^2 \Lambda/\partial x_j \partial x_i\) for all \(i, j = 1, 2, 3\) everywhere\(^3\).

\(^2\)The frequent naming of \(V\) as a ‘scalar’ potential originating in the non-relativistic context
does not consider that \(V\) will soon turn out to be 0\(^\text{th}\) component of a 4-vector. Thus we will
not use this name in this book other than this one time.

\(^3\)For aficionados of the Gauss-unit system we note that to go back from SI to Gauss one
needs to restore in the partial time derivative in Eq. (23.5), Eq. (23.6), and Eq. (23.8) \(\partial t \rightarrow c \partial t\)
to convert from SI to Gauss system. This \(c\) is canceled out by the middle rule in Eq. (23.2), by
which \(\vec{A}\) is as much affected as \(\vec{B}\) in view of Eq. (23.6).
An example how the choice of gauge works is provided by ‘transverse gauge condition’. We demand
\[ \nabla \cdot \mathbf{A} = 0. \] (23.9)
Should Eq. (23.9) not apply, we choose a scalar field \( \Lambda \) such that a new vector field \( \mathbf{A}' \) satisfies Eq. (23.9)
\[ 0 = \nabla \cdot \mathbf{A}' = \nabla \cdot \mathbf{A} - \Delta \Lambda, \quad \Delta \equiv \nabla \cdot \nabla = \nabla^2. \] (23.10)
For a ‘given source’ \( \mathbf{\nabla} \cdot \mathbf{A} \) with given boundary conditions Eq. (23.9) can be solved for \( \Lambda \) using the fundamental solution of the time independent Poisson equation
\[ \Delta \frac{1}{|\mathbf{r} - \mathbf{r}'|} = -4\pi \delta^3(\mathbf{r} - -\mathbf{r}'). \] (23.11)
We obtain the required gauge transformation function
\[ \Lambda = -\frac{1}{4\pi} \int d^3r' \frac{\mathbf{\nabla} \cdot \mathbf{A}}{|\mathbf{r} - \mathbf{r}'|}. \] (23.12)

Given the choice of transverse gauge our relations potentials and fields are invertible: given the EM-potentials, Eq. (23.6) shows how to compute the fields. The inverse problem is solved in the following two exercises where given the EM-fields and choosing the transverse gauge we evaluate the corresponding EM-potential. The first exercise IX–3 addresses the case of EM-fields that are constant, while the second exercise IX–4 solves the general problem, which requires integrability of EM-fields, a condition not satisfied by the simpler special case considered in exercise IX–3.

The transverse gauge condition Eq. (23.9) is non-covariant. Hence once a transverse gauge is introduced finding potentials when changing the frame of reference is a tedious exercise. However the advantages of transverse gauge are numerous. In particular this is the gauge which we use to compute the quasi-static field of slowly moving charges, in order to obtain the Coulomb field potential, and hence some call this gauge ‘Coulomb gauge’. This is also the gauge in which one often describes the light (radiation) plane wave, since the transverse nature of light naturally fits the gauge condition, see section 24.1. Therefore this gauge is also sometimes called ‘radiation gauge’. There are, of course, well studied gauges that are Lorentz-covariant; that is, all observers agree to the constraint that one imposes on the potentials, and we will introduce a covariant gauge named after L. Lorenz (not ‘our’ H.A. Lorentz) further below, see Eq. (27.21) on page 409.

Exercise IX–3: Potentials for homogeneous & uniform \( \mathbf{E}, \mathbf{B} \)-fields

Obtain in transverse gauge the potentials corresponding to a set of homogeneous in space and uniform in time \( \mathbf{E}, \mathbf{B} \)-fields.

Solution
Since we do not want $\vec{B}$ in Eq. (23.6) to depend on time $\vec{A}$ must also be independent of time and thus the gradient of a potential $V$ alone determines the electrical field $\vec{E}$. The choice that satisfies Eq. (8) is

$$V = -\vec{r} \cdot \vec{E}, \quad \Leftrightarrow \quad \vec{E} = -\nabla V,$$

which satisfies the transverse gauge conditions Eq. (8). Since the 3-vector potential $\vec{A}$ producing a magnetic field $\vec{B}$ has to be a linear in $\vec{r}$ vector function, there is just one form we can try, with the coefficient determined in the verification process

$$\vec{A} = -\frac{1}{2} \vec{r} \times \vec{B}, \quad \Leftrightarrow \quad \vec{B} = -\frac{1}{2} \nabla \times (\vec{r} \times \vec{B}).$$

Clearly $\vec{A}$ is a divergence-free vector and thus Eq. (23.9) is satisfied. To verify the form as is shown on right we will need to compute a double cross product, a special case of the ‘bac-cab’ rule which reappears frequently in the remaining pages

$$\vec{a} \times (\vec{b} \times \vec{c}) = \vec{b}(\vec{a} \cdot \vec{c}) - \vec{c}(\vec{a} \cdot \vec{b}),$$

To prove this rule note that the right-hand side is the only expression orthogonal to both $\vec{a}$ and $(\vec{b} \times \vec{c})$. This leaves the sign undetermined which we confirm by choosing the Cartesian unit vectors for abc: $\vec{b} = \hat{i}$ and $\vec{c} = \hat{j}$ with either $\vec{a} = \hat{i}$ or $\vec{a} = \hat{j}$. By direct computation we obtain

$$\hat{i} \times (\hat{i} \times \hat{j}) = \hat{i} \times \hat{k} = -\hat{j}, \quad \hat{j} \times (\hat{i} \times \hat{j}) = \hat{j} \times \hat{k} = \hat{i},$$

in agreement with the result we obtain using Eq. 3.

In order to confirm Eq. (2) we choose in the bac-cab rule $\vec{a} \rightarrow \nabla$, $\vec{b} \rightarrow \vec{r}$ and $\vec{c} \rightarrow \vec{B}$. Keeping in mind that $\vec{B}$ is homogeneous we shift $\vec{B}$ to the left and obtain

$$-\frac{1}{2} \nabla \times (\vec{r} \times \vec{B}) = -\frac{1}{2} \left( (\vec{B} \cdot \nabla)\vec{r} - \vec{B}(\nabla \cdot \vec{r}) \right) = -\frac{1}{2} (\vec{B} - 3\vec{B}) = \vec{B}.$$

Thus Eq. 1 and Eq. 2 are the potentials which in the transverse gauge produce a set of homogeneous in space and constant in time $\vec{E}, \vec{B}$-fields.

---

End IX–3: Potentials for homogeneous & uniform $\vec{E}, \vec{B}$-fields

---

Exercise IX–4: EM-potentials in transverse gauge

---

Given any integrable and regular EM-fields $\vec{E}$ and $\vec{B}$, obtain analytical solutions for potentials $V, \vec{A}$ in transverse gauge.

Solution
We consider first the time independent second equation in Eq. (23.6) subject to \( \vec{\nabla} \cdot \vec{B} = 0 \), see Eq. (23.5) to obtain in transverse gauge

\[
\vec{A} = \vec{\nabla} \times \frac{1}{4\pi} \int d^3 r' \frac{\vec{B}(r', t)}{|\vec{r} - \vec{r}'|}.
\]

We see that by construction Eq. (1) satisfies the transverse gauge condition Eq. (23.9) since for any regular vector function \( \vec{W}(x) \) such that \((\partial_i \partial_j - \partial_j \partial_i) \vec{W} = 0\), \(i, j = x, y, z\) we have

\[
\vec{\nabla} \cdot (\vec{\nabla} \times \vec{W}) = \begin{vmatrix}
\partial_x & \partial_y & \partial_z \\
\partial_x & \partial_y & \partial_z \\
W_x & W_y & W_z
\end{vmatrix} = 0.
\]

We now show that Eq. (1) is also a solution we seek to the second equation in Eq. (23.6). For this we must evaluate \( \vec{\nabla} \times \vec{A} \) which requires understanding of \( \vec{\nabla} \times (\vec{\nabla} \times \vec{W}) \). We will use bac-cab rule from Eq. 3 in exercise IX–3 as follows

\[
\vec{a} \times (\vec{b} \times \vec{c}) = \vec{b}(\vec{a} \cdot \vec{c}) - (\vec{a} \cdot \vec{b})\vec{c} \Rightarrow \vec{\nabla} \times (\vec{\nabla} \times \vec{W}) = \vec{\nabla} (\vec{\nabla} \cdot \vec{W}) - (\vec{\nabla} \cdot \vec{\nabla}) \vec{W}.
\]

Differentiating Eq. (1) we obtain

\[
\vec{\nabla} \times \vec{A} = \vec{\nabla} \frac{1}{4\pi} \int d^3 r' \left( \vec{B}(r', t) \cdot \vec{\nabla} \frac{1}{|\vec{r} - \vec{r}'|} \right) - \frac{1}{4\pi} \int d^3 r' \frac{\vec{B}(r', t)}{|\vec{r} - \vec{r}'|} \frac{1}{|\vec{r} - \vec{r}'|}.
\]

In the first term we change inside the integral from not primed to primed differentiation and we integrate by parts and in the second we use the Poisson equation Eq. (23.11) to obtain

\[
\vec{\nabla} \times \vec{A} = -\vec{\nabla} \frac{1}{4\pi} \int d^3 r' \frac{1}{|\vec{r} - \vec{r}'|} \vec{\nabla} \cdot \vec{B}(r', t) + \int d^3 r' \frac{\vec{B}(r', t)}{|\vec{r} - \vec{r}'|} \delta^3(\vec{r} - \vec{r}') = \vec{B}(\vec{r}, t).
\]

The first term vanishes in consideration of Eq. (23.5). We further need to assure that in the integration by parts described above the surface term vanishes, which is the case if the field \(|\vec{B}|\) decreases faster than \(1/r\) at large \(r\).

Next we turn to the time dependent first equation in Eq. (23.6). We form the divergence to obtain in transverse gauge \( \vec{\nabla} \cdot \vec{A} = 0 \)

\[
\vec{\nabla} \cdot \vec{E} = -\frac{1}{c} \frac{\partial \vec{\nabla} \cdot \vec{A}}{\partial t} - \vec{\nabla}^2 V = -\triangle V.
\]

Note that the time dependence of \( V \) is now a direct consequence of the possible time dependence of \( \vec{\nabla} \cdot \vec{E} \). We solve Eq. (5) using the fundamental solution of Poisson equation Eq. (23.11) for \( V \)

\[
V = \frac{1}{4\pi} \int d^3 r' \frac{\vec{\nabla}' \cdot \vec{E}(r', t)}{|\vec{r} - \vec{r}'|}.
\]
Note that one sees this equation written replacing $\nabla' \cdot \vec{E}(\vec{r}', t) = 4\pi \rho(\vec{r}', t)$, which is correct but we are seeking now to find in transverse gauge the potentials as a functional of the EM-fields. Therefore we now perform integration by parts – obtaining also a surface term that vanishes in general if the field $|\vec{E}|$ decreases faster than $1/r$ at large $r$. In 2nd step we change from derivative with respect to $\vec{r}'$ to derivative with respect to $\vec{r}$, the signs from each step cancel and we obtain

$$ V = \nabla \cdot \frac{1}{4\pi} \int d^3r' \frac{\vec{E}(\vec{r}', t)}{|\vec{r} - \vec{r}'|}. $$

With Eq. 1 and Eq. 8 we obtained in the transverse gauge Eq. (23.9) a unique definition of potentials $V, \vec{A}$ given regular and integrable fields $\vec{E}, \vec{B}$.

End IX–4: EM-potentials in transverse gauge

### 23.3 The Lorentz-force: from fields to potentials

When one inserts Eq. (23.6) into Eq. (23.1), the term $\vec{v} \times (\nabla \times \vec{A})$ arises, requiring the use of the bac-cab rule to evaluate the double cross product, see Eq. 3 in exercise IX–3

$$ \vec{a} \times (\vec{b} \times \vec{c}) = \vec{b}(\vec{a} \cdot \vec{c}) - (\vec{a} \cdot \vec{b})\vec{c} \Rightarrow \vec{v} \times (\nabla \times \vec{A}) = \nabla(\vec{v} \cdot \vec{A}) - (\vec{v} \cdot \nabla)\vec{A}. \quad (23.13) $$

We have sequence the terms such that $\vec{b} \to \nabla$ is always to the left of $\vec{c} \to \vec{A}$ keeping the operator $\nabla$ operating on $\vec{A}$. However the location of $\vec{a} \to \vec{v}$ can be right of $\nabla$ since the velocity $\vec{v}$ is not explicitly dependent on the position of the particle. Eq. (23.13) allows us to rewrite Eq. (23.1) in terms of potentials $V, \vec{A}$, Eq. (23.6)

$$ \vec{F}_L = -\nabla eV - \frac{\partial e\vec{A}}{\partial t} + \nabla(\vec{v} \cdot e\vec{A}) - (\vec{v} \cdot \nabla)e\vec{A}. \quad (23.14) $$

We further note that the total derivative of $\vec{A}$ with respect to time is:

$$ \frac{d\vec{A}}{dt} = \frac{\partial \vec{A}}{\partial t} + 3 \sum_{i=1} dx_i \frac{\partial}{\partial x_i} \vec{A} = \frac{\partial \vec{A}}{\partial t} + (\vec{v} \cdot \nabla)\vec{A}. \quad (23.15) $$

We now rewrite Eq. (23.14) using Eq. (23.15) to obtain for $\vec{F}_L$:

$$ \vec{F}_L = -e\nabla \left( V - \vec{v} \cdot \vec{A} \right) - \frac{d e\vec{A}}{dt}. \quad (23.16) $$

We should remember that the last term in Eq. (23.16) comprises the change as a function of time that a particle experiences due to both a direct time-dependence
of the vector field $\vec{A}(\vec{r},t)$ and the (proper) time-dependent change in position of the particle $\vec{r}(\tau(t))$.

Moving the last term in Eq. (23.16) to left side in Newton’s equation Eq. (23.1) we obtain
\[
d\left( m\gamma\vec{v} + e\vec{A} \right) = -e\nabla \left( V - \vec{v} \cdot \vec{A} \right).
\] (23.17)

We recognize that when the gradient on right hand side of Eq. (23.16) vanishes in some spatial direction, the generalized ‘canonical’ momentum that appears on the left side of Eq. (23.17)
\[
\vec{p} = m\vec{v}/\sqrt{1 - v^2/c^2} + e\vec{A},
\] (23.18)
pointing in this direction is conserved. Conservation of canonical momentum components can be a helpful feature in determining charged particle motion which is best exploited in Lagrangian formulation of classical mechanics.

### 23.4 Lorentz-force from variational principle

In classical nonrelativistic mechanics we learn that it is possible to obtain Newton’s force law by considering a variation of the action $I$ along a path a particle takes in space as a function of time. The Lagrange function $L$ integrated along the path between events $P_1$ and $P_2$ provides the action
\[
I = \int_{P_1}^{P_2} L(\vec{r},\dot{\vec{r}}) \, dt, \quad L = T - U.
\] (23.19)

As indicated, the Lagrange function $L$ and hence the action $I$ depends on space coordinated $\vec{r}$ where at time $t$ the considered particle can be found, and, of the particle velocity $\dot{\vec{r}} = \vec{v}$, an independent dynamical property of the particle. Only after we have obtained dynamical equations of motion can $\vec{r}$ be related to $\vec{r}$, as it is the case considering Newton’s force law.

The search for stationary point conditions (also called critical point condition) of the action Eq. (23.19) establishes these dynamical equations that can be solved for $\vec{r}(t)$ and $\dot{\vec{r}}(t)$. When seeking to find the best path $\vec{r}(t)$ for which the action $I$ is smallest we require that the particle passes through the same initial and final coordinates $P_1, P_2$. To find the best path we perform a variation
\[
\vec{r}(t) \to \vec{r}(t) + \delta\vec{r}, \quad \delta\vec{r}|_{P_1} = 0, \quad \delta\vec{r}|_{P_2} = 0,
\] (23.20)

which results in
\[
I = \int_{P_1}^{P_2} L(\vec{r},\dot{\vec{r}}) \, dt \to I + \delta I, \quad \delta I = \int_{P_1}^{P_2} \delta\vec{r} \cdot \vec{S}.
\] (23.21)
Thus if \( \delta \vec{r} \) is arbitrary, the ‘least action’ path \( \vec{r}(t) \) is found solving the dynamical equation

\[
\vec{S}(\vec{r}, \dot{\vec{r}}) = 0 .
\]

(23.22)

In this procedure borrowed from nonrelativistic formulation, \( t \) is a parameter of motion. Searching for the minimum of \( I \) we do not incorporate a change in elapsed time for each different path considered. Thus the unity of space and time is not (yet) implemented.

The action \( I \) includes as indicated in Eq. (23.19) both kinetic \( T \) and potential \( U \) energy. The motion related \( T \) is typically only a function of \( \dot{\vec{r}} \) and not of \( \vec{r} \). The action-potential \( U \) is seen as an ‘external’ field generated by all other particles of the system. The presumption of this approach is that the one particle we study does not perturb any of the external particles, and for that matter, neither does it perturb the field configuration significantly. In principle one can attempt to incorporate the ‘back-reaction’ effect, that is modify \( U \) by the dynamics of the considered particle. However, causality can be compromised by attempting such improvements in an ad-hoc way\(^4\). A consistent description of multiparticle relativistic dynamics allowing for mutual particle influence may require a manifestly covariant approach in the context of relativistic Hamiltonian mass function, see section 25.3 on page 375.

To obtain \( \delta I \) Eq. (23.21) we consider

\[
\delta I = \int_{P_1}^{P_2} \left[ L \left( \vec{r} + \delta \vec{r}, \frac{d(\vec{r} + \delta \vec{r})}{dt} \right) - L \left( \vec{r}, \dot{\vec{r}} ; t \right) \right] dt = \int_{P_1}^{P_2} \left( \delta \vec{r} \frac{\partial L}{\partial \vec{r}} + \left( \frac{d \delta \vec{r}}{dt} \right) \frac{\partial L}{\partial \dot{\vec{r}}} \right) dt.
\]

(23.23)

We integrate by parts the second term to obtain

\[
\delta I = \int_{P_1}^{P_2} \delta \vec{r} \left( \frac{\partial L}{\partial \vec{r}} - \frac{d}{dt} \frac{\partial L}{\partial \dot{\vec{r}}} \right) dt + \left. \left( \delta \vec{r} \cdot \frac{\partial L}{\partial \dot{\vec{r}}} \right) \right|_{P_1}^{P_2} .
\]

(23.24)

Since according to Eq. (23.20) we consider arbitrary variations \( \delta \vec{r} \) vanishing at the end points of the integration domain, we identify the usual Lagrange equation of motion

\[
\frac{\partial L}{\partial \vec{r}} - \frac{d}{dt} \frac{\partial L}{\partial \dot{\vec{r}}} = 0 .
\]

(23.25)

The differentiation with respect to \( \vec{r} \) has, for a function that depends on \( \vec{r} \), the same meaning as \( \nabla \). Since the Lagrange equation is fixing the velocity of a particle \( \dot{\vec{r}} = \vec{v} \), from this point forward we use \( \vec{v} \) or \( |\vec{v}| = v \) instead of \( \dot{\vec{r}} \).

The generalization of the kinetic term \( T = \frac{mv^2}{2} \) to relativistic formulation must produce the modification of the \( m \to m\gamma \) on the left hand side of Newton’s equations, Eq. (22.21). When used in the first term in Eq. (23.25), this

must produce the inertial part of Newton’s equation. We recall the relativistic generalization of $T \rightarrow T_r$ as

$$T = m\frac{v^2}{2} \rightarrow T_r = -mc^2\sqrt{1-v^2/c^2} = -mc^2 + m\frac{v^2}{2} - \ldots$$ (23.26)

where the constant term $-mc^2$ does not impact the result of variation. We find for the inertia term of the Lorentz-force

$$\frac{d}{dt} \frac{\partial T_r}{\partial \vec{v}} = \frac{d\vec{p}_I}{dt}, \quad \vec{p}_I = \frac{\partial T_r}{\partial \vec{v}} = \frac{m\vec{v}}{\sqrt{1-v^2/c^2}}.$$ (23.27)

Here we introduced the inertial momentum $\vec{p}_I$. It differs from the generalized canonical momentum $\vec{p}$ we introduced in Eq. (23.18)

$$\vec{p}_I = \vec{p} - e\vec{A}.$$ (23.28)

The potential term $U$ leading to the Lorentz-force Eq. (23.16) has the same form

$$U = e\left(V - \vec{v} \cdot \vec{A}\right).$$ (23.29)

We verify using Eq. (23.25): with

$$L = T_r - U,$$ (23.30)

we obtain for the canonical momentum

$$\vec{p} \equiv \frac{\partial L}{\partial \vec{v}} = \frac{\partial T_r}{\partial \vec{v}} - \frac{\partial U}{\partial \vec{v}} = \vec{p}_I + e\vec{A},$$ (23.31)

and we obtain the resultant form for the force

$$\frac{d\vec{p}_I}{dt} = \frac{d}{dt} \frac{\partial U}{\partial \vec{v}} - \frac{e}{c} \frac{dA}{dt} - \vec{\nabla} \left(eV - \vec{v} \cdot e\vec{A}\right),$$ (23.32)

which agrees with $\vec{F}_L$, Eq. (23.16). This confirms the choice Eq. (23.29) for the potential. Thus we have found for the relativistic particle dynamics in the presence of electromagnetic fields described by the Lagrange function

$$L = -mc^2\sqrt{1-v^2/c^2} - e\left(V - \vec{v} \cdot \vec{A}\right),$$ (23.33)

which leads upon variation to the Lorentz-force, Eq. (23.1). We can further use the canonical momentum Eq. (23.31) to obtain the equivalent form of the Lorentz force

$$\frac{d\vec{p}}{dt} = \frac{d}{dt} \left(\vec{p}_I - \frac{d}{dt} \frac{\partial U}{\partial \vec{v}}\right) = -\vec{\nabla} \left(eV - \vec{v} \cdot e\vec{A}\right).$$ (23.34)
This establishes the conservation of canonical momentum in special cases of interest, i.e. when the right hand side in Eq. (23.34) vanishes.

Given that the energy of a particle is connected to momentum to form a 4-vector, it seems appropriate to explore the form of the Hamilton $H(\vec{p}, \vec{r}; t)$, to assure that this connection remains valid. We recall

$$H(p_i, r_i; t) \equiv \sum_i \dot{r}_i p_i - L(\vec{r}_i, \dot{\vec{r}}_i; t), \quad p_i = \frac{\partial L}{\partial \dot{r}_i}.$$  \hspace{1cm} (23.35)

It is not our present intent to develop Hamiltonian mechanics, e.g. to demonstrate that the Legendre transform that defines $H$, (see Eq. (23.35)), implies the variable dependence as shown$^5$. To obtain Hamilton’s dynamical equations from a variational principle we vary the action

$$I = \int dt \, L(r_i, \dot{r}_i; t) = \int dt \left( \sum_i \dot{r}_i p_i - H(p_i, r_i; t) \right).$$  \hspace{1cm} (23.36)

Note that the subscript ‘i’ indicates now all independent components, these can be the vector components of one or many particles. The variation of $I$, Eq. (23.36), leads to, see Ref.$^5$

$$\delta I = \int dt \sum_i \left\{ \delta p_i \left( r_i - \frac{\partial H}{\partial \dot{r}_i} \right) + \delta r_i \left( -p_i - \frac{\partial H}{\partial q_i} \right) \right\} = 0.$$  \hspace{1cm} (23.37)

Since the variations are all independent we obtain the canonical equations of motion

$$\dot{r}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial r_i}.$$  \hspace{1cm} (23.38)

A direct consequence of is

$$\frac{dH(p_i, r_i; t)}{dt} = \sum_i \left( \dot{r}_i \frac{\partial H}{\partial r_i} + \dot{p}_i \frac{\partial H}{\partial p_i} \right) + \frac{\partial H}{\partial t} = \frac{\partial H}{\partial t},$$  \hspace{1cm} (23.39)

in consideration of the canonical equations of motion Eq. (23.38).

We thus find, Eq. (23.39), that the total system energy changes only when the action potential is explicitly time-dependent. Setting $H(p, r; t) \rightarrow E$

$$\frac{dE}{dt} = -\frac{\partial L}{\partial t} = \frac{\partial U}{\partial t}.$$  \hspace{1cm} (23.40)

To understand better the meaning of Eq. (23.40), we recall that when $L$ is not explicitly time-dependent we have particle motion where there is change in velocity $\vec{v}$ when kinetic and potential components of particle energy shift. However,

even for a complicated many-body system as these shifts occur and the speed of each particle changes, the system total energy – and thus the system rest mass – never changes, as we discussed in section 16. However, if we study the dynamics of a system that is only partially described, then the action potential $U$ can be manifestly time-dependent. As a consequence we would expect the sum of kinetic and potential energies not to be a constant, as energy could now be provided by the external action potential $U$. In the Hamiltonian formulation of particle dynamics one finds that the particle energy changes as shown in Eq. (23.40).

We now evaluate explicitly the Hamiltonian Eq. (23.35) for one particle. We use Eq. (23.31) to compute $\mathbf{p} \cdot \mathbf{\dot{r}}$ we thus obtain for the Hamiltonian

$$H = mc^2 \frac{-v^2/c^2}{\sqrt{1 - v^2/c^2}} + \bar{v} \cdot e\bar{A} + mc^2 \sqrt{1 - v^2/c^2} + e \left( V - \bar{v} \cdot \bar{A} \right). \quad (23.41)$$

We see that the second and last term cancel and we combine the first and third term to obtain

$$H - eV = \frac{mc^2}{\sqrt{1 - v^2/c^2}}, \quad (23.42)$$

We need to replace the speed in Eq. (23.42) by canonical momentum. In Eq. (23.31) we move $e\bar{A}$ to left and we square

$$(\mathbf{p} - e\bar{A})^2 = \frac{m^2c^2(v^2/c^2 - 1 + 1)}{1 - v^2/c^2} = -m^2c^2 + \frac{m^2c^2}{1 - v^2/c^2} \quad (23.43)$$

which when combined with Eq. (23.42) yields the relativistic Hamiltonian

$$H = eV + \sqrt{(mc^2)^2 + c^2(\mathbf{p} - e\bar{A})^2}, \quad (23.44)$$

Finally, we write Eq. (23.44) in the suggestive invariant format

$$\begin{align*}
(H - eV)^2 - c^2(\mathbf{p} - e\bar{A})^2 &= (mc^2)^2, \\
(23.45)
\end{align*}$$

showing $H - eV$ as the timelike and $c(\mathbf{p} - e\bar{A})$ as the spacelike component of a 4-vector. This result confirms the relativistic relationship between all quantities. We keep in mind that $H$ is the energy of the particle; $\mathbf{p}$ is the canonical momentum not simply related to the measured velocity $\bar{v}$ of a particle, in that it also includes the potential term $\bar{A}$.

---

**Exercise IX–5: Relativistic motion in $1/r$ potential**

Obtain periodic orbits of a relativistic particle in an attractive $1/r$ potential: determine a) the angular relativistic orbits $r(\phi)$; b) the orbit energy and c) the orbit size
parameter, as a function of angular momentum and orbit eccentricity. Using Bohr quantization condition adapt your solution to the special case an electron in the Coulomb field of a point nucleus of charge $|e|Z$.

**Solution**

We consider particle motion in a radial force field according to Eq. (22.21)

$$\frac{dm\gamma \vec{v}}{dt} = eE_r(r)\hat{r},$$

where $\hat{r}$ is the unit vector in the radial direction. For the motion of an electron in the Coulomb field of a point nucleus we have specifically

$$eV_c = -\frac{e^2}{4\pi \epsilon_0} \frac{Z}{r}, \quad e\vec{E} = -\nabla eV \rightarrow eE_r = -\frac{\partial eV_c}{\partial r} \rightarrow eE_r(r) = -\frac{e^2}{4\pi \epsilon_0} \frac{Z}{r^2}.$$

Note that the Coulomb potential $eV_c$ is negative (attractive), since the unit charges associated with a proton and an electron are of opposite sign. For some more details see exercise XI–2 on page 405.

The relativistic angular momentum

$$\vec{l} \equiv \vec{r} \times m\gamma \vec{v},$$

is a constant of motion:

$$\frac{d\vec{l}}{dt} = \vec{v} \times m\gamma \vec{v} + \vec{r} \times \frac{dm\gamma \vec{v}}{dt} = 0 + \vec{r} \times qE_r(r)\hat{r} = 0.$$

The constancy of $\vec{l}$ assures motion in the plane normal to $\vec{l}$, allowing use of polar coordinates $(r, \phi)$, and

$${|l|} = l = m\gamma r^2 \frac{d\phi}{dt}.$$

Another constant of motion is given in Eq. (22.36), or, equivalently, Eq. (23.44), which we write in form

$$\left(E + \frac{e^2}{4\pi \epsilon_0} \frac{Z}{r}\right)^2 = c^2 \vec{p}^2 + m^2 c^4.$$

The conventional vector algebra can be used to find for the momentum $\vec{p}$

$$\vec{p} \equiv \hat{r} p_r + \hat{\phi} p_\phi = m\gamma \left(\hat{r} \frac{dr}{dt} + \hat{\phi} r \frac{d\phi}{dt}\right) = \hat{r} m\gamma \frac{dr}{dt} + \hat{\phi} \frac{l}{r},$$

where we used Eq. [5] in the last equality. Thus we identify

$$p_r = m\gamma \frac{dr}{dt}, \quad p_\phi \equiv \frac{l}{r}, \quad \vec{p}^2 = p_r^2 + p_\phi^2.$$
This allows us to write for Eq. 6

\[ 9 \quad m^2c^4 \left( \frac{E}{mc^2} + \frac{Zr_e}{r} \right)^2 = m^2c^4 + \frac{l^2c^2}{r^2} + p_r^2c^2, \]

where to short-hand the notation we have introduced the classical electron radius \( r_e \), see Eq. (19.42) and Eq. (23.4).

We are interested in the orbit equation, that is to determine \( r(\phi) \), and this is also allowing us to recast the last term in terms of angular momentum. The procedure is akin to the non-relativistic case since the relativistic Lorentz-factors cancel. We consider

\[ 10 \quad \frac{dr}{d\phi} = \frac{dr}{dt} \frac{1}{d\phi/dt} = m^\gamma \frac{dr}{dt} \frac{1}{m^\gamma d\phi/dt} = p_r \frac{r}{p_\phi}, \quad \Rightarrow p_r = \frac{dr}{d\phi} \frac{r}{p_\phi} = \frac{l}{r^2} \frac{dr}{d\phi}. \]

Using Eq. 10 in Eq. 9 produces the relativistic orbit equation

\[ 11 \quad \left( \frac{E}{mc^2} + \frac{Zr_e}{r} \right)^2 = 1 + \left( \frac{l}{mca} \right)^2 \left[ 1 + \left( \frac{1}{r} \frac{dr}{d\phi} \right)^2 \right] \]

which in principle can be solved for \( r(\phi) \). As a first step we simplify the form by introducing the orbit variable \( s \)

\[ 12 \quad s = \frac{a}{r}, \quad s' = \frac{ds}{d\phi} = -\frac{a}{r^2} \frac{dr}{d\phi}, \]

and we find sequencing the terms in powers of \( s \) and \( s' \)

\[ 13 \quad \left( \frac{E}{mc^2} \right)^2 - 1 = -2 \frac{Zr_e}{a} \frac{E}{mc^2} s + \left[ \left( \frac{l}{mca} \right)^2 - \left( \frac{Zr_e}{a} \right)^2 \right] s^2 + \left( \frac{l}{mca} \right)^2 s'^2. \]

Like in the non-relativistic case the simplest way to find a solving function is to differentiate with respect to \( \phi \). After dividing by \( 2s'(l/mca)^2 \) we obtain

\[ 14 \quad s'' + \left[ 1 - \left( \frac{Zmcr_e}{l} \right)^2 \right] s = \frac{ZEmar_e}{l^2}. \]

The choice of scaling factor \( a \) is now made such that

\[ 15 \quad \frac{ZEmar_e}{l^2} = \left[ 1 - \left( \frac{Zmcr_e}{l} \right)^2 \right] \rightarrow a = \frac{Zmc^2r_e}{E} \left[ \left( \frac{l}{Zmcr_e} \right)^2 - 1 \right]. \]

This allows us to write for Eq. 14

\[ 16 \quad (s - 1)'' + \left[ 1 - \left( \frac{Zmcr_e}{l} \right)^2 \right] (s - 1) = 0, \]
and hence we find the general periodic i.e. orbiting particle solution
\[ s - 1 = \epsilon_x \cos(\kappa \phi + \delta) \]
which implies
\[ a \equiv \frac{1}{r} \rightarrow r = \frac{a}{1 + \epsilon_x \cos(\kappa \phi + \delta)} , \quad |\epsilon_x| < 1 , \quad 0 < \kappa^2 = 1 - \left( \frac{Zmc^2r_e}{cl} \right)^2 < 1 . \]

The choice of initial value \( \delta \rightarrow 0 \) is common. However, we can absorb the sign of the eccentricity \( \epsilon_x \) by choice \( \delta \rightarrow \pi \). This shows that orbits with positive and negative \( \epsilon_x \) are identical. Like in nonrelativistic mechanics \( \epsilon_x \) describes the deviation of the orbit from circular shape. Together with energy \( E \), and angular momentum \( l \), the parameter \( \epsilon_x \) is determined for each orbit observed. Note that in the non-relativistic limit \( c \rightarrow \infty \) since \( mc^2r_e = \text{Const.} \) we obtain \( \kappa = 1 \). The key feature of relativistic orbits with \( 0 < \kappa < 1 \) is that the orbit in general does not close, this effect is called apsidal precession we describe in next exercise IX–6. In this computation it will be useful to recall the orbit size parameter \( a \), Eq. (15) written as a function of \( \kappa^2 \)

\[ a = \frac{Zmc^2r_e}{E} \left[ \left( \frac{l}{Zmc^2r_e} \right)^2 - 1 \right] = \frac{Zmc^2r_e}{E} \frac{\kappa^2}{1 - \kappa^2} . \]

Inserting the solution Eq. (17) in Eq. (11) we not only check the math, but also obtain constraints between the three parameters \( E, l, \epsilon_x \) characterizing the orbit. The required last term in Eq. (11) follows differentiating the orbit solution Eq. (17)

\[ \frac{1}{r} \frac{dr}{d\phi} = \kappa \epsilon_x \sin \kappa \phi , \quad \rightarrow \frac{1}{r} \frac{dr}{d\phi} = \frac{\kappa \epsilon_x \sin \kappa \phi}{1 + \epsilon_x \cos \kappa \phi} . \]

To shorten notation we hence will use

\[ \epsilon_x \cos \kappa \phi \equiv c_{\kappa} , \quad \epsilon_x \sin \kappa \phi \equiv s_{\kappa} , \quad s_{\kappa}^2 + c_{\kappa}^2 = \epsilon_x^2 . \]

Inserting Eq. (19) and substituting for \( r \) Eq. (17) we obtain for Eq. (11)

\[ \left( \frac{E}{mc^2} + \frac{Zr_e}{a} (1 + c_{\kappa}) \right)^2 = 1 + \left( \frac{l}{mca} \right)^2 \left[ (1 + c_{\kappa})^2 + \kappa^2 s_{\kappa}^2 \right] . \]

On left we regroup terms and on right we can simplify using the form of \( \kappa^2 \), Eq. (17)

\[ \left[ \left( \frac{E}{mc^2} + \frac{Zr_e}{a} \right) + \frac{Zr_e}{a} c_{\kappa} \right]^2 = 1 + \left( \frac{l}{mca} \right)^2 \left( 1 + \epsilon_x^2 + 2c_{\kappa} \right) - \left( \frac{l}{mca} \frac{Zmc^2r_e}{l} \right)^2 s_{\kappa}^2 . \]

Where on right we note the ensuing simplification. We now group odd, and even terms in \( \epsilon_x \), on left hand side, and on right hand side respectively and we obtain

\[ \text{Lhs} = 2 \left[ \left( \frac{E}{mc^2} + Zr_e \right) \frac{Zr_e}{a} - \left( \frac{l}{mca} \right)^2 \right] c_{\kappa} = \text{Rhs} \]

\[ \text{Rhs} = 1 + (1 + \epsilon_x^2) \left( \frac{l}{mca} \right)^2 - \left( \frac{Zr_e}{a} \right)^2 (s_{\kappa}^2 + c_{\kappa}^2) - \left( \frac{E}{mc^2} + \frac{Zr_e}{a} \right)^2 \]
Since there is no $\phi$-dependence in Rhs, both sides of the equation must independently vanish. On the left hand side we note that since $1/a \propto E$, see Eq. (15), the factor $E^2$ cancels and only $l$ remains from among the three orbital parameters, thus this term is zero as a trivial identity. This is verified easily using the relation following from Eq. (15)

\[
\frac{Zre}{a} = \frac{E}{mc^2} \frac{1}{(1/Zmcr)e^2 - 1} = \frac{E}{mc^2} \frac{\tilde{Z}^2}{1 - \tilde{Z}^2}, \quad \tilde{Z} \equiv \frac{Zmcr_e}{l}.
\]

Hence given that Lhs=0, we are left with the one single constraint, Rhs=0, which defines the energy of the orbit as a function of $l$ and $\epsilon_x^2$

\[
0 = \text{Rhs} = 1 + (1 + \epsilon_x^2) \left( \frac{Zre}{a} \frac{l}{Zmcr_e} \right)^2 - \epsilon_x^2 \left( \frac{Zre}{a} \right)^2 - \left( \frac{E}{mc^2} + \frac{Zre}{a} \right)^2.
\]

Using Eq. (25) and dividing by the common factor $(E/mc^2)^2$ this simplifies to

\[
0 = \left( \frac{mc^2}{E} \right)^2 + (1 + \epsilon_x^2) \frac{\tilde{Z}^2}{(1 - \tilde{Z}^2)^2} - \epsilon_x^2 \frac{\tilde{Z}^4}{(1 - \tilde{Z}^2)^2} - \frac{1}{(1 - \tilde{Z}^2)^2}.
\]

This algebraic equation has a simple solution

\[
E = mc^2 \sqrt{\frac{1 - \tilde{Z}^2}{1 + \epsilon_x^2 \tilde{Z}^2}}.
\]

In order to make contact with relativistic quantum mechanics we now introduce Bohr’s quantization condition for the angular momentum $l$

\[
l = n \hbar, \quad \frac{\hbar}{mc} = \lambda_C, \quad \frac{r_e}{\lambda_C} = \frac{e^2}{4\pi e_0 mc^2} \frac{mc^2}{\hbar c} = \frac{e^2}{4\pi e_0 \hbar c} \equiv \alpha = \frac{1}{137 035 999}
\]

where we also show the quantum-related natural constants. Checking the definition of $\tilde{Z}$ and $\kappa$ we find

\[
\tilde{Z} \equiv \frac{Zmcr_e}{l} = \frac{Z\alpha}{n}, \quad \kappa = \sqrt{1 - \left( \frac{Z\alpha}{n} \right)^2}.
\]

This leads to

\[
E = mc^2 \sqrt{\frac{1 - (Z\alpha/n)^2}{1 - (\epsilon_x Z\alpha/n)^2}}, \quad r = \frac{a}{1 + \epsilon_x \cos(\kappa \phi)}
\]

The energy of the bound particle assumes in case of an circular orbit $\epsilon_x = 0$ the value

\[
E = mc^2 \sqrt{1 - \left( \frac{Z\alpha}{n} \right)^2}, \quad \text{as} \rightarrow mc^2 \left( 1 - \frac{1}{2} \left( \frac{Z\alpha}{n} \right)^2 + \ldots \right).
\]
342 23. The Lorentz Force

Figure 23-6: The classical relativistic energy of electrons bound by \( V = -Z\alpha/r \) Coulomb potential as a function of \( Z \), for angular momentum \( l = n\hbar, n = 1, 2 \).

The small-\( Z \) expansion produces the Schrodinger equation spectrum as is indicated to the right. In the limit of strong coupling the singular behavior of the classical binding of electrons (\( m_e = 0.511 \text{ MeV} \)) for \( Z\alpha \to 1 \), is depicted for \( n = 1, 2 \) as a function of \( Z \) in figure 23-6.

We now determine the value of \( a \), the orbit size parameter, Eq. (15). In the notation previously introduced we obtain

\[
33 \quad a = \frac{lc\tilde{Z}}{E} \left( \frac{1}{\tilde{Z}^2} - 1 \right) = \frac{l}{mc} \sqrt{1 - \left(\frac{\epsilon_x\tilde{Z}}{1 - \tilde{Z}^2} \right)^2} \quad \frac{1 - \tilde{Z}^2}{\tilde{Z}} = \frac{l}{mc\tilde{Z}} \sqrt{(1 - \tilde{Z}^2)(1 - (\epsilon_x\tilde{Z})^2)}.
\]

In the Bohr quantization case we have again \( l \to h_n, \tilde{Z} \to Z\alpha/n \) and we find

\[
34 \quad a = a_B \frac{n^2}{Z} \sqrt{\left(1 - \left(\frac{Z\alpha}{n}\right)^2\right) \left(1 - \left(\frac{\epsilon_x Z\alpha}{n}\right)^2\right)}, \quad a_B = \frac{r_e}{\alpha^2} = \frac{\lambda_C}{\alpha},
\]

where we have introduced the Bohr radius \( a_B = 0.529177211 \times 10^{-10} \text{ m} \). In Eq. (34) we note the scaling with \( n^2/Z \) of the orbit radius \( a \), familiar from quantum mechanics.

This completes the solution of the relativistic Coulomb problem: we have obtained the orbit shape of the particle, and the energy of the orbit as a function of angular momentum and orbit eccentricity. We have seen for increasing coupling strength \( Z \) that the binding turns to consume the entire mass of a particle. We therefore addressed here solely the case \( \tilde{Z} \equiv Z\alpha/n < 1 \), i.e. \( \kappa^2 > 0 \) for which periodic quantum physics
Figure 23-7: Apsidal precession: shown is the precession of the aphelion by $\Delta \phi$.

solutions exist$^6$. Classical solutions were presented for $Z\alpha/n > 1$ showing, however, considerable pathologies$^7$.

---

End IX–5: Relativistic motion in $1/r$ potential

---

Exercise IX–6: Apsidal precession in $1/r$ potential

The SR motion in $1/r$ potential shows an apsidal precession made famous as a GR test. Show that this is true by comparing the magnitude of SR-Coulomb and GR effects which differ substantially.

Solution

Since $\kappa < 1$, see Eq.[17] above, the periodicity in $r$ requires $\kappa \phi_c = 2\pi$ and since according to Eq.[17] $\kappa < 1$ we have $\phi_c > 2\pi$: the elliptic $0 < \varepsilon < 1$ orbits do not close in exact fashion, an example is seen in figure [23-7]. We find

\[ \Delta \phi \equiv \phi_c - 2\pi = 2\pi \frac{1 - \kappa}{\kappa} = 2\pi \frac{1 - \kappa^2}{\kappa(1 + \kappa)} = \frac{2\pi \kappa}{1 + \kappa} \frac{1 - \kappa^2}{\kappa^2}. \]

Using the value of $a$ from Eq.[18]

\[ \Delta \phi = \frac{2\pi \kappa}{1 + \kappa} \frac{Zmc^2r_e}{E} \frac{1}{a}. \]

---


$^7$T.H. Boyer, “Unfamiliar trajectories for a relativistic particle in a Kepler or Coulomb potential”, Am. J. Phys. 72, 992 (2004). Another resource for classical relativistic Coulomb potential solution is: J.D. Garcia, “Quantum solutions and classical limits for strong Coulomb fields”, Phys. Rev. A 34, 4396 (1986); both these works overlap and complement each other, clearly do not know of each other, and in their references are fully orthogonal, an example of how science is inadvertently duplicated in two parallel developments.
One calls $\Delta \phi$ apsidal precession remembering the observation of aphelion and perihelion precession of Mercury. The aphelion (opposite of perihelion) is the farthest point from the location of the Sun in an elliptic orbit. Note that in figure [23-7] we show aphelion precession more distinct for the choice of illustration parameters.

In the case of Mercury, all known nonrelativistic perturbations (e.g. the influence of other planets, Sun asphericity) combine to 532”/century, with a discrepancy of 43”/century compared to the experimental value 574”/century. Explanation of this discrepancy was of pivotal importance in establishing Einstein’s GR\(^8\). Our SR Coulomb bound system is, however, not an equivalent model of relativistic effects in planetary motion in a gravitational field.

The apsidal precession motion of Mercury according to SR Eq.\(^4\) is made explicit introducing the gravitational coupling

\[
\frac{Z e^2}{4\pi \varepsilon_0} = mc^2 (Z \tau_e) \rightarrow \frac{mc^2}{2} \frac{2GM}{c^2} = \frac{mc^2}{2} R_S, \quad R_S = 2.953 \text{ km}
\]

where $R_S$ is the solar Schwartzschild radius $\simeq 3 \text{ km}$, and $m$ planetary mass. Given a typical planetary speed, $10^{-5}c$ and eccentricity $\epsilon^2 \leq 0.1$ it is appropriate in a first evaluation to use in Eq.\(^4\) $\kappa \simeq 1$, $E = mc^2$

\[
\Delta \phi \simeq \frac{\pi}{2} \frac{R_S}{R}, \quad \frac{2}{R} \equiv \frac{1}{R_-} + \frac{1}{R_+}.
\]

In the last relation the orbit parameter $R$ value in terms of aphelion ($R_-$) and aphelion ($R_+$) of the orbit enters. It is important to note the $1/R$-dependence showing the largest effect for the innermost planet, Mercury. Comparing to the GR result we note that Weinberg\(^9\) finds using post-Newtonian characterization of GR a result that is 6 times greater

\[
\Delta \phi_{\text{GR}} = 3\pi \frac{R_S}{R}.
\]

Weinberg’s result agrees with Einstein’s, presented in Ref.\(^8\) and repeated at the end of the pivotal GR manuscript\(^{10}\).

A disagreement by factor 6 between SR-Coulomb and GR clearly shows that in specific features SR can be a poor model for effect of GR. However, the $1/R$-behavior is specific to relativity, both SR and GR. This $1/R$-behavior is unlike any other effect e.g. due to planetary perturbations or solar deformation. Therefore the observation of the scaling with $1/R$ is evidence for relativity, SR or GR, and the larger apsidal precession effect is a clear confirmation of GR.

---


We note that we obtained here the effect of the apsidal precession per orbit while experimental data is usually presented per 100 revolutions of the Earth (per century); for Mercury there are 415 revolutions per century and thus the aphelion precession effect is amplified by a factor 4.15. Similarly the effect on Venus is amplified by a factor 1.49. The $1/R$-dependence predicts for Venus and Earth compared to Mercury a ratio

$$\frac{\Delta \phi_{\text{Merc}}}{\Delta \phi_{\text{Venus}}} = 4.15 \frac{R_{\text{Venus}}}{R_{\text{Merc}}} = 5.43, \quad \frac{\Delta \phi_{\text{Merc}}}{\Delta \phi_{\oplus}} = 4.15 \frac{R_{\oplus}}{R_{\text{Merc}}} = 11.2. $$

Here we used $R_{\text{Mercury}} = 5.55 \times 10^7 \text{km}$, $R_{\text{Venus}} = 1.08 \times 10^8 \text{km}$, and $R_{\oplus} = 1.50 \times 10^8 \text{km}$. Using as base the GR effect on Mercury, the above scaling is observed and it confirms the relativity origin of the unexplained part of the planetary apsidal precession.

---

### 24 Electrons Riding a Plane Wave

#### 24.1 Fields and potentials for a plane wave

The study of dynamics of electron motion in the electromagnetic field of a plane wave lay foundations for principles that lead to the understanding of relativistic particle dynamics in presence of ultra-intense laser pulses. We address this problem now using non-covariant approach exploiting the conservation laws we have established. The first step is to gain familiarity with the electromagnetic plane waves describing the propagation of light, a solution of Maxwell’s equations without sources here considered as a prescribed ‘external’ field. We show in exercise XI–3 on page 406 that such waves can be generated.

For a light plane wave the $\vec{E}, \vec{B}$-fields are time-dependent and transverse to direction of propagation. We choose in transverse gauge

$$\vec{A} = \Re \left[ \vec{A}_{\|0} e^{i\chi} \right], \quad \chi = \vec{k} \cdot \vec{r} - \omega t, \quad V = 0. \quad (24.1)$$

where $\vec{A}_{\|0}$ is a constant amplitude. In order to satisfy the transverse gauge condition we must have

$$0 = \vec{\nabla} \cdot \vec{A} = \Re \left[ i \vec{k} \cdot \vec{A}_{\|0} e^{i\chi} \right] \Rightarrow \vec{k} \cdot e \vec{A}_{\|0} = 0 \quad (24.2)$$

Thus we find that the ‘polarization vector’

$$\bar{e}_{\vec{A}} \equiv \frac{\vec{A}_{\|0}}{A_{\|0}}, \quad A_{\|0} = \sqrt{A_{\|0}^* A_{\|0}} \quad (24.3)$$
Electrons Riding a Plane Wave

(where $A^*$ is complex conjugate of $A$) must be always orthogonal to the direction of propagation $\vec{k}$ of the wave. The reader should distinguish the polarization vector $\vec{e}_A$ from elementary charge $e$. For a plane light-wave propagating along the light front with the light speed $c = |d\vec{r}|/dt = r/t$, we have according to Eq. (24.1)

$$k = \frac{\omega}{c} = \frac{2\pi}{\lambda},$$

and we note the last relation to the wavelength $\lambda$.

Differentiating the 3-vector $\vec{A}$ we obtain

$$\vec{E} = -\frac{\partial \vec{A}}{\partial t} = \omega \Im \left[ \vec{A}_0 e^{i\chi} \right],$$

$$\vec{B} = \vec{\nabla} \times \vec{A} = \vec{k} \times \Im \left[ \vec{A}_0 e^{i\chi} \right].$$

In view of Eq. (24.4) we have $|\vec{E}| = |\vec{B}|$, and $\vec{B} \perp \vec{E} \perp \vec{k}$, where the last orthogonality follows from Eq. (24.2).

If the polarization vector $\vec{e}_A$, Eq. (24.3), is real, it does not interfere with the the phase $\chi$ of the wave. In this case the direction of $\vec{E} \perp \vec{B}$ remains fixed in space, and we have a ‘linearly polarized’ (LP) light plane wave. We can choose for $\vec{e}_A$ any orientation transverse to the direction of wave propagation $\vec{k} \parallel \vec{e}_3$

$$\vec{e}_A = \cos \delta \, \vec{e}_1 + \sin \delta \, \vec{e}_2, \quad \vec{e}_A^* \cdot \vec{e}_A = 1, \quad \text{LP}.$$  

(24.6)

Here $\delta$ is a (fixed) phase angle.

To obtain circularly polarized (CP) light, we choose a complex (and constant) polarization vector

$$\vec{e}_A = \frac{\vec{e}_1 \pm i\vec{e}_2}{\sqrt{2}}, \quad \vec{e}_A^* \cdot \vec{e}_A = 1, \quad \text{CP},$$

(24.7)

where $\vec{k} \perp \vec{e}_1 \perp \vec{e}_2$ assures Eq. (24.2) and two signs represent two possible circular polarizations. From Eq. (24.5) we obtain the unit vectors and magnitudes of the fields $\vec{E} \perp \vec{E}^*$

$$\vec{e}_E = \frac{\vec{e}_1 \sin \chi \pm i \vec{e}_2 \cos \chi}{\sqrt{2}}, \quad E = \omega A_0, \quad \text{CP},$$

$$\vec{e}_B = \vec{e}_k \times \frac{\vec{e}_1 \sin \chi \pm i \vec{e}_2 \cos \chi}{\sqrt{2}}, \quad B = kA_0, \quad \text{CP}.$$  

(24.8)

Thus observing $\vec{e}_E$, $\vec{e}_B$ at the same position as a function of time; that is choosing $\chi = -\omega t$, we see that these unit vectors rotate along the unit circle. Note that $e\vec{A}$ also rotates

$$\vec{A} = A_0 \left( \cos \chi \vec{e}_1 \mp \sin \chi \vec{e}_2 \right).$$

(24.9)
PART IX: MOTION OF CHARGED PARTICLES

Here $\chi$ is the time- and space-dependent plane wave phase. At a fixed point in space $z = 0$ we have $\chi = -\omega t$ and thus

$$\vec{A}(z = 0, t) = A_{|0}\left(\cos \omega t \vec{e}_1 \pm \sin \omega t \vec{e}_2\right), \quad (\vec{A})^2 = (A_{|0})^2, \quad \text{CP}, \quad (24.10)$$

rotates in a mathematically positive sense for the $(+)$ sign. For the right-handed coordinate system the plane wave moves out of the plane of this book towards you.

We will be interested in evaluating the intensity of the plane wave which by definition is the magnitude of the Poynting vector Eq. (28.3) which can be written in several equivalent ways in vacuum

$$\vec{S} = c^2 \varepsilon_0 \vec{E} \times \vec{B} \equiv c^2 \vec{D} \times \vec{B} \equiv \frac{1}{\mu_0} \vec{E} \times \vec{B} \equiv \vec{E} \times \vec{H}, \quad (24.11)$$

we will look at the meaning of $\vec{S}$ in section 28.1 on page 425, where the reader will find some additional explanations addressing the meaning of each variant seen in Eq. (24.11). Here, we focus on evaluation of $\vec{S}$ for the plane wave. The cross product $\vec{E} \times \vec{B}$ points according to Eq. (24.5) in the direction of the plane wave vector $\vec{k}$. We obtain

$$\vec{S} = \frac{\vec{k} \omega}{4\pi} \frac{4\pi \varepsilon_0}{e^2} \left(\Im \left[ e \vec{A}_{|0} \, e^{i\chi}\right]\right)^2. \quad (24.12)$$

Note the appearance of $c$ as prefactor of $\vec{A}_{|0}$ in Eq. (24.12). In Gauss-units this factor $c$ is absent as it has the same origin as the factor $c$ in $\mathcal{B}$ in Eq. (23.2); in addition the factor $4\pi \varepsilon_0/e^2$ is $1/e^2$ in Gauss-units. This is consistent with the last rule in Eq. (23.2).

We further have

$$\Im \left[ e \vec{A}_{|0} \, e^{i\chi}\right] = \begin{cases} eA_{|0} \vec{e}_1 \sin \chi & \text{LP}, \\ \frac{1}{2}eA_{|0}\left(\vec{e}_1 \sin \chi \pm \vec{e}_2 \cos \chi\right) & \text{CP}, \end{cases} \quad (24.13)$$

and hence

$$\left(\Im \left[ e \vec{A}_{|0} \, e^{i\chi}\right]\right)^2 = \frac{eA_{|0}^2}{2} \begin{cases} 2 \sin^2 \chi & \text{LP}, \\ 1 & \text{CP}, \end{cases} \quad (24.14)$$

---

11 John Henry Poynting (1852-1914) a collaborator of Maxwell and an eminonct physics educator.

In consideration of the fact that we have for time and/or space average
\[
2\langle \sin^2 \chi \rangle = 1 \tag{24.15}
\]
the final result for the averaged Poynting vector of a plane wave is
\[
\langle \vec{S} \rangle = \frac{\hbar \omega}{8\pi} (ceA_0)^2 \frac{4\pi \epsilon_0}{e^2} = \frac{\hbar}{2\alpha^2} (ceA_0)^2 \frac{4\pi \epsilon_0}{e^2} . \tag{24.16}
\]
The last equality follows in view of the Eq. (24.4). Since our unit vectors were normalized the same way, see Eq. (24.6) and Eq. (24.7), the result Eq. (24.16) applies equally to the two polarization cases we considered.

Exercise IX–7: Power transfer: check potential form of Lorentz-force

Show using Eq. (23.16) for the Lorentz-force in terms of potential, that the power (energy per time) a plane wave delivers to a particle is the same as the one found from the usual format of Lorentz-force written in terms of fields, Eq. (23.1).

Solution

We recall that the power that an electromagnetic field imparts on a particle of charge \( q \) is, according to Eq. (22.33),
\[
1 \quad P_q \equiv \frac{d\vec{r} \cdot \vec{F}_L}{dt} = \vec{v} \cdot \vec{F}_L = e\vec{v} \cdot \vec{E} .
\]
Using Eq. (24.5) we obtain
\[
2 \quad P_q = \frac{eA_0\omega}{c} \Im \left[ \vec{v} \cdot \vec{e}Ae^{i\chi} \right] .
\]

Now considering the result in terms of potentials, we use Eq. (23.16), and set \( V = 0 \) and use Eq. (24.1) to obtain
\[
3 \quad P_q = \frac{e}{c} \Re \left[ (\vec{v} \cdot \vec{\nabla})(\vec{v} \cdot \vec{A}) - \vec{v} \cdot \frac{d\vec{A}}{dt} \right] .
\]
The last term contains two contributions
\[
4 \quad \frac{d\vec{A}}{dt} = \frac{\partial \vec{A}}{\partial t} + \left( \frac{d\vec{r}}{dt} \cdot \vec{\nabla} \right) \vec{A} .
\]
Inserting Eq. (4) into Eq. (3) we see that among the resulting three terms two cancel, since we can pull \( \vec{v} \) across \( \vec{\nabla} \). The remaining term
\[
5 \quad P_q = \frac{e}{c} \Re \left[ -\vec{v} \cdot \frac{\partial \vec{A}}{\partial t} \right] = \frac{e}{c} \Re \left[ i\omega A_0 \vec{v} \cdot \vec{e}Ae^{i\chi} \right] ,
\]
which is the same as Eq. [2].

We see that the energy transfer between a particle and the field cannot be avoided since only if \( \vec{v} \cdot \vec{e}_A = 0 \) we find \( P_q \to 0 \), there is no energy transfer from a plane wave to a charged particle. We note that the sign of power transfer is not fixed. Therefore the energy transfer can be small or even vanish as an average over time, in which case a particle is shaken but not moved. Of some interest are the (strong) field conditions in which \( P_q \) is large and unidirectional, accelerating particles. We will study the dynamics of particles riding a wave in order to understand this very interesting limit in section 24.2 and section 24.3.

End IX–7: Power transfer: check potential form of Lorentz-force

### 24.2 Role of conservation laws

A particle riding the field of a plane wave is subject to conservation laws arising from several symmetries of the problem. These allow us to constrain the particle dynamics considerably, creating an opportunity to verify that particle motion is subject to the effect of collective action of the coherent field rather than individual particle scattering processes.

For plane waves we have \( V = 0 \) and in transverse gauge the Lagrangian is

\[
L(\vec{v}, z - ct) = -mc^2 \sqrt{1 - v^2/c^2} + \vec{v} \cdot \vec{A}(z - ct),
\]

where for wave propagating along the \( z \)-axis we made explicitly visible that the dependence on the ‘light-cone’, coordinate \( z - ct \), see section [24.1].

\[
e\vec{A} = mca_0 \Re[e\vec{e}_A e^{i\chi}], \quad \chi = k(z - ct),
\]

where dimensionless variable \( a_0 \) characterizes the strength of the 3-vector potential, see exercise IX–8 below.

Two conservation laws emerge considering the explicit format Eq. (24.18):

I The Lagrangian is independent of \( \vec{r}_\perp \) and thus we have

\[
0 = \frac{\partial L}{\partial \vec{v}_\perp} \rightarrow \frac{d}{dt} \frac{\partial L}{\partial \vec{v}_\perp} = 0, \quad \rightarrow \quad mc\frac{\vec{v}_\perp/c}{\sqrt{1 - v^2/c^2}} + e\vec{A}_\perp = \vec{C},
\]

(24.19)

II For the change of energy of a particle we have according to Eq. (23.40) and using Eq. (23.19)

\[
\frac{dE}{dt} = -\frac{\partial L}{\partial t} = c\frac{\partial L}{\partial z} = c\frac{d}{dt} \frac{\partial L}{\partial v_z}.
\]

(24.20)
Since in transverse gauge $A_z = 0$, the only contribution to the last term is from the kinetic energy term $T_r$. We therefore can obtain the first integral

$$E - mc^2 = mc^2 \sqrt{1 - v^2/c^2},$$  \hspace{1cm} (24.21)$$

where the integration constant was chosen for a particle initially at rest, picking up a ride on the plane wave.

We use the two conservation laws to study constraints on particle dynamics. We use Eq. (24.21) to write

$$(E - mc^2)^2 = m^2 c^4 \frac{v_z^2}{c^2 + v^2/c^2 - v^2/c^2}{1 - v^2/c^2} = m^2 c^4 \frac{v^2/c^2 - 1 + 1}{1 - v^2/c^2} - m^2 c^4 \frac{v^2/c^2}{1 - v^2/c^2},$$  \hspace{1cm} (24.22)$$

and substitute $E^2 = m^2 c^4/(1 - v^2/c^2)$

$$E^2 - 2Emc^2 + m^2 c^4 = -m^2 c^4 + E^2 - m^2 c^4 \frac{v^2/c^2}{1 - v^2/c^2}.$$  \hspace{1cm} (24.23)$$

Solving for $E$ we obtain

$$E = mc^2 + mc^2 \frac{\vec{v}_\perp^2/c^2}{2} \frac{1}{1 - v^2/c^2} = mc^2 \frac{(q\vec{A}_\perp - \vec{C})^2}{2mc},$$  \hspace{1cm} (24.24)$$

where the last equality follows by Eq. (24.19). According to last equality in Eq. (24.24) the magnitude of the Lorentz factor for the electron riding the wave is

$$\frac{E}{mc^2} = \gamma \to 1 + \frac{1}{2} a_0^2.$$  \hspace{1cm} (24.25)$$

We can relate the longitudinal-to-transverse velocity combining Eq. (24.21) with Eq. (24.24)

$$\frac{v_z/c}{\sqrt{1 - v^2/c^2}} = \frac{1}{2} \frac{\vec{v}_\perp^2/c^2}{1 - v^2/c^2}.$$  \hspace{1cm} (24.26)$$

Condition Eq. (24.26) is valid for an electron initially at rest capturing a ride on the plane wave. We now show that this condition amounts to a relation between the angle $\theta$ of particle motion with reference to the wave axis and particle total energy. We have

$$v_\perp = v \sin \theta, \quad v_z = v \cos \theta,$$  \hspace{1cm} (24.27)$$

and thus

$$\tan^2 \theta = \left( \frac{v_\perp}{v_z} \right)^2 = 2 \frac{1 - v^2/c^2}{v_z^2} \frac{(v_z/c)c^2}{\sqrt{1 - v^2/c^2}} = 2 \frac{\sqrt{1 - v^2/c^2}}{v_z/c}.$$  \hspace{1cm} (24.28)$$
Using the conservation law Eq. (24.21) we obtain
\[
\tan^2 \theta = \frac{2mc^2}{E - mc^2} = \frac{2}{\gamma - 1}.
\] (24.29)

This qualitative statement was tested experimentally, as shown in figure 24-1. The results of measuring the energy of the accelerated electron and the angle \(\theta\) of electron observation against the direction defined by the propagation of the laser pulse are shown. We recall that in a realistic situation one obtains a relativistic strength ‘plane wave’ by focusing a relatively homogeneous laser pulse and capturing the particle in the focus. Thus this wave defocuses leaving behind the surfing electron which can be later observed as shown in figure 24-1.

Exercise IX–8: Natural units for plane waves

Connect the plane wave amplitude \(A_0\) and related fields \(\mathcal{E}, \mathcal{B}\) chosen in natural units to unit values in standard units.

Solution

The potential \(\vec{A}\) multiplied with a unit charge \(e\) and speed of light \(c\) we see in Eq. (24.12) and in reduced form in Eq. (24.16) has in the SI-unit system the unit of energy. In the context of a particle that rides the wave shown in section 24.1 it would be appropriate

to use as reference the rest energy of that particle to define the unit wave height. Thus
we scale the magnitude of the plane wave amplitude $A_0$ and introduce dimensionless
potential amplitude $a_0$

$$ceA_0 \equiv a_0 mc^2.$$  

In the following we chose to consider an electron as the reference particle, and the
appropriate ratio of masses allows the scaling of these expressions for other elementary
particles. Thus $a_0 = 1$ corresponds to the value

$$cA_0 \equiv cA_0(a_0 = 1) \frac{mc^2}{e} = 0.5110 \times 10^6 \text{ V},$$

where we canceled the $e$ from the MeV energy unit and made the ‘mega’-visible. Let us
look at a light wave with a quantum of light energy

$$\hbar \omega = 1 \text{ eV} = \hbar c \lambda,$$

The corresponding electrical field strength according to Eq. (24.5) is

$$\mathcal{E}(a_0 = 1, \hbar \omega = 1 \text{ eV}) = \frac{\hbar \omega}{\hbar c} cA_0|_1 = \frac{1Vmc^2}{\hbar c \lambda C} = \frac{1V}{\lambda C}, \quad \lambda C = \frac{\lambda_C}{2\pi} = \frac{\hbar}{mc},$$

where the reduced Compton wavelength appears, Eq. (22.3). Written in SI-units the
electrical field we are now considering is

$$\mathcal{E}(a_0 = 1, \hbar \omega = 1 \text{ eV}) = \frac{1V}{386.16 \text{ fm}} = 2.590 \times 10^{12} \text{ V/m}.$$

Even though this appears to be a ‘strong’ field, it is not on elementary scale. The
natural unit-one value of electrical field strength requires a natural frequency $\omega_e$, that
is choosing $\hbar \omega_e = mc^2$,

$$\mathcal{E}_e \equiv \frac{mc^2}{\hbar c} cA_0|_1$$

corresponding to the Schwinger critical field $E_c$ we introduce later, see Eq. (29.3) on
page 444. In order to achieve this critical field using visible light with the typical energy
of $\hbar \omega = 1 \text{ eV}$ we need to compensate the low photon energy by a large value of laser
wave amplitude, $a_0 = mc^2/\hbar \omega = 511,000$.

Turning now to the value of magnetic field: We recollect that to establish the
equivalence of the relation between SI unit of electrical field $[\mathcal{E}] = \text{V/m}$ (Volt/meter)
and SI unit of magnetic field, $[B] = \text{T}$ (Tesla), we need to divide out the magnitude of
light velocity $c = 2.998 \times 10^8 \text{ m/s}$ from the SI value of electrical field, see Insight on page 320. Using Eq. (5) we obtain

$$\mathcal{B}(a_0 = 1, \hbar \omega = 1 \text{ eV}) = \frac{2.590 \times 10^{12}}{2.998 \times 10^8} \text{T} = 0.8639 \times 10^4 \text{T}.$$
As before with the electrical field, the true unit-1 magnetic field is 511,000 times bigger, as seen in Eq. (29.4) on page 444.

End IX–8: Natural units for plane waves

Exercise IX–9: Natural units and laser community units

Connect the natural plane wave unit $a_0$ to laser units: power units GW and 1000 GW = TW (tera W) and W/cm$^2$.

Solution

The intensity $I$ of a laser is by definition measured in terms of the magnitude of the (time averaged for linear polarization) Poynting vector, see Eq. (24.16) and Eq. (1) in exercise IX–8

$$I \equiv |\langle \vec{S} \rangle| = \left| c^2 \epsilon_0 \langle \vec{E} \times \vec{B} \rangle \right| = \frac{\hbar |\omega| (c \epsilon_0 A_0)}{8\pi} 4\pi \epsilon_0 \frac{a_0^2 c}{\lambda^2} \langle mc^2 \rangle^2 \frac{4\pi \epsilon_0}{e^2}.$$

We note that intensity scales with $a_0^2$. The achievable power of a laser is obtained multiplying the intensity by the minimal geometric focal surface

$$P_{\text{Laser}} \equiv \pi \lambda^2 I = \frac{\pi^2}{2} a_0^2 8.7105 \times 10^9 \text{W} = a_0^2 43.0 \text{ GW}.$$

The SI-unit coefficient we see in Eq. (2) arises using mass-energy equivalent $mc^2$ of the electron, charge of the electron $e$, and classical electron radius $r_e$, Eq. (23.4)

$$\frac{mc^2}{e} = 0.5110 \times 10^6 \text{V}, \quad e = 1.602177 \times 10^{-19} \text{C}, \quad mc^2 \frac{4\pi \epsilon_0}{e^2} = \frac{1}{r_e}.$$

We obtain for the prefactor in Eq. (1)

$$c(mc^2)^2 \frac{4\pi \epsilon_0}{e^2} = \left( 2.998 \times 10^8 \text{s} \right) \left( 1.6022 \times 10^{-19} \text{C} \right) \left( 0.5110 \times 10^6 \text{V} \right) \frac{10^{15}}{2.8179 \text{m}}.$$

Canceling the unit m(eter) and using C/s=A, AV=W we obtain

$$5 \ c(mc^2)^2 \frac{4\pi \epsilon_0}{e^2} = \frac{2.998 \times 1.6022 \times 5.110}{2.8179} 10^9 \text{W} = 8.7105 \text{GW},$$

the numerical result stated in Eq. (2). Thus a $P_{\text{Laser}} = \text{TW} (10^{12} \text{W})$ power implies a relativistic regime with $a_0 = 4.82$. Petawatt (PW, 10^{15} W) scale laser can be focused to reach $a_0 > 150$. At the time of writing exawatt (EW, 10^{18} W) is on the visible horizon of technology, permitting $a_0 \rightarrow 5000$. We further note that when the energy in the pulse is a Joule and the pulse length $\Delta t = 10^{-12} \text{s}$, then when focused to geometric limit the power of TW is achieved as a time and space average.
Another popular laser power unit arises as follows: one multiplies Eq. 1 by the ratio of a typical wavelength; normally this is a micron (squared) to cm\(^2\), that amounts to the introduction of an additional factor \(10^8\) in the value of \(I\). In consequence one sees in literature statements about laser intensity \(I\) in units of W/cm\(^2\). These are thus understood according to the following relation

\[
6 \quad I = a_0^2 \times 1.368 \times 10^{18} \frac{\text{W}}{\text{cm}^2} \left(\frac{[\mu\text{m}]}{\lambda}\right)^2.
\]

We recognize that a laser above peak intensity \(10^{18} \text{ W/cm}^2\) can be focused to achieve \(a_0 > 1\). The relation Eq.6 can be inverted to state the value of \(a_0\) in terms of physical qualities of the laser pulse

\[
7 \quad a_0 = 0.855 \times 10^{-9} \lambda [\mu\text{m}] I^{1/2} [\text{W/cm}^2]
\]

where \(I\) is laser peak intensity, \(\lambda\) is the wavelength.

From the point of view of relativistic physics the quantity that matters foremost in the study of particle dynamical problems is the value of \(a_0\). One refers to lasers that achieve \(a_0 > 1\) as operating in the relativistic regime\(^{14}\). This name reflects the profound change in particle dynamics when \(a_0 > 1\). As we have seen in exercise 24.2 on page 349 and exercise 24.3 in this condition a particle catching a laser wave can be nearly linearly accelerated and achieves relativistic dynamics.

It is of interest to the reader to recognize that the technology that allows relativistic optics with \(a_0 > 0\) relies on non-monochromatic light wave: superposition of nearly monochromatic light waves allows formation of a finite length in time wave-trains of light. Using optical devices to change the distribution of wave-lengths one can define a long wave train for purpose of amplification, and later compress the wave train for the purpose of achieving high intensity. This method introduced for this purpose to optics by Strickland and Mourou\(^{15}\) is called chirped pulse amplification (CPA).

---

**24.3 Surfing the plane wave**

We now solve the Lorentz-force equations of motion for an electron in the field of an electromagnetic plane wave. We use the results of section 24.2 to set-up the dynamical equations as follows: Eq. (24.19) can be written in the form

\[
\gamma \vec{v}_\perp / c = \gamma \frac{d \vec{r}_\perp}{dt} = \frac{d \vec{r}_\perp}{d\tau} = \vec{a}_0 + \vec{C}' , \tag{24.30}
\]


where (possible confusion with acceleration should be avoided)

\[ \vec{a}_0 \equiv -\frac{eA}{mc}, \]  

(24.31)

and \( \vec{C}' = \vec{C}/(mc^2) \) is an integration constant. Equation (24.21) becomes, using the last form of Eq. (24.24) and Eq. (24.26),

\[ \gamma v_z/c = \gamma \frac{dz}{dct} = \frac{dz}{dc\tau} = \frac{1}{2} \left( \vec{a}_0 + \vec{C}' \right)^2. \]  

(24.32)

As noted in the opening of exercise 24.2, the argument of \( \vec{a}_0 \) is \( ct - z \). This variable satisfies

\[ \frac{d(ct - z(\tau))}{d\tau} = \gamma - \frac{dz}{d\tau} = \gamma - \frac{1}{2} \left( \vec{a}_0 + \vec{C}' \right)^2, \]  

(24.33)

where the last relation follows using Eq. (24.32). For \( \gamma \) we obtain using the last form in Eq. (24.24)

\[ \gamma = 1 + \frac{1}{2} \left( \vec{a}_0 + \vec{C}' \right)^2, \]  

(24.34)

which shows that

\[ \frac{d(ct - z)}{d\tau} = 1, \Rightarrow \]  

\[ ct - z = c\tau, \]  

(24.35)

where we chose the coordinate system origin coinciding with the location of the particle for \( \tau = 0 \). Eq. (24.35) is the key result allowing direct integration of the equations of motion since the exponential phase of the plane wave is according to Eq. (24.35) only a function of \( \tau \), the particle proper time.

We now can write all the components of the 4-velocity the particle. According to Eq. (24.34) we have

\[ u^0 = \frac{cdt}{d\tau} = c\gamma = c + \frac{c}{2} \left( \vec{a}_0 + \vec{C}' \right)^2, \]  

(24.36)

and the spatial components are according to Eq. (24.32), Eq. (24.30)

\[ u_\parallel = \frac{dz}{d\tau} = \gamma v_z = \frac{1}{2} \left( \vec{a}_0 + \vec{C}' \right)^2, \quad \vec{u}_\parallel = \frac{d\vec{x}_\parallel}{d\tau} = \gamma \vec{v}_\parallel = c\vec{a}_0 + c\vec{C}', \]  

(24.37)

and we recall Eq. (24.31). As expected, the 4-velocity satisfies

\[ u^2 = \left( \frac{cdt}{d\tau} \right)^2 - \left( \frac{dy}{d\tau} \right)^2 - \left( \frac{dx}{d\tau} \right)^2 - \left( \frac{dz}{d\tau} \right)^2 = c^2 \left( 1 + \frac{1}{2} \left( \vec{a}_0 + \vec{C}' \right)^2 \right)^2 - c^2 \left( \vec{a}_0 + \vec{C}' \right)^2 - \left( \frac{c}{2} \left( \vec{a}_0 + \vec{C}' \right)^2 \right)^2 = c^2. \]  

(24.38)
We now consider in turn the motion of a charged particle riding a linear (LP), and circular (CP), polarized plane waves.

**LP:** For a linear polarized plane wave with the polarization vector oriented in $y$-direction we obtain, using Eq. (24.30) and Eq. (24.18)

\[
\frac{1}{c} \frac{dy}{d\tau} = a_0 \cos \omega \tau + C', \quad \Rightarrow \quad \frac{dy}{d\tau} = \frac{c}{a_0} a_0 (\cos \omega \tau - 1) \approx \frac{c}{2} a_0 \omega^2 \tau^2 ,
\]

where $C' = -a_0$ is chosen for the case that the particle is initially at rest, and the start up of motion for $\tau \approx 0$ is shown explicitly. The particle transverse position

\[
y = \frac{c a_0}{\omega} (\sin \omega \tau - \omega \tau) ,
\]

drifts with time to a large and negative $y$.

The motion along the propagation direction follows from Eq. (24.32)

\[
\frac{1}{c} \frac{dz}{d\tau} = \frac{1}{2} \left( a_0 \cos \omega \tau + C' \right)^2 , \quad \Rightarrow \quad \frac{dz}{d\tau} = \frac{c a_0^2}{2} (\cos \omega \tau - 1)^2 \approx \frac{c a_0^2}{8} \omega^4 \tau^4 ,
\]

and

\[
z = \frac{c a_0^2}{2 \omega} \left( \frac{3 \omega \tau^2}{2} - 2 \sin \omega \tau + \frac{1}{4} \sin 2 \omega \tau \right) ,
\]

where integration constants are chosen so that the coordinate origin is with the particle at $\tau = 0$. Therefore the solutions we present for $dy/d\tau$ Eq. (24.39) and $dz/d\tau$ Eq. (24.41) are constrained by Eq. (24.26)

\[
2c \frac{dz}{d\tau} = \left( \frac{dy}{d\tau} \right)^2 \Leftrightarrow \frac{2c v_z}{\sqrt{1 - (v_z^2 + v_y^2)/c^2}} = \left( \frac{v_y}{\sqrt{1 - (v_z^2 + v_y^2)/c^2}} \right)^2 .
\]

We see that the particle moves forward but initially this is an exceedingly small effect $\propto \tau^5$. However, since the motion in the direction of wave propagation is proportional to $a_0^2$, for relativistic plane waves with $a_0 > 1$ the motion along with the wave wins over the transverse motion, and particles are pushed forward, but retain an angular asymmetry moving off the original axis in opposite direction to the polarization direction, we do not have azimuthal symmetry. Conversely, for $a_0 < 1$ the transverse motion dominates. This clarifies the importance of achieving $a_0 > 1$ for the purpose of direct acceleration of electrons in the plane wave field.

According to Eq. (24.34) we also have

\[
\frac{dt}{d\tau} = \gamma = \frac{E}{mc^2} = 1 + \frac{a_0^2}{2} (\cos \omega \tau - 1)^2 ,
\]
which shows that the energy of a particle riding a linearly polarized plane wave is bounded by

\[ E < mc^2 (1 + 2a_0^2) . \]  

\[(24.45)\]

CP: For the circular polarized plane wave we return to Eq. (24.10) in section 24.1, noting that now \( \vec{A} \) is simply a function of \( \tau \)

\[ \frac{1}{c} \frac{dx}{d\tau} = a_0 \cos \omega \tau + C'_x = a_0(\cos \omega \tau - 1) , \Rightarrow \omega x = c a_0(\sin \omega \tau - \omega \tau) , \]

\[ \frac{1}{c} \frac{dy}{d\tau} = \pm a_0 \sin \omega \tau + C'_y = \pm a_0 \sin \omega \tau , \Rightarrow \omega y = \mp c a_0 \cos \omega \tau , \]

\[ \frac{1}{c} \frac{dz}{d\tau} = \frac{1}{2} (a_0^2 + C''_x + 2a_0 \cos \omega \tau) = a_0^2 (1 - \cos \omega \tau) , \Rightarrow \omega z = a_0^2 (c_\omega \tau - \sin \omega \tau) , \]  

\[(24.46)\]

where \( C'_x, C'_y \) are chosen such that the particle is initially at rest. According to Eq. (24.34) we also have

\[ \frac{dt}{d\tau} = \gamma = \frac{E}{mc^2} = 1 + a_0^2(1 - \cos \omega \tau) , \quad E < mc^2 (1 + 2a_0^2) . \]  

\[(24.47)\]

We note that the limit on kinetic energy in this case is identical to what we found with linear polarization. The motion described by Eq. (24.46) and Eq. (24.47) is checked by forming the sum of squares of all components

\[ u^2 = \left( \frac{c}{d\tau} \frac{dt}{d\tau} \right)^2 - \left( \frac{dx}{d\tau} \right)^2 - \left( \frac{dy}{d\tau} \right)^2 - \left( \frac{dz}{d\tau} \right)^2 = c^2 . \]  

\[(24.48)\]

We now consider an interesting special case where the velocity along with the wave direction is a constant in time and in transverse direction there is only a circular motion. Such specific CP solution emerges setting \( C'_y = 0 \) that is,

\[ \frac{1}{c} \frac{dx}{d\tau} = a_0 \cos \omega \tau , \Rightarrow \omega x = c a_0 \sin \omega \tau , \]

\[ \frac{1}{c} \frac{dy}{d\tau} = \pm a_0 \sin \omega \tau , \Rightarrow \omega y = \mp c a_0 \cos \omega \tau , \]

\[ \frac{1}{c} \frac{dz}{d\tau} = \frac{a_0^2}{2} , \Rightarrow z = \frac{a_0^2}{2} c \tau . \]  

\[(24.49)\]

We also have

\[ \frac{dt}{d\tau} = \gamma = \frac{E}{mc^2} = 1 + \frac{a_0^2}{2}(\cos^2 \omega \tau + \sin^2 \omega \tau) = 1 + \frac{a_0^2}{2} , \]  

\[(24.50)\]

which is constant, and actually assumes the value that is the average in \( \tau \) of prior results. In consideration of Eq. (24.44) which shows a constant value of \( \gamma \), we have

\[ \tau = \frac{t}{\gamma} . \]  

\[(24.51)\]
Using Eq. (24.51) in Eq. (24.49) allows us to describe the motion in laboratory variables alone. The position and longitudinal velocity of the particle as a function of laboratory time thus are

\[ z = \frac{a_0^2/2}{1 + a_0^2/2} \cdot c t, \Rightarrow v_z \equiv \frac{dz}{dt} = c \frac{a_0^2/2}{1 + a_0^2/2} < c. \]  

(24.52)

Since the particle speed along the z-axis is ultra-relativistic for values of \( a_0 \rightarrow 1000 \) available in foreseeable future, we evaluate the particle rapidity, Eq. (17.19)

\[ y = \frac{1}{2} \ln(1 + a_0^2). \]  

(24.53)

According to Eq. (24.49) and using also Eq. (24.50)

\[ \left( \frac{dx}{dt} \right)^2 + \left( \frac{dy}{dt} \right)^2 = c^2 \left( \frac{a_0}{1 + a_0^2/2} \right)^2, \quad \rho \equiv \sqrt{x^2 + y^2} = \lambda a_0, \]  

(24.54)

the motion is helical as shown before in figure 23-1 displaying the motion in a longitudinal magnetic field.

The radius of circular motion is now governed by the light wavelength, and thus a comparison with motion in a constant magnetic field finds its limit: according to Eq. 13 in exercise IX–1 if we setup using a magnetic field a rotation with \( \omega = \omega_c \) the orbital speed is

\[ v_0 = \rho \omega_c \rightarrow v_0^{eq} = \rho \omega = \lambda a_0 \frac{c}{\lambda} = c a_0. \]  

(24.55)

We see that for \( a_0 > 1 \), in the relativistic optics domain, we would need a speed faster than light. Thus the helical particle motion in the CP plane wave becomes too extreme to be compared to particle motion in a constant magnetic field of any strength. It is further of interest to consider the strength of the magnetic field that could simulate the frequency of circular motion

\[ \omega = \omega_c \rightarrow \frac{c}{\lambda} = \frac{eB^{eq}}{\gamma m}. \]  

(24.56)

We find using for \( \gamma \) the value for the particle motion in the laser field, Eq. (24.50)

\[ B^{eq} = \left( 1 + \frac{a_0^2}{2} \right) \frac{mc}{e\lambda} = \frac{m^2c^2}{e\hbar} \left( 1 + \frac{a_0^2}{2} \right) \frac{\hbar/mc}{\lambda} = B_{cr} \left( 1 + \frac{a_0^2}{2} \right) \frac{\lambda_C}{\lambda}. \]  

(24.57)

where we introduced the Compton wavelength \( \lambda_C \) and the critical field strength \( B_{cr} = 4.414 \times 10^9 \text{T} \), see Eq. (29.4). For optical wavelength \( \lambda = 1 \mu m \) we have \( \lambda_C/\lambda = 2.425 \times 10^{-6} \). This means that for \( a_0 = 1.1 \times 10^3 \) the laser plane wave induces particle motion features that are comparable to those generated by magnetic critical field.
To close this extensive discussion of relativistic charged particle motion in presence of relativistic strength \( a_0 > 1 \) light wave, we should remember the tacit requirements allowing the use of the above idealized solutions:

i) the growing with \( a_0 \) particle orbit means that the plane wave should not taper off in the transverse directions over a distance \( a_0 \lambda \), (where \( \lambda \) is the wavelength), from the focal point;

ii) in the longitudinal direction for the linear polarized wave the particle oscillates over distance \( a_0^2 \lambda/\gamma \) around its mean location and thus the laser pulse that imitates a plane wave needs to be an approximate plane wave over this pulse time-length. For circular polarization the same result follows since we need to make a full circle which requires \( \tau_0 = 2\pi/\omega \). Thus the pulse length should be \( \Delta z = a_0^2 \lambda \).

iii) due to the strong acceleration accompanying the extreme conditions we described, the inclusion of radiation friction phenomena is necessary, see section 29.5 for an introduction.

The first two remarks clash with laser pulse reality: to achieve a high value of \( a_0 \) one must both compress in time the pulse to as few as possible wavelengths, and simultaneously focuses towards geometric limit in the focal spot where the condition \( a_0 \gg 1 \) is achieved.

Clearly the solution we described cannot be pushed to these conditions. Moreover, we did not consider the process of how a particle encountering the pulse joins, or better, ‘catches’ or ‘jumps’, onto the wave. Therefore the study of motion in plane wave fields we present is but a rough orientation that must be considered with some caution when comparing to real life situations. The actual behavior of charged particles accelerated by intense laser pulses the connection with experimental results is in general obtained in terms of numerical simulation of charged particle dynamics riding realistic light pulses. Because of the transforming impact of laser driven particle acceleration this is a very active present day research domain.

---

**Exercise IX–10: Plane Wave Acceleration**

Obtain the magnitude of invariant acceleration of charged particles riding plane waves, both for linear and for circular polarization.

**Solution**

We consider now the magnitude of acceleration that charged particles described in exercise 24.3 on page 354 experience

\[
1 \quad b^\mu = \frac{du^\mu}{d\tau}
\]
In this problem $b$ stands for acceleration, as $a_0$ is the amplitude of the wave. For LP (linearly polarized) waves we use Eq. (24.39), Eq. (24.41), and Eq. (24.44)

$$\begin{aligned}
 b^0 &= c \frac{d^2 t}{d\tau^2} = -\frac{c \omega a_0^2}{2} \sin 2\omega \tau, \\
 b^x &= 0, \\
 b^y &= -c \omega a_0 \sin \omega \tau, \\
 b^z &= -\frac{c \omega a_0^2}{2} \sin 2\omega \tau, \\
 b_{LP}^2 &= (b^0)^2 - (b^x)^2 - (b^y)^2 - (b^z)^2 = -\left( c \omega a_0 \right)^2 \sin^2 \omega \tau,
\end{aligned}$$

we see that $b_{LP}^2 < 0$ is spacelike, and that between timelike and longitudinal spacelike components the $a_0^2$-term cancels exactly. Thus even though the value of $\gamma \propto a_0^2$, Eq. (24.44), the invariant acceleration achieved is $b \propto a_0$. We further note that the time-averaged value is

$$\frac{|b_{LP}|}{c \sqrt{2}} = \frac{c^2 a_0}{\lambda \sqrt{2}}.$$

Turning our attention to CP (circular polarization) we obtain the 4-acceleration differentiating the 4-velocity as presented in Eq. (24.49) and Eq. (24.50)

$$\begin{aligned}
 b^0 &= 0, \\
 b^x &= -\frac{c \omega a_0}{\sqrt{2}} \sin \omega \tau, \\
 b^y &= \pm \frac{c \omega a_0}{\sqrt{2}} \cos \omega \tau, \\
 b_{CP}^2 &= (b^0)^2 - (b^x)^2 - (b^y)^2 - (b^z)^2 = -\left( c \omega a_0 \right)^2 \cos^2 \omega \tau
\end{aligned}$$

which is the same as in Eq. (3) without need for time average; $b_{CP}^\mu$ is a spacelike vector. Note that the $a_0^2$ terms being constant cancel-out in differentiation.

For the case of idealized plane wave when forming the invariant quantity $b^2$ we see that the contributions in $a_0^2$ cancel. Thus to achieve laser pulse accelerated particle energies scaling with $a_0^2$, the pulse will need to last a long enough time $\tau \propto a_0$.

End IX–10: Plane Wave Acceleration
Part X

Covariant Force and Field
Introductory remarks to Part X

We begin with discussion of the covariant Lorentz-force in terms of the 4-potential vector. The word ‘vector’ implies that the 4-potential transforms under LT like e.g. the 4-momentum. The 4-Lorentz-force is by construction fully covariant, and we show that it satisfies the established constraints: it is a spacelike 4-vector and it is 4-orthogonal to the 4-velocity. The 0th component of the 4-Lorentz-force is recognized as describing the energy transfer from the EM-field to the particle motion.

We next seek a variational principle from which one can derive the Lorentz-force with focus on finding a covariant method. We present and discuss several approaches found in literature: a) the non-covariant relativistic generalization of the usual 3-dimensional action as discussed earlier in section 23.4 b) the covariant generalization based on particle proper time as the evolution parameter; c) a variational principle proposed by Landau-Lifshitz where the variation of the particle proper times is also introduced, making it a part of dynamics for the kinetic energy only; and d) extension of the case c) to be fully consistent, with both kinetic and potential term in the action treated in same fashion.

None of these four approaches is convincing as the speed of light constraint \( u^2 = c^2 \) is a result that follows from imposing a specific inertia component in the action of which the form is not unique. The light speed as maximum speed is therefore an imposed result rather than a naturally built-in part of particle dynamics. This situation has generated interest in Hamiltonian-like 4-dimensional alternative based on a mass-function which also helps explain how the mass of a bound system accounts for the binding effect.

The introduction of the covariant form for the electric and magnetic field completes the development of covariant EM theory. We show how the EM-fields can be obtained from the 4-vector potential. The six EM-field components are shown not to be components of a vector but elements of an antisymmetric \( 4 \times 4 \) matrix, which on account of the prescribed Lorentz-transformation condition is called a ‘tensor’. This allows us to obtain the form of Lorentz-transformed fields and to construct the two Lorentz-invariants of electromagnetic (EM) fields both in covariant and non-covariant notation.

The covariant format of the fields allows the presentation of the Lorentz-force in a covariant form that relies on the field tensor. An integral representation of the field-based covariant Lorentz-force allows us to obtain solutions for the 4-velocity in an iterative fashion in powers of the proper time \( \tau \) of the accelerated particle. We show in this approach how new Poynting-vector directional force appears.
25 Covariant Formulation of Force

25.1 Lorentz-force in terms of 4-potential

Our objective now is to recognize Eq. (23.16) as the three spatial components of a 4-vector. By multiplying Eq. (23.16) with the Lorentz-factor $\gamma$ we obtain the spatial components of the Lorentz four-force, compare Eq. (22.20)

$$\vec{K}_L = \gamma \vec{F}_L = -e \left( \vec{\nabla} \gamma c \frac{V}{c} - \sum_{i=1}^{3} \vec{\nabla} \gamma v^i A^i \right) + \gamma \frac{d\vec{A}}{dt}, \quad (25.1)$$

where particle velocity $\vec{v}$ is time-dependent and thus $\gamma(v)$ can be commuted with the spatial derivatives.

We now introduce the four-potential $A^\mu$

$$A^\mu \equiv \{ V/c, \vec{A} \} = \{ V/c, A^1, A^2, A^3 \}. \quad (25.2)$$

Note that the division by $c$ is necessary in SI units according to rule 2 in Eq. (23.2) to assure that the entire 4-vector has the same dimension. The rationale for proceeding as shown in Eq. (25.2) is:

i) to unite in the force $\vec{K}_L$ the two terms in round bracket, see Eq. (25.1); and

ii) to unite the two quantities we called potentials $V$ and $\vec{A}$ into one 4-vector.

The evidence that this approach is correct will grow as we proceed. Note that $V/c = A^0$ is the 0th component of a 4-vector. The key property of $A^\mu$ is that it transforms under LT akin to 4-momentum $p^\mu$.

We use the four-potential $A^\mu$, Eq. (25.2) combining the first two terms in Eq. (25.1) into a single quantity

$$\vec{\nabla} \gamma c \frac{V}{c} - \sum_{i=1}^{3} \vec{\nabla} \gamma v^i A^i = \vec{\nabla} u \cdot A. \quad (25.3)$$

Using $dt = \gamma d\tau$ and Eq. (25.3) in Eq. (25.1) we obtain

$$\vec{K}_L = e \left( (-\vec{\nabla}) u \cdot A - \frac{d\vec{A}}{d\tau} \right) = e \left( u_\nu(-\vec{\nabla}) A^\nu - e \frac{d\vec{A}}{d\tau} \right), \quad (25.4)$$

with the total derivative with respect to proper time in the last term. We present two forms, both are equivalent since the 4-vector of velocity commutes with spatial derivatives. However, as we will show only one of these admist a suitable 4-dimensional generalization.

We seek a 4-vector generalization of Eq. (25.4) – this requires a deeper look at how covariant and contravariant 4-vectors relate to differentiation. We recall that
in section [20] we introduced two types of 4-vectors, with index *up* contravariant and those with index *down* covariant. We now extend this to consider differentiation. The key feature is that a ‘covariant’ derivative $\partial_\mu$ involves a differentiation with respect to ‘contravariant’ $x^\mu$, and vice-versa. In the immediately following exercise X–2 examples and rules that govern the up-down indices in differentiation will be further considered. We have

$$\frac{\partial}{\partial x^\mu} \equiv \partial_\mu \equiv \left\{ \frac{1}{c} \frac{\partial}{\partial t}, \vec{\nabla} \right\}, \quad \frac{\partial}{\partial x_\mu} \equiv \partial^\mu \equiv \left\{ \frac{1}{c} \frac{\partial}{\partial t}, -\vec{\nabla} \right\}.$$  (25.5)

In presence of non-flat space geometry this distinction between covariant and contravariant differentiation and vectors is necessary, while in our present context we are practicing a method that simplifies our notation and guides us in search for 4-notation of 3-equations.

The generalization of the second form in Eq. (25.4) to a four-vector equation is

$$K^\mu_L = e \left( u_\nu \frac{\partial A^\nu}{\partial x_\mu} - \frac{dA^\mu}{d\tau} \right).$$  (25.6)

Note that to obtain on the right side of Eq. (25.6) in the first term the index $\mu$ up, we need to differentiate with index $\mu$ down. This hides an extra minus-sign seen on right in Eq. (25.5) needed to show that the spatial $\vec{K}_L$ part is reproduced. We consider this sign further in exercise X–2.

We chose the second form in Eq. (25.4) as only this form of the 4-Lorentz-force Eq. (25.6) satisfies the constraint Eq. (22.30) $u \cdot K_L = 0$, compare exercise VIII–17

$$u_\mu K^\mu_L = e \left( u_\nu u_\mu \frac{\partial A^\nu}{\partial x_\mu} - u_\mu \frac{dA^\mu}{d\tau} \right) = e \left( u_\nu \frac{dA^\nu}{d\tau} - u_\mu \frac{dA^\mu}{d\tau} \right) = 0. \quad (25.7)$$

where we used

$$u_\mu \frac{\partial A^\nu}{\partial x_\mu} = \frac{dx_\mu}{d\tau} \frac{\partial A^\nu}{\partial x_\mu} = \frac{dA^\nu}{d\tau}. \quad (25.8)$$

If instead we were to consider generalizing the first form in Eq. (25.4) we would find for the 4-force

$$K^\mu_{(1)} = e \left( \frac{\partial u \cdot A}{\partial x_\mu} - \frac{dA^\mu}{d\tau} \right). \quad (25.9)$$

and thus using Eq. (25.8)

$$u_\mu K^\mu_{(1)} = e \left( \frac{d u \cdot A}{d\tau} - u \cdot \frac{dA^\mu}{d\tau} \right) = e \frac{d u}{d\tau} \cdot A \neq 0. \quad (25.10)$$

As indicated, despite the gauge freedom, see section [23.2], it is for a given set of EM fields in general not possible to find $A^\mu$ such that Eq. (25.10) vanishes. Thus
the requirement that \( c^2 = \text{Const.} \), see Eq. (22.10), allows for the 4-dimensional Lorentz force only the form Eq. (25.6).

We next show that the force Eq. (25.6) is as required spacelike, \( K^2_L < 0 \), compare exercise VIII–17 on page 315. All we need to show is that the 0th component \( K^0_L \) vanishes in some frame of reference and this is the case for the rest-frame of the particle \( u(0) = \{c, \vec{0}\} \). We obtain

\[
K^0_L\big|_{(0)} = e \left( \frac{\partial A^0}{\partial t} - \frac{dA^0}{d\tau}\right)_{(0)} = 0.
\] (25.11)

This is so since \( d\tau = \gamma dt \rightarrow dt \) in the rest-frame

\[
\frac{d}{d\tau} A^0\big|_{(0)} = \frac{d}{dt} A^0\big|_{(0)} = \left( \frac{\partial}{\partial t} A^0 + \frac{d\vec{x}}{dt} \cdot \vec{\nabla} A^0\right)_{(0)} = \frac{\partial}{\partial t} A^0\big|_{(0)},
\] (25.12)

where \( d\vec{x}/dt = \vec{v} = 0 \) in rest frame. More generally we have for the new component \( K^0 \) in Eq. (25.6), reverting to \( A^0 \rightarrow V/c \)

\[
K^0 = e \left( \gamma \left( \frac{\partial V}{c \partial t} - \frac{\vec{v}}{c} \cdot \frac{\partial \vec{A}}{\partial t} \right) - \frac{\gamma dmc}{c dt} V \right).
\] (25.13)

Using Eq. (25.12) and the definition of electric field in terms of potentials, Eq. (23.6), we obtain

\[
K^0 = \gamma e \left( -\frac{\vec{v}}{c} \cdot \frac{\partial \vec{A}}{\partial t} - \frac{\vec{v}}{c} \cdot \vec{\nabla} V \right) = \gamma \vec{v} \cdot e\vec{E}/c.
\] (25.14)

Furthermore we have

\[
\frac{dmc}{d\tau} = \gamma K^0_L, \quad \rightarrow \frac{d(mc^2\gamma)}{dt} = e\vec{v} \cdot \vec{E}, \quad \rightarrow \quad d(mc^2\gamma) = e d\vec{x} \cdot \vec{E},
\] (25.15)

which agrees with the general result Eq. (22.33). This confirms that the added 0th component in the force Eq. (25.6) has the expected physical meaning, it describes the work done by the field on the particle.

---

**Exercise X–1: Check of Lorentz 4-force**

Derive the expression Eq. (25.14) for the 0th-component of the Lorentz-force using Eq. (25.7) \( u \cdot K = 0 \).

**Solution**

Eq. (25.7) implies that there are only three independent components. Writing out the Einstein sum convention in Eq. (25.7) explicitly we have

\[
1 \quad \gamma cK^0_L - \gamma \vec{v} \cdot \vec{K}_L = 0.
\]
Solving for $K^0$ we find

\[ K^0_L = \frac{1}{u_0} \bar{u} \cdot \vec{K}_L = \frac{1}{\gamma_c} \gamma \bar{v} \cdot \vec{K}_L = \gamma c \cdot \vec{F}_L. \]

We insert the expression in Eq. (23.1) for the Lorentz-force $\vec{K}_L$:

\[ K^0_L = \gamma \bar{v} \cdot \left( \vec{E} + \bar{v} \times \vec{B} \right) = \gamma \bar{v} \cdot \vec{E}. \]

in agreement with Eq. (25.14).

End X–1: Check of Lorentz 4-force

Exercise X–2: Up and down indices in differentiation

We saw in the derivation of the covariant form of Lorentz-force Eq. (25.6) that a differentiation with respect to quantity with index down produced another quantity with index up. We study this in more detail by example here.

Solution

We need to convince ourselves that covariant differentiation has the outcome

\[ \frac{\partial A^\nu}{\partial x_\mu} = B^{\mu\nu}, \]

noting that the index $\mu$ is up. The fact that the function we differentiate is a 4-vector is of no relevance to our argument and thus we consider by example the simpler case

\[ \frac{\partial f}{\partial x_\mu} = h^\mu. \]

Once again, note that the function $h^\mu$ has index up.

The simplest non-trivial example is (making the Einstein summation convention explicit here)

\[ f = x^2 = x_\nu x^\nu = g^{\nu\kappa}x_\nu x_\kappa, \]

which produces

\[ \frac{\partial x^2}{\partial x_\mu} = g^{\mu\kappa}x_\kappa + g^{\nu\mu}x_\nu = 2x^\mu. \]

The next example concerns the plane wave with phase

\[ \chi = \vec{k} \cdot \vec{r} - (\omega/c)t = -x_\nu k^\nu, \]
where we introduced the 4-vector
\[ k^\nu = \{ \omega/c, \vec{k} \} . \]

We consider the function \( f = e^{i\chi} \)
\[ \frac{\partial e^{i\chi}}{\partial x_\mu} = i \frac{\partial \chi}{\partial x_\mu} e^{i\chi} = i k^\mu e^{i\chi} . \]
which shows again that differentiating with index down produced index up result.

It is therefore common to write
\[ \frac{\partial f}{\partial x_\mu} \equiv \partial^\mu f , \]
and similarly
\[ \frac{\partial A^\nu}{\partial x_\mu} = \partial^\mu A^\nu . \]

End X–2: Up and down indices in differentiation

### 25.2 Covariant variation principle

To find the covariant form of the Lorentz-force we seek an extrema of a relativistic action \( L_L \) under variation in space-time of the particle path understood to be a function of an evolution parameter, typically particle proper time
\[ x^\mu \rightarrow x^\mu + \delta x^\mu , \quad \delta x^\mu|_{p_1,p_2} = 0 . \] (25.16)

In the action \( I \), Eq. (23.19), the action potential \( U \), Eq. (23.29) takes the form
\[ \int_{p_1}^{P_2} U \, dt = \int_{p_1}^{P_2} e \left( V - \vec{v} \cdot \vec{A} \right) \frac{dt}{d\tau} \, d\tau = \int_{p_1}^{P_2} e \left( u^0 V/c - \vec{u} \cdot \vec{A} \right) \, d\tau . \] (25.17)

We use the four-potential \( A^\mu \), Eq. (25.2) to combine the two terms in the potential \( U \) into one
\[ \int_{p_1}^{P_2} U^* \, d\tau \equiv \int_{p_1}^{P_2} U \, dt = e \int_{p_1}^{P_2} u_\mu A^\mu \, d\tau = e \int_{p_1}^{P_2} u \cdot A \, d\tau . \] (25.18)

Similarly, the relativistic kinetic energy term \( T_r \), Eq. (23.26), can be transcribed into an integral over the proper time
\[ \int_{p_1}^{P_2} L_T \, dt = \int_{p_1}^{P_2} mc^2 \sqrt{1 - v^2/c^2} \, dt = -\int_{p_1}^{P_2} mc^2 \sqrt{dt^2 - dx^2/c^2} = -\int_{p_1}^{P_2} mc^2 \, d\tau . \] (25.19)
We thus obtained manifestly covariant expressions for both the potential in new form $U^r$, Eq. (25.18) and the kinetic term Eq. (25.19).

Despite the simplicity of the transition to a new relativistic format there are unresolved issues. Therefore we now review different ways often used to develop the relativistic Lorentz-force from variation of the path. The three first cases A:, B:, C: are seen scattered in different references, and the fourth case D: which does not exactly reproduce the Lorentz-force, is justified by the approach C:. Rejection of D: amounts to elimination of the frequently used C: as well.

A: Relativistic generalization of nonrelativistic action principle
The spatial position and velocity of the particle can be studied in an approach that parallels the nonrelativistic ideas, introducing as the only modification the relativistic kinetic energy term. This is the approach shown by the Lagrangian $L$, Eq. (23.33). This approach to particle dynamics relies on the well tested Lorentz-force format, and for this reason some advanced texts that otherwise cherish relativistic formulation of just about anything else, revert to this approach when addressing the variational principle formulation of the Lorentz-force. This approach breaks the unity of time and space but it assures that the dynamics of charged particles is not modified and that the evolution parameter is, like in non-relativistic format, the time $t$. The tacit assumption made is that there is such universal $t$ that can be used as an evolution parameter, which thus cannot be varied searching for the least action path of a particle.

B: Proper-time as evolution parameter
The proper time of a particle, $\tau$, is introduced as an evolution parameter. This is justified by proper time being scored by the body clock, and the definition of velocity as a measure of change in particle space-time position per unit of proper time. Given the relation with time $\tau = t\sqrt{1 - v^2/c^2}$ this means that particle dynamics could be different from case A:. An expression of the circumstance is seen in the covariant format of kinetic energy Eq. (25.19): the kinetic energy action is a constant.

Therefore if $\tau$ is an evolution parameter, the format of Eq. (25.19) must be amended to preserve the form of Lorentz-force. In this process the constraint $u^2 = c^2$ for the 4-dimensional particle motion is implemented, for example we adopt

$$T^r = -\frac{1}{2}m(c^2 + u^2).$$  \hspace{1cm} (25.20a)

The rationale for this choice is that the actual value of the action is not modified since the dynamical motion following from Eq. (25.20a) as we show satisfies $u^2 = c^2$, including the case $m \to m(x)$, motion in a mass-potential, see exercise X–3. Other approaches include

$$T^B = -mc\sqrt{u^2},$$  \hspace{1cm} (25.20b)

---

\[ T_R = -\frac{1}{2}m\left(u^2 - c^2\right). \] (25.20c)

Eq. (25.20b) is used by Barut, and Rohrlich\(^2\), the latter work amplifies that any scalar parameter, not only proper time is appropriate to characterize the particle path. One sees Eq. (25.20c) frequently and it was used in another relativity book by this author. The rational is that \( m \) is a Legendre parameter fixing the constraint \( u^2 - c^2 = 0 \). Since the kinetic term Eq. (25.20c) differs from the other seen Eq. (25.20a) by a constant, the only difference between the two is how we explain the choice. The dynamical equations that we find are the same. The choice Eq. (25.20b) is in that regard entirely equivalent.

We now obtain the dynamical equation that follow with \( \tau \) as evolution parameter and the choice Eq. (25.20a) for the kinetic energy. Given that \( \tau \) as an evolution parameter is not part of dynamics, we study

\[
\delta \left( \int_{P^1}^{P^2} L^r_L \ d\tau \right) = \int_{P^1}^{P^2} (\delta L^r_L) \ d\tau . \tag{25.21}
\]

Following the same argument line as in nonrelativistic approach we obtain

\[
\int_{P^1}^{P^2} (\delta L^r_L) \ d\tau = \int_{P^1}^{P^2} \delta x^\mu \left( \frac{\partial L^r_L}{\partial x^\mu} - \frac{d}{d\tau} \frac{\partial L^r_L}{\partial u^\mu} \right) d\tau . \tag{25.22}
\]

Setting the coefficients of variation \( \delta x^\mu \) to zero we find the Lagrange equations of particle motion for relativistic dynamics

\[
I_L = \int_{\tau^1}^{\tau^2} L^r_L(x,u)d\tau , \quad L^r_L = T^r - U^r , \quad 0 = \frac{\partial L^r_L}{\partial x^\mu} - \frac{d}{d\tau} \frac{\partial L^r_L}{\partial u^\mu} , \tag{25.23a}
\]

where from Eq. (25.20a) and Eq. (25.18) our choice for the action is

\[
L^r_L = -\frac{1}{2}m(u^2 + c^2) - e u \cdot A . \tag{25.23b}
\]

The Lagrange equation of motion for the covariant form of the Lorentz-force using the action Eq. (25.20a) takes the form

\[
\frac{d(mu^\mu)}{d\tau} = K^r_L = -\frac{\partial U^r}{\partial x^\mu} + \frac{d}{d\tau} \frac{\partial U^r}{\partial u^\mu} . \tag{25.24}
\]

The differentiation with respect to $x_\mu$ has the same meaning as the covariant gradient and since $u^\mu(\tau)$ is the proper time-dependent 4-velocity of the particle, the gradient commutes. We thus obtain

$$\frac{d(mu^\mu)}{d\tau} = K^\mu_L = eu_\nu \partial^\mu A^\nu - e \frac{d}{d\tau} A^\mu,$$

which indeed is the 4-Lorentz-force as introduced in Eq. (25.6). Upon multiplication with $u^\mu$ given that $u \cdot K_L = 0$ we find that any solution satisfies $d(mu^2)/d\tau = 0$, that is, the rest energy of a particle cannot change, a result interpreted to mean $u^2 = c^2$, that is, $c$ is maximum speed of the particle.

The result Eq. (25.25) is the expected format of the relativistic Lorentz-force. Thus adopting Eq. (25.20a) (or for that matter Eq. (25.20b), Eq. (25.20c)) and treating $\tau$ as the evolution parameter reproduces the dynamics that we expect to derive. All seems to be in order. However, there is the matter of the ad-hoc kinetic energy modification, which is not unique, see the alternate forms Eq. (25.20a), Eq. (25.20b) and Eq. (25.20c). Another difficulty of this formulation is that we assume that all four components of the 4-velocity are dynamically independent, and yet we know that principles of relativity require the constraint $u^2 = c^2$ without concern for how the force is chosen. Here $u^2 = c^2$ is a dynamical result and not a principle. Thus we return in section 25.3 to discuss an alternate approach that is equivalent to this case B where $u^2 = c^2$ arises as general principle.

C: The evolution parameter (proper time) variation in the kinetic term only

The above remark about demerits of B: explains why in many contemporary texts a 'hybrid' approach is taken\(^3\). In the study of the kinetic energy these authors consider the proper time as being part of the particle dynamics and seek to minimize the proper time needed to move between events $P^1$ and $P^2$ according to Eq. (25.19). With any variation of the path $x^\mu(\tau)$, Eq. (25.16), the proper time integral varies, in case of Eq. (25.19) we seek the path of least proper time connecting the two events.

The variation of the proper time defining the kinetic term proceeds as follows:

$$\delta c\tau = \delta \sqrt{g_{\mu\nu}dx^\mu dx^\nu} = \left(\frac{g_{\mu\nu}(\delta dx^\mu)dx^\nu + g_{\mu\nu}dx^\mu(\delta dx^\nu)}{2\sqrt{g_{\mu\nu}dx^\mu dx^\nu}}\right) = \frac{dx_\mu \delta dx^\mu}{cd\tau}.$$  \hspace{1cm} (25.26)

We recognize $dx_\mu/d\tau = u_\mu$, and that we can exchange the infinitesimal increments $\delta dx^\mu = d\delta x^\mu$ which results in the kinetic energy Eq. (25.19) variation taking the form

$$\delta \left( \int_{P^1}^{P^2} d\tau \right) = \int_{P^1}^{P^2} \delta d\tau = \int_{P^1}^{P^2} u_\mu \frac{d(\delta x^\mu)}{c^2} d\tau = -\int_{P^1}^{P^2} \delta x^\mu \frac{du^\mu}{c^2} d\tau,$$  \hspace{1cm} (25.27)

where we used the constraint in Eq. (25.16) in order to drop the endpoint terms arising in partial integration.

The variation of $\tau$ is a convenient tool for avoiding somewhat arbitrary redefinition of the kinetic energy in the action. The problem with this approach is that variation of $\tau$ is not done for the potential energy, $U^r$, explaining our naming of the method ‘hybrid’. Instead one writes

$$\int_{\tau_1}^{\tau_2} U^r d\tau = e \int_{\tau_1}^{\tau_2} A \cdot ud\tau \equiv e \int_{\tau_1}^{\tau_2} A \cdot dx , \quad (25.28)$$

thus adopting in terms of the last expression without explicit reference to $\tau$ a path integral in 4-dimensional space as the quantity of interest. When we side-step the variation of $\tau$ in the potential energy term, nothing has changed compared to case B in regard to the variation of the potential term $U^r$ and thus the relativistic Lorentz-force equations of motion follow.

However, there is no way known to give meaning to the last expression in Eq. (25.28) other than by way of the proper time integral. For this reason to hide the presence of $\tau$ from view in order to avoid variation of $\tau$ seems inconsistent: writing the potential energy action as a path integral stripped of an evolution parameter is just a device for hiding an inconsistency in the treatment of the kinetic energy term compared to the potential energy term. We remember that we describe the dynamics of a particle as a function of $\tau$ and obtain from the variational principle the dynamical equations of the particle as a function of $\tau$. Therefore it must be agreed that omitting $\tau$ in a study of the variation of $U^r d\tau$, when we vary it in the kinetic term, is blatantly inconsistent.

This case is discussed since it is widely present in literature. Below we attempt to repair its inconsistency and we encounter another variational variant D: which we state for completeness, pending further consideration of its (de)merits.

**D: Consistent variation of action, and particle proper time**

In this last case in comparison to case C we explicitly acknowledge that to define action integral along the particle path there is always an integration along the path in terms of the proper time of the particle. With any variation of the path $x'^\mu(\tau)$, Eq. (25.16), the proper time integral varies as we seek the path of least action connecting the two events.

The rationale for varying the evolution parameter $\tau$ is two-fold: we already know from the study of the case C that this works for kinetic energy delivering the expected inertial term dynamics; we also know that $\tau$ is not an independent evolution parameter since the proper age of a particle depends on the path taken. This is clearly unlike case A where the particle dynamics did not impact the laboratory time.
In Eq. (25.21) we had just one dynamical contribution to consider, and now we have two

\[ \delta \left( \int_{P_1}^{P_2} L_\mu^r \, d\tau \right) = \int_{P_1}^{P_2} \left( \delta L_\mu^r \, d\tau + L_\mu^r \delta d\tau \right). \]  

(25.29)

To evaluate the second term we use the variation of the proper time, Eq. (25.26),

\[ \delta \tau = \delta \sqrt{g_{\mu\nu} dx^\mu dx^\nu} = \frac{g_{\mu\nu}(\delta x^\mu)dx^\nu + g_{\mu\nu}dx^\mu(\delta x^\nu)}{2\sqrt{g_{\mu\nu} dx^\mu dx^\nu}} = \frac{dx_\mu \delta x^\mu}{d\tau}. \]  

(25.30)

We recognize \( dx_\mu/d\tau = u^\mu \), and that we can exchange the infinitesimal increments \( \delta x^\mu = d\delta x^\mu \) which results in the second term in Eq. (25.29) taking the form

\[ \int_{P_1}^{P_2} L_\mu^r d\tau = \int_{P_1}^{P_2} L_\mu^ru^\mu d\tau = -\int_{P_1}^{P_2} \delta x^\mu \frac{d(L_\mu^ru^\mu)}{c^2 d\tau} d\tau. \]  

(25.31)

Here we used the constraint in Eq. (25.16) in order to drop the end-point terms arising in partial integration. For the first term in Eq. (25.29) we use Eq. (25.22). Assembling all terms we have for Eq. (25.29)

\[ \delta \left( \int_{P_1}^{P_2} L_\mu^r \, d\tau \right) = \int_{P_1}^{P_2} \delta x_\mu \left( \frac{\partial L_\mu^r}{\partial x_\mu} - \frac{d}{d\tau} \frac{\partial L_\mu^r}{\partial u_\mu} - \frac{d(L_\mu^ru^\mu)}{c^2 d\tau} \right) d\tau \]  

(25.32)

The variation \( \delta x_\mu \) being arbitrary we obtain

\[ \frac{d(L_\mu^ru^\mu)}{c^2 d\tau} = \frac{\partial L_\mu^r}{\partial x_\mu} - \frac{d}{d\tau} \frac{\partial L_\mu^r}{\partial u_\mu}, \]  

(25.33a)

where

\[ L_\mu^r = -mc^2 - U^r, \quad U^r = eu \cdot A, \]  

(25.33b)

and hence we find

\[ \frac{d[(m + U^r/c^2)u^\mu]}{d\tau} = K_\mu^r = eu_\nu\partial^\mu A^\nu - ce \frac{d}{d\tau} A^\mu. \]  

(25.33c)

The mass comprises in the presence of EM-potential explicitly an additional component

\[ m(x) \equiv m + \frac{1}{c^2} eu \cdot A, \]  

(25.34)

which makes the resulting force different from e.g. case B, Eq. (25.25). The new dynamical term appears not to be negligible, and it makes the inertia of a particle respond to potential rather than field, a situation that is in general not acceptable.

We further find multiplying exercise 25.33c with \( (mc^4 + U^r)u_\mu \) that

\[ \frac{d[(mc^2 + U^r)^2u^2]}{d\tau} = 0 \rightarrow (mc^2 + U^r)^2u^2 = \text{Const.} = m^2e^6, \]  

(25.35)
where the value of the \( \text{Const.} \) is arrived at considering the domain of space where \( U \to 0 \). Solving for \( u^2 \) we obtain
\[
\begin{align*}
    u^2 &= \frac{c^2}{1 + \frac{e u \cdot A}{mc^2}}. \\
\end{align*}
\]
(25.36)

The value of \( u^2 > c^2 \) seems allowed, though the ultimate answer requires full solution of Eq. \((25.33c)\) so that we can actually determine if \( eu \cdot A < 0 \) is possible. If this can happen, it implies that in domain filled with potential of great strength faster than light motion is allowed.

Thus it seems that allowing for variation of proper time in both kinetic and potential energy terms finding the dynamical equations of motion is both inconsistent with the theoretical principles and with experimental physics reality.

**To summarize:**
Case **A** works but lacks symmetry between space and time. Case **B**, with \( \tau \) as an evolution parameter is what we use in the following, however there is ambiguity in regard to the form of the inertial term. Case **C** extends the variation principle to minimize particle proper time \( \tau \) – however only for the kinetic energy term. This treats kinetic and potential energy inconsistently. Case **D** arises repairing this inconsistency. In presence of EM-potentials this approach introduces modification of the Lorentz-force in that the mass is acquiring addition inertial contribution from the 4-potential according to Eq. \((25.34)\) and there is possibility of faster than light motion, see Eq. \((25.36)\). For cases **A**, **B**, and **C**, any spatial dependence of mass \( m \) could only arise from an additional externally prescribed ‘mass’ potential \( m \to m(x) = m + h(x) \), see exercise \([X-3]\), while in case **D** presence of EM-potential \( A^\mu \) suffices to introduce such a dependence. An entirely different formulation that translates Hamiltonian dynamics, see section \([23.4]\) into the 4-dimensional context will be presented in the next section \([25.3]\). Even though it has been very little used this maybe the only fully consistent non-arbitrary method to connect relativistic particle dynamics presented in covariant formulation to a variational principle.

---

**Exercise X–3: Particle in a mass-potential**

Explore particle dynamics assuming the presence of a mass potential \( m = m + h(x) \equiv m(x) \). There is no EM-potential, \( A^\mu = 0 \).

**Solution**

The relativistic action that describes the motion of a particle in space where mass can vary as a function of position will be considered according to case **B** following Eq. \((25.23a)\) and case **D** following Eq. \((25.33a)\).
**Dynamics according to case B:** According to Eq. (25.23b) the action is

\[ L^B_L = -\frac{1}{2} m(x)(u^2 + c^2). \]

which leads according to Eq. (25.23a) to the equation of motion

\[ \frac{d(mu^\mu)}{d\tau} = \frac{\partial m}{\partial x_\mu} \left( \frac{c^2 + u^2}{2} \right). \]

Upon multiplication with \( u_\mu \), and summing over \( \mu \) we obtain

\[ 0 = \frac{d}{d\tau} \left[ \frac{m(c^2 - u^2)}{2} \right] - \frac{m}{2} \frac{d}{d\tau} \left[ \frac{1}{2} d[m(c^2 - u^2)] \right], \]

where we used the chain rule for the total derivative of mass with respect to proper time,

\[ \frac{d}{d\tau} = \frac{dx^\nu}{d\tau} \frac{\partial}{\partial x^\nu} = u^\nu \partial_\nu, \]

For arbitrary \( m(x) \) constraint Eq. 3 can only be satisfied by

\[ u^2 = c^2 = \text{Const}. \]

Thus as the particle moves according to Eq. 2 it maintains the condition Eq. 5 that guarantees adherence to special relativity.

Using Eq. 5 to simplify Eq. 2 we obtain

\[ \frac{d(mu^\mu)}{d\tau} = c^2 \frac{\partial m}{\partial x_\mu}, \]

which at first sight is only valid for solutions that satisfy the stated constraint Eq. 5 obtained from the full dynamical form Eq. 2. However, we now consider what happens when we multiply Eq. 6 by \( u_\mu \)

\[ \frac{1}{2} m \frac{d}{d\tau} \left[ \frac{1}{u^2} \frac{d}{d\tau} \right] + u_\mu \frac{dm}{d\tau} = c_3 \frac{dm}{d\tau}, \]

where we used the relation Eq. 4 on the right hand side. Upon multiplication with \( m \) we see that Eq. 7 is a total differential

\[ 0 = \frac{1}{2} \frac{d}{d\tau} \left[ \frac{m^2}{u^2 - c^2} \right]. \]

Thus all unconstrained solutions of simplified dynamical equation Eq. 6 automatically satisfy constraint Eq. 5 and provide all solutions to Eq. 2.

**Dynamics according to case D:** According to Eq. (25.33b) the action is

\[ L^D_L = -m(x)c^2, \]
which leads according to Eq. (25.33a) to

\[ \frac{d(mu^\mu)}{c^2 d\tau} = \frac{\partial m}{\partial x_\mu}, \]

which is identical to Eq. 6 showing in absence of EM-potential the equivalence of the dynamics according to B and D.

End X–3: Particle in a mass-potential

### 25.3 Covariant Hamiltonian for action principle

The relativistic dynamics when described in terms of 4-position \( x^\mu \) and 4-velocity \( u^\mu = dx^\mu/d\tau \) of a particle is subject to the constraint \( u^2 = c^2 = \text{Const.} \), since the velocity of light is constant. This means that not all four relativistic components \( u^\mu \) are independent dynamical variables. On the other hand in consideration of the dynamically fixed binding effect on the unbound particle mass we have for a bound body \( p^2 = c^2 M^2 \neq \text{Const.} \). Thus a Hamiltonian covariant formulation that employs 4-momentum would not be subject to any constraint.

A relativistic generalization of the Hamiltonian approach to relativistic dynamics was proposed by Bergmann and elaborated by Tauber and Weinberg\(^4\). We base our approach on some yet to be determined covariant Hamiltonian \( \mathcal{H}(p^\mu, x^\mu) \) and construct by means of a Legendre transform, compare Eq. (23.35), the action \( I_H \)

\[ I_H = \int d\tau \left( \frac{dx^\mu}{d\tau} p_\mu - \mathcal{H}(x^\mu, p^\mu) \right) \quad (25.37) \]

This form closely resembles Hamilton’s form of the variational principle in non-relativistic mechanics as developed in second part of section 23.4.

The Hamilton equations of motion, compare Eq. (23.38), in covariant form are obtained by varying \( I_H \) independently with respect to the four canonical momenta \( p_\mu(\tau) \) and the four particle location coordinates \( x^\mu(\tau) \) and are:

\[ \frac{dx^\mu}{d\tau} = \frac{\partial \mathcal{H}(x,p)}{\partial p_\mu}, \quad \frac{dp_\mu}{d\tau} = -\frac{\partial \mathcal{H}(x,p)}{\partial x^\mu}. \quad (25.38) \]

These equations have canonical form and hence, compare Eq. (23.39)

\[ \frac{d\mathcal{H}(x,p)}{d\tau} = \frac{\partial \mathcal{H}(x,p)}{\partial \tau} = \frac{dx^\mu}{d\tau} \frac{\partial \mathcal{H}(x,p)}{\partial x^\mu} + \frac{dp_\mu}{d\tau} \frac{\partial \mathcal{H}(x,p)}{\partial p_\mu} = \frac{\partial \mathcal{H}(x,p)}{\partial \tau}. \quad (25.39) \]

In absence of explicit (proper) time dependence, the conserved and invariant \( \mathcal{H}(x, p) \) energy can only be the mass-energy equivalent, \( \mathcal{H}_{\text{body}} = Mc^2 \), of a (bound) body in its rest-frame. The sign in Eq. (25.37) is chosen in such a way that the result agrees with our finding Eq. (25.19) and Eq. (23.33).

We choose the form of \( \mathcal{H} \) as described to be the mass-energy equivalent of the system, for free particle

\[
\mathcal{H}_0(x, p) = c\sqrt{p_\mu p^\mu} = Mc^2.
\]  

(25.40)

Which is just the expression for the mass of a particle, as expected. In absence of EM 4-potential both inertial and canonical momentum are the same. However in above Eq. (25.40) we must in the general case use inertial momentum as otherwise we would not be able to recover the invariant bound mass-energy equivalent in presence of the 4-potential. Thus in the presence of the 4-potential Eq. (25.40) reads

\[
\mathcal{H}(x, p) = c\sqrt{(p_\mu - eA_\mu(x))(p^\mu - eA^\mu(x))},
\]  

(25.41)

where \( p^\mu \) is now the canonical momentum, compare our discussion in section 23.4.

The dynamical equations Eq. (25.38) following from Eq. (25.41) are

\[
\frac{dx^\mu}{d\tau} = c^2 \frac{p^\mu - eA^\mu(x)}{\mathcal{H}(x, p)},
\]

(25.42a)

\[
\frac{dp^\mu}{d\tau} = c^2 \frac{(p_\nu - eA_\nu(x))e\partial^\mu A^\nu(x)}{\mathcal{H}(x, p)}.
\]

(25.42b)

To obtain the Lorentz-force we complement on both sides of Eq. (25.42b) by the term \(-eA_\mu\) to obtain

\[
\frac{d(p^\mu - eA^\mu(x))}{d\tau} = -\frac{edA^\mu(x)}{d\tau} + c^2 \frac{(p_\nu - eA_\nu(x))e\partial^\mu A^\nu(x)}{\mathcal{H}(x, p)}.
\]

(25.43)

Next we use Eq. (25.42a) on the right and left of Eq. (25.43)

\[
\frac{d}{d\tau} \left( \mathcal{H}(x, p) \frac{dx^\mu}{c^2} \right) = -e \frac{dA^\mu}{d\tau} + e \frac{dx_\nu}{d\tau} \partial^\mu A^\nu.
\]  

(25.44)

This is the 4-Lorentz-force as introduced in Eq. (25.6) and rederived from a different form of variational principle, see Eq. (25.25). \( \mathcal{H} \) is recognized as the rest-energy of the body. Thus in the case considered now, a particle moving in potential \((e/c)A_\mu\), the relativistic Hamiltonian \( \mathcal{H}/c^2 \) is the invariant mass, accounting for any binding effect; for example exercise IX–5 on page 337 illustrated by figure 23-6 on page 342 shows the rest energy of Coulomb bound electron for a given strength of the binding potential.
We multiply Eq. (25.44) from left with 
\[ \frac{\mathcal{H}(x,p) \, dx_\mu}{c^2} \frac{d}{d\tau}. \]
The right hand side vanishes as is always the case for the 4-Lorentz-force, while 
the left hand side provides the mass-energy equivalent conservation law
\[ \frac{d}{d\tau} \left( \frac{\mathcal{H}^2(x,p) \, dx}{c^4} \frac{dx}{d\tau} \cdot \frac{dx}{d\tau} \right) = 0. \] (25.45)

To simplify this relation further we square equation Eq. (25.42a) to find
\[ \frac{dx^\mu}{d\tau} \cdot \frac{dx_\mu}{d\tau} = c^2 cp^\mu - eA^\mu = c^2 \frac{\mathcal{H}^2(x,p)}{\mathcal{H}(x,p)} = c^2. \] (25.46)

We use this result in Eq. (25.45) to obtain
\[ \frac{d}{d\tau} \left( \frac{\mathcal{H}(x,p)}{c^2} \right) = 0, \quad \rightarrow \quad \mathcal{H}(x,p) = mc^2. \] (25.47)

which does not depend on form of the 4-potential \( A^\mu \), and allows to introduce 
conserved quantity we call mass \( m \). Note that this conservation law is stronger 
compared to Eq. (25.39). As formulated a particle moving in a the applied 4-
potential cannot change mass-energy equivalent, and we also find that \( u^2 = c^2 \), 
see Eq. (25.46).

We now obtain the explicit value of the action \( I_H(x,dx/d\tau) \), Eq. (25.37), which 
can be evaluated once the path of a particle is known
\[ I_H = \int d\tau \left( -\mathcal{H} + \left[ \frac{dx^\mu}{d\tau} \left( p_\mu - eA_\mu \right) + e \frac{dx^\mu}{d\tau} A_\mu \right] \right), \] (25.48)

where we recognize the cancellation between the last terms that restores the form 
Eq. (25.37). We now eliminate all four canonical momenta using Eq. (25.42a) in 
favor of \( dx^\mu/d\tau \) and obtain
\[ I_H = \int d\tau \left( e \left. \frac{dx^\mu}{d\tau} \right| A_\mu + m \left[ \frac{dx^\mu}{d\tau} \cdot \frac{dx_\mu}{d\tau} - c^2 \right] \right). \] (25.49)

The substitution of \( m \) for \( \mathcal{H}/c^2 \) is allowed considering Eq. (25.47). We see that 
for any dynamical solution \( I_H \) has a value that is different from our earlier consider-
ations: the inertial mass term differs from Eq. (25.20c) by a factor 2. Another 
difference that is more subtle is that mass \( m \) that appears in Eq. (25.49) is not 
a natural input constant, but a dynamically determined system mass. These 
differences suggest that the covariant Hamiltonian approach will differ in some 
applications and results from the Lagrangian formalism where all the four velocities 
\( u^\mu = dx^\mu/d\tau \) appear as dynamically independent which is inconsistent with 
the dynamical constraint Eq. (25.46).
The discussion of covariant variational principle based on covariant Hamiltonian developed in Ref. [4] is offered here after being lost for half century in the hope that it can rekindle interest in the 4-dimensional Hamiltonian description of relativistic (multi-body) systems or/and could help improve of our understanding of the EM-force which must include physics that the Lorentz-force does not fully capture. There are clear conceptual advantages of this approach which is not beset with the problems we encounter in the other covariant approaches that we presented where one aims to obtain charged particle dynamics from a variation principle.

26 Covariant Fields and Invariants

26.1 EM-fields: relativistic form

In exercise X–2 on page 366 the rules that govern the up-down indices were considered. We use these insights now in order to introduce covariant form of EM-fields. As a first step we form the generalized divergence

\[ \partial_\mu A^\mu = \frac{\partial A^\mu}{\partial x^\mu} = \frac{\partial A^0}{\partial x^0} + \frac{\partial A^i}{\partial x^i} = \frac{\partial V/c}{\partial t} + \mathbf{\nabla} \cdot \mathbf{A}. \] (26.1)

We now seek another first order derivative of 4-vector potential \( A^\mu \) that we can use to express the six fields, \( E_i, B_i; \ i = 1, 2, 3 \) in terms of derivatives of the 4-potential \( A^\mu \). We consider the antisymmetric 4 × 4 field matrix as it has naturally just six independent components (see below exercise X–7 on page 390)

\[ F^{\mu\nu} \equiv \frac{\partial A^\nu}{\partial x_\mu} - \frac{\partial A^\mu}{\partial x_\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu. \] (26.2)

It is important to remember that in Eq. (26.2) the differentiation is with respect to covariant space-time variables is taken and we will point below to signs that arise.

There are three groups of terms to consider:

1. Because of antisymmetry \( F^{\mu\nu} = -F^{\nu\mu} \) we have

\[ F^{00} = F^{11} = F^{22} = F^{33} = 0. \] (26.3)

2. Setting \( \nu = 0 \) we obtain three non-vanishing terms

\[ F^{i0} \equiv \frac{\partial A^0}{\partial x_i} - \frac{\partial A^i}{\partial x_0} = - (\mathbf{\nabla})^i V/c - \frac{1}{c} \frac{\partial (\mathbf{A})^i}{\partial t} \equiv (\mathbf{E})^i/c, \ i = 1, 2, 3, \] (26.4a)
where the last identity follows from Eq. (23.6). Note that a sign appears in the first term to accommodate the covariant differentiation. To assure that there cannot be doubt that we are looking at a 3-vector component which by definition is a contravariant (positive) component we included here the superscript-$i$. Considering the antisymmetry of $F^{\mu\nu}$ we also have

$$F^{0i} = -F^{i0} = -(\vec{E})^i / c, \quad i = 1, 2, 3.$$  \hspace{1cm} (26.4b)

3. The other 3+3 spatial independent components of $F^{\mu\nu}$ are

$$F^{\mu=i\nu=j} \equiv \left( \frac{\partial A^j}{\partial x_i} - \frac{\partial A^i}{\partial x_j} \right) = - \epsilon^{ij3} \frac{\partial A^i}{\partial x^j} \equiv -B^3, \quad \epsilon^{ij2} \frac{\partial A^i}{\partial x^j} \equiv B^2, \quad \epsilon^{ij1} \frac{\partial A^i}{\partial x^j} \equiv B^1.$$  \hspace{1cm} (26.5a)

where

$$(i, j) = (1, 2), (1, 3), (2, 3) \quad \& \quad (i, j) \rightarrow (j, i).$$ \hspace{1cm} (26.5b)

The extra prefactor ‘$\text{−}$’ arises from the last term in Eq. (25.5), the contravariant derivatives with respect to spatial components introduce the minus sign. Considering this result term by term and remembering that the components shown on right are now those of the 3-vector we obtain

$$F^{12} = \frac{\partial A^1}{\partial x^2} - \frac{\partial A^2}{\partial x^1} = \sum_{ij} \epsilon^{ij3} \frac{\partial A^i}{\partial x^j} \equiv -B^3,$$

$$F^{13} = \frac{\partial A^1}{\partial x^3} - \frac{\partial A^3}{\partial x^1} = -\sum_{ij} \epsilon^{ij2} \frac{\partial A^i}{\partial x^j} \equiv B^2,$$

$$F^{23} = \frac{\partial A^2}{\partial x^3} - \frac{\partial A^3}{\partial x^2} = \sum_{ij} \epsilon^{ij1} \frac{\partial A^i}{\partial x^j} \equiv -B^1.$$  \hspace{1cm} (26.6a)

and the antisymmetric complement

$$F^{21} = B^3, \quad F^{31} = -B^2, \quad F^{23} = B^1.$$  \hspace{1cm} (26.6b)

We now combine the results shown in Eq. (26.3), Eq. (26.4a) & Eq. (26.4b) and Eq. (26.6a) & Eq. (26.6b) to write explicitly the matrix $(F^{\mu\nu})$
26.2 LT of electromagnetic fields and field invariants

In the following we explore the Lorentz transformation rules for the EM-fields $\vec{E}$ and $\vec{B}$. To this end we study how the field matrix $F$ Eq. (26.7) transforms. Both Lorentz indices in Eq. (26.7) need to be transformed using the boost matrix shown in Eq. (20.19); using the matrix multiplication notation we have

$$F' = \Lambda^T F \Lambda, \quad \hat{F}' = g \Lambda^T F g .$$  \hspace{1cm} (26.8)

Here we make explicit the distinction between the contravariant $F \equiv (F^{\mu\nu})$ and covariant fields

$$\hat{F} \equiv (F_{\mu\nu}) = g F g .$$  \hspace{1cm} (26.9)

We need to include the metric $g$ on the right in Eq. (26.8) when transforming covariant indices since our matrix $\Lambda$ was set up for contravariant indices.

Similar to discussion of the Lorentz invariance of the metric $g$ see Eq. (20.33) we can focus our attention on the LT boosts for which $\Lambda^T = \Lambda$. We obtain using the $z$-boost

$$F' = \begin{pmatrix} \gamma & 0 & 0 & -\beta \gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\beta \gamma & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} \gamma & 0 & 0 & -\beta \gamma \\ 0 & -\mathcal{E}^1/c & -\mathcal{E}^2/c & -\mathcal{E}^3/c \\ \mathcal{E}^1/c & 0 & -B^3 & B^2 \\ \mathcal{E}^2/c & B^3 & 0 & -B^1 \\ \mathcal{E}^3/c & -B^2 & B^1 & 0 \end{pmatrix} \begin{pmatrix} \gamma & 0 & 0 & -\beta \gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\beta \gamma & 0 & 0 & \gamma \end{pmatrix} = \begin{pmatrix} \beta \mathcal{E}^3/c & -\mathcal{E}^1/c & -\mathcal{E}^2/c & -\gamma \mathcal{E}^3/c \\ \gamma \mathcal{E}^1/c - \beta \gamma B^2 & 0 & -B^3 & -\beta \gamma \mathcal{E}^1/c + \gamma B^2 \\ \gamma \mathcal{E}^2/c + \beta \gamma B^1 & B^3 & 0 & -\beta \gamma \mathcal{E}^2/c - \gamma B^1 \\ \gamma \mathcal{E}^3/c & -B^2 & B^1 & -\beta \gamma \mathcal{E}^3/c \end{pmatrix}$$

$$= \begin{pmatrix} 0 & -\mathcal{E}^1/c + \beta \gamma B^2 & -\gamma \mathcal{E}^2 - \beta \gamma B^1 & -\mathcal{E}^3/c \\ -\gamma \mathcal{E}^1/c - \beta \gamma B^2 & 0 & -B^3 & \gamma B^2 - \beta \gamma \mathcal{E}^1/c \\ \gamma \mathcal{E}^2/c + \beta \gamma B^1 & B^3 & 0 & -\gamma B^1 - \beta \gamma \mathcal{E}^2/c \\ \mathcal{E}^3/c & -\gamma B^2 + \beta \gamma \mathcal{E}^1/c & \gamma B^1 + \beta \gamma \mathcal{E}^2/c & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & -\mathcal{E}'^1/c & -\mathcal{E}'^2/c & -\mathcal{E}'^3/c \\ \mathcal{E}'^1/c & 0 & -B'^3 & B'^2 \\ \mathcal{E}'^2/c & B'^3 & 0 & -B'^1 \\ \mathcal{E}'^3/c & -B'^2 & B'^1 & 0 \end{pmatrix} .$$  \hspace{1cm} (26.10)
This can be done in the same fashion for the $x$- and $y$-boosts and in a comparison of new components in $F'$ we read off a simple result

\[
\begin{align*}
\vec{E}'_\perp &= \gamma \left( \vec{E}_\perp + \vec{v} \times \vec{B}_\perp \right), \\
\vec{B}'_\perp &= \gamma \left( \vec{B}_\perp - \frac{\vec{v}}{c^2} \times \vec{E}_\perp \right), \\
\vec{E}'_\parallel &= \vec{E}_\parallel, \\
\vec{B}'_\parallel &= \vec{B}_\parallel,
\end{align*}
\]

(26.11)

where $\perp$ denotes the components of the field orthogonal to the boost velocity $\vec{v}$, and $\parallel$ denotes the components parallel to the boost velocity. The asymmetry seen in Eq. (26.11) in regard of appearance of speed of light is specific to the SI-unit system, see Insight on page 320. A symmetric expression arises using $\vec{E}$ and $c\vec{B}$ as variables.

We now highlight several significant differences from the LT we have so far considered: 1) the normally LT modified components of a vector parallel to the LT axis are now left unchanged; 2) the normally LT invariant transverse components are now changed; 3) while the other transformations we have seen have mixed the $0^{th}$ component (time / energy) with a vector quantity (position / momentum), the transformation of the fields mixes two vector quantities, namely the electric field $\vec{E}$ and the magnetic field $\vec{B}$. Property 3) is key to understanding the physics of EM-phenomena observed in the laboratory. A particle moving in an external purely electric or purely magnetic field experiences in general a combination of both electric and magnetic fields in its own rest frame.

We now look for Lorentz-invariants made of the EM-fields. We proceed in a manner similar to the manner we identified earlier the Lorentz-invariants of four-vectors considering a ‘square’ of Eq. (26.7)

\[
S \equiv \frac{\varepsilon_0 e^2}{4} F_{\mu\nu} F^{\mu\nu} = \frac{\varepsilon_0 e^2}{4} \text{Tr} \, \hat{F} F = \frac{\varepsilon_0 e^2}{4} \text{Tr} g F g F = \frac{\varepsilon_0 e^2}{4} \text{Tr}(g F)^2.
\]

(26.12)

We chose through introduction of the coefficient the dimension of energy density for $S$. The letter $S$ reminds us that this quantity is invariant under Lorentz-transformations, thus it is a Lorentz-scalar. Since the field strengths appear quadratic, $S$ a true scalar, it is invariant under axis-reversal discrete transformation called $P$ (for parity). That $P$ can change the sign of fields is manifest considering for example Eq. (23.6). Since $\vec{V}$ flips sign under axis reversal and so does any true vector $\vec{A}$, we see that $\vec{E}$ flips sign under $P$ having a sense of direction (polar ‘true’ vector), while $\vec{B}$ does not, being an ‘axial’ vector, that is a vector defined by sense of rotation (sometimes also called ‘pseudo vector’).
It can be easily shown $S$ is indeed a Lorentz-invariant. Keeping in mind Eq. (20.34), the invariance property of $g' = \Lambda g \Lambda = g$ we have

$$\text{Tr} \hat{F}'F' = \text{Tr} gF'gF' = \text{Tr} g\Lambda F\Lambda g\Lambda F\Lambda . \quad (26.13)$$

The well-known property of the trace, $\text{Tr} AB = \text{Tr} BA$ allows the movement of the matrices from one end to the other end of a trace equation, one by one, and we find

$$\text{Tr} \hat{F}'F' = \text{Tr} F(\Lambda g\Lambda)F(\Lambda g\Lambda) = \text{Tr} FgFg = \text{Tr} gFgF = \text{Tr} \hat{F}F' . \quad (26.14)$$

We now evaluate $S$, Eq. (26.12)

$$\hat{F}F = \left( \begin{array}{ccc} 1 & 0 & -\varepsilon^1/c \\ -1 & \varepsilon^1/c & -\varepsilon^2/c \\ -1 & -\varepsilon^2/c & -\varepsilon^3/c \end{array} \right)^2$$

$$= \left( \begin{array}{ccc} 0 & -\varepsilon^1/c & -\varepsilon^2/c \\ -\varepsilon^1/c & 0 & -B^3 \\ -\varepsilon^2/c & -B^3 & 0 \end{array} \right) \left( \begin{array}{ccc} 0 & -\varepsilon^1/c & -\varepsilon^2/c \\ -\varepsilon^1/c & 0 & -B^3 \\ -\varepsilon^2/c & -B^3 & 0 \end{array} \right)$$

$$= \left( \begin{array}{ccc} \hat{\varepsilon}^2 & \ldots & \ldots \\ \ldots & \varepsilon^1/c^2-B^3 \cdot B^2 & \ldots \\ \ldots & \ldots & \varepsilon^2/c^2-B^1 \cdot B^2 \end{array} \right) \text{,} \quad (26.15)$$

where we have only identified diagonal elements in the matrix. We thus find

$$\text{Tr} \hat{F}F = \varepsilon^2/c^2 + \varepsilon^1/c^2 - B^3 \cdot B^2 + \varepsilon^2/c^2 - B^1 \cdot B^2 + \varepsilon^3/c^2 - B^2 \cdot B^1 , \quad (26.16)$$

and restoring the factor defining $S$ we obtain

$$S = \epsilon_0 \frac{c^2}{2} \left( \hat{\varepsilon}^2 - c^2 B^2 \right) = \epsilon_0 \frac{c^2}{2} \left( \hat{\varepsilon}'^2 - c^2 B'^2 \right) . \quad (26.17)$$

There is another Lorentz-invariant we can form from the EM-fields, as can be seen by trial and error inspecting Eq. (26.11) and forming the vector product (see also exercise X–4 that follows)

$$\mathcal{P} \equiv c\epsilon_0 \hat{\varepsilon} \cdot \hat{B} = c\epsilon_0 \hat{\varepsilon}' \cdot \hat{B}' . \quad (26.18)$$
The letter $P$ reminds us that this quantity can be a pseudo-scalar; that is, it changes sign under axis reversal ‘parity’ $P$-transformation. This is so since this invariant is linear in fields $\vec{E}, \vec{B}$. $\vec{B}$ generated with the use of a cross product, Eq. (23.6), is an axial-vector, it does not flip sign, while $\vec{E}$ does under $P$-transformation.

To construct $P$ from $F$ we introduce the dual field tensor

$$F^*_{\mu \nu} = \frac{1}{2} \varepsilon_{\mu \nu \delta \kappa} F^{\delta \kappa},$$

where $\varepsilon_{\mu \nu \delta \kappa}$ is the 4-dimensional generalization of the totally antisymmetric 3-dimensional $\varepsilon_{ijk}$. Like $F$, the dual tensor $F^*$ is also antisymmetric. Thus we have to image 6 components from $F \rightarrow F^*$. One finds explicitly

$$F^* \equiv (F^*_{\mu \nu}) = (\varepsilon_{\mu \nu \delta \kappa} F^{\delta \kappa}) = \mu \nu \rightarrow
\begin{pmatrix}
0 & -B^1 & -B^2 & -B^3 \\
B^1 & 0 & -E^3/c & E^2/c \\
B^2 & E^3/c & 0 & -E^1 \\
B^3 & -E^2/c & E^1/c & 0
\end{pmatrix}.$$  \hspace{2cm} (26.20)

To show Eq. (26.20), it will be enough to to consider two representative elements:

i) For $\{\mu \nu \delta \kappa\} = \{0123\}$ the 01-component of $F^*_{01} = (F^{23} - F^{32})/2 = -B^1$ since $\varepsilon_{0123} = -\varepsilon_{0132} = +1$. We recall for comparison that $F^{01} = -E^1/c$.

ii) For $\{\mu \nu \delta \kappa\} = \{1203\}$ the 12-component of $F^*_{12} = (F^{03} - F^{30})/2 = -E^3/c$ since $\varepsilon_{1203} = -\varepsilon_{1230} = +1$ (even number of permutations moves ‘0’ to the new location). We recall for comparison that $F^{12} = -B^3$.

One can show that for all the other cases the same situation emerges; we therefore can rewrite the matrix $(F) \leftrightarrow (F^*)$, exchanging the entries $E_i/c \leftrightarrow B_i$. Thus

$$F^{\mu \nu}(\vec{E}/c \rightarrow \vec{B}, \vec{B} \rightarrow \vec{E}/c) = F^*_{\mu \nu}, \quad \left( F(\vec{E}/c \rightarrow \vec{B}, \vec{B} \rightarrow \vec{E}/c) \right) = (F^*).$$ \hspace{2cm} (26.21)

In our notation $(F^*)$ as shown in Eq. (26.20) has indices down, which explains why in this book the exchange Eq. (26.21) is with a positive sign.
We now evaluate the product

\[ c F_{\nu \mu}^* F^{\nu \mu} = c \text{Tr}(-F^* F) \]

\[ = c \text{Tr} \left( \begin{array}{ccc}
0 & B^1 & B^2 & B^3 \\
- B^1 & 0 & \mathcal{E}^3/c & - \mathcal{E}^2/c \\
- B^2 & - \mathcal{E}^3/c & 0 & \mathcal{E}^1/c \\
- B^3 & \mathcal{E}^2/c & - \mathcal{E}^1/c & 0
\end{array} \right) \left( \begin{array}{cccc}
0 & -\mathcal{E}^1/c & -\mathcal{E}^2/c & -\mathcal{E}^3/c \\
\mathcal{E}^1/c & 0 & -B^3 & B^2 \\
\mathcal{E}^2/c & B^3 & 0 & -B^1 \\
\mathcal{E}^3/c & -B^2 & B^1 & 0
\end{array} \right) \]

\[ = \text{Tr} \left( \begin{array}{ccc}
\mathcal{E} \cdot \mathcal{B} & \ldots & \ldots \\
\ldots & \mathcal{E} \cdot \mathcal{B} & \ldots \\
\ldots & \ldots & \mathcal{E} \cdot \mathcal{B}
\end{array} \right) \]

where we only show the needed diagonal elements. Thus we see that

\[ \frac{c^2 \epsilon_0}{4} F_{\nu \mu}^* F^{\nu \mu} = c \epsilon_0 \mathcal{E} \cdot \mathcal{B} \equiv \mathcal{P} \quad \text{(26.23)} \]

In exercise X–4 below we show the LI of \( \mathcal{P} \), and \( \mathcal{S} \) by implementing the explicit Lorentz-transformation of the fields, Eq. (26.11).

There is one more linearly independent bilinear in fields we can construct using two 3-vectors \( \mathcal{E}, \mathcal{B} \)

\[ \varepsilon \equiv \frac{\epsilon_0 c^2}{2} \left( \mathcal{E}^2/c^2 + \mathcal{B}^2 \right) \quad \text{(26.24)} \]

This quantity is not Lorentz-invariant being the energy density of the field in the reference frame in which the source of the field is at rest, see section 27.4 on page 422 for further discussion.

Exercise X–4: EM-field invariants

Show by explicit evaluation that \( \mathcal{S} = \frac{\epsilon_0}{2} (\mathcal{E}^2 - c^2 \mathcal{B}^2) \) and \( \mathcal{P} = c \epsilon_0 \mathcal{E} \cdot \mathcal{B} \) are invariant under Lorentz-transformations.

Solution

We separate components parallel and orthogonal to direction of the transformation and consider:

\[ \frac{1}{\epsilon_0} \frac{2}{\varepsilon} \mathcal{S} = \mathcal{E}'^2 - c^2 \mathcal{B}'^2 = \mathcal{E}_{\parallel}^2 + \mathcal{E}_{\perp}^2 - c^2 \mathcal{B}_{\parallel}^2 - c^2 \mathcal{B}_{\perp}^2 , \]
and

\[ 2 \quad \frac{1}{\epsilon_0} \mathcal{P} = \vec{E}' \cdot c\vec{B}' = \vec{E}'_\parallel \cdot c\vec{B}'_\parallel + \vec{E}'_\perp \cdot c\vec{B}'_\perp , \]

Now, we first consider Eq. 1. Using the transformation equations Eq. (26.11), where for symmetry we rewrite \( B \rightarrow c\vec{B} \), we obtain

\[ 3 \quad \vec{E}''_\parallel^2 = \vec{E}_\parallel^2 + \gamma^2 \vec{E}_\perp^2 + \gamma^2 (\vec{\beta} \times c\vec{B}_\perp)^2 + 2\gamma^2 \vec{E}_\perp \cdot (\vec{\beta} \times c\vec{B}_\perp), \]

\[ 4 \quad c^2 \vec{B}''_\parallel^2 = c^2 \vec{B}_\parallel^2 + \gamma^2 c^2 \vec{B}_\perp^2 + \gamma^2 (\vec{\beta} \times \vec{E}_\perp)^2 - 2\gamma^2 c\vec{B}_\perp \cdot (\vec{\beta} \times \vec{E}_\perp), \]

where as usual \( \vec{\beta} = \frac{\vec{v}}{c} \), \( \gamma = 1/\sqrt{1-\vec{\beta}^2} \). We can then substitute into Eq. 1 and use the cyclic property of the triple product

\[ 5 \quad \vec{a} \cdot (\vec{b} \times \vec{c}) = \vec{b} \cdot (\vec{c} \times \vec{a}) = \vec{c} \cdot (\vec{a} \times \vec{b}), \]

to cancel the last term of Eq. 3 with that of Eq. 4, and we obtain

\[ 6 \quad \frac{2}{\epsilon_0} S = \vec{E}''_\parallel^2 - c^2 \vec{B}''_\parallel^2 + \gamma^2 (\vec{E}_\perp^2 - c^2 \vec{B}_\perp^2 - (\vec{\beta} \times \vec{E}_\perp)^2 + (\vec{\beta} \times c\vec{B}_\perp)^2) . \]

Because \( \vec{E}_\perp \) and \( \vec{B}_\perp \) are by definition perpendicular to \( \vec{\beta} \), we can simplify the cross products, yielding

\[ 7 \quad \frac{2}{\epsilon_0} S = \vec{E}_\parallel^2 - c^2 \vec{B}_\parallel^2 + \gamma^2 \left(1 - \vec{\beta}^2\right) \left(\vec{E}_\perp^2 - c^2 \vec{B}_\perp^2\right) . \]

As \( \gamma^2 (1 - \vec{\beta}^2) = 1 \), we have verified that

\[ 8 \quad \frac{2}{\epsilon_0} S = \vec{E}^2 - c^2 \vec{B}^2 = \vec{E}'_\parallel^2 - c^2 \vec{B}'_\parallel^2 . \]

Thus, the quantity \( S \) is invariant under Lorentz-transformations.

Moving on to Eq. 2, we have

\[ 9 \quad \frac{1}{\epsilon_0} \mathcal{P} = \vec{E}' \cdot c\vec{B}' = \vec{E}'_\parallel \cdot c\vec{B}'_\parallel + \vec{E}'_\perp \cdot c\vec{B}'_\perp . \]

Again using the transformation equations Eq. (26.11), along with the identity \( \vec{a} \cdot (\vec{b} \times \vec{a}) = 0 \), we obtain

\[ 10 \quad \frac{1}{\epsilon_0} \mathcal{P} = \vec{E}_\parallel \cdot c\vec{B}_\parallel + \gamma^2 \left(\vec{E}_\perp \cdot c\vec{B}_\perp - (\vec{\beta} \times \vec{E}_\perp) \cdot (\vec{\beta} \times c\vec{B}_\perp)\right) . \]
Using Eq. 5 we have

\[ \frac{1}{\epsilon_0} \mathcal{P} = \vec{E}_\parallel \cdot c\vec{B}_\parallel + \gamma^2 \left( \vec{E}_\perp \cdot c\vec{B}_\perp - c\vec{B}_\perp \cdot (\vec{\beta} \times \vec{E}_\perp \times \vec{\beta}) \right). \]

At last, with the bac-cab rule, see Eq. 3 in exercise IX–3 we obtain

\[ \frac{1}{\epsilon_0} \mathcal{P} = \vec{E}_\parallel \cdot c\vec{B}_\parallel + \gamma^2 (1 - \vec{\beta}^2) \left( \vec{E}_\perp \cdot c\vec{B}_\perp \right). \]

Since \( \gamma^2 (1 - \vec{\beta}^2) = 1 \), we see that \( \mathcal{P} \) is indeed a Lorentz-invariant

\[ \frac{1}{\epsilon_0} \mathcal{P} = \vec{E} \cdot c\vec{B} = \vec{E}' \cdot c\vec{B}' . \]

End X–4: EM-field invariants

26.3 Constraints on field invariants

Considering the two Lorentz-invariants \( S = \frac{\epsilon_0}{2} (\vec{E}^2 - c^2 \vec{B}^2) \) and \( P = \epsilon_0 \vec{E} \cdot c\vec{B} \) we find the following important properties of the Lorentz-transformation of the EM-field. Some are trivial; others we demonstrate in the following exercises.

a) \( \vec{E}^2 - c^2 \vec{B}^2 = \vec{E'}^2 - c^2 \vec{B}'^2 \) is true in all frames of reference. Thus the sign of this invariant is preserved. Therefore, if \( |\vec{E}| > |c\vec{B}| \) or \( |\vec{E}| < |c\vec{B}| \) in one frame of reference, then the same is true for all other Lorentz observers; i.e. \( |\vec{E}| > |c\vec{B}| \) implies \( |\vec{E}'| > |c\vec{B}'| \) and \( |\vec{E}| < |c\vec{B}| \) implies \( |\vec{E}'| < |c\vec{B}'| \).

b) If \( \vec{E} \cdot \vec{B} = 0 \) in any frame of reference, then the same is true in all other frames of reference. If \( \vec{E} \cdot \vec{B} > 0 \) or \( \vec{E} \cdot \vec{B} < 0 \) this again remains true for all observers.

c) When \( \vec{E} \cdot \vec{B} = 0 \) but \( S \neq 0 \) one can find a frame of reference in which one of the fields vanishes; see exercise X–5 for explicit demonstration.

d) As long as \( \vec{E} \cdot \vec{B} = 0 \) and \( |\vec{E}|^2 - |c\vec{B}|^2 = 0 \) are not both true, we can always find a frame of reference in which the fields are parallel to each other; that is,

\[ \vec{E} \times \vec{B} = 0 , \quad (26.25) \]

for demonstration see exercise X–6.
For the case $\vec{E} \cdot \vec{B} = 0$, determine the Lorentz-transformations for which $\vec{E} = 0$ when $(\vec{E}^2 - c^2\vec{B}^2) < 0$, or for which $\vec{B} = 0$ when $(\vec{E}^2 - c^2\vec{B}^2) > 0$.

Solution

Since $\vec{E} \cdot \vec{B} = 0$, either one of the fields is already zero and there is nothing more to do, or we have fields that are orthogonal and a third normal direction is then $\vec{E} \times \vec{B}$.

Revisiting Eq. (26.11), we see that to transform both fields, $v$ cannot be parallel to the fields. The only way to change both fields is to have a transformation velocity parallel to $\vec{E} \times \vec{B}$:

1. $\vec{\beta} = \frac{\vec{v}}{c} = a(\vec{E} \times c\vec{B})$.

Choosing $a = 1/c^2\vec{B}^2$, and transforming the electric field with Eq. (26.11) we obtain

2. $\vec{E}' = \gamma \left( \vec{E} + \frac{\vec{E} \times \vec{B}}{\vec{B}^2} \times \vec{B} \right) = \gamma \left( \vec{E} - \frac{\vec{E} \vec{B}^2 - \vec{B}(\vec{E} \cdot \vec{B})}{\vec{B}^2} \right) = \vec{0}$,

where we have used $\vec{E} \cdot \vec{B} = 0$.

Choosing instead $a = 1/\vec{E}^2$ and transforming the magnetic field with Eq. (26.11) we obtain, similarly to Eq. 2

3. $\vec{B}' = \gamma \left( \vec{B} - \frac{\vec{E} \times \vec{B}}{\vec{E}^2} \times \vec{E} \right) = \vec{0}$.

We now consider the magnitude of $\vec{\beta}$, to ensure that we are considering physical transformations. For $a = 1/\vec{B}^2$ and nearly orthogonal fields we have

4. $|\vec{\beta}| = \frac{|\vec{E}| |c\vec{B}|}{c^2\vec{B}^2}$.

However, $|\vec{\beta}| < 1$ for any physical Lorentz-transformation. We see that a transformation that makes $\vec{E}' = \vec{0}$ is only possible when $(\vec{E}^2 - c^2\vec{B}^2) < 0$. Likewise, for $a = 1/\vec{E}^2$ we obtain

5. $|\vec{\beta}| = \frac{|\vec{E}| |c\vec{B}|}{\vec{E}^2}$.

Thus, a transformation for which $\vec{B}' = \vec{0}$ is only possible when $(\vec{E}^2 - c^2\vec{B}^2) > 0$.

End X–5: The zero field frame

Exercise X–6: Reference frame with parallel fields
Find the transformation that renders Lorentz-transformed fields $E'$ and $B'$ parallel to each other.

**Solution**

By assumption the condition Eq. (26.25) is not fulfilled and we seek the transformation that will render fields that are parallel and satisfy Eq. (26.25). We recall that a Lorentz-transformation leaves the field components parallel to direction of transformation carried out with $\vec{\beta}$ unchanged. Thus to find transformed fields that satisfy condition Eq. (26.25), we must alter field components that are normal so that Eq. (26.25) can be satisfied. To achieve this we seek a function $N(\vec{E}, c\vec{B})$ where we transform with

$$1 \quad \vec{\beta} = N(\vec{E} \times c\vec{B}) .$$

The transformed fields are, according to Eq. (26.11)

$$2 \quad \vec{E}' = \gamma(\vec{E} + \vec{\beta} \times c\vec{B}) = \gamma(\vec{E} + N(\vec{E} \times c\vec{B}) \times c\vec{B}) = \gamma[(1 - Nc^2\vec{B}^2)\vec{E} + N\vec{E} \cdot c\vec{B} \ c\vec{B}] ,$$

$$c\vec{B}' = \gamma(c\vec{B} - \vec{\beta} \times \vec{E}') = \gamma(c\vec{B} - N(\vec{E} \times c\vec{B}) \times \vec{E}') = \gamma[(1 - N\vec{E}^2) c\vec{B} + N\vec{E} \cdot c\vec{B} \ \vec{E}'] .$$

The condition Eq. (26.25) for the transformed fields reads, omitting the overall factor $\gamma^2(\vec{E} \times c\vec{B}) \neq 0$

$$3 \quad 0 = \vec{E}' \times c\vec{B}' = (1 - Nc^2\vec{B}^2)(1 - N\vec{E}^2) - N^2(\vec{E} \cdot c\vec{B})^2 .$$

We obtain a quadratic equation for $1/N$

$$4 \quad 0 = [c^2\vec{B}^2 \vec{E}^2 - (\vec{E} \cdot c\vec{B})^2] - \frac{1}{N}(c^2\vec{B}^2 + \vec{E}^2) + \frac{1}{N^2} .$$

with the solution

$$5 \quad \frac{1}{N} = \frac{c^2\vec{B}^2 + \vec{E}^2}{2} \pm \sqrt{\frac{(c^2\vec{B}^2 - \vec{E}^2)^2}{4} + (\vec{E} \cdot c\vec{B})^2} .$$

We seek the positive and smallest value of $N$ which assures that in Eq.1 we satisfy $0 < \vec{\beta}^2 < 1$; this corresponds to ‘+’ in Eq.5. This is so since when squaring $\frac{1}{N}$ we have all contributions manifestly positive and we can say that $\frac{1}{N^2} > \vec{E}^2 c^2\vec{B}^2$ and hence using Eq.1 we obtain

$$6 \quad \vec{\beta}^2 < \frac{(\vec{E} \times c\vec{B})^2}{\vec{E}^2 \vec{B}^2} \leq 1 .$$

We now look at special cases in terms of invariants that appear in Eq.5:

i) When the root in Eq.5 vanishes the value of $\beta = 1$ indicates that our solution is
meaningless. In fact the vanishing of this root means that the fields are for all observers always orthogonal and equal in magnitude. This also says that for no observer can a light wave be made of parallel fields.

ii) When $\vec{E} \cdot \vec{B}$ vanishes, the parallel condition is achieved by making one of the fields vanish, see exercise [X–5]. We also see these solutions in Eq. (2).

iii) Case $c^2B^2 - E^2 = 0$ does not create difficulties.

End X–6: Reference frame with parallel fields

### 26.4 Covariant form of the Lorentz-force in terms of fields

In the non-relativistic notation, the Lorentz-force involved fields and particle velocity. We now aim to cast the covariant form of the 4-Lorentz-force Eq. (25.6) and Eq. (25.25) into a form that also involves, by extension, the particle 4-velocity and spatial derivatives of potentials, i.e., the fields.

All the results we obtained imply that the covariant 4-Lorentz-force $K^\mu_L$ can be written in the form

$$\frac{d(m u^\mu)}{d\tau} = K^\mu_L = e u_\nu \partial^\mu A^\nu - e \frac{d}{d\tau} A^\mu ,$$  

(26.26)

We use the chain rule for the total derivative with respect to proper time, akin to what we used in exercise [X–3]

$$\frac{d}{d\tau} = \frac{dx_\nu}{d\tau} \frac{\partial}{\partial x_\nu} = u_\nu \partial^\nu .$$  

(26.27)

Using Eq. (26.27) in Eq. (25.25),

$$K^\mu_L = e u_\nu (\partial^\mu A^\nu - \partial^\nu A^\mu) .$$  

(26.28)

This expression has the format we seek, a 4-velocity multiplying a first derivative of potentials which thus must be a covariant representation of the fields $\vec{E}$ and $\vec{B}$. Combining Eq. (25.25) with Eq. (26.26) we obtain the covariant form of the 4-Lorentz-force written in terms of the $4 \times 4$ matrix of EM-fields

$$\frac{d(m u^\mu)}{d\tau} = e F^{\mu\nu} u_\nu , \quad F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu .$$  

(26.29)

The antisymmetric matrix $F^{\mu\nu}$, the ‘field tensor’ contains the fields $\vec{E}$ and $\vec{B}$.

We next reconsider the role of the constraint on particle dynamics, $u_\mu a^\mu = 0$:

$$u_\mu \frac{d(m u^\mu)}{d\tau} = e u_\mu F^{\mu\nu} u_\nu = 0 .$$  

(26.30)
The proof relies on antisymmetry of $F^{\mu\nu}$ under exchange of the indices

$$F^{\mu\nu} = -F^{\nu\mu}, \quad \text{e.g., } F^{01} = -F^{10}, \quad (26.31)$$

which is self evident seen the definition in Eq. (26.29). On the other hand, the product of two 4-velocities is symmetric in the two indices, $u_\mu u_\nu = u_\nu u_\mu$. Therefore we obtain

$$u_\mu F^{\mu\nu} u_\nu = -u_\mu F^{\nu\mu} u_\nu = -u_\nu F^{\mu\nu} u_\mu = -u_\mu F^{\mu\nu} u_\nu = 0, \quad (26.32)$$

where we for clarity indicated explicit the sums and we have used: first the antisymmetry of the field tensor; after that we renamed indices; and in the last step we changed the sequence of commuting variables; see exercise X–7 below for more related comments. Since we now have shown that the right hand side in Eq. (26.30) is negative of itself, it must be zero.

We have now obtained the general constraint Eq. (26.30) on the motion of charged particles subject to Lorentz-force Eq. (26.26). This leads further to

$$0 = m u_\mu \frac{d m u^\mu}{d\tau} = \frac{1}{2} \frac{d (m^2 u^2)}{d\tau}, \quad m^2 u^2 = \text{Const.} = m^2 c^2, \quad (26.33)$$

where the value of the integration constant is recognized considering a particle at rest and remote from all force fields. $u^2 = \text{Const.} = c^2$ follows here from $u \cdot F_L = 0$. This result includes the consideration of dynamics where the mass is modified $m \rightarrow m(x)$.

We now check if the covariant form of writing the Lorentz-force agrees with the usual format, and this inquiry addresses spatial components. We obtain for $\mu = i = 1, 2, 3$, and using now $i$ as component of the 3-vector

$$\frac{d t}{d\tau} \frac{d (m u_i)}{d t} = \gamma \frac{d (m \gamma v_i)}{d t} = e(F^{0i} \gamma c - F^{ij} \gamma v_j) = e\gamma \left( \mathbf{E}_i + \sum_{j,k=1}^{3} \epsilon_{ijk} v_j B_k \right). \quad (26.34)$$

We cancel the common factor $\gamma$ and recall Eq. (26.6a) to recognize in the last equality the Lorentz-force in the usual form Eq. (23.1).

**Exercise X–7: Constraint on Lorentz 4-force with fields I**

Show that from Eq. (26.29) follows $u_\mu K^\mu_L = 0$.

**Solution**

We first briefly review some simple matrix algebra properties. A real $N_{ij}$ $4 \times 4$ matrix has 16 components and can be decomposed into symmetric and antisymmetric parts

$$N_{ij} = \frac{N_{ij} + N_{ji}}{2} + \frac{N_{ij} - N_{ji}}{2} = S_{ij} + A_{S\,ij}. \quad (26.34)$$
The antisymmetric part
\[ 2 \frac{N_{ij} - N_{ji}}{2} = A_{ij} = -A_{ji}, \]
has as we see in the definition no diagonal elements \( i = j \) and among 16-4=12 off-diagonal elements constrained by the antisymmetry condition Eq.2 there are only 6 independent components. The symmetric part
\[ 3 \frac{N_{ij} + N_{ji}}{2} = S_{ij} = S_{ji}, \]
has 4 diagonal elements \( i = j \) and also 6 off-diagonal independent components for a total of 10 components.

A product of a symmetric matrix \( S \) with an antisymmetric matrix \( A_S \) is always vanishing
\[ 4 \sum_j S_{ij} A_{S jk} = \sum_j (-A_{S kj}) S_{ji} = 0, \]
since zero is the only number that is a negative of itself. Finally, a product of two 4-tuplets always forms a symmetric matrix
\[ 5 a_i a_j = a_j a_i \equiv S_{ij}. \]

We now recognize the matrix \( F^{\mu\nu} \) in Eq. (26.29) as the antisymmetric part of the matrix \( \partial^\mu A^\nu \).

We check the consistency of Eq. (26.29) proceeding as we did in exercise X–1
\[ 6 u_\mu K^\mu_L = u \cdot K_L = 0, \quad \rightarrow \quad u \cdot K_L = e u_\mu u_\nu F^{\mu\nu} = 0, \]
since the first factor \( u_\mu u_\nu \) is symmetric in both indices, while the second factor is antisymmetric, since the matrix \( F^{\mu\nu} \) in Eq. (26.29) is the antisymmetric part of the matrix \( \partial^\mu A^\nu \).
To show that the Lorentz-force is spacelike we can choose any convenient frame of reference since the magnitude of a 4-vector is invariant under LT. The natural frame of reference is the rest-frame, thus we aim to show

$$K_{L} \cdot K_{L} = K_{L}(0) \cdot K_{L}(0), \quad K^{\mu}_{(0)} = eF^{\mu 0},$$

where we used $u_{(0)} = \{c, \vec{0}\}$. Noting Eq. (26.7) we thus have

$$K^{\mu}_{(0)} = \mathcal{E}_{(0)} = \{0, -e\vec{E}_{(0)}\}.$$

where we alert that Eq. 2 uses the electric field in the rest frame of the particle and not rest frame of the laboratory. Using Eq. 2 it is evident that

$$K_{L} \cdot K_{L}(0) = -0^2 - [-e\vec{E}_{(0)}]^2 = -e^2 \vec{E}_{(0)}^2 < 0.$$

End X–8: Constraint on Lorentz 4-force with fields II

Exercise X–9: Lorentz-force in matrix format

Present the covariant Lorentz-force in matrix format developed in Part VIII

Solution

In order to be able to multiply the matrix $F$ Eq. (26.7) with a 4-velocity vector matrix we assume a single column $4 \times 1$ matrix $u$ as a representation of $u^{\mu}$ and a $4 \times 4$ matrix as representation of $g_{\mu\nu}$

$$u = (u^{\mu}) = \mu \begin{pmatrix} u^0 \\ u^1 \\ u^2 \\ u^3 \end{pmatrix}, \quad g = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

To lower the index of the 4-velocity we need to include the metric matrix $g$

$$u_{\nu} = g_{\nu\mu}u^{\mu}, \quad \Rightarrow (u_{\nu}) = (gu)_{\nu}, \quad (\tilde{u}_{\nu}) = (\tilde{ug})_{\nu}$$

since for diagonal $g$ its transpose is the same $\tilde{g} = g$; $\tilde{u}$ is a single row $4 \times 1$ transposed matrix of $u$. This allows us to write for

$$u^{\mu}g_{\mu\nu}u^{\nu} = c^2, \quad \Rightarrow \text{Tr} ugu = c^2, \quad \text{or} \quad \tilde{ug}u = c^2.$$
With Eq. (26.7) we have established the precise meaning of the first order in proper
time dynamical equation Eq. (26.29) which we now write in matrix form as follows

\[ \left( \frac{du}{d\tau} \right)_\mu = \frac{e}{m} (Fgu)_\mu , \quad \Rightarrow \quad \frac{du}{d\tau} = \frac{e}{m} Fgu , \]

where we also wrote the key result in more concise format, where

\[ F_g = \begin{pmatrix}
  0 & -E^1/c & -E^2/c & -E^3/c \\
  E^1/c & 0 & -B^3 & B^2 \\
  E^2/c & B^3 & 0 & -B^1 \\
  E^3/c & -B^2 & B^1 & 0 \\
\end{pmatrix} \]

Since the transpose \( \tilde{F} = -F \) due to antisymmetry of \( F_{\mu\nu} \) we also have

\[ \frac{d\tilde{u}}{d\tau} = -\frac{e}{m} \tilde{u}gF . \]

This allows us to recognize

\[ \frac{dc^2}{d\tau} = \frac{d(\tilde{u}gu)}{d\tau} = \frac{d\tilde{u}}{d\tau} gu + \tilde{u}g \frac{du}{d\tau} = -\frac{e}{m} \tilde{u}gFgu + \frac{e}{m} \tilde{u}gFgu = 0 , \]

which shows in matrix notation how the usual antisymmetry property of \( F_{\mu\nu} \) leads to
\[ u^2 = c^2 . \]

End X–9: Lorentz-force in matrix format

26.5 The Poynting force

The covariant Lorentz-force written in form of an integral equation using
matrix form reads

\[ u(\tau) = u(0) + \frac{e}{m} \int_0^\tau d\tau' Fgu(\tau') . \tag{26.35} \]

When Eq. (26.35) is differentiated with respect to \( \tau \) we obtain the form presented
in Eq. 4 in exercise X–9 on page 392. Note that in Eq. (26.35) the past determines
the present; thus this is the causal solution, and for \( \tau \to 0 \) the left and right hand
side agree, thus the initial condition is implemented correctly.

We iterate Eq. (26.35) by inserting the form of \( u(\tau') \) as defined by the right
hand side of Eq. (26.35) into the last term in Eq. (26.35) to obtain

\[ u(\tau) = u(0) + \frac{e}{m} \int_0^\tau d\tau_1 Fgu(0) + \left( \frac{e}{m} \right)^2 \int_0^\tau d\tau_1 (Fg) \int_0^{\tau_1} d\tau_2 (Fgu)_2 . \tag{26.36} \]
where the subscript stands for dependence on respectively $\tau_1$ and $\tau^2$. Note that $\tau \geq \tau_1 \geq \tau^2 \geq 0$.

Such time-ordered series we see emerging from iteration of Eq. (26.35) in Eq. (26.36) are familiar to many. One can continue the process, retaining exactness by iterating and building-up a series solution with $\tau \geq \tau_1 \geq \tau^2 \geq \ldots \geq \tau_N \geq 0$. After $N$ steps the proper time $\tau_N$ has effectively a small enough range to assume that one can in appropriate situations terminate the series by inserting on right the unchanged $u(\tau_N) \to u(0)$. When the field is constant the iteration process produces a power series in $\tau$ as well as in powers of the field strength.

We now seek to identify the 4-force that is related to particle trajectory that Eq. (26.36) describes. We thus differentiate with respect to $\tau$ and obtain

$$a(\tau) = \frac{du}{d\tau} = \frac{e}{m} (Fg) u(0) + \left( \frac{e}{m} \right)^2 (Fg) \tau \int_0^\tau d\tau_1 (Fgu)_1 .$$  \hspace{1cm} (26.37)

The first term is just like the Lorentz-force but for the fact that the particle 4-velocity $u \to u(0)$ is not changing. As the particle accelerates under the influence of the Lorentz-force we see this effect accounted for in the second term in Eq. (26.37).

We now make the approximation that the EM-field is sufficiently constant so that through some short period in proper time we can ignore the variation of the field along the path of the particle. Under this assumption the last term in Eq. (26.37) can be integrated and we find, without additional approximation a remarkably simple result

$$a(\tau) = \frac{e}{m} (Fg) u(0) + \left( \frac{e}{m} \right)^2 (FgFg) \Delta x , \quad \Delta x = x(\tau) - x(0) ,$$  \hspace{1cm} (26.38)

where $a, u, x$ are one column 4 \times 1 matrices as presented in Eq. 1 in exercise X–9.

We now evaluate the matrix defining this secondary force using results of Eq. 5 in exercise X–9

$$FgFg = \begin{pmatrix} 0 & \mathcal{E}^1/c & \mathcal{E}^2/c & \mathcal{E}^3/c \\ \mathcal{E}^1/c & 0 & B^3 & -B^2 \\ \mathcal{E}^2/c & -B^3 & 0 & B^1 \\ \mathcal{E}^3/c & B^2 & -B^1 & 0 \end{pmatrix} \begin{pmatrix} 0 & \mathcal{E}^1/c & \mathcal{E}^2/c & \mathcal{E}^3/c \\ \mathcal{E}^1/c & 0 & B^3 & -B^2 \\ \mathcal{E}^2/c & -B^3 & 0 & B^1 \\ \mathcal{E}^3/c & B^2 & -B^1 & 0 \end{pmatrix} \begin{pmatrix} \mathcal{E}^2/c^2 & -(\mathcal{E}^2 B^3 - \mathcal{E}^3 B^2)/c & -(\mathcal{E}^3 B^1 - \mathcal{E}^1 B^3)/c & -(\mathcal{E}^1 B^2 - \mathcal{E}^2 B^1)/c \\ (\mathcal{E}^2 B^3 - \mathcal{E}^3 B^2)/c & \mathcal{E}^1 2/c - B^2 - B^3/2 & \mathcal{E}^1 \mathcal{E}^2/c + B^1 B^2 & \mathcal{E}^3 \mathcal{E}^3/c + B^1 B^3 \\ (\mathcal{E}^2 B^3 - \mathcal{E}^3 B^2)/c & \mathcal{E}^1 \mathcal{E}^2/c + B^1 B^2 & \mathcal{E}^2 2/c - B^1 2 - B^3 2 & \mathcal{E}^2 \mathcal{E}^2/c + B^2 B^2 \\ (\mathcal{E}^1 B^2 - \mathcal{E}^2 B^1)/c & \mathcal{E}^1 \mathcal{E}^3/c + B^1 B^3 & \mathcal{E}^2 \mathcal{E}^3/c + B^2 B^3 & \mathcal{E}^3 2/c - B^2 2 - B^1 2 \end{pmatrix} .$$ \hspace{1cm} (26.39)
We note the antisymmetry of \('i0' and '0i' elements
\[
(FgFg)^{0i} = \frac{1}{c} (\vec{E} \times \vec{B})_i = -(FgFg)^{0i}, \quad i = 1, 2, 3,
\]
while the remaining 3 \( \times \) 3 matrix elements are symmetric.

We now look at the acceleration in the rest frame of the particle \( u^\mu = \{c, \vec{0}\} \) and we ignore terms that are initially small given that spatial displacement \( |\Delta \vec{x}| \ll c\Delta t \)
\[
ma^0 = c\Delta t \frac{e^2}{mc^2} \left[ \vec{E}^2 - (\vec{E} \times c\vec{B}) \cdot \frac{\Delta \vec{x}}{c\Delta t} \right],
\]
\[
ma = e\vec{E} + c\Delta t \frac{e^2}{mc^2} \left[ (\vec{E} \times c\vec{B}) + \mathcal{O}\left( \frac{\Delta \vec{x}}{c\Delta t} \right) \right].
\]

We have learned that in its rest frame a particle experience a 3-force that aside of the usual electrical field term also includes a ‘Poynting’ term that acts in direction of
\[
\vec{S} = \epsilon_0 \vec{E} \times c\vec{B}.
\]
This Poynting force can dominate the usual Lorentz-force when the intensity of the field as defined in exercise IX–9 on page 353 is sufficiently large. Direct acceleration of a particle in direction of \( \vec{S} \) can be achieved with present day high intensity lasers. This corroborates the study of particle acceleration by light plane waves which we found to carry the particle predominantly along the Poynting vector direction, see exercise 24.3 on page 354.

---

**Exercise X–10: Acceleration of short-lived particle**

Evaluate as a function of proper time \( \tau \) the motion of a particle accelerated in a constant EM-field.

**Solution**

Considering that a short-lived particle has a mean lifespan \( \tau_0 \) we seek to determine particle motion as a function of particle proper time. Given the near constancy of applied EM-fields, and the limitation to a small value of \( \tau \) it is appropriate to terminate the iteration loop in Eq. (26.36) by setting in the last term \( u(\tau^2) \rightarrow u(0) \)

1. \( u(\tau) = u(0) + \frac{e\tau}{mc} cFg u(0) + \frac{1}{2} \left( \frac{e\tau}{mc} \right)^2 cFg cFg u(0) + \mathcal{O}(\tau^3) \).

For a particle initially at rest, \( \vec{u}(0) = \{c, \vec{0}\} \) we obtain for the ‘0’ component

2. \( u^0 = c \frac{dt}{d\tau} = c\gamma(\tau) = c + \frac{e\tau}{mc} (cFg)^{00} + \frac{1}{2} \left( \frac{e\tau}{mc} \right)^2 (cFg cFg)^{00}. \)
Due to antisymmetry of $F$ the second term on right vanishes. Thus using Eq. (26.39)

$$
\frac{dt}{d\tau} \equiv \gamma(\tau) = 1 + \frac{1}{2} \left( \frac{e|\vec{E}|\tau}{mc} \right)^2.
$$

To achieve highly relativistic motion and practical utility we must assure that the argument in parenthesis is significantly greater than unity. This is normally the case since the proper distance $c\tau$ which could be fixed by particle lifespan or by duration of acceleration period can be very large on the scale of other quantities of interest.

Turning now to spatial components, we find for the component ‘$i$’

$$
\frac{dx^i}{d\tau} = \gamma(\tau) v_i(t) = \frac{e\tau}{mc} (cFg)^0 + \frac{1}{2} \left( \frac{e\tau}{mc} \right)^2 (cFg cFg)^0.
$$

The value $(cFg)^0 = \mathcal{E}^1$ is easily read off Eq. (26.7) while the non-linear terms are shown in Eq. (26.40). Thus

$$
\frac{\vec{u}}{c} = \frac{d\vec{x}}{d\tau} = \gamma \frac{d\vec{x}}{dt} = \frac{e\vec{E}}{mc} + \frac{\tau^2}{2} \frac{e\vec{E} \times c\vec{B}}{(mc)^2}.
$$

A notable feature of our solution is the appearance of motion along the Poynting vector $\vec{S} = \epsilon_0 \vec{E} \times c\vec{B}$, which appears and dominates the motion for field intensity large as measured in units of particle mass.

We now check the level of consistency of our solution by forming

$$
u \cdot u = c^2 + \frac{e^2}{4} \left( \frac{\tau e}{mc} \right)^4 \left[ |\vec{E}|^2 - \left( \frac{e\vec{E} \times c\vec{B}}{|\vec{E}|^2} \right)^2 \right];
$$

we recognize the last term $O(\tau^4)$ to be incomplete, since the next term in expansion of $u^\mu$ at order $O(\tau^3)$ when combined with the first term $O(\tau)$ would contribute to this term. We conclude that at the level of precision our solution satisfies $u \cdot u = c^2$ — further terms in powers of $\tau^3$ need to be obtained to assure that through order $\tau^4$ the 4-velocity satisfies relativity constraint.

We can integrate the 4-velocity $u$, Eq. (2) and Eq. 4, with respect to $\tau$ in order to obtain the location of the particle as a function of particle proper time:

$$
x^0 = ct = c\tau + \frac{ct}{6} \left( \frac{e|\vec{E}|\tau}{mc} \right)^2,
$$

$$
\vec{x} = \frac{ct e\vec{E} \tau}{2 mc} + \frac{ct}{6} \left( \frac{e|\vec{E}|\tau}{mc} \right)^2 \frac{\vec{E} \times c\vec{B}}{|\vec{E}|^2}.
$$

This allows us to characterize the distance traveled by unstable particle in its lifespan $\tau$ under influence of EM-fields. The scheme here presented can be extended to obtain solutions in higher powers of $\tau$. 


We note that Eq. (7) provides the relation between laboratory time $ct = x^0$ and proper time $c\tau$. Applying Cardan’s formula we obtain the sole real solution

$$
c\tau = \left( \sqrt{\left( \frac{3m^2c^4}{|e\vec{E}|^2} \right)^2 (ct)^2 + \left( \frac{2m^2c^4}{|e\vec{E}|^2} \right)^3 + \frac{3m^2c^4}{|e\vec{E}|^2} (ct) } \right)^{1/3}
$$

Thus using Eq. (8) we can obtain from Eq. (7) the laboratory position $\vec{x}$ explicitly as a function of laboratory time $ct$.

We see that through order $\tau^4$ the location of the particle originally in coordinate origin at zero velocity remains for all times within the future light cone,

$$
x^2 = x^0^2 - \vec{x}^2 = (c\tau)^2 + \frac{(c\tau)^2}{12} \left( \frac{e|\vec{E}|\tau}{mc} \right)^2 > 0.
$$

End X–10: Acceleration of short-lived particle
Part XI

Dynamics of Fields and Particles
The Lorentz-covariant EM action leads to the covariant format of dynamical Maxwell field equations with sources. We complement these four equations with another four covariant equations that allow us to relate fields to potentials once a gauge is chosen. We describe the Lorentz covariant gauge condition and obtain covariant 4-vector format for the EM potential.

We study the energy-mass equivalent locked in the EM-field distribution. Based on the energy-momentum tensor we introduce the dynamical equations that describe the combined dynamics of EM-fields and particles. This formulation shows that the energy locked in the particle generated field is a part of particle inertial mass. We demonstrate the reality of delocalized EM part of inertial mass considering EM-field of atomic nuclei, a component of relevance in the study of binding of heavy nuclei.

For the lightest charged elementary particle, the electron, the classical EM-mass component vastly dominates the inertial mass. To resolve this problem we call on quantum delocalization of the classical point-particle. We show how a charged particle alters the energy-mass equivalent locked in the field of the other particle — that is action at a distance potential energy. We introduce the nonlinear Born-Infeld (BI) modification of Maxwell theory with a finite point-particle EM-field energy. In BI-EM there is a limit to the strength of electrical field, and thus a limit to the strength of the force that we can impart on a particle: there is maximum acceleration.

Acceleration is a concept that reaches beyond the ideas of SR. How are we able to tell if the body or the observer is accelerated or inertial? How do we know that there is acceleration at all? Without acceleration that takes bodies from one inertial frame of reference to another many arguments in this book would make little sense. We argue that acceleration is defined locally and that it manifests itself by emission of radiation; accelerated particles experience radiation-friction.

We describe how the emission of EM radiation and the associated loss in energy by accelerated bodies generate for non-inertial particles radiation-friction. Radiation vacuum-friction requires acceleration, in that it differs from regular matter-friction. We describe the proposed improvements to the Lorentz-force that take account of the effect of this vacuum friction effect. The ad-hoc descriptions of radiation reaction and vacuum friction mean that a consistent theory of electromagnetism in the regime of strong acceleration awaits discovery.

27 Variational Principle for EM-fields

27.1 Maxwell equations with sources

It is generally believed that just as the Lorentz-force follows from the principle of least action, the shape of the EM-fields generated by charged particles follows demanding that a field action assumes a minimum value. The field action must sum over all field-penetrated space-time and thus it will be a 4-volume integral. The form of the field action

\[ I_F = \int d^4x \mathcal{L}_{\text{field}}, \quad d^4x \equiv dt d^3x = \frac{1}{c} dx^0 dx^3 = \frac{1}{c} d^4\tilde{x} \]  

(27.1)

is further strongly constrained by the requirement that it should be gauge invariant, meaning that it is expected to have the format of a functional of fields and not potentials. The action \( I_F \) must be an invariant under Lorentz coordinate transformations – according to exercise [VIII–3] the 4-volume element is LI. This means that the field action density \( \mathcal{L}_{\text{field}} \) over which we integrate must itself be invariant; it must be constructed from the invariants of the EM-field, \( S \) and \( P \) Eq. (26.17), and Eq. (26.18).

Since the components in the action that we obtained in the study of the Lorentz-force have units of action (energy \( \times \) time, or, momentum \( \times \) length), see Eq. (25.18) and Eq. (25.20a), we expect this to be the case for the field action as well. Therefore, the action density, the Lagrangian \( \mathcal{L} \), must be, by dimensional considerations, bilinear in the EM-fields. In the absence of new dimensioned constants we can only consider linear powers of invariants \( S \) and \( P \).

We demand that the action is invariant under coordinate transformations that reflect the direction of spatial axis (parity), we eliminate linear terms in \( P \) and the outcome is one and only possible form of the field action density

\[ \mathcal{L}_{\text{field}} = S , \]  

(27.2)

where no proportionality factor is needed to assure the dimension of \( I_F \) given how we normalized \( S \) in Eq. (26.12). Note that we defined \( S \propto F_{\mu\nu}F^{\mu\nu} \) which absorbs a sign that sometimes appears if one writes \( S \propto F_{\mu\nu}F^{\mu\nu} \). Also, when using \( d^4\tilde{x} \), see Eq. (27.1) an extra factor \( 1/c \) should be included to assure that \( I_F \) has correct dimension. This factor appears below whenever we need to use \( d^4\tilde{x} \) instead of \( d^4x \).

Since the field equations that follow by variation reduce the power in which fields enter, the field equations will be linear in the fields and thus we also have by the choice of Eq. (27.2) implemented the experimentally known superposition principle, that the field of two sources will be the sum of fields generated by each source. Conversely, breaking the superposition principle requires nonlinearity in
field invariants $\mathcal{S}$ and $\mathcal{P}$ which requires new dimensioned constants in the action, a theoretical option we will discuss below in section 28.3.

To recapitulate, the requirement of LI, parity invariance, combined with the superposition principle of EM-fields constrains the form of action $\mathcal{L}_{\text{field}}$ sufficiently to determine a unique form. The variation of the action will be performed with respect to the 4-potential $A^\mu$. This means that the choice of gauge in choosing the potential that corresponds to a field, see section 23.2, is a variational constraint.

The dynamics of EM-fields combine with the dynamics of material bodies in the action

$$ I = \frac{\varepsilon_0 c^2}{4} \int_{\text{volume}} d^4x \, F_{\mu\nu} F^{\mu\nu} - e \int_{\text{path}} d\tau \frac{dx}{d\tau} \cdot A - \int_{\text{path}} d\tau \, L_{\text{Iner}} + G_F. \tag{27.3} $$

where we see in sequence from left to right the proposed field term Eq. (27.2), the field-matter interaction, see Eq. (25.18), the matter inertial term, see discussion in section 25.2, and the gauge fixing term which is only needed when we compute in a particular gauge the form of the potential $A^\mu$. In the first term we used $d^4\tilde{x} = (c^2/c)d^4x$ with factors $c$ written in this manner for future convenience.

As now written the action Eq. (27.3) comprises just one particle as the source of the field. In consideration of superposition principle of fields it is common to extend the field-matter interaction term by summing over many charged particles that source the field. In the first step we convert the path integral using the Dirac $\delta$-function into a volume integral and in the second we extend to a large number of charged particles ‘i’ seen evolving in each particle’s proper time along each particle individual path

$$ -e \int_{\text{path}} d\tau \frac{dx}{d\tau} \cdot A(x) = - \int_{\text{vol}} d^4x \, A_\mu(x) \int_{\text{path}} d\tau \, q \frac{dz^\mu}{d\tau} \delta^4(x - z(\tau)) $$

$$ \rightarrow - \int_{\text{vol}} d^4x \, A_\mu(x) \sum_i \int_{\text{path}_i} d\tau_i \, q_i \frac{dz^\mu_i}{d\tau_i} \delta^4(x - z_i(\tau)) \tag{27.4} $$

$$ \rightarrow - \int_{\text{vol}} d^4x \, A(x) \cdot J(x). $$

In the final form we introduced an average of all individual charged sources, a space-time 4-current-density distribution

$$ J^\mu(x) = \{c\rho(x), \vec{j}(x)\}. \tag{27.5} $$

The 4-current is implicitly defined in Eq. (27.4) to be

$$ J^\mu \equiv \sum_i \int_{\text{path}_i} d\tau_i \, q_i \frac{dz^\mu_i}{d\tau_i} \delta^4(x - z_i(\tau)) \tag{27.6} $$

Since each of the component particles in Eq. (27.6) moves under the influence of the field generated by all other charged particles, the distribution $J^\mu(x)$ must
itself be derived from a dynamical equation, where the EM-fields shape $J^\mu(x)$. The equation that leads to this self-consistent generalization of how EM-fields interact with the 4-current is the Vlasov transport equation normally used in the plasma physics context; further discussion is beyond the scope of this book.

We will proceed assuming that the distribution of the field source $J^\mu(x)$ is prescribed. In order to describe the local motion of a group of particles in a ‘cloud’ with particle distribution $\rho$ we note that at each location $x$ we can find a local frame of reference ‘(0, x)’ in which the 3-vector $\vec{j}(x)$ of the 4-current in Eq. (27.6) averages out. This is the local rest-frame where

$$J^\mu(x)|_{(0,x)} = q\{c\rho(0,x)(x), \vec{0}\} .$$

Upon back-Lorentz-transformation into a more general frame of reference which is typically the frame in which the entire cloud of particles is on average at rest, we find that the local 4-current can be written as

$$J^\mu = q\{c\rho(x), \vec{u}(x)\rho(x)\} ,$$

where both $\rho(x)$ and $\vec{u}(x)$ are the local density and velocity distributions within the cloud of particles.

We now perform variation with respect to $A^\mu$ of the field action and source action Eq. (27.4), see exercise [XI–1], and obtain the Maxwell equations with sources in covariant form

$$\frac{\partial F^{\mu\nu}}{\partial x^\mu} - \frac{1}{c^2\epsilon_0} J^\nu = 0 .$$

The Maxwell equations can be written in non-covariant notation. For $\nu = 0$ we have $F^{\nu 0} = \mathcal{E}^i/c$ and thus

$$\vec{\nabla} \cdot \mathcal{E} = \frac{1}{\epsilon_0} \rho , \quad \vec{\nabla} \cdot \mathcal{D} = \rho ,$$

For each spatial component $\nu = i, i = 1, 2, 3$ we find accordingly

$$-\frac{1}{c^2} \frac{\partial \mathcal{E}}{\partial t} + \vec{\nabla} \times \vec{B} = \frac{1}{c^2\epsilon_0} \vec{j} , \quad -\frac{\partial \mathcal{D}}{\partial t} + \vec{\nabla} \times \vec{H} = \vec{j}$$

where in Eq. (27.10) and Eq. (27.11) we used on the right hand side Eq. (23.3).

We now differentiate Eq. (27.9) and obtain due to antisymmetry of $F^{\mu\nu}$ and symmetry of the derivatives, $\partial_\mu \partial_\nu = \partial_\nu \partial_\mu$ (which requires absence of singularities)

$$\frac{\partial^2 F^{\mu\nu}}{\partial x^\nu \partial x^\mu} = 0 = \frac{\partial J^\mu}{\partial x^\mu} .$$
One refers to Eq. (27.12) as current conservation law. The non-covariant form of Eq. (27.12) reads, according to our definition Eq. (27.5) and recalling Eq. (25.5),

\[ 0 = \frac{\partial \rho}{\partial t} + \nabla \cdot \vec{j}. \]  

(27.13)

The 4-current-density \( J^\mu(x) \) defined in Eq. (27.5) is thus constrained by current conservation as follows

\[
\frac{\partial J^\mu}{\partial x^\mu} = - \sum_i \int_{\text{path}_i} d\tau_i \, q_i \frac{dz_i}{d\tau_i} \frac{\partial \delta^4(x-z_i(\tau_i))}{\partial x^\mu} = \sum_i \int_{\text{path}_i} d\tau_i \, q_i \frac{dz_i}{d\tau_i} \frac{\partial \delta^4(x-z_i(\tau_i))}{d\tau_i} = \sum_i q_i \delta^4(x-z_i(\tau_i)) \bigg|_{\text{path}_i} = 0.
\]

(27.14)

The boundary \( \delta \)-function assures balance of the flow-in of charges against the flow-out. Thus Eq. (27.12) is true only if all charges traveling along a path do not disappear, thus when charge is conserved. We referred here to charges rather than particles as the broader statement we made includes the possibility that particles undergo reactions converting one into the other in which charge is conserved.

Exercise XI–1: Variation with respect to \( A^\mu \)

Derive the dynamical covariant EM-field equation Eq. (27.9) by seeking extrema of action functional as a function of 4-potential \( A^\mu \).

Solution

For \( \mathcal{L}(A^\alpha, \partial^\beta A^\alpha, x_\mu) \), the potential is varied as \( A^\alpha \rightarrow \tilde{A}^\alpha = A^\alpha + \epsilon \eta^\alpha(x_\mu) \), where \( \eta \) is an arbitrary function that vanishes at the boundary of the space-time domain of interest. \( \epsilon \) is an infinitesimal parameter for variation. We find the change in \( \mathcal{I} \) differentiating with respect to \( \epsilon \) and keeping lowest order terms

\[
\frac{d\mathcal{I}[\tilde{A}]}{d\epsilon} = \int d^4x \left( \frac{\partial \mathcal{L}}{\partial A_\nu} \frac{\partial \tilde{A}_\nu}{\partial \epsilon} + \frac{\partial \mathcal{L}}{\partial (\partial_\mu A_\nu)} \frac{\partial (\partial_\mu \tilde{A}_\nu)}{\partial \epsilon} \right)
\]

\[
= \int d^4x \left( \frac{\partial \mathcal{L}}{\partial A_\nu} \frac{\partial \tilde{A}_\nu}{\partial \epsilon} + \frac{\partial \mathcal{L}}{\partial (\partial_\mu A_\nu)} \frac{\partial \tilde{A}_\nu}{\partial \epsilon} \right) - \frac{1}{2} \int d^4x \left[ \frac{\partial \mathcal{L}}{\partial A_\nu} - \partial_\mu \frac{\partial \mathcal{L}}{\partial (\partial_\mu A_\nu)} \right] \frac{\partial \tilde{A}_\nu}{\partial \epsilon} + \partial_\mu \left[ \frac{\partial \mathcal{L}}{\partial (\partial_\mu A_\nu)} \frac{\partial \tilde{A}_\nu}{\partial \epsilon} \right] \right)
\]

\[
= \int d^4x \left[ \frac{\partial \mathcal{L}}{\partial A_\nu} \eta_\nu(x_\mu) - \partial_\mu \frac{\partial \mathcal{L}}{\partial (\partial_\mu A_\nu)} \right] \eta_\nu(x_\mu) = 0,
\]

where the last equality is true if in the third line the last integral, a surface term, vanishes. This is true by assumption since \( \partial \tilde{A}_\nu / \partial \epsilon = \eta_\nu \) and the variation function \( \eta_\nu \) vanishes at the 3-surface of the 4-integral over the 4-divergence.
Other than the surface constraint, \( \eta(x) \) is an arbitrary function within the domain of \( x \) and thus the square bracket in the last line in Eq. (1), the Euler-Lagrange equations, can be pulled out of the integral. For

\[
\mathcal{L} = \frac{c^2 \epsilon_0}{4} F_{\alpha \beta} F^{\beta \alpha} - A_\alpha J^\alpha,
\]

we obtain

\[
\frac{\partial \mathcal{L}}{\partial A_\nu} = -J^\nu
\]

\[
- \frac{1}{c^2 \epsilon_0} \frac{\partial \mathcal{L}}{\partial (\partial_\mu A_\nu)} = \frac{1}{4} \frac{\partial}{\partial (\partial_\mu A_\nu)} F^{\alpha \beta} F_{\alpha \beta} = \frac{1}{4} \frac{\partial}{\partial (\partial_\mu A_\nu)} F_{\alpha \beta} F^{\alpha \beta}
\]

\[
= \frac{1}{2} F^{\alpha \beta} \left( \frac{\partial}{\partial (\partial_\mu A_\nu)} \partial_\alpha A_\beta - \frac{\partial}{\partial (\partial_\mu A_\nu)} \partial_\beta A_\alpha \right)
\]

\[
= \frac{1}{2} F^{\alpha \beta} \frac{\partial}{\partial (\partial_\mu A_\nu)} \partial_\alpha A_\beta + \frac{1}{2} F^{\beta \alpha} \frac{\partial}{\partial (\partial_\mu A_\nu)} \partial_\beta A_\alpha
\]

\[
= \frac{1}{2} F^{\alpha \beta} \delta_{\alpha \mu} \delta_{\beta \nu} + \frac{1}{2} F^{\beta \alpha} \delta_{\alpha \nu} \delta_{\beta \mu} = F^{\mu \nu}.
\]

Thus multiplying by \( c^2 \epsilon_0 \),

\[
\frac{\partial \mathcal{L}}{\partial A_\nu} - \partial_\mu \frac{\partial \mathcal{L}}{\partial (\partial_\mu A_\nu)} = -J^\nu + c^2 \epsilon_0 \frac{\partial F^{\mu \nu}}{\partial \mu x} = 0,
\]

in agreement with the dynamical covariant EM-field equation Eq. (27.9).

End XI–1: Variation with respect to \( A^\mu \)

---

Exercise XI–2: Coulomb solution and the coupling constant

When considering the relativistic electron orbit we introduced the force between the nucleus and the electron in exercise IX–5 on page 337, Eq. 2. Obtain this force using Maxwell equations.

Solution

For a point nucleus of charge \( Z \) protons of charge \( e_p \) we are solving Eq. (27.10) for the ‘displacement field’ \( \vec{D} \)

\[
\vec{\nabla} \vec{D} = Z e_p \delta^3(x).
\]
In view of the spherical symmetry of the nuclear charge we are seeking a radial solution. Integrating Eq. (1) over the volume and converting to surface integral using Gauss theorem on a sphere of radius \(|x|\) one immediately finds the radial Coulomb displacement field \(D_r\):

\[
2 \int_{|x|} \delta^3 x \nabla \cdot \vec{D} = \int_{|x|} \delta^2 S \vec{n} \cdot \vec{D} = 4\pi |x|^2 D_r = Ze_p, \quad \rightarrow \quad D_r = \frac{e_p Z}{4\pi |x|^2}.
\]

We now recall the relation with the electric field \(E_r\):

\[
3 \quad E_r = \frac{1}{\epsilon_0} D_r = \frac{e_p Z}{4\pi \epsilon_0 |x|^2}.
\]

The Coulomb-force that acts between electron of charge \(e_e\) and the \(Z\) proton charges \(e_p\) is accordingly:

\[
4 \quad F = eE_r = \frac{e_e e_p Z}{4\pi \epsilon_0 |x|^2} = -\frac{e^2}{4\pi \epsilon_0} \frac{Z}{|x|^2},
\]

where the sign is due to the opposite charge of the proton compared to the electron.

There are two specific to SI-units features visible in the ‘coupling constant’, that is the elementary coefficient characterizing the force strength, \(e^2/4\pi \epsilon_0\). Since there is no \(4\pi\) in Eq. (1), we see it in the solution Eq. (4). The solution of Maxwell equations with sources produces the field \(D\) and thus to obtain the force field \(E\) we see in the Lorentz-force there appears the SI-unit vacuum permittivity \(\epsilon_0\). In the cgs-Gauss system of units both these factors are absent: \(4\pi\) is included in Eq. (1) and \(\epsilon_0 = 1\) in vacuum. Since the force has the same strength irrespective of how we define the system of units, the last rule in Eq. (23.2) follows.

---

Exercise XI–3: Light wave solutions of Maxwell equations

When considering an electron surfing on a light plane wave we assumed a form of the wave in section 24.1 on page 345; show that this wave is a solution of Maxwell equations.

Solution

The EM-fields are given according to Eq. (24.5) in the here appropriate format

\[
1 \quad \vec{E} = \Im \left[ e^{i\chi} \omega \vec{A}_0 \right], \quad \vec{B} = \Im \left[ e^{i\chi} \vec{k} \times \vec{A}_0 \right], \quad \chi = \vec{k} \cdot \vec{r} - \omega t.
\]

\(^1\)Charles-Augustin de Coulomb (1736-1806) engineer by training, and physicist by vocation. He developed methods to measure magnetic and electric forces showing before 1791 that the electrostatic forces obey Newton’s inverse-square law.
These solutions are constructed to satisfy naturally

\[ \mathbf{\nabla} \cdot \mathbf{B} = 0 , \quad \mathbf{\nabla} \cdot \mathbf{E} = \frac{1}{\epsilon_0} \rho = 0 , \]

and

\[ \mathbf{\nabla} \cdot \mathbf{E} = 0 , \quad \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 , \]

in consideration of the transverse gauge \( \mathbf{k} \cdot \mathbf{A}_0 = 0 \).

We need to assure that the two vectorial Maxwell equations are satisfied

\[ \mathbf{\nabla} \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 , \]

and

\[ - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} + \mathbf{\nabla} \times \mathbf{B} = \frac{1}{c^2 \epsilon_0} \mathbf{j} = 0 . \]

Proceeding as before we obtain respectively

\[ \Im \left[ e^{i\chi} \left( \omega i\mathbf{k} \times \mathbf{A}_0 - i\omega \mathbf{k} \times \mathbf{A}_0 \right) \right] \]

and

\[ \Im \left[ e^{i\chi} \left( -\frac{2\omega}{c} \frac{\omega}{c} \mathbf{A}_0 + \mathbf{k} \times \left( \mathbf{k} \times \mathbf{A}_0 \right) \right) \right] = \Im \left[ e^{i\chi} \left( \frac{\omega^2}{c^2} - \mathbf{k}^2 \right) \mathbf{A}_0 \right] = 0 . \]

where we recall the bac-cab rule to obtain \( \mathbf{k} \times \left( \mathbf{k} \times \mathbf{A}_0 \right) = -\mathbf{k}^2 \mathbf{A}_0 \) and considering that photon are massless \( \omega^2 - c^2 \mathbf{k}^2 = 0 \).

One single plane wave as presented is a special case of any linear superposition of plane waves which also will satisfy the linear Maxwell equations. Thus many continuous wave (CW) light beams of diverse and complex superposition structure will satisfy Maxwell equations. In a recent review Iwo and Zofia Bialynicki-Birula \(^2\) show how one can obtain intricate solutions of Maxwell’s equations employing the complex EM-field method attributed to Riemann, and Silberstein that was under consideration in the early days of EM-theory.

---

27.2 Covariant gauge condition

Since the covariant forms of the Lorentz-force and Maxwell’s equations depend on fields and not potentials we cannot measure $A^\mu$. Therefore all forms of $A^\mu$ that produce the same $F^{\mu\nu}$ are acceptable and in general they are related by a ‘gauge’ transformation. We discussed the non-covariant transverse gauge in section 23.2.

The gauge transformation presented in Eq. (23.8) can be now cast into the covariant form

$$A'\mu = A\mu + \partial^\mu f(x^2) .$$

(27.15)

To show invariance of the fields under transformation Eq. (27.15) we evaluate

$$F'\mu\nu = \partial_\nu A_\mu - \partial_\mu A_\nu + \partial^\rho \partial^\sigma f(x^2) - \partial^\mu \partial^\sigma f(x^2) = F^{\mu\nu} .$$

(27.16)

The last equality in Eq. (27.16) follows if $f$ is differentiable allowing the sequence of derivatives to be exchanged, and showing that under any transformation Eq. (27.15) EM-fields remain invariant.

This result implies that when we search for gauge invariant quantities, we can default to expressions written in terms of EM-fields. However we note an exception to this rule, which already appears in the covariant action term Eq. (25.18) where we see a path integral of a particle involving the 4-potential. We consider the effect of a gauge transformation Eq. (27.15) on this expression

$$U' \equiv e\int_{P_1}^{P_2} u \cdot A' d\tau = e\int_{P_1}^{P_2} u \cdot A d\tau + e\int_{P_1}^{P_2} \frac{dx^\mu}{d\tau} \cdot \frac{\partial \Lambda}{\partial x^\mu} d\tau = U + e\int_{P_1}^{P_2} \frac{d\Lambda}{d\tau} d\tau .$$

(27.17)

Thus we conclude

$$U' = U + e\Lambda(x_2) - e\Lambda(x_1) .$$

(27.18)

The gauge invariance of Eq. (27.18) and thus Eq. (27.15) is assured when the gauge transformation satisfies

$$\Lambda(x_2) = \Lambda(x_1) .$$

(27.19)

For a particle that enters and exits a space-time domain of an EM-field this is the case since $0 = \Lambda(x_2) = \Lambda(x_1)$. Similarly this condition is satisfied for a particle that performs an exactly closed path.

A similar argument applies when we consider the gauge transformation of the last form in Eq. (27.4)

$$U' \equiv -\int_{vol} d^4x \, A' \cdot J = -\int_{vol} d^4x \, (A + \partial f) \cdot J = U + \int_{vol} d^4x \, [f \partial \cdot J - \partial \cdot (fJ)] = U .$$

(27.20)

In the square bracket the first term vanishes in view of current conservation Eq. (27.12), the last surface term vanishes for any $J^\mu$ with bounded support.
The best known covariant gauge is the Lorenz-gauge constraint named after Ludvig Lorenz (not ‘Relativity’ Hendrik A. Lorentz)

\[ \partial_\mu A^\mu = 0 = \partial V \frac{c^2}{\partial t} + \vec{\nabla} \cdot \vec{A}. \] (27.21)

If a potential \( A^\mu \) does not satisfy this gauge condition we perform the transformation Eq. (27.15) such that \( A'^\mu \) does satisfy Eq. (27.21)

\[ \partial_\mu A'^\mu = \partial_\mu A^\mu + \Box f(x^2) = 0, \] (27.22)

with

\[ \Box \equiv \partial_\mu \partial^\mu = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial c^2} \equiv \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \Delta. \] (27.23)

The required gauge function \( f(x^2) \) is obtained solving Eq. (27.22)

\[ f(x^2) = \int d^4x' D_{ret}(x - x') \left[ \partial_\mu A^\mu(x') \right], \quad \Box D_{ret}(x - x') = \delta^4(x - x'). \] (27.24)

We characterize the distribution \( D_{ret}(x - x') \), the fundamental solution of 1+3-dimensional d’Alambert equation in the following exercise XI–4.

**Exercise XI–4:** The function \( D_{ret}(x - x') \)

Determine the fundamental solution of the d’Alambert equation in 1+3 dimensions that satisfies the retarded (causal) boundary condition. Use this solution to obtain the Liénard-Wichert potential of a charged particle on a trajectory.

**Solution**

We seek a solution of the d’Alambert equation with one timelike and three spacelike derivatives

\[ \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \Delta \right] f(x) = s(x), \]

the \( \Delta \) Laplace-operator appeared in Eq. (27.23) just above, with the causal boundary condition, which relies on knowing the elementary function \( D_{ret} \) which satisfies, see Eq. (27.24)

\[ \Box D_{ret}(x - x') = \delta^4(x - x') \equiv \delta(ct - ct')\delta^3(x - x'). \]

A general form of the solution is guessed by counting dimensions: since there is no dimensioned (length or energy) constant seen in Eq. [2] the function \( D_{ret} \) must be inversely proportional to length-squared, be covariant and singular so that the \( \delta \)-function
can be generated, and in many of the limits to be considered it needs to contain the Coulomb equation fundamental solution

\[ \Delta D_{\text{Coul}}(\vec{x} - \vec{x'}) = \delta^3(\vec{x} - \vec{x'}) , \quad D_{\text{Coul}}(\vec{x} - \vec{x'}) = -\frac{1}{4\pi|\vec{x} - \vec{x'}|} . \]

These requirements lead us to generalize \( D_{\text{Coul}} \) to its retarded time dependent form

\[ D_{\text{ret}}(x-x') = \frac{1}{8\pi} \delta ((x-x')^2) |_{\text{ret}} = \frac{1}{4\pi} \frac{\delta(ct-ct' - |\vec{x} - \vec{x'}|)}{|\vec{x} - \vec{x'}|} . \]

The equivalence of both forms of the solution \( D_{\text{ret}} \) stated in Eq. 4 is established by recalling the elementary property of the \( \delta \)-function

\[ \delta(f(s)) = \sum_z \frac{\delta(s-s_z)}{|f'(s_z)|} , \quad f(s_z) = 0 . \]

We find

\[ \delta((x-x')^2) = \frac{2\delta(ct-ct' - |\vec{x} - \vec{x'}|)}{|\vec{x} - \vec{x'}|} + \frac{2\delta(ct-ct' + |\vec{x} - \vec{x'}|)}{|\vec{x} - \vec{x'}|} . \]

We are keeping only the zero of \((x-x')^2\) that satisfies the retardation condition \( t-t' > 0 \); that is, the first term in Eq. 5. Since ‘proper’ Lorentz-transformations cannot change time sequence the imposition of the retardation constraint does not violate Li of \( D_{\text{ret}} \).

To check the validity of Eq. (4) we insert the second form of solution Eq. 4 into Eq. 1 and carry out the integral over \( t' \). This yields

\[ f = \frac{1}{4\pi} \int d^3r' s(\vec{r}', ct' = ct - R) / R , \quad R = |\vec{x} - \vec{x'}| . \]

The source function \( s \) is, in mathematical language, a smooth test-function. Next we evaluate

\[ \Delta f = \frac{1}{4\pi} \int d^3r' \left( \frac{\Delta s(\vec{r}', ct - R)}{R} + 2\nabla s(\vec{r'}, ct - R) \cdot \nabla \frac{1}{R} - s 4\pi \delta^3(x-x') \right) . \]

where we used, see Eq. 3 \( \Delta R^{-1} = -4\pi \delta^3(x-x') \). Since

\[ \nabla s(\vec{r'}, ct - R) = -\nabla R \frac{\partial s(\vec{r'}, ct - R)}{\partial ct} , \]

we obtain

\[ \Delta f + s = \frac{1}{4\pi} \int d^3r' \left( a_1 \frac{\partial^2 s(\vec{r}', ct - R)}{c^2 ct^2} - a_2 \frac{\partial s(\vec{r'}, ct - R)}{\partial t} \right) , \]

where a straightforward computation shows

\[ a_1 = (\nabla R)^2 = 1 , \quad a_2 = 2\nabla R \cdot \nabla R^{-1} + R^{-1} \Delta R = 0 . \]
The first term in parenthesis in Eq. (11) is thus the second time derivative of \( f \) Eq. (7) while the second term vanishes. This completes the demonstration that \( f \) as defined in Eq. (7) is a solution of Eq. (1). Therefore \( D_{\text{ret}}(x - x') \), Eq. (4), is the elementary causal solution of Eq. (2).

An important example of use of the retarded fundamental solution is the Liénard-Wichert solutions for a charged particle moving along a trajectory, a problem we mention in next exercise XI–5 and return to in exercise XI–6 on page 412.

End XI–4: The function \( D_{\text{ret}}(x - x') \)

Exercise XI–5: Maxwell’s equation in covariant Lorenz gauge

Obtain covariant Maxwell equations in terms of 4-potential and simplify the form using a Lorenz-gauge condition, Eq. (27.21). Describe new properties of the solution for \( A^\mu \) that differ from those seen in non-covariant transverse gauge.

Solution

Inserting the definition of the fields in terms of potentials, Eq. (26.29) into the Maxwell equation Eq. (27.9) we obtain

\[
\frac{1}{c^2 \epsilon_0} J^\nu = \frac{\partial F^{\mu \nu}}{\partial x^\mu} = \partial^\mu A^\nu - \partial^\nu A^\mu .
\]

For the first term on the right-hand side we have

\[
\partial^\mu \partial_\nu A^\nu = \left( \frac{\partial^2}{c^2 \partial t^2} - \Delta \right) A^\mu \equiv \Box A^\mu.
\]

The d’Alambert operator \( \Box \) was introduced in Eq. (27.23). We assume as we have done before that the potential \( A^\mu \) is not singular and allows us to exchange the sequence of differentiation. Thus we obtain for the last term in Eq. (1)

\[
\partial^\nu \partial_\mu A^\mu = \partial^\nu (\partial_\mu A^\mu) = \partial^\nu \left( \frac{\partial V}{c^2 \partial t} + \vec{\nabla} \cdot \vec{A} \right) .
\]

In the non-covariant transverse gauge (also called Coulomb or radiation gauge), see section 23.2 the last term is set to zero. Here we choose Lorenz-gauge, see Eq. (27.22), such that the entire argument in the bracket in Eq. (3) vanishes,

\[
\partial_\mu A^\mu = 0 . \quad \text{Lorentz gauge condition}
\]

Inserting Eq. (4) and Eq. (2) in Eq. (1) we obtain

\[
\frac{1}{c^2 \epsilon_0} J^\nu = \Box A^\nu . \quad \text{Lorentz gauge}
\]
We want to clarify the prefactor $1/c^2$ – the 4-potential is defined by $A^\nu = \{V/c, \vec{A}\}$ and the 4-current by $J^\nu = \{c\rho, \vec{j}\}$. Thus for the timelike and spacelike components respectively Eq. (5) reads

$$\begin{align*}
\frac{1}{\epsilon_0} \rho &= \Box V, \\
\mu_0 \vec{j} &= \Box \vec{A}.
\end{align*}$$

Lorentz gauge

where to obtain last relation we used Eq. (23.3). The two equations Eq. (6) make perfect sense considering the form of Maxwell equations.

The Lorenz-gauge was at first considered because it renders as seen in Eq. (5) the equation defining the potentials both simple, and as we know today, covariant. Moreover, the solutions of Eq. (5) show a new physics content that the Lorentz-gauge makes accessible: the solution of the covariant Eq. (5) can be chosen such that the action of the source function $J^\mu$ propagates with light speed to any observation point. This is causal i.e. ‘retarded’ boundary condition for the solution of the d’Alambert equation Eq. (5).

$$A^\nu(x) = \int d^4x' D_{ret}(x-x') \frac{1}{c^2\epsilon_0} J^\mu(x').$$

For $J^\mu$ of a point charged particle moving along a trajectory this is Liénard-Wichert solution.

The function $D_{ret}$ assures that the effect of a relativistic motion of source $J^\nu$ is felt at the observation point $x$ retarded in time. For comparison note that in the transverse gauge, see exercise IX–4 on page 330, the action of charge is instantaneous, the transverse gauge thus is suitable if the potential configuration is nearly stationary, or if the retardation effect is not relevant for other reasons. For example in case of motion in the field of transverse plane wave we have considered in section 24 we have $A^0 = 0$ and a periodic structure of $\vec{A}$ thus use of transverse gauge is justified.

End XI–5: Maxwell’s equation in covariant Lorentz gauge

Exercise XI–6: Liénard-Wichert potential

Present analytic covariant and non-covariant explicit forms for the EM potential $A^\mu$ in Lorenz gauge for a point charge moving along an arbitrary path $z^\mu(\tau)$.

Solution

The elements that we need to combine to solve this problem were developed previously:

- Maxwell’s equations in Lorenz gauge, Eq. (4) and Eq. (5) in exercise XI–5 on page 411

$$\Box A^\nu = \frac{1}{c^2\epsilon_0} J^\nu, \quad \partial_\mu A^\mu = 0.$$
The fundamental solution for Eq. 1 presented in exercise XI–4 in Eq. 4:

\[ D_{\text{ret}}(x - x') = \frac{1}{8\pi} \delta \left((x - x')^2\right)_{\text{ret}} = \frac{1}{4\pi} \frac{\delta(ct - ct' - |\vec{x} - \vec{x}'|)}{|\vec{x} - \vec{x}'|}. \]

The subscript ‘ret’ reminds us that we retain the retarded causal component as indicated on the right. We further must remember that we obtained and presented \( D_{\text{ret}} \) in a symmetric way between time \( ct \) and space \( \vec{x} \) thus we must remember when using it to take care in integration of the additional factor \( c \) in 4-volume, compare Eq. (27.1).

The current of a moving point particle presented in Eq. (27.6)

\[ J^\nu(x') = e c \int_{\text{path}} d\tau \frac{dz^\nu}{d\tau} \delta^4(x' - z(\tau)). \]

The formal solution to Eq. 1 is

\[ A^\nu(x) = \int d^4x' D_{\text{ret}}(x - x') \frac{1}{\epsilon_0 c^2} J^\nu(x'), \]

and using Eq. 2 and Eq. 3 we obtain, switching for ease of integration to the volume element \( d^4\vec{x} \), see Eq. (27.1)

\[ A^\nu(x) = \frac{e}{8\pi\epsilon_0 c} \int d^4\vec{x}' \left( \delta \left((x - x')^2\right)_{\text{ret}} \int_{\text{path}} d\tau \frac{dz^\mu}{d\tau} \delta^4(x' - z(\tau)) \right). \]

We can carry out the \( x' \)-integral using the \( \delta^4 \)-function to obtain

\[ A^\nu(x) = \frac{e}{8\pi\epsilon_0 c} \int_{\text{path}} d\tau \frac{dz^\nu}{d\tau} \delta \left((x - z(\tau))^2\right)_{\text{ret}}. \]

We check the gauge condition

\[ \frac{8\pi\epsilon_0 c}{e} \frac{\partial A^\nu(x)}{\partial x^\nu} = \int_{\text{path}} d\tau \frac{dz^\nu}{d\tau} \delta \left((x - z(\tau))^2\right)_{\text{ret}} = \int_{\text{path}} d\tau \frac{d}{d\tau} \delta \left((x - z(\tau))^2\right)_{\text{ret}}, \]

thus

\[ \frac{8\pi\epsilon_0 c}{e} \frac{\partial A^\nu(x)}{\partial x^\nu} = \delta \left((x - z(\tau))^2\right)_{\text{ret}} \bigg|_{-\infty}^{\infty} = 0. \]

For a path which begins in a distant past and ends in a distant future of the observation point \( x \) there is one causal crossing of the light cone at some finite retardation time which cannot coincide with the boundary of the path for any material particle which has to move slower than the speed of light.

We return now to evaluate Eq. (6). We use

\[ \delta(f(\tau)) = \sum_k \frac{\delta(\tau - \tau_k)}{|f'(\tau_k)|}, \quad f(\tau_k) = 0. \]
where in present case takes the form

\[ f(\tau) = (\mathcal{L} \cdot \mathcal{L})^2, \quad |f'(\tau)| = \left| \frac{\partial (\mathcal{L} \cdot \mathcal{L})}{\partial \tau} \right| = \left| -2 \frac{\partial z^\mu(\tau)}{\partial \tau} \cdot (x_\mu - z_\mu(\tau)) \right| . \]

We retain the retarded value of \( \tau_r \) for which

\[ ct_x - ct_z(\tau_r) = |\vec{x} - \vec{z}(\tau_r)| > 0, \]

and reject the advanced ‘zero’ at \( \tau_a \) of the \( f \)-function for which

\[ ct_x - ct_z(\tau_a) = -|\vec{x} - \vec{z}(\tau_a)| < 0. \]

Using Eq. 10 to integrate Eq. 6 we obtain the covariant Liénard-Wichert potential

\[ A^\nu(x) = \frac{e}{4\pi\epsilon_0 c} \left| \frac{dz^\nu}{d\tau} \right| \cdot (x - z_\tau), \quad \frac{dz^\nu}{d\tau} = \frac{dz^\nu(\tau)}{d\tau} \bigg|_{\tau=\tau_r}. \]

This solution offers and addition in its non-covariant presentation as in this case we recover the expected Coulomb-like format. We use

\[ \frac{dz^\nu}{d\tau} = c\gamma_r \{\mathbf{1}, \vec{\beta}_r\}, \quad \vec{\beta}_r = \frac{d\vec{z}_r}{c d\tau} = \frac{d\vec{z}_r}{c d\tau} \bigg|_{\text{ret}}, \quad \gamma_r = \frac{1}{\sqrt{1 - \vec{\beta}_r^2}}. \]

For the denominator of Eq. 13 we obtain

\[ \left| \frac{dz_\tau}{d\tau} \cdot (x - z_r) \right| = c\gamma_r \left| ct_x - ct_z - \vec{\beta}_r \cdot (\vec{x} - \vec{z}_r) \right|. \]

We now use Eq. 11

\[ \left| \frac{dz_\tau}{d\tau} \cdot (x - z_r) \right| = c\gamma_r |\vec{x} - \vec{z}_r| \left| 1 - \vec{\beta}_r \cdot \vec{n}_r \right| \equiv \frac{\vec{x} - \vec{z}_r}{|\vec{x} - \vec{z}_r|}. \]

We insert in Eq. 13 both the nominator Eq. 14 and denominator Eq. 16 to obtain the non-covariant form of the Liénard-Wichert potential

\[ A^\nu(x) = \frac{e}{4\pi\epsilon_0} \frac{\{1, \vec{\beta}_r\}}{|\vec{x} - \vec{z}_r| \left| 1 - \vec{\beta}_r \cdot \vec{n}_r \right|}. \]

This result shows that the potential observed at location \( x \) is created by a charged particle crossing the past light cone at position \( z_r \) with an observation time ahead of this event, allowing the signal to travel at speed of light to the observation point. We further note that the difference between the Coulomb potential, i.e. the solution in transverse gauge for \( A^0 \) differs from the Lorenz gauge solution by the factor \( 1/(1 - \vec{\beta}_r \cdot \vec{n}_r) \), which vanishes for a particle at rest and is negligible for non-relativistic motion.

End XI–6: Liénard-Wichert potential
27.3 Homogeneous Maxwell equations

The set of four homogeneous (source-less) Maxwell equations Eq. (23.5) can be re-written in a covariant format using the dual fields $F^*_{\mu\nu}$, Eq. (26.19). Even if we did not yet know about Eq. (23.5), when seeking covariant form of Maxwell equations, it would seem natural to evaluate the consequences of the covariant first order equation

$$\frac{\partial F^*_{\mu\nu}}{\partial x_\mu} = 0 \ , \quad (27.25)$$

with the nuance that there are no sources on the right-hand side. We can confirm Eq. (27.25) directly by using the definition of $F^*_{\mu\nu}$, Eq. (26.19)

$$\frac{\partial F^*_{\mu\nu}}{\partial x_\mu} = \varepsilon_{\mu\nu\delta\kappa} \partial^\mu F_{\delta\kappa} = \varepsilon_{\mu\nu\delta\kappa} \left( \partial^\mu \partial^\delta A^\kappa - \partial^\mu \partial^\kappa A^\delta \right) = 0 \ . \quad (27.26)$$

Due to total antisymmetry property of $\varepsilon_{\mu\nu\delta\kappa}$, the two terms on the right vanish independently as long as we can exchange the sequence of second derivatives of the potentials.

We draw attention to a possible modification of this situation: Dirac\(^3\) presented\(^4\) a generalization introducing singular field configurations that in string-like space-time domains do not allow the exchange of the two derivative sequence in Eq. (27.26). These singular solutions describe the effect of magnetic monopoles. A consequence of discovery of even one single magnetic monopole is justification for the quantization of electric charge: since, according to Dirac, the product of magnetic and electric charge would be quantized. After extensive search for a magnetic monopole such charges $\rho_m$ have not been discovered. Accordingly, the quantization of electric charge remains unexplained today.

We now seek the non-covariant form of Eq. (27.25). We recall that transition of $F^\mu\nu \leftrightarrow F^*_{\mu\nu}$ requires the exchange of fields seen in Eq. (26.21) $\vec{E}/c, \vec{B} \leftrightarrow \vec{B}, \vec{E}/c$ (there is no sign in this rule as we chose to compare index up $F^\mu\nu$ with index down $F^*_{\mu\nu}$). Thus we turn to Maxwell’s Eq. (27.10) and Eq. (27.11) exchanging $\vec{E} \leftrightarrow \vec{B}$ and setting sources to zero. We also recall that with the differentiation index being down in Eq. (27.25) there is an extra sign in the spatial differentiation. We thus obtain

$$- \vec{\nabla} \cdot \vec{B} = 0 \ , \quad (27.27)$$

\(^3\)Paul A.M. Dirac (1902 – 1984) theoretical physicist contributing at creation of relativistic quantum physics, Nobel Prize in Physics (1933) for ‘Dirac equation,’ i.e. the formulation of the relativistic quantum wave theory of the electron. Dirac made many theoretical and mathematical physics contributions, in the classroom his name is associated most often with the ‘Dirac’$\delta$-function.

and
\[- \frac{\partial \vec{B}}{\partial t} - \vec{\nabla} \times \vec{E}/c = \vec{0}. \quad (27.28)\]

We recognize Eq. (27.27) and Eq. (27.28) to be identical to Eq. (23.5), a condition obtained to assure that fields \( \vec{E}, \vec{B} \) can be written as derivatives of potentials. This in fact is how in Eq. (27.26) we show the validity of Eq. (27.25) and thus Eq. (27.27) and Eq. (27.28).

In absence of a magnetic monopole source in Eq. (27.27), it would seem that only motion of charged particles can generate a magnetic field. We will show that this is not quite true. In exercise XI–8 where we demonstrate the solution of Maxwell’s equation for a static magnetic point dipole.

**Exercise XI–7: Other forms of homogeneous Maxwell eq.**

Reconsider the covariant form of homogeneous Maxwell equation Eq. (27.25) and present other ways to write these four component equations by constraining the fields in a way that allows the introduction of 4-potential determining the fields.

**Solution**

The second form in Eq. (27.26) shows that for each of the 4 possible values of \( \nu = 0, 1, 2, 3, \) the other three indices can only choose cyclic values. In each case there are three terms, for example for \( \nu = 0 \) we have \( \mu = 1, \delta = 2, \kappa = 3 \) and for the other two cyclic permutations, we have \( \mu = 2, \delta = 3, \kappa = 1 \) and \( \mu = 3, \delta = 1, \kappa = 2 \). Explicitly

\[
\begin{align*}
\nu &= 0 : \quad \partial^1 F^{23} + \partial^2 F^{31} + \partial^3 F^{12} = 0, \\
\nu &= 1 : \quad \partial^0 F^{23} + \partial^2 F^{30} + \partial^3 F^{02} = 0, \\
\nu &= 2 : \quad \partial^1 F^{03} + \partial^0 F^{31} + \partial^3 F^{10} = 0, \\
\nu &= 3 : \quad \partial^1 F^{20} + \partial^2 F^{01} + \partial^0 F^{12} = 0.
\end{align*}
\]

These four equations can be stated in a more elegant format as follows

\[
\begin{align*}
\nu &= 0 : \quad \partial^1 F^{23} + \partial^2 F^{31} + \partial^3 F^{12} = 0, \\
\nu &= 1 : \quad \partial^0 F^{23} + \partial^2 F^{30} + \partial^3 F^{02} = 0, \\
\nu &= 2 : \quad \partial^1 F^{03} + \partial^0 F^{31} + \partial^3 F^{10} = 0, \\
\nu &= 3 : \quad \partial^1 F^{20} + \partial^2 F^{01} + \partial^0 F^{12} = 0.
\end{align*}
\]

That this form is the same as Eq. (1) is seen by choosing any set of three different \( \mu \delta \kappa \). The one number not chosen is the value of \( \nu \), and the three terms that can be written are exactly those seen in Eq. (1).

Eq. (2) is nice, useful and also a trivial identity. We insert the definition of \( F^{\delta \kappa} \) in terms of potential derivatives

\[
\begin{align*}
\partial^\mu \partial^\delta A^\kappa - \partial^\mu \partial^\kappa A^\delta + \partial^\delta \partial^\kappa A^\mu - \partial^\delta \partial^\mu A^\kappa + \partial^\kappa \partial^\mu A^\delta - \partial^\kappa \partial^\delta A^\mu &= 0.
\end{align*}
\]
There are all together 6 terms. When and if we can exchange the first and second
derivative we see that the first term cancels the fourth, the second term the fifth, and
the third term the sixth.

Exercise XI–8: Magnetic dipole moment: fields and forces

Obtain the EM-fields of an elementary magnetic point-dipole moment \( \vec{\mu}_0 \) and characterize the related force.

Solution

The existence of elementary magnetic point-dipole \( \vec{\mu}_0 \) is undisputed. For example an
electron has both charge and magnetic dipole moment and no discernible size. However,
there is some confusion on how one can write fields and potentials, and the form of the
force remains in a rudimentary format. Given the incomplete stage of understanding
we proceed with caution. When we speak in the following of a ‘distribution’it is in the
mathematical sense a generalized function. Appearance of distributions is part of the
problem when solving equations that include point sources, and in the present context
we encounter point-currents.

The magnetic point-dipole is consistent with the properties of Maxwell’s equation.
In particular we seek a solution satisfying Eq. (27.25), here specifically Eq. (27.27), \textit{i.e.}
the magnetic field \( \vec{B} \) remains source-free, there is no magnetic dipole-charge-density
\( \rho_{md} \)

\[ \nabla \cdot \vec{B}_{md} \equiv \rho_{md} = 0. \]

We check next Maxwell’s equations with sources, Eq. (27.9), considering here their
vectorial form Eq. (27.10) and Eq. (27.11). If and when a particle has a charge in its
rest-frame we obtain the usual \( \vec{E} \) solution solving Eq. (27.10). This time independent
and radial \( \vec{E} \) solves Eq. (27.28) and in its presence the first term in Eq. (27.11) vanishes.
Thus to find the magnetic field generated by the magnetic static point-dipole current
\( \vec{j}_{md} \) we need to satisfy

\[ \nabla \times \vec{H}_{md} \equiv \vec{j}_{md}. \]

To fix a suitable form of \( \vec{j}_{md} \) we require that the integral of the circulating point-current
creates the point-dipole magnetic charge

\[ \vec{\mu}_0 \equiv \frac{1}{8\pi} \int d^3x \vec{x} \times \vec{j}_{md}. \]
The distribution that satisfies this requirement is

\[ \vec j^{\text{ind}} = 3 \vec \mu_0 \times \hat x \frac{\delta^3(\vec x)}{|\vec x|}. \]

We check Eq. 4

\[ \frac{1}{8\pi} \int d^3x \vec x \times \vec j^{\text{ind}} = \int_0^\infty dr \delta_+(r) \frac{1}{4\pi} \int d\Omega \frac{3}{2} \hat x \times (\vec \mu_0 \times \hat x) = \vec \mu_0, \quad r = |\vec x|, \quad \hat x = \frac{\vec x}{|\vec x|}. \]

The last equality follows since \( \hat x \times (\vec \mu_0 \times \hat x) = \vec \mu_0 - \hat x(\vec \mu_0 \cdot \hat x) \) and choosing the 3-axis of the coordinate system along the direction of \( \vec \mu_0 \) the spherical integral involves the angular average \( (1 - \cos^2 \theta) = 2/3 \).

We introduced in Eq. 5 the distribution \( \delta_\pm(r) \): the subscript implies: `-': approaching zero from negative domain of \( r \); and `+': approaching zero from positive domain of \( r \). Introduction of these functions is inspired by the observation

\[ \operatorname{sgn}(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0 \end{cases}, \quad \operatorname{sgn}(|x|) = \begin{cases} 1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0 \end{cases}, \]

which has as consequence

\[ \frac{d}{dx} \operatorname{sgn}(x) = \delta_-(x) + \delta_+(x), \quad \frac{d}{dx} \operatorname{sgn}(|x|) = -\delta_-(x) + \delta_+(x). \]

We note that Ref. 5 proposes instead the use of zero-up-step Heaviside function

\[ \operatorname{Hup}(r) = \begin{cases} \text{undefined} & r < 0 \\ 0 & r = 0 \\ 1 & r > 0 \end{cases}, \quad \frac{d}{dr} \operatorname{Hup}(r) = \delta_+(r). \]

Either of the forms Eq. 6 & Eq. 7, or Eq. 8 is helpful in the computation of singular derivatives involving a non-Cartesian set of coordinates, e.g. radial variables, which are defined in half space \( r \geq 0 \).

Turning now to solutions of Maxwell equations Eq. 1 and Eq. 2, the distribution for the magnetic field is provided in Ref. 5 in radial coordinates

\[ \vec B^{\text{ind}} = 3 \vec \mu_0 \cdot \hat r \frac{\hat r}{r^3} \operatorname{Hup}(r) + \hat r \times (\vec \mu_0 \times \hat r) \frac{\delta_+(r)}{r^2}, \quad \hat r = \frac{\vec r}{r}. \]

The above distribution follows in transverse gauge from the potential distribution

\[ \vec A^{\text{ind}} = \frac{\vec \mu_0 \times \vec r}{r^3} \operatorname{Hup}(r), \quad \vec \nabla \times \vec A^{\text{ind}} = \vec B^{\text{ind}}, \quad \vec \nabla \cdot \vec A^{\text{ind}} = 0. \]

The form of magnetic field presented in Eq. 9 is often verified and tested experimentally. The greatest precision is achieved by means of the hyperfine splitting where

the effect observed depends decisively on second, i.e. δ-function, term not always noted in pertinent textbooks. Note that without this term experimental results would disagree with theory, and Eq. (1) would not hold (if computed with test function helping recognize a singularity) in violation of theoretical principles.

Now that we have understood a field that a magnetic point-dipole generates there is another challenge, is there a force when we place a magnetic moment in an external electromagnetic field? The famous Stern-Gerlach experiment showed that this force exists in 1922. A magnetic dipole moment $\vec{\mu}_0$ placed at rest in the domain of any magnetic field $\vec{B}$ contributes to the potential $U_m$ and thus experiences a force $\vec{F}_m$.

$$ U_m = -\vec{\mu}_0 \cdot \vec{B}, \quad \vec{F}_m = \vec{\nabla} (\vec{\mu}_0 \cdot \vec{B}). $$

We keep in mind that Eq. (11) is valid: a) in the rest-frame of the magnetic dipole moment and, b) requires a variation of the $\vec{B}$-field at particle location. We note that Eq. (11) contribution to the force is not obtainable from the Lorentz-force written in its conventional form. For more comprehensive discussion of magnetic and electric dipole moment interaction with external fields we refer to a recent review of Kholmetskii and collaborators\textsuperscript{6} A full relativistic formulation of charged particle dynamics incorporating magnetic (and electric) dipole moments has to best of the author’s knowledge not been presented at the time of writing.

It is of interest to record the magnitude of the magnetic moment $\mu_p$ of a particle ‘p’ with particle spin $s_p$ is written in terms of particle’s Bohr magneton $\mu_{B,p}$.

$$ \mu_p = \mu_{B,p} s_p g_p, $$

where for the electron we have

$$ \mu_{B,p} = \frac{e\hbar}{2M_p}, \quad \mu_{B,e} = 5.788 \ 381 \ 8 \ 10^{-5} \ \text{eV} \ T. $$

The other factors in Eq. (12) are the spin, typically $s_e = 1/2$, and the particle $g$-factor $g_p$, for an electron

$$ g_e = 2 \left(1 + 0.001 \ 159 \ 652 \ 181\right), \quad s = 1/2. $$

There are no errors stated for the values in Eq. (13) and Eq. (14) since to the precision stated the results are exact.

For particles made of several constituents, such as three quarks making a proton, the $g$-factor can be relatively large, for the proton $g_{\text{prot}} = 5.585 \ 695$, while for the neutron where the sign of constituent electric charges is reversed and charges are smaller, the value is $g_{\text{neut}} = -1.913 \ 043$, error is invisible to the stated precision.

Exercise XI–9: Relativistic Spin Dynamics

We introduce the relativistic dynamics of an internal particle spin \( \vec{s} \) in an arbitrary external field – the dynamics of particle proper angular momentum in the presence of fields was presented by and is called Bergmann-Michel-Telegdi (BMT) equation\(^7\). Show that the BMT covariant equation corresponds to the nonrelativistic form.

Solution

BMT obtained relativistic dynamics based on empirical relation

\[
\vec{T}_0 \equiv \frac{d\vec{s}}{dt} = \vec{\mu}_0 \times \vec{B} = g\mu_B \vec{s} \times \vec{B}.
\]

where \( \vec{T}_0 \) is the torque, and Eq.\(^1\) is valid in the rest-frame of the magnetic dipole. In the last form we entered the magnetic moment as proportional to proper angular momentum (spin) \( s \)

\[
\vec{\mu}_0 = g\mu_B \vec{s}, \quad \mu_B = \frac{e\hbar}{2m},
\]

as introduced in exercise XI–8. According to BMT the proper angular momentum is a 4-vector \( L^\mu \) with a rest-frame component \( \vec{s} \); thus in the rest-frame of the particle

\[
L^\mu \equiv \{0, \vec{s}\}.
\]

Considering that the 4-velocity in the rest-frame is \( u^\mu = \{c, \vec{0}\} \), we thus obtain the invariant condition

\[
L \cdot u = 0 , \quad \frac{dt}{d\tau} \left( L^0 c - \vec{s} \cdot \vec{v} \right) = \frac{dt}{d\tau} \left( 0 c - \vec{s} \cdot \vec{0} \right) = 0.
\]

Any relativistic covariant generalization of Eq.\(^1\) must assure validity of Eq.\(^4\). BMT presents the covariant form

\[
\frac{dL^\mu}{d\tau} + \frac{u^\mu}{c^2} \frac{du}{d\tau} \cdot L = g\mu_B \left( F^{\mu\delta} g_{\delta\nu} L^\nu - \frac{u^\mu u^\alpha}{c^2} g_{\alpha\beta} F^{\beta\delta} g_{\delta\nu} L^\nu \right).
\]

We see that upon multiplication with \( u^\alpha g_{\alpha\mu} \) and summation, the right-hand side vanishes as an identity since \( u^2 = c^2 \). The left hand side becomes

\[
\frac{dL}{d\tau} + \frac{u^2}{c^2} \frac{du}{d\tau} \cdot L = \frac{d(L \cdot u)}{d\tau} = 0.
\]

We recognized Eq.\(^6\) as the proper time derivative of the covariant condition Eq.\(^4\), assuring that any solution of Eq.\(^6\) is consistent with the Eq.\(^4\). We also now understand that in Eq.\(^5\) both the second term on the left and on the right were included to assure that the constraint of Eq.\(^4\) is satisfied.

We now proceed to show by a direct computation that in the particle rest-frame, Eq. 5 reduces to Eq. 1. The rest-frame $\vec{v} \to 0, u \to \{c, \vec{0}\}$ is characterized by

- $\gamma \to 1$, $d\tau \to dt$.

However, we need to pay attention to derivatives of both $u^\mu$ and $J^\mu$ which need not vanish in the rest-frame.

- $u^\mu = \gamma \{c, \vec{v}\} \to \{1, \vec{0}\}$, $\frac{du^\mu}{d\tau} \to \left\{ \frac{1}{2} \frac{d v^2}{dct}, \frac{d\vec{v}}{dt} \right\}$.

- $L^\mu = \{L^0, \vec{s}\} \to \{0, \vec{s}\}$, $\frac{dL^\mu}{d\tau} \to \left\{ \frac{d L^0}{dt}, \frac{d\vec{s}}{dt} \right\} = \left\{ \vec{s} \cdot \frac{d\vec{v}}{dt}, \frac{d\vec{s}}{dt} \right\}$

where we used the second form Eq. 4 in the last two equalities.

The timelike component of the left hand side of Eq. (5) vanishes

- $\frac{dL^0}{d\tau} + u^0 \frac{du}{dct} \cdot S \to \vec{s} \cdot \frac{d\vec{s}}{dt} - \vec{s} \cdot \frac{d\vec{v}}{dt} = 0$.

For the right-hand side of the timelike component (omitting prefactor), we find

- $0 = -\sum_{k=1}^{3} F^{0k} s^k + \sum_{k=1}^{3} F^{0k} s^k$,

which also vanishes. Thus only the spacelike components of BMT equation have non-vanishing contributions in the rest-frame, as we expect in consideration of Eq. (1).

Turning now to the spacelike component of Eq. 5, we obtain, omitting terms with powers of $v \to 0$.

- $\frac{ds^k}{dt} = -g_\mu B \sum_{n=1}^{3} F^{kn} s^n$.

Due to antisymmetry of $F^{kn}$ there are two terms contributing in the sum in Eq. 12. For example, for $k = 1$ we have

- $\frac{ds^1}{dt} = -g_\mu B \left( F^{12} s^2 + F^{13} s^3 \right)$.

In consideration of Eq. (26.6a) we thus find

- $\frac{ds^1}{dt} = -g_\mu B \left( -\mathcal{B}^3 s^2 + \mathcal{B}^2 s^3 \right) = g_\mu B \left( \vec{s} \times \vec{B} \right)^1$.

The procedure can be repeated for $k = 2$ and $k = 3$, completing the demonstration that the BMT equation Eq. 5 in the rest-frame of the particle is Eq. 1.

End XI–9: Relativistic Spin Dynamics
27.4 Energy, momentum and mass of the EM-field

We seek to establish the spatial field energy-momentum distribution $T_{EM}^{\mu\nu}$ of the EM-field. For the case of Maxwell theory where the superposition principle applies exactly, it is uniquely (up to a multiplicative constant) defined by the requirements that it is:

i) quadratic in the EM-fields;
ii) a covariant expression;
iii) divergence-free (conserved) outside of a space-time domain containing a charge distribution.

Requirement i) arises from the quadratic form of the action, the first term in Eq. (27.3). Requirement ii) assures that all inertial observers agree to the energy-momentum content of space-time filled with fields. The last, requirement iii), assures that in any given space-time volume as much energy-momentum flows in as flows out. This condition is broken where charges are present since an electric field does work on charges, sharing the field energy-momentum with the motion energy of a charge.

In passing we note that the form of $T_{EM}^{\mu\nu}$ is found naturally in the context of general relativity (GR), and this result confirms what we find by constructive approach in SR. However, one can add to SR expression any covariant, bilinear in fields, and divergence free term, thus the definition of $T_{EM}^{\mu\nu}$ is also constrained by consistency with GR, so that $T_{EM}^{\mu\nu}$ we study is also a source of gravitating energy obtained as response of the EM-action to variation of the space-time metric.

We therefore consider the specific energy-momentum tensor $T_{EM}^{\mu\nu}$ (‘tensor’ means here as before a symmetric 2nd rank matrix subject to specific LT with regard to each index)

$$T_{EM}^{\mu\nu} = c^2\varepsilon_0 \left( F^{\mu\kappa} g_{\kappa\delta} F_{\delta\nu} - \frac{1}{4} g^{\mu\nu} F_{\delta\kappa} F^{\kappa\delta} \right). \quad (27.29)$$

Note that the last term is $-g_{\mu\nu} L_{\text{field}}$, Eq. (27.2). This is a covariant, and bilinear in EM-fields form. We need to verify that Eq. (27.29) is divergence free in empty space-time domain. We obtain

$$\frac{1}{c^2\varepsilon_0} \partial_{\mu} T_{EM}^{\mu\nu} = (\partial_{\mu} F^{\mu\kappa}) g_{\kappa\delta} F_{\delta\nu} + F^{\mu\kappa} g_{\kappa\delta} (\partial_{\mu} F_{\delta\nu}) - \frac{1}{4} (\partial^{\mu} F_{\delta\kappa}) F^{\kappa\delta} - \frac{1}{4} F_{\delta\kappa} (\partial^{\nu} F^{\kappa\delta})$$

$$= \frac{1}{c^2\varepsilon_0} J_{\delta} F^{\delta\nu} + F^{\mu\kappa} g_{\kappa\delta} (\partial_{\mu} F_{\delta\nu}) - \frac{1}{2} F_{\delta\kappa} (\partial^{\nu} F^{\kappa\delta})$$

$$= \frac{4\pi}{c} J_{\delta} F^{\delta\nu} - F_{\delta\mu} \left( \partial^{\mu} F_{\delta\nu} + \frac{1}{2} \partial^{\nu} F^{\mu\delta} \right), \quad (27.30)$$

where in the first term we used the Maxwell equation with sources Eq. (27.9), and in last term we recombined the two 1/4-terms and wrote the content by renaming indices to obtain a common factor of the field derivative.
We now recall Eq. \ref{2} in exercise XI–7 on page 416, multiply it with the factor $F_\delta^\mu$ and rename $\kappa \rightarrow \nu$ to obtain

$$F_\delta^\mu \partial^\mu F_\delta^\nu + F_\nu^\mu \partial^\nu F_\delta^\mu + F_\delta^\mu \partial^\nu F_\nu^\mu = 0,$$

(27.31)

upon renaming the middle term $\mu \leftrightarrow \delta$ we obtain upon division by 2

$$F_\delta^\mu \left( \partial^\mu F_\delta^\nu + \frac{1}{2} \partial^\nu F_\nu^\mu \right) = 0.$$

(27.32)

Inserting Eq. (27.32) in Eq. (27.30) we see that

$$\partial_\mu T^{\mu\nu}_{\text{EM}} + F_\nu^\mu J_\mu = 0.$$

(27.33)

The zero on the right-hand side is due to the fact that we have only looked at the EM-field dynamics, the inertial resistance to motion by a massive particles will be introduced in exercise XI–11 on page 428. This result establishes the change in energy-momentum of electromagnetic field:

a) In all space-time domains free of electromagnetic charges the EM energy-momentum tensor $T^{\mu\nu}_{\text{EM}}$ is divergence free;

b) In space-time domains where charges are present the divergence of $T^{\mu\nu}_{\text{EM}}$ contains the product of EM-field with the current-density as we would expect of the differential form of the Lorentz-force.

In exercise XI–10 we obtain the explicit form

$$T^{\mu\nu}_{\text{EM}} = \frac{1}{\epsilon_0} \left( \begin{array}{cccc}
\frac{1}{2}(\vec{E}^2+c^2\vec{B}^2) & (\vec{E} \times c\vec{B})^2 & (\vec{E} \times c\vec{B})^3 \\
(\vec{E} \times c\vec{B})^1 & -c^2B^1B^2 -c^2B^1B^3 & -c^2B^1B^3 -c^2B^3B^1 \\
(\vec{E} \times c\vec{B})^2 & -\vec{E} \times c\vec{B}^2 & -c^2B^2B^3 -c^2B^3B^2 \\
(\vec{E} \times c\vec{B})^3 & -c^2B^1B^3 -c^2B^3B^1 & \frac{1}{2}(\vec{E}^2+c^2\vec{B}^2) -\vec{E}^2 -c^2B^2B^2 -c^2B^3B^3 \end{array} \right).$$

(27.34)

Inspecting Eq. (27.34), we see that $(T^{\mu\nu}_{\text{EM}})$ is symmetric. Inspection of Eq. (27.29) shows that this can be the case only when $(g_{\mu\nu})$ is symmetric and $(F^{\mu\nu})$ has a specific symmetry, here it is antisymmetric, providing 6 field components. A symmetric $4 \times 4$ matrix has 10 linearly independent components as we have discussed in exercise X–7 on page 390. On the other hand, given 6 distinct EM-field components in the antisymmetric field $(F^{\mu\nu})$ there are 36 different bilinear field combinations. Thus the covariance principles which lead us to consider the particular form Eq. (27.29) reduce the number of allowable bilinear terms from 36 to 10.
Actually, there are just 9 independent terms since we recognize another constraint in Eq. (27.34)

\[ T_{\nu}^{\nu} \equiv g_{\nu\mu} T_{EM}^{\mu\nu} = T_{EM}^{00} - \sum_{i=1}^{3} T_{EM}^{ii} = 0. \]  

(27.35)

We say \( T_{EM}^{\mu\nu} \) is ‘traceless’. This result holds in all Lorentz-frames of reference since as shown in Eq. (27.35), it is an algebraic identity. In fact the trace of a tensor is always a LI. Symmetry considerations beyond the context of this book allow us to understand the vanishing of the trace in terms of the absence of a dimensioned constant (energy or length scale) in \( T_{EM}^{\mu\nu} \).

Exercise XI–10: Explicit form of \( T_{EM}^{\mu\nu} \)

Obtain the explicit form of the energy-momentum distribution \( T_{EM}^{\mu\nu} \) for a given electromagnetic field.

Solution

To obtain the explicit form of \( T_{EM}^{\mu\nu} \) we consider in Eq. (27.29) the two terms we call in sequence

\[ 1 \quad T_{EM}^{\mu\nu} = T_{1}^{\mu\nu} + T_{2}^{\mu\nu}, \]

with

\[ 2 \quad \frac{1}{c^2 \epsilon_0} T_{1}^{\mu\nu} = F_{\mu \kappa} g_{\kappa \delta} F_{\delta \nu} , \quad \frac{1}{c^2 \epsilon_0} T_{2}^{\mu\nu} = -\frac{1}{4} g^{\mu\nu} F_{\delta \kappa} F_{\kappa \delta} , \quad \frac{1}{c^2 \epsilon_0} T_{1}^{\mu} = g_{\nu \mu} F^{\nu \kappa} g_{\kappa \delta} F_{\delta \nu} , \quad \frac{1}{c^2 \epsilon_0} T_{2}^{\mu} = -F_{\delta \kappa} F^{\kappa \delta} , \]

where the last equality uses \( \epsilon_{\mu \nu} g^{\mu \nu} = \text{Tr} gg^{-1} = \text{Tr} 1_4 = 4. \) Note that summing the two terms on the right in Eq. (2) we have shown that Tr \( gT = 0 \)

To evaluate the ten non-trivial components of the symmetric energy momentum tensor \( T_{EM}^{\mu\nu} \) we employ the matrix notation

\[ 3 \quad T_1 = c^2 \epsilon_0 (FG) , \quad T_2 = g \frac{\epsilon_0}{2} \left( -\vec{E}^2 + c^2 \vec{B}^2 \right). \]

The last term \( T_2 \) form follows considering Eq. (26.12) and Eq. (26.17).
Given the explicit format of the field tensor \( F \) seen in Eq. (26.7) we proceed as in Eq. (26.15) to evaluate \( T_{1}^{\mu\nu} \) in Eq. 3

\[
\frac{1}{c^2\epsilon_0} T_1 = F g F
\]

\[
= \begin{pmatrix}
0 & -\mathcal{E}^1/c & -\mathcal{E}^2/c & -\mathcal{E}^3/c \\
\mathcal{E}^1/c & 0 & -B^3 & B^2 \\
\mathcal{E}^2/c & B^3 & 0 & -B^1 \\
\mathcal{E}^3/c & -B^2 & B^1 & 0
\end{pmatrix}
\begin{pmatrix}
1 & 0 & -\mathcal{E}^1/c & -\mathcal{E}^2/c & -\mathcal{E}^3/c \\
-1 & -1 & 0 & -B^3 & B^2 \\
0 & -B^3 & B^3 & 0 & -B^1 \\
0 & -B^2 & B^1 & 0
\end{pmatrix}
\begin{pmatrix}
0 & -\mathcal{E}^1/c & -\mathcal{E}^2/c & -\mathcal{E}^3/c \\
\mathcal{E}^1/c & 0 & -B^3 & B^2 \\
\mathcal{E}^2/c & B^3 & 0 & -B^1 \\
\mathcal{E}^3/c & -B^2 & B^1 & 0
\end{pmatrix}
\]

\[
= \begin{pmatrix}
\mathcal{E}^2/c^2 & \mathcal{E} \times \mathcal{B}/c & \mathcal{E} \times \mathcal{B}^2/c & \mathcal{E} \times \mathcal{B}^3/c \\
\mathcal{E} \times \mathcal{B}/c & -\mathcal{E}^1/c^2 + B^3 B^2 & -\mathcal{E}^2/c^2 - B^1 B^2 & -\mathcal{E}^3/c^2 - B^3 B^1 \\
\mathcal{E} \times \mathcal{B}^2/c & -\mathcal{E}^1/c^2 + B^3 B^2 & -\mathcal{E}^2/c^2 + B^3 B^2 & -\mathcal{E}^3/c^2 + B^3 B^1 \\
\mathcal{E} \times \mathcal{B}^3/c & -\mathcal{E}^1/c^2 + B^3 B^2 & -\mathcal{E}^2/c^2 + B^3 B^2 & -\mathcal{E}^3/c^2 + B^3 B^1
\end{pmatrix}
\]

(27.36)

Combining with Eq. 3 for the diagonal \( T_2 \) we obtain the full energy-momentum tensor presented in Eq. (27.34).

---

End XI–10: Explicit form of \( T^{\mu\nu}_{EM} \)

28 \hspace{1em} EM-field Inertia

28.1 Field energy-momentum dynamics

For the observer \( R \) having in her rest-frame the 4-velocity \( u_{R,0} = (c, \vec{0}) \) and measuring the energy momentum carried by fields, the covariant 4-vector of interest is

\[
\left. T^{\mu\nu}_{EM} \left( \frac{u_{\nu}}{c} \right) \right|_0 = \left\{ T^{00}_{EM}, T^{k0}_{EM} \right\} \equiv \left\{ \varepsilon_{R,0}, \frac{1}{c} S^k_{R,0} \right\}.
\]

(28.1)

The factor \( 1/c \) is included in the definition of the Poynting vector \( S^k_{R,0} \) in Eq. (28.1) so that it describes the field momentum density, while \( \varepsilon_{R,0} \) is the energy density, both observed by an observer characterized by her 4-velocity \( u_{R,0} \). Henceforth we drop the subscript characterizing the observer by definition corresponding to rest
frame of the laboratory. Explicitly, the field energy density $\varepsilon$ and field momentum density $\bar{S}$ (Poynting vector)

$$
\varepsilon \equiv T^{00}_{EM} = \frac{\varepsilon_0}{2} \left( \bar{E}^2 + c^2 \bar{B}^2 \right) , \quad S^k \equiv cT^{k0}_{EM} = c^2 \varepsilon_0 \left( \bar{E} \times \bar{B} \right)^k .
$$

(28.2)

To understand better the meaning of $\bar{S}$ we look at timelike component of Eq. (27.33)

$$
\partial_0 T^{00}_{EM} + \nabla_k T^{k0}_{EM} + \bar{j}_i F^{i0} = f^0_V - \partial_\mu T^{\mu 0}_V .
$$

(28.3)

Where we included in this line on the right the particle inertial terms seen in following exercise [XI-11], see Eq. 5. This reminds us that any EM-field dynamics is compensated by the motion of material charged particles.

For the present we follow only the electromagnetic components. Using definitions Eq. (28.2), canceling the common factor $1/c$

$$
\frac{\partial \varepsilon}{\partial t} + \nabla \cdot \bar{S} + \bar{j} \cdot \bar{E} = 0 ,
$$

(28.4)

which is the differential form of the energy-momentum conservation. Outside of the range of the distribution of the 3-current, that is where $\bar{j} = 0$, the change of the local energy density $\varepsilon$ is directly compensated by the outflow in energy carried by the Poynting vector $\bar{S}$. We recall now that there are a few other ways to write the Poynting vector, see Eq. (24.11) and in particular we have

$$
\bar{S}_{energy \ flux} = \bar{D} \times \bar{B} ,
$$

(28.5)

the only expression in Eq. (24.11) that does not have a dimensioned coefficient. Therefore according to Eq. (28.4) $\bar{D} \times \bar{B}$ is the energy flux, forming a four-vector with the energy density

$$
\varepsilon_{energy \ density} = \frac{1}{2} \left( \bar{D} \cdot \bar{E} + \bar{H} \cdot \bar{B} \right) .
$$

(28.6)

The situation is just like with charge current $\bar{j}$ which describes the flow of charge density $\rho$, compare Eq. (27.13), and prior discussion Eq. (27.8). For Maxwell theory in vacuum this identification Eq. (28.5) and Eq. (28.6) may appear not essential, but whenever we encounter more complex systems, including the case of nonlinear electromagnetism the above expressions become handy, the use of correct energy flow is important. Similarly the momentum flux is identified to be, see Ref. 12 on page 347

$$
\bar{S}_{momentum \ flux} = c^2 \bar{E} \times \bar{H} .
$$

(28.7)

In the domain where we find the non vanishing charged particle current $\bar{j}$, the energy is further shared with the dynamical motion of charged particles driven
by the fields. To see this we inspect Eq. (28.4) in local rest-frame of the charge distribution where we can write according to Eq. (27.8)

\[ \frac{\partial \varepsilon}{\partial t} + \nabla \cdot \vec{S} + \rho_{(0,x)} \frac{d\vec{x}}{d\tau} \cdot \vec{E} = 0, \]  

(28.8)

which reminds of the energy conservation we addressed in section 22.4 and in section 25.1, comparing in particular to Eq. (25.15).

We have learned to appreciate now the role of four terms Eq. (28.2) in \((T_{EM}^{\mu\nu})\). There are 6 more terms defining the \(3 \times 3\) symmetric matrix \(T_{EM}^{ij} ; i, j = 1, 2, 3\) which is called the ‘stress’ tensor as it describes the internal forces of the field distribution. This becomes evident when we consider the spacelike component of Eq. (27.33)

\[ \partial_0 T_{EM}^{0i} + \nabla_k T_{EM}^{ki} = \rho c F_0^i - \vec{j}_k F^{ki}, \]  

(28.9)

or

\[ \frac{1}{c^2} \frac{\partial \vec{S}}{\partial t} + \vec{T} + \rho \vec{E} + \vec{j} \times \vec{B} = 0, \quad \vec{T}_i \equiv \nabla_k T^{ki}. \]  

(28.10)

The first term is the inertial term of the field which balances against the differential form of the Lorentz-forces we are familiar with, while the new middle term \(\vec{T}\) is the self-force due to inhomogeneity of the field.

A a Lorentz-invariant laboratory frame mass density \(\mathcal{M}\) can also be constructed

\[ \mathcal{M} c^2 \equiv \frac{1}{c} \sqrt{u_\mu T_{EM}^{\mu\nu} T_{EM \nu \alpha} u^\alpha}_{|_{R,0}} = \sqrt{(T_{EM}^{00})^2 - (T_{EM}^{0k})^2}, \]  

(28.11)

where the right-hand side is measured by the observer at rest \(R\). Elementary evaluation inserting Eq. (28.2) in Eq. (28.11) yields

\[ (\mathcal{M} c^2)^2 = \frac{\epsilon_0^2}{4} \left( \vec{E}^4 + c^4 \vec{B}^4 \right) + \frac{1}{2} \vec{E}^2 c^2 \vec{B}^2 - \vec{E}^2 c^2 \vec{B}^2 + \left( \vec{E} \cdot c \vec{B} \right)^2. \]  

(28.12)

We thus find the manifestly LI result

\[ \mathcal{M} c^2 = \epsilon_0 \sqrt{\frac{1}{4} \left( \vec{E}^2 - c^2 \vec{B}^2 \right)^2 + \left( \vec{E} \cdot c \vec{B} \right)^2} = \sqrt{\mathcal{S}^2 + \mathcal{P}^2}. \]  

(28.13)

By definition \(\mathcal{M}\) can be considered the local mass density of the field distribution accompanying the energy density \(\varepsilon\) and the momentum density flow \(\vec{S}\). Returning to Eq. (28.11) we note that for any timelike 4-vector \(u^\mu\) one can show by direct computation using the explicit form Eq. (27.34)

\[ \frac{1}{c^2} u_\mu T_{EM}^{\mu\nu} T_{EM \nu \alpha} u^\alpha = (\mathcal{M} c^2)^2 \frac{u^2}{c^2} = (\mathcal{M} c^2)^2, \]  

(28.14)
which is to be expected since this is true for the observer at rest in laboratory and the expression is LI. This result Eq. (28.14) shows that

$$T_{\mu\nu}^{EM} T_{\nu\alpha}^{EM} = \delta^\mu_\alpha (Mc^2)^2.$$  

(28.15)

Examples

a) Light waves: a perfect EM plane wave as introduced in section 24.1 has everywhere $M = 0$. Both energy and momentum of the field do not vanish, thus material bodies never can catch up with this wave.
b) A laser energy pulse is a space-time field configuration that in view of focusing must have aside of an energy content also a distribution $M \neq 0$. A many photon system will in general always carry an invariant mass except for the case that all photons have exactly parallel momentum, the case a) above, see also exercise VIII–9 on page 297. Therefore focused laser pulses travel with slightly subluminal speed.
c) A static ‘Coulomb’ field distribution sourced by a charge distribution at rest in the laboratory has an energy and mass density that will be (up to proportionality factor $c^2$) equal, since $\vec{B} = 0$.

Exercise XI–11: Vacuum inertial part $T^{\mu\nu}_V$

Include in the energy-momentum distribution the term that describes the particle resistance to motion, the ‘vacuum’ mass – the mass that exists in limit of vanishing electron charge.

Solution

All particles carry a mass component that for most part is not directly related to electromagnetic field distribution. We will denote this mass as $m_V$, the vacuum mass. For strongly interacting particles such as protons or neutrons the dominant part of this mass derives from the way strong interactions relate to the quantum vacuum properties, i.e. the phenomenon of quark-confinement. For particles without strong interactions the vacuum Higgs field provides a mechanism to generate $m_V$.

We can construct a Newton 4-force density as a 4-vector by following the example of Eq. (27.6) and summing all particles ‘$i$’

$$f^{\mu}_V(x) \equiv \sum_i \int_{\mathbf{path}_i} d\tau_i \frac{dp_i^\mu}{d\tau_i} \delta^4 (x - z_i(\tau_i)) = - \sum_i m_{V,i} \int_{\mathbf{path}_i} d\tau_i u_i^\mu \frac{d\delta^4 (x - z_i(\tau_i))}{d\tau_i}.$$  

The inertial energy-momentum tensor belonging to this cloud of particles can be guessed by counting dimensions. It must be velocity $\times$ momentum and have two Greek indices, so we try

$$T^{\mu\nu}_V(x) \equiv \sum_i \int_{\mathbf{path}_i} d\tau_i u_i^\mu p_i^\nu \delta^4 (x - z_i(\tau_i)).$$
To check Eq. 2 we evaluate 4-divergence of $T_{\nu}^{\mu
u}$. We obtain changing variables $\partial_x \rightarrow -\partial_z$

$$\partial_{\mu} T_{\nu}^{\mu \nu} = - \sum_i \int_{\text{path}_i} d\tau_i \ p^\mu_i u^\nu_i \frac{\partial \delta^4 (x - z_i(\tau_i))}{\partial z_i^\mu} \ d\tau = - \sum_i \int_{\text{path}_i} d\tau_i \ p^\mu_i \frac{d\delta^4 (x - z_i(\tau_i))}{d\tau_i} \ d\tau .$$

Performing on the right the $\tau$-integration we find

$$\partial_{\mu} T_{\nu}^{\mu \nu} = f^\mu_V \ - \sum_i p^\mu_i \delta^4 (x - z_i(\tau_i)) \bigg|_{\tau = \tau_1}^{\tau = \tau_2} ,$$

where aside of the vacuum force density Eq. 1 we see the change of momentum of the cloud of particles. This is the contribution to particle dynamics by their interaction with surrounding fields. For a particle cloud not subject to any external force this term would vanish. For a particle cloud comprising charged particles, there is an electromagnetic field these particles generate. Therefore, we would need, as presented in Eq. (27.33), to allow for the electromagnetic component of the energy momentum tensor to balance the change in particle momentum Eq. 4.

We therefore consider the joint EM+V energy-momentum conservation law to be

$$\partial_{\mu} \left( T_{\mu \nu}^{\text{EM}} + T_{\mu \nu}^{\text{V}} \right) + F^{\nu \mu} J_{\mu} = f^{\nu}_V .$$

The last two terms are the Lorentz-force written in differential format. Only when omitting the energy-momentum tensors we find

$$\sum_i \int_{\text{path}_i} d\tau_i \left( e_i F^{\nu \mu} u^\mu_i \ - \ m_i \frac{d u^\nu_i}{d\tau_i} \right) \delta^4 (x - z_i(\tau_i)) \ d\tau_i = 0 .$$

Thus without consideration of a possible change in the energy-momentum density $T^{\mu \nu} = T_{\mu \nu}^{\text{EM}} + T_{\mu \nu}^{\text{V}}$, the Lorentz-force determines the dynamics completely. This is not the case in general, a modification arises when the field part $T^{\mu \nu}$, the first term in Eq. 5, varies significantly, e.g. due to radiation emission effect. The energy loss by particles to radiation is balanced by growth in the field term. When looking only at particle dynamics this we can account for this friction by adding suitable effective force, the radiation-reaction force. We return to consider this phenomenon in section 29.5 on page 452.

End XI–11: Vacuum inertial part $T_{\nu}^{\mu \nu}$

### 28.2 Mass of the electric field

According to the Maxwell equations, charged particles at rest in the laboratory are surrounded by a static EM-field. This field contains in the particle rest-frame
energy, and thus an EM-mass. We presented the invariant definition of EM-field mass density in Eq. (28.13). Since the charge of a particle sources the field, the EM-mass is a component of the total particle inertial mass. According to Einstein: “elementary constituents of matter are in their nature nothing else but condensations of the EM-fields” (see section 1.3 on page 9). The only difference between this early comment with the present day understanding is that we know, aside of electro-magnetism, that other interactions can and in general dominate the mass of elementary particles.

An example is the EM-mass component associated with the charge $Z|e|$ of an atomic nucleus. The field energy of a charged nucleus with radius $R \propto A^{1/3}$ is known quantitatively from fits to nuclear masses according to the semi-empirical Bethe-Weizsäcker mass formula. This provides the Coulomb field energy component

$$c^2 \delta M^\text{EM-field}_A = 0.691 \text{MeV} \frac{Z(Z-1)}{A^{1/3}}. \quad (28.16)$$

Each of the protons in the atomic nucleus is shown as experiencing the EM-force of the other $Z-1$ protons. For the lead nucleus $^{208}_{82}$Pb we find a mass excess $c^2\delta M = 776 \text{MeV}$, a contribution at 80% level of the mass-energy content of one nucleon to the total $^{208}_{82}$Pb nuclear mass. Since the EM-energy Eq. (28.16) scales with $Z^2$ for large $Z$, this normally small term overwhelms the nuclear binding, which scales with number of nucleons $A \simeq 2.5Z$. High-$Z$ atomic nuclei are therefore unstable and only elements $Z \leq 118$ have been discovered or created at the time of writing.

There is no doubt whatsoever that the EM field mass is: i) as just described a finite, well-defined component of the nuclear mass; and ii) not localized in space where the atomic nucleus is present. Experiments such as those based on Fermi’s equivalent photon method\footnote{For a recent review, see: C. A. Bertulani, S. R. Klein and J. Nystrand, Physics of ultra-peripheral nuclear collisions, Ann. Rev. Nucl. Part. Sci. 55 (2005) 271.} show that this EM-field part of the nuclear mass is largely separate in space from the 99.6% of the energy-mass localized over the small spatial domain of the atomic nucleus.

The energy and mass density associated with the Coulomb field is delocalized over the large volume filled by $(\vec{E}_c)^2$

$$m^c c^2 \equiv \frac{e_0}{2} \int d^3 x \vec{E}_c^2,$$  \quad (28.17)

which is following from Eq. (28.13) for the case $\vec{B} = 0$. There is sometimes a misunderstanding which originates in the following: the Maxwell equation for the Coulomb field Eq. (27.10) has as the general solution

$$\vec{E}_c = -\int d^3 x' \rho(x') \frac{1}{4\pi\epsilon_0} \nabla \frac{1}{|\vec{x} - \vec{x}'|}, \quad (28.18)$$
compare also with the special case of point-source we discussed in exercise XI–2.
We verify Eq. (28.18)

\[
\vec{\nabla} \cdot \vec{E}_c = -\int d^3x' \frac{\rho(\vec{x}')}{4\pi\epsilon_0} \frac{1}{|\vec{x} - \vec{x}'|} = -\int d^3x' \frac{\rho(\vec{x}')}{4\pi\epsilon_0} \left(-4\pi\delta^3(\vec{x} - \vec{x}')\right) = \frac{\rho(\vec{x})}{\epsilon_0},
\]

(28.19)

which is the Maxwell equation Eq. (27.10). Insertion of the solution Eq. (28.18) for \(\vec{E}_c\) into Eq. (28.17) leads to a form also describing the Coulomb field energy content

\[
m^C c^2 \equiv \frac{\epsilon_0}{2} \int d^3x \vec{E}_c^2 \rightarrow \frac{1}{8\pi\epsilon_0} \int \int d^3x'd^3x \frac{1}{|\vec{x}' - \vec{x}|} \rho(x) .
\]

(28.20)

While this expression on the right yields the same numerical value for the energy content of the field, it does not provide insight about where the energy density that creates the integral value \(m^C c^2\) originates\(^9\). This result that has as sole objective simplification of the evaluation effort by converting a 3-dimensional delocalized integrand on the left to the 6-dimensional but ‘localized’ in space form on the right.

Of special interest to us is the field energy contribution to the mass-energy equivalent of a nearly point-like electron – the computation proceeds just like the example presented in exercise XI–12 below for the proton

\[
\delta m^C c^2 \equiv \frac{3}{5} \frac{r_e}{R_e} m_e c^2
\]

(28.21)

where \(R_e\) is the yet unknown electron uniform charge distribution radius. If we argue that the origin of electron mass is entirely in the EM-field, we find setting \(\delta m^C c^2 \rightarrow m_e c^2\)

\[
R_m = \frac{3}{5} r_e = 0.6 \times 2.82 = 1.7 \text{ fm} = \frac{3}{5} \lambda_C \alpha ,
\]

(28.22)

where for later convenience we also show the answer in terms of the Compton wavelength \(\lambda_C \equiv h/mc \simeq 386 \text{ fm}\), and \(e^2/4\pi\epsilon_0hc \equiv \alpha \simeq 1/137\), the fine-structure constant, compare Eq. (19.43).

Many experiments were carried out to determine the electron charge distribution. A charge distribution electron radius of magnitude \(r_e = 1.7 \text{ fm}\) is untenable by at least two orders of magnitude. Conversely, given that the classical field energy scales inversely with electron radius, for such a small electron a field energy 100s times greater than electron mass would follow. This clearly shows that the classical size radius of electron charge distribution constrained by known charge structure limits is incompatible with the magnitude of the field energy.

The difference between the EM-field energy of a heavy bound particle, such as a nucleus, or even a proton, and the light electron is the quantum de-localization. In qualitative terms one can argue that the Heisenberg uncertainty principle implies uncertainty in the measurement of the location of a (charged) particle at the scale of the Compton wavelength, \( \lambda_c = \frac{\hbar}{mc} \). For a heavy nucleus, and even for a proton with \( \lambda_{p,c} = 0.21 \text{ fm} \) compared to the charge distribution size \( R_p = 0.85 \text{ fm} \) is sufficiently large. We can compute the field energy of the heavy atomic nucleus reliably in the classical limit because the particle is well localized and its material charge density sources the field.

For an electron \( \lambda_c = 386 \text{ fm} \), we see that quantum size is more than 100 times greater than the classical electron radius \( r_e \). This means that the EM-mass component of the electron must be considered in the context of quantum physics. Taking for the electron as the relevant charge radius the quantum delocalization \( \lambda_{e,c} \), we recognize that the EM-mass component of the electron for a uniform charge distribution as in Eq. (28.22), should have the magnitude

\[
m^e E_c^2 \simeq \frac{3}{5} \frac{e^2}{\lambda_{e,c}} = \frac{3\alpha m_e c^2}{5} \simeq 2 \text{keV}.
\]

Isolation of this finite quantum EM-field-energy mass component of the electron in the context of quantum-electrodynamics (QED), the theoretical framework that in principle should be capable of accomplishing this task has not been presented: The EM-field mass remains divergent in the realm of perturbative QED. This context provides an interesting foundational and unresolved question, both in theory and in experiment, about the EM-field mass component of an electron.

---

**Exercise XI–12: Coulomb field energy of the proton**

Consider a proton, a particle of charge \( e \), mass \( M c^2 = 938.3 \text{ MeV} \), and quark-structure radius of \( \simeq 1 \text{ fm} \), hence the equivalent uniformly charged sphere of a Radius \( R_p \simeq \sqrt{3/5} = 0.81 \text{ fm} \); the actual experimental charge mean square radius is reported between \( R_p \in (0.84, 0.87) \text{ fm} \)\(^{10}\). Obtain the Coulomb field energies associated with the spatial domains outside, and inside, the proton.

**Solution**

A proton can be localized in space with the precision

\[
\Delta r = \frac{\hbar c}{m_p c^2} = 0.21 \text{ fm} < R_p,
\]

and thus an estimate of the classical field energy based on a fixed-size classical object is possible. We consider a radial electrical field

$$\vec{E}^c = \mathcal{E}_r(r)\hat{r}.$$ 

The solution of Eq. (27.10) for the charge distribution of a homogeneous charged sphere of radius $R$ is

$$\mathcal{E}_r = \begin{cases} \frac{e_p}{4\pi\epsilon_0} \frac{Z}{R^2} \frac{r}{R}, & \text{for } 0 \leq r \leq R, \\ \frac{e_p}{4\pi\epsilon_0} \frac{Z}{r^2}, & \text{for } R \leq r. \end{cases}$$

The result outside the range of charge distribution is as described in exercise [XI-2] inside the charge distribution the Gaussian sphere will enclose less of the charge so the radial field must be smaller, for the normalized uniform charge distribution there is a factor $(|\vec{x}|/R)^3$ which upon division by the Gaussian surface $4\pi|\vec{x}|^2$ produces Eq. 3.

The field energy density according to Eq. (28.18) thus is

$$\frac{\epsilon_0}{2} \mathcal{E}_r^2 = \begin{cases} \frac{e_p^2}{4\pi\epsilon_0} \frac{Z^2}{R^6} \frac{r^2}{8\pi}, & \text{for } 0 \leq r \leq R, \\ \frac{e_p^2}{4\pi\epsilon_0} \frac{Z^2}{8\pi r^4}, & \text{for } R \leq r. \end{cases}$$

Upon elementary integration with $\int d^3xf \rightarrow 4\pi \int r^2 f(r)dr$ of the energy density indicated in Eq. 4 we find the two contributions, from within the charge sphere ($<$) containing $Z$-elementary charges, and from outside (>) the charge distribution to be

$$m^Cc^2 = \frac{e_p^2}{4\pi\epsilon_0} \frac{Z^2}{2R} \left[ \left( \frac{1}{5} \right)_< + 1 > \right], \quad \frac{e_p^2}{4\pi\epsilon_0} = m_ec^2 \frac{e^2}{4\pi\epsilon_0 m_ec^2} = m_ec^2 r_e.$$

Taking the values for the proton as $Z = 1, R_p = 0.85 \text{ fm}$, with $m_ec^2 = 0.511 \text{ MeV}$ and $r_e = 2.82 \text{ fm}$ the proton Coulomb field energy is

$$m_<^Cc^2 = 0.170 \text{ MeV} \quad m_>^Cc^2 = 0.848 \text{ MeV} \quad m_p^C = 1.02 \text{ MeV}.$$ 

This is an upper estimate, since the quantum delocalization may increase by 10-20% the effective radius $R$ of the charge distribution to be used in Eq. 6. There is no such Coulomb energy originating mass shift for the sibling of the proton, the charge neutral neutron. However instead of being heavier, the proton is $1.293 \text{ MeV/}c^2$ lighter compared to the neutron. This means that the two other effects which are both present for protons and neutrons, a) magnetic EM-field contributions, and b) structural quark mass effects, contribute to the neutron - proton mass difference a mass increment $\delta (m_n - m_p) c^2 \simeq 2.3 \text{ MeV}.$
Exercise XI–13: The EM-field energy of a bound system

Consider e.g. a neutral hydrogen atom, a bound state of a proton and an electron. The electron neutralizes the electrical field of the proton. What is the effect of the field screening – is there a polarization effect that changes the mass of the proton? How do we explain the two particle interaction potential?

Solution

Since the Maxwell equations in the vacuum are linear in the fields, the superposition principle holds, meaning that the field of two sources superposes linearly.

1 \[ \vec{E}_{12} = \vec{E}_1 + \vec{E}_2. \]

Upon insertion of Eq. 1 in Eq. (28.20), and the integration of particle ‘1’ and ‘2’ terms, we are left with bound system mass

2 \[ m_{12}c^2 = m_1c^2 + m_2c^2 + \epsilon_0 \int d^3x \vec{E}_1 \cdot \vec{E}_2, \]

where the terms in order \( \int d^3x_0 (\vec{E})^2_{1,2}/2 \) are a part of the inertial masses \( m_{1,2} \).

Thus we find that the effect of the field screening is

3 \[ \delta mc^2 = m_{12}c^2 - m_1c^2 - m_2c^2 = -\epsilon_0 \int d^3x \vec{V}_1 \cdot \vec{E}_2 = \int d^3x \vec{V}_1 \vec{\nabla} \cdot \vec{D}_2, \]

where to obtain the last term we used Eq. (23.6) integrating by parts and substituting \( \epsilon_0 \vec{E} \rightarrow \vec{D} \). Up to this point the expressions are equivalent to the original form, showing that the energy is localized in the space filled with the Coulomb field and potential.

In the next step we use the Maxwell equation (Coulomb law) Eq. (27.10)

4 \[ \delta mc^2 = \int d^3x V_1 \rho_2 \quad \Rightarrow \quad \delta mc^2 = \int d^3x V_1 \psi_2^\ast \psi_2, \]

where in last step we show how the Coulomb potential of the proton (particle ‘1’) combines with the electron density (particle ‘2’) expressed by quantum wave function \( \psi \). This is the term that we use in quantum mechanics solving the Schrodinger equation to obtain atomic orbits. The electron binding to the proton by \( E_{1s} = -13.6 \text{ eV} \) is thus originating in the reduction in the electrical field mass due to screening of the field lines of the proton by the opposite charge of the electron.

This shows how the superposition principle of fields in the vacuum allows us to separate the polarization-screening effect recognized as the potential we use to evaluate the electron binding to the proton in the realm of quantum theory. Beyond the term Eq. 4, the screening effect of one particle on another does not affect the EM-mass of any of the two particles.
This argument can be extended to any number of charged particles: all charges present contribute to the electrical field according to the superposition

$$\vec{E}_n = \sum_{i=1}^{n} \vec{E}_i, \quad \epsilon_0 \nabla \cdot \vec{E}_i = \rho_i,$$

with the total contribution to mass energy equivalent of all particles being

$$\delta m c^2 = \frac{\epsilon_0}{2} \int d^3x (\vec{E}_n)^2 = \sum_{i=1}^{n} \frac{\epsilon_0}{2} \int d^3x (\vec{E}_i)^2 + \sum_{i \neq j} \epsilon_0 \int d^3x \vec{E}_i \cdot \vec{E}_j.$$

The first term on the right is the Coulomb-self energy of each particle incorporated already in the particle mass, and the last term describes how particle $i$ responds to particle $j$ (and vice-versa) according to

$$\delta m_{ij} c^2 = -\epsilon_0 \int d^3x \vec{V}_i \cdot \vec{E}_j = \int d^3x \vec{V}_i \nabla \cdot \vec{D}_j = \int d^3x \vec{V}_j \nabla \cdot \vec{D}_i.$$

28.3 Limiting field/force electromagnetism

Max Born\textsuperscript{[11]} proposed a nonlinear theory of electromagnetism by considering the possibility that akin to a maximum velocity in special relativity, there should be a maximum field strength and thus maximal force and maximum acceleration a charged particle can experience. The Born-Infeld\textsuperscript{[12]} (BI) theory of electromagnetism\textsuperscript{[13]} is a nonlinear generalization of the Maxwell theory that further renders the EM-field mass of charged classical point-particles finite – the electron mass can be entirely the finite energy content of the EM-field.

In order to be able to implement these ambitious objectives the field action $\mathcal{I}_F$ must assume a nonlinear form similar to the situation in a medium. A particularly elegant form that is also connected to gravity was invented, where the usual EM-field action is modified according to

$$L_{\text{field}} = \frac{\epsilon_0}{2} (\mathcal{E}^2 - c^2 \mathcal{B}^2) \rightarrow L_{\text{BI}}^{\text{BI}} = -\epsilon_0 E_{\text{BI}}^2 \left( \sqrt{-\det G} - \sqrt{-\det g} \right), \quad (28.24)$$

\textsuperscript{11}Max Born (1882 –1970) has been among the most influential theoretical physicists rocking the cradle of quantum theory. Einstein on Born: “Theoretical physics will flourish wherever you happen to be...”. Nobel prize 1954 for interpretation of quantum wave function.

\textsuperscript{12}Leopold Infeld (1898 – 1968) associate of Einstein and founder of Warsaw school of theoretical physics.

where we see the maximum allowed EM-field $E_{\text{BI}}^2$, the metric $g = (g_{\mu\nu})$ and its extension $G = (G_{\mu\nu})$

$$G_{\mu\nu} = g_{\mu\nu} + \frac{F_{\mu\nu}}{E_{\text{BI}}^2/c}, \quad -\det G = 1 - \frac{\mathcal{E}^2 - c^2\mathcal{B}^2}{E_{\text{BI}}^2} - \frac{\mathcal{E} \cdot c\mathcal{B}^2}{E_{\text{BI}}^4}. \quad (28.25)$$

We evaluated in Eq. (28.25) the determinant of $G_{\mu\nu}$ in flat space where $\sqrt{-\det g} = 1$, see Eq. (20.7).

Like in any other theory of nonlinear electromagnetism, introduced in the conventional description of material bodies, the Maxwell equations with sources apply to the electric displacement fields $\vec{D}$ and magnetic field $\vec{H}$. These are partial derivatives with respect to the electric field $\vec{E}$, and the magnetic induction field $\vec{B}$, respectively, of the nonlinear action density $\mathcal{L}^{\text{BI}}$. On the other hand, the source-less equations between the fields $\vec{E}, \vec{B}$, Eq. (23.7) are unchanged. These are consequence of the definition of these fields in terms of potentials. Similarly, the Lorentz-force is not modified as it arises from the particle part of the action that is not impacted by introduction of the non-linear in EM-fields action.

The nonlinear relation between $\vec{D}, \vec{H}$ and $\vec{E}, \vec{B}$ within the Born-Infeld theory imposes a limit on the field strength prescribed by the quantity $E_{\text{BI}}$. This is an upper limit on the force acting on a charged particle, and thus on the acceleration a charged particle can experience. In the simplified situation of $\vec{H} = 0 = \vec{B}$ one obtains from Eq. (28.24) for the radial component of the electrical field $\mathcal{E}_r$ in terms of the radial displacement field $D_r$ driven by a spherical charge distribution, in one line

$$\frac{\partial \mathcal{L}^{\text{BI}}}{\partial \mathcal{E}_r} = \frac{\epsilon_0 \mathcal{E}_r}{\sqrt{1 - \mathcal{E}^2/E_{\text{BI}}^2}} \equiv D_r \Rightarrow \mathcal{E}_r = \frac{D_r/\epsilon_0}{\sqrt{1 + D_r^2/\epsilon_0^2E_{\text{BI}}^2}}$$

where for a point charge

$$D_r = \frac{e}{4\pi|\vec{x}|^2}, \quad \text{for} \quad \rho = e\delta^3(\vec{x}). \quad (28.26)$$

We see that as $D_r$ increases without bound, there is a limit $\mathcal{E}_r \leq E_{\text{BI}}$. We will use a further well known result of nonlinear EM theory: the energy-momentum tensor Eq. (27.29) is modified to include the nonlinear effects

Born proposed that the value of the limiting field $E_{\text{BI}}$ should be determined by requiring that the mass of the electron is equal to its field energy. Using the generalized Eq. (28.20) this condition reads for $\vec{H} = 0 = \vec{B}$

$$m_{\text{BI}}c^2 \equiv \frac{1}{2} \int d^3x \left\{ \epsilon_0 E_{\text{BI}}^2 \left( \sqrt{1 - \mathcal{E}^2/E_{\text{BI}}^2} - 1 \right) + \mathcal{E} \cdot \mathcal{D} \right\} = m_e c^2. \quad (28.28)$$

Using Eq. (28.26)

\[ m^{\text{BI}} c^2 = \epsilon_0 E_{\text{BI}}^2 \int d^3 x \left( \sqrt{1 + \frac{\mathcal{D}_r^2}{\epsilon_0 E_{\text{BI}}^2}} - 1 \right). \]  

(28.29)

It is now convenient to define a length parameter \( r_{\text{BI}} \) defining the BI limiting field

\[ E_{\text{BI}} = \frac{e}{4\pi \epsilon_0} \frac{1}{r_{\text{BI}}^2}, \quad \epsilon_0 E_{\text{BI}}^2 = \frac{1}{4\pi} \frac{e^2}{r_{\text{BI}}^4}. \]  

(28.30)

Using Eq. (28.27) we see that

\[ \frac{\mathcal{D}_r^2}{\epsilon_0 E_{\text{BI}}^2} = \left( \frac{r_{\text{BI}}}{|\vec{x}|} \right). \]  

(28.31)

We now obtain for Eq. (28.29) entering Eq. (28.30), Eq. (28.31) and substituting \(|\vec{x}| = s r_{\text{BI}}\)

\[ m^{\text{BI}} c^2 = \frac{1}{4\pi} \frac{e^2}{4\pi \epsilon_0} \frac{1}{r_{\text{BI}}^4} 4\pi r_{\text{BI}}^3 \int_0^\infty s^2 ds \left( \sqrt{1 + 1/s^4} - 1 \right). \]  

(28.32)

We recall that \( e^2/4\pi \epsilon_0 = m_e c^2 r_e \) and obtain

\[ m^{\text{BI}} c^2 = m_e c^2 \frac{r_e}{r_{\text{BI}}} \int_0^\infty ds \left( \sqrt{s^4 + 1} - s^2 \right) = m_e c^2 \frac{r_e}{r_{\text{BI}}} \int_0^\infty ds \frac{\sqrt{s^4 + 1} + s^2}{s^4 + 1 + s^2}. \]  

(28.33)

where the last equality follows by multiplication and division of the integrand by \( \sqrt{s^4 + 1} + s^2 \). The integral can be also related to known tabulated elliptic integrals

\[ m^{\text{BI}} c^2 = m_e c^2 \frac{r_e}{r_{\text{BI}}} 1.23605 \ldots . \]  

(28.34)

Setting this to the electron mass-energy equivalent we find

\[ m^{\text{BI}} c^2 = m_e c^2 \frac{r_e}{r_{\text{BI}}} \rightarrow r_{\text{BI}} = 1.23605 r_e = 3.483 \text{ fm}, \]  

(28.35)

To obtain the actual value of \( E_{\text{BI}} \) we use the first form of Eq. (28.30) to obtain using Eq. (23.4)

\[ E_{\text{BI}} = \frac{1.4403 \text{ MV fm}}{(3.483 \text{ fm})^2} = 1.187 \times 10^{20} \text{ V/m}. \]  

(28.36)

When we try to interpret the mass of the electron as being the EM-field mass we find the smallest value of \( E_{\text{BI}} \) possible, as this is the elementary particle with smallest mass and largest (elementary) charge.

Such value of a limiting field Eq. (28.36) is impossible for two reasons: i) An electron has the effective electron charge distribution inherent in the definition

\[ \rho \equiv \epsilon_0 \vec{\nabla} \cdot \vec{E} \Rightarrow \rho_r = \frac{1}{r^2} \frac{\partial}{\partial r} \frac{r^2 D_r}{\sqrt{1 + \frac{D_r^2}{\epsilon_0 E_{\text{BI}}^2}}} = \frac{e}{2\pi r_{\text{BI}}^3} \frac{1}{(1 + (r/r_{\text{BI}})^4)^{3/2}}. \]  

(28.37)
where we used Eq. (28.26). This charge distribution is characterized by size by $r_{BI} = 3.483 \text{ fm}$. This conflicts with experiments failing to see any charge distribution. ii) The electrical fields at the surface of nuclear charge distribution that govern the computed EM energy content of the atomic nuclei for $Z \geq 80$, see Eq. (28.16), are $20=80/2^2$ times greater. This value arises considering the nuclear surface field

$$E_r = \frac{|e|}{4\pi\epsilon_0} \frac{Z}{R^2} \to \frac{1.44 \text{ MV m} \times 80}{(7 \times 10^{-18} \text{ m})^2} = 2.35 \times 10^{21} \frac{\text{V}}{\text{m}}, \quad R \approx 1.1 A^{1/3} < 2r_0.$$

(28.38)

As is well known, muons (heavy electrons) when bound to heavy nuclei probe the EM interaction near the nuclear surface. This means that in order to continue considering the BI-model with a finite classical EM field energy we must explore values of the limiting field $E_{BI}$ which are significantly higher\textsuperscript{15}. In order to avoid conflict with the nuclear EM-field energy we would need $E_{BI}$ to be at least a factor 100 times larger than the value shown in Eq. (28.36). Note that the value of the field integral in Eq. (28.34) is scaling as $E_{BI}^{1/2}$. This implies that requiring a 100 times bigger field to accommodate nuclear EM energy yields an EM field energy content that is $10 \times m_e c^2$, thus well above the observed mass of the electron – however allowing for quantum delocalization, see Eq. (28.23), the situation for the quantum electron remains unchanged.

According to Eq. (28.35) given the point nature of many elementary particles, the scale of effective charge distribution size is about a factor 100 or more too large. To resolve this we would need to increase the $E_{BI}$ field strength by factor $10^4$, to the scale of $10^{24} \text{ V/m}$, which also means that the EM field energy of a point particle would rise to be at the scale of 50 MeV. However, for most elementary particles the quantum delocalization reduces this value and there is no evident conflict of a BI theory with $E_{BI} \simeq 10^{24} \text{ V/m}$ with any experimental data.

The issue is how we can create a quantum model to evaluate the EM field energy, and continuing this thought, is there really need to assure that the classical EM field energy is finite? We keep in mind that a limit on the strength of the electrical field implies that a charged particle cannot be accelerated with arbitrarily high force, a topic that is of importance in the following section where we consider the impact of radiation friction on particle dynamics.

29 Afterword: Acceleration

29.1 Can there be acceleration in SR?

Let us briefly recapitulate a few insights we have presented in this book. We recognized that an accelerated material body differs from any inertial reference observer, and there is no relativity principle that applies to accelerated motion:

- It is the traveling-accelerated ‘twin’ who upon rejoining her base notices a shorter increment of proper time measured by her clock compared to the proper time ticking in the base station;
- It is the train departing a station that will upon reaching relativistic speed fit into the usually not long enough tunnel.

The consistency of any changes of the body properties with the Lorentz transformation emerges when we implement the measurement process appropriate for the circumstance. We recall that the spatial separation between two events measured at equal time in observer’s frame of reference agrees with the Lorentz-FitzGerald contracted body that can be placed between these events, see section 9. Similarly, we can measure that a moving clock with $\Delta x' \neq 0$ ticks slower by observing it from a fixed $\Delta x = 0$ base location, see section 12.3.

Upon making these statements we need to recognize that the special theory of relativity does not incorporate explicitly the notion of accelerated material bodies. Acceleration is added in when we postulate the existence of Lorentz-force. Introduction of this ad-hoc force is on first sight inconsistent with two accomplished theoretical frameworks, general relativity (GR) and quantum mechanics (QM), which both are foundational theories without acceleration. Here is how this happens:

**GR:** By recognizing the proportionality of inertial and gravitational mass (equivalence principle), Einstein could describe the force of gravity as quasi-inertial: a free-falling observer in a gravity field does not locally experience any force at all. In a second step, Einstein’s GR equations introduce space-time deformation by the action of the energy-momentum of all particles, allowing this back-reaction modification of the space-time geometry. The free-falling particles follow trajectories in a curved space-time showing the properties we associate with motion executed according to the effect of gravity.

Note that the force of gravity we feel is due to the ‘surface normal’ force that keeps us from free-falling. We therefore report being subject to Earth’s force. In classical point-particle model of matter all particles are free-falling and there is no force.

**QM:** Quantum mechanics describes particles on elementary scale. The electron in a quantum orbit is stationary, and not accelerated; hence there is no
continuous emission of radiation; an electron does not spiral into the nucleus. A modification needed to assure the stability of atoms against acceleration induced radiation had motivated in the first place the discovery of the atomic quantum theory. To clarify, when an excited atomic state decays emitting a photon, this is like an unstable particle radioactive decay, not a result of continuous ‘synchrotron’ radiation due to accelerated motion.

29.2 Evidence for acceleration

Theories are built without acceleration and yet our daily experience contradicts this. One is justified to ask if there is acceleration at all – and this is synonymous with the question - how do we know if a body is accelerated? This is not a new question. Newton explored the meaning of acceleration in his widely quoted water bucket experiment and more than 300 years later this is still a good place to start. Consider a suspended rotating water bucket – at first the water in the bucket does not rotate with the bucket. When the water begins to rotate, the surface of the water becomes concave.

Newton concluded that by observing the curving of water surface in a rotating bucket, one can determine that the bucket is being rotated and not the rest of the world around the bucket. If you have doubts about Newton’s conclusions, place a water bucket in the immobile center of a merry-go-round and while enjoying a ride see if the water surface turns concave. However, some will still claim that this demonstration is flawed since the experiment was not carried out in an inertial frame in empty space.

For this reason Newton supplemented a real bucket experiment with what we today call ‘thought experiments’. He considered two rocks at the end of a rope in empty space, far from everything. When they rotate around, the rocks will create an outward force pulling the rope tight. In empty space there is nothing to define and measure the rotation as occurring with respect to, except for space itself. Since we expect the rope will record a tension, space itself provides a reference. Such (thought) experiments led Newton to propose the existence of absolute space.

This was the situation when nearly 150 years ago Ernst Mach\textsuperscript{16} returned to the question raised by Newton. Mach proposed the ‘Universe at rest’, defined by the fixed stars, to be Newton’s absolute space. Adapting these ideas to the post-relativity context of this book we would say that we always measure acceleration against any known inertial frame, and the fixed star frame is a suitable inertial reference frame. Therefore choice of a preferred inertial frame such as that proposed by Mach is not in conflict with special relativity. This differs from a

\textsuperscript{16}Ernst Mach (1838 – 1916), Professor at Graz, Salzburg, Prague (for most of his life), and Vienna; remembered for Mach number, shock waves, and Mach’s principle.
preferred origin in Newton’s absolute space required in the heliocentric ‘Weltbild’ from the age of Newton, a notion that is manifestly incompatible with SR.

The choice of an inertial frame made by Mach makes good sense for the merry-go-round experiment in space, far from anything. By looking at starlight we can determine who is rotating and who is not. Now the entire Universe dictates the answer to who is accelerated. However, this also prompts us to reflect on the question: “How does the water bucket know about the universal inertial frame?” Especially in a book where we often check if a causal sequence of events is guaranteed, it seems doubtful that this information is transferred from far and away location of fixed stars, ‘informing’ the water bucket to curve its surface. This information about who or what is accelerated has to be available locally. Within the laws of classical physics this is a more difficult to fulfill requirement compared to quantum physics where we have accepted that the quantum vacuum state is structured, a topic that is now of importance but transcends this book’s scope.\(^{17}\)

One can see the situation as follows: the same quantum vacuum state properties that assure quark confinement, and predominantly define the inertial mass of matter, also define locally at a very microscopic scale the inertial frame of reference. In this way the structured quantum vacuum supplants Einstein’s view of the universal relativistically invariant æther we have described in this book’s first pages. Taking this argument further, the reader should wonder in what way the gravitational radiation friction we introduce below relies on concepts transcending the ideas of free-falling mass in that we accept the universal presence of a class of reference frames we call inertial that allow an objective characterization of motion as being non-inertial.

The author of this book has not found one student who would disagree with the fact that acceleration exists. But which experiment shows this unequivocally? All will agree that charged particles in inertial motion cannot emit radiation. The argument that will be developed in these pages is that emission of radiation by charged particles is the experimental evidence for presence of accelerated motion. I fact I will first argue that emission of gravitational radiation by accelerated massive bodies provides decisive evidence in recognizing that local acceleration exists. I turn to gravity first since everybody is in agreement that gravitational phenomena are in the classical domain that this books addresses. For the interested reader we refer to discussion of the contextual story about gravitational radiation emission presented recently by Poisson and Will.\(^{18}\)


Today we have convincing experimental evidence that gravitational radiation-friction exists:

i) Based on long term analysis of relativistic radio pulsars in binary systems there is agreement that the orbit change observed is a consequence of radiative energy loss. The experimental data showing how excess orbital phase (relative to an unchanging orbit) has accumulated is displayed in figure 29-1 based on 30 years of observation of a relativistic binary pulsar B1913+16. This binary system will spiral inward and crash in about 250 million years.

ii) The LIGO-VIRGO collaboration announced in February 2016 the event GW150914, the first detection of gravitational waves emitted during the last stage of inspiraling in a very massive binary – it seems there is no sufficient data evidence to call the compact objects in the event GW150914 (September 14, 2015 event) a black hole. A second GW151226 event confirms the evidence for compact massive object mergers.

Here are a few important insights that we can draw from these experiments:

i) Acceleration (just like deceleration) cannot be relative for if it were, how could gravitational radiation-friction arise? ii) Any two orbiting objects in space have a well defined acceleration with regard to the inertial Universe. iii) Emission of gravitational radiation experiments demonstrate that tiny radiation-friction effects accumulate. iv) Radiation-friction can further reinforce the deceleration radiation-friction, with the final ultra-strong collapse becoming observable over a cosmic scale distance.

In summary, the insights about SR and acceleration are:

1. Acceleration can be measured against any inertial reference frame. The cosmic reference frames, such as Mach’s fixed star frame of reference, or microwave background (CMB) frame – i.e. the frame in which the CMB spectrum is isotropic – are usually considered.

2. The 4-vector of acceleration section provides local invariant measure describing the magnitude of the spacelike acceleration. This allows all inertial observers to agree to a common way of establishing the magnitude of acceleration. Any inertial observer recognizes the accelerated observer to be different. There is no relativity of acceleration.

---


3. Acceleration is defined at the body location in space-time, in a process that manifestly does not depend on the global Universe reference frame. The structured quantum vacuum frame provides the required local point of reference. This is a more general reference point compared to the cosmic reference frames: Just like in Einstein’s æther, the concept of local velocity cannot be associated with the structured quantum vacuum reference frame\textsuperscript{23}.

\section*{29.3 Strong acceleration}

In section 22.2 we introduced ‘small acceleration’ into SR comparing to natural unit-1 strong acceleration Eq. (22.4). Since we just considered gravitational radiation-friction it is natural to ask when and how can such acceleration arise,

\textsuperscript{23}It is generally believed that dominant component in the gravitating energy of the Universe is originating in the vacuum structure. This ‘dark’ energy cannot thus be moved and/or concentrated.
beginning with Newton’s force. Since this force increases as distance decreases we need some very short and elementary unit of length. The shortest known elementary length is the Planck length

\[
\ell_P = \sqrt{\frac{\hbar G_N}{c^3}} \equiv 1.6162 \times 10^{-35} \text{ m}, \tag{29.1}
\]

twenty orders of magnitude below the proton size. Today it is thought that near \(\ell_P\) gravity connects with quantum physics. Planck noted introducing \(\ell_P\) in an added-in appendix\(^{24}\) that such a new elementary scale is of interest as it arises in consequence of the introduction of the quantum of radiation \(\hbar\) and connects with Newton’s Gravity.

We consider the Newton gravity self-acceleration by any particle of mass \(m\) at the Planck distance \(\ell_P\):

\[
a_P = \frac{G_N m}{\ell_P^2} = mc^2 \frac{c}{\hbar} \equiv a_{cr}. \tag{29.2}
\]

We note the cancellation of Newton’s constant \(G_N\) and recognize \(a_P\) to be the unit acceleration \(a_{cr}\) we have introduced in Eq. (22.4). Thus where quantum and gravity phenomena are expected to meet, a unit-1 acceleration is present. In this sense the unit-1 acceleration is connected to Planck’s natural scales but it does not require \(G_N\). Moreover it is in realm of the possible that when we achieve unit-1 acceleration, some new phenomena appear, reminding us of the deeper connection that Eq. (29.2) brings to mind.

Thus there is profound foundational interest in achieving unit-1 acceleration. An electrical field required to generate this critical acceleration for the lightest elementary particle, the electron, would have the so-called ‘Schwinger critical’ field strength\(^{25}\)

\[
E_{cr} = \frac{ma_{cr}}{e} = \frac{m c^2}{e \hbar} = \frac{m c^2}{e \lambda_C} = 1.323 \times 10^{18} \text{ V/m}. \tag{29.3}
\]

The corresponding critical magnetic field is\(^{26}\)

\[
B_{cr} \equiv \frac{ma_{cr}}{ce} = \frac{m c^2}{e \hbar} = 4.414 \times 10^9 \text{ T}. \tag{29.4}
\]

\(^{24}\)M. Planck, “Über irreversible Strahlungsvorgänge,” (translated: “On irreversibility of radiation processes”) Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin 5 440 (1899), see last page 480

\(^{25}\)To evaluate the value \(E_{cr}\) in SI units i.e. V/m we insert for the elementary constants in Eq. (29.3) the SI values \(m[\text{kg}]\), \(c[\text{m/s}]\), \(e[\text{C}]\), \(\hbar[\text{Js}]\) and compute (note ‘C’ stands here for ‘Coulomb’). However it is by far simpler to remember that \(mc^2 = 0.511\text{ MeV}\) and to cancel the \(e\) between numerator and denominator. The numerical coefficient in Eq. (29.3) is recognized as \(511/386\), the last number being \(\lambda_C\) in fm, while the power comes from remembering that ‘M’ stands for mega=\(10^6\) and that fm=\(10^{-15}\text{ m}\).

\(^{26}\)We need to divide the SI value of \(E_{cr}\) by \(c = 3 \times 10^8\text{ m/s}\) to find the SI value of the magnetic field, see Insight on page 320.
At this point we note that there is actually a hierarchy of critical accelerations. For example we could ask what force do we need to apply to rip apart an atom? The atom size is larger by a factor corresponding to the inverse fine-structure constant, $\alpha^{-1} \simeq 137.036$. Thus the acceleration, and hence the field required to ‘critically’ accelerate atoms is of magnitude

$$E_{\text{Atom}}^{\text{cr}} = \frac{m\alpha a_{\text{cr}}}{e} = \frac{m^2 c^3}{e\hbar} = \frac{mc^2}{e(\alpha^{-1}\lambda_C)} = 0.965 \times 10^{16} \frac{V}{m},$$

(29.5)

and

$$B_{\text{Atom}}^{\text{cr}} \equiv \frac{m\alpha a_{\text{cr}}}{ec} = \frac{m^2 c^2}{e\hbar} = 3.208 \times 10^7 \text{T}.$$  

(29.6)

On the other hand fields that can rip apart an elementary particle must address distances at the scale of classical electron radius $r_e$, a length scale that is by the same factor $\alpha^{-1} \simeq 137.036$ smaller, requiring stronger fields

$$E_{\text{Part}}^{\text{cr}} = \frac{m\alpha^{-1} a_{\text{cr}}}{e} = 1.813 \times 10^{20} \frac{V}{m}, \quad B_{\text{Part}}^{\text{cr}} = 6.05 \times 10^{11} \text{T}.$$  

(29.7)

Achieving such fields maybe impossible by laboratory devices including pulsed lasers as we encounter on the way vacuum instabilities related to the critical fields as presented in Eq. (29.3). One can imagine that a magnetized neutron star, a ‘magnetar’, could offer an astrophysical laboratory where such magnetic fields Eq. (29.7) are possibly present, and thus protons and neutrons are ripped apart into quarks and gluons. Alternatively, in the context of strong interaction physics and the formation of quark-gluon plasma in relativistic heavy ion collisions, this estimate produces and order of magnitude for the required fields in the context of ‘chromodynamics’.

The Schwinger critical electric field Eq. (29.3) characterizes in quantum electrodynamics the condition when spontaneous pair production of particles of mass $m$ is so abundant that the field is rapidly neutralized. Said differently, the field energy ‘materializes’ – achieving such an electric field in static condition in laboratory is thus, in principle, not possible. However the equivalent static critical magnetic field can be in principle created.

The laboratory acceleration achieved today is unobservable seen from the perspective of the scale considerations above: the largest bending magnetic field in accelerators is typically only a few Tesla, and certainly in the foreseeable future below 10s of Tesla. The laboratory bending magnetic field is 8 orders of magnitude below the critical acceleration field Eq. (29.4). The only reason that we can measure and actually in certain circumstances use radiation friction is that we accumulate a tiny effect on an elementary scale over a macroscopic time and distance scale. What this teaches us is that the perturbative description of radiation friction works well, and we return to this situation in following section.
Let us, however, mention a few experimental environments in which it is perhaps possible to achieve and study critical acceleration today and near future:

**Supercharged (quasi) nuclei:** Compared to the field strength we discussed in the context of the Born-Infeld model of the electron in section 28.3, the value $E_{cr}$ Eq. (29.3) appears small, and even smaller compared to fields available at the nuclear surface, see Eq. (28.38). However, atomic electrons are not near to the nuclear surface and are delocalized over the atomic Bohr radius. Moreover, the Schwinger field at the nuclear surface is found in a small volume only. Clearly, just reaching $E_{cr}$ is not a key criterion for appearance of new physics phenomena. We ask the question: what nuclear charge $Z\alpha e$ is needed so that at distance $\lambda_C$ the field becomes critical? This condition can be also written as

$$\frac{Ze}{\lambda_C^2} = E_{cr} = \frac{m^2 c^3}{e\hbar} \rightarrow Z\alpha = 1, \quad \alpha = \frac{e^2}{\hbar c} = \frac{1}{137.036}. \quad (29.8)$$

Thus if we can assemble a nuclear charge of $Z > 137$ we could test in an atomic physics experiment the condition of critical acceleration. This situation has been thoroughly explored and described in the context of relativistic quantum mechanics\[\text{[27]}\]. Since the nuclei have finite size, the actual value of ‘critical charge’ $Z_{cr} \approx 171$ is recognized. At this point the quantum vacuum, the ground state, becomes charged. The experimental realization of this situation is in the study

---

of slow, $v \ll c$, large nuclei called ‘heavy ion’ collisions, such that $Z_1 + Z_2 > Z_{cr}$. The study of positron production in such heavy ion collisions carried out in the 1980s was hampered by large nuclear reaction backgrounds.

**Ultra-Intense laser pulse collision with relativistic electrons:** It has been noted that when an ‘observer’ riding a relativistic electron is hit by a laser pulse, the required characteristics of supercritical force can occur. In figure 29-2 the domain where the required conditions are achieved\(^{28}\) is located above the line. We need a combination of a) a relatively large electron Lorentz factor $\gamma$, with b) intense laser pulses with a large factor $a_0$, see exercise IX–8. What comes handy is that such intense laser pulses are capable of accelerating electrons\(^{29}\) to energies in GeV range ($\gamma > 1000$). Thus one can envisage as experimental environment the collisions of laser pulses with laser pulse generated electron pulses. Laser pulse laboratories where the required conditions and experimental beam lines are prepared are at the time of writing under construction\(^{30}\). At these facilities experimental study of a radiation-reaction effect will be possible, a physics phenomenon we return to discuss in a qualitative manner in section 29.4 – charged particle interacting with strong EM-fields experiences ‘vacuum friction’, dissipating its energy in form of radiation, a foundational consistent description of this novel physics frontier has not been achieved\(^{31}\).

**Ultrarelativistic heavy ion collisions:** We have established in exercise VIII–14 on page 308 that the 4-acceleration magnitude is

$$a \equiv c \frac{dy_p}{d\tau}, \quad \text{or} \quad a \equiv c \frac{dy_p}{dt} \cosh y_p. \quad (29.9)$$

When big nuclei collide head-on in a relativistic heavy ion collider experiment, the duration time of the collision in the laboratory frame is $dt \to \Delta t = R/c$, where $R$ is the nuclear radius. This is the maximum time; the actual time $dt$ could be a fraction of $\Delta t$, but we want to be conservative in estimating the value


\(^{30}\)p71 of *European Strategy Forum Report 2016 on Research Infrastructures*, prepared by the StR-ESFRI project. I quote: “the Extreme Light Infrastructure (ELI) is a Research Infrastructure of Pan-European interest for experiments on extreme light-matter interactions at the highest intensities, shortest time scales and broadest spectral range. ELI is based on three sites (known as pillars, located in the Czech Republic, Hungary and Romania) with complementary scientific profiles, and the possible implementation of a fourth pillar, the highest intensity pillar, dependent on on-going laser technology development and validation. The fourth pillar laser power is expected to exceed that of the current ELI pillars by another order of magnitude, allowing for an extended scientific program in particle physics, nuclear physics, gravitational physics, nonlinear field theory, ultrahigh-pressure physics, astrophysics and cosmology (generating intensities exceeding $10^{23}$ W/cm$^2$).”

New State of Matter created at CERN

CERN press office 10 Feb 2000

Figure 29-3: CERN announces the discovery of quark-gluon plasma in a press-release on 10 February 2000 [press.web.cern.ch/press-releases/2000/02/new-state-matter-created-cern]

of $a$

$$a \simeq c \frac{\Delta y_p}{\Delta t} \cosh \Delta y_p \Rightarrow a_{\text{RHIC}} \Delta y_p \cosh \Delta y_p,$$

(29.10)

where the scale on which we measure acceleration in these collisions is

$$a_{\text{RHIC}} \equiv \frac{c}{\Delta t} = \frac{c^2}{R} \simeq 2 \times 10^{31} \text{ m s}^{-2}.$$

(29.11)

This result is by a factor 100 greater than the values shown in Eq. (22.5) and corresponds to fields we also obtained in somewhat different manner, see Eq. (29.7). As these considerations suggest when nucleons from the incoming nucleus are stopped so that rapidity shifts significantly from a value prior to collision, critical acceleration must have been achieved ripping nucleons apart. In such process there is creation of a large particle multiplicity. This finding coincides with the formation of quark-gluon plasma (QGP) at CERN in the year 2000, see figure 29-3. The abundant particle production accompanying achievement of critical acceleration is probably the cause of formation of this new state of matter.

29.4 EM radiation-friction

We presented in figure 29-1 the effect of gravitational radiation friction: since the binary system spirals in as energy is carried out by gravitational radiation, one can say the motion is damped by radiation friction. Similarly, non-inertial moving charged particles will emit EM radiation and thus be subject to radiation vacuum-friction effect. The emitted EM radiation is in a particular context of radial circular motion called synchrotron radiation. Synchrotron radiation is a
parasitic friction effect in particle accelerators. It can be greatly enhanced in special devices in order to make the emitted radiation a research tool. There are many practical applications of intentional synchrotron radiation. This is a very well studied domain of physics with its own and diverse literature.

Our objective here is different: in order to understand the modification of charged particle dynamics we seek to establish the energy loss that accompanies non-inertial motion of a charged particle. It is sufficient in a first step to consider the vacuum-friction effect in the instantaneous particle rest-frame $\vec{v}(t) = 0$ but

$$\vec{a} \equiv \frac{d\vec{v}}{dt} = \vec{\dot{v}}.$$  \hspace{1cm} (29.12)

The key point is that aside of the point charge $\rho \equiv e \delta^3(\vec{x})$ there is another source of Maxwell equations

$$\frac{dj}{dt} = \vec{a} \delta^3(\vec{x}), \hspace{1cm} (29.13)$$

and in principle there are higher derivatives with respect to time we could consider. We restrict present study to the case of uniform acceleration $\vec{a} = \text{Const.}$, neglecting ‘acceleration’ of acceleration radiation effects.

Taking the time derivative of Maxwell’s equation Eq. (27.11) and replacing the time derivative of $\vec{B}$ using Maxwell’s equation Eq. (27.28) we obtain in the particle rest-frame

$$-\frac{1}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} + \vec{\nabla} \times \frac{\partial \vec{B}}{\partial t} = -\frac{1}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} + \vec{\nabla} \times (\vec{\nabla} \times \vec{E}) = \frac{1}{\epsilon_0 c^2} \vec{a} \delta^3(\vec{x}) \hspace{1cm} (29.14)$$

Since both sources of $\vec{E}$ are static, as we see in Eq. (29.14), and inspecting Maxwell’s Eq. (27.10) restated here for convenience

$$\vec{\nabla} \cdot \vec{E} = \frac{e}{\epsilon_0} \delta^3(\vec{x}) \hspace{1cm} (29.15)$$

the solution is static; thus we can ignore time derivative in Eq. (29.14).

We denote acceleration by $\dot{\vec{v}} = \vec{a} = a \hat{a}$ and the radial unit vector $\hat{r}$. Since both Eq. (29.14) and Eq. (29.15) are linear we decompose the solution into a radial and ‘rotational’ part

$$\vec{E} = \hat{r} \vec{E}_c + \vec{E}_a, \hspace{1cm} \hat{r} \cdot \vec{E}_a = 0 \hspace{1cm} (29.16)$$

and hence

$$\vec{E}_a = (\hat{r} (\hat{r} \cdot \hat{a}) - \hat{a}) \vec{E}_a = \hat{r} \times (\hat{r} \times \hat{a}) \vec{E}_a \hspace{1cm} (29.17)$$

where we used the ‘bac-cab’rule backwards. Note that $\vec{E}_a$ component vanishes when inserted into the Eq. (29.15), while $\hat{r} \vec{E}_c$ component in $\vec{E}$ vanishes when inserted into Eq. (29.14). Hence, the solution in the instantaneous rest-frame of a uniformly accelerated charge $e$ is

$$\vec{E} = \frac{e}{4\pi \epsilon_0} \left( \frac{\hat{r}}{r^2} + \frac{a}{c^2} \frac{\hat{r} \times (\hat{r} \times \hat{a})}{r} \right) \hspace{1cm} (29.18)$$
The result Eq. (29.18) is identical to the one following when taking the $\vec{\beta} = \vec{v}/c = 0$ limit of Liénard-Wichert fields obtained differentiating the potential presented below in exercise XI–6 and keeping only first time derivative terms in $\vec{v}$. A systematic expansion of the Liénard-Wichert potentials exploring both higher order derivative of the velocity and multipole moments has been presented by Kijowski and collaborators.\(^{32}\)

Since both contributions in Eq. (29.16) are orthogonal, the square of the field comprises these two terms without interference

$$\vec{E}^2 = \vec{E}_c^2 + \vec{E}_a^2 = \frac{e^2}{(4\pi\epsilon_0)^2} \left( \frac{1}{r^4} + \frac{a^2}{c^4 r^2} (\hat{r} \times \hat{a})^2 \right), \quad (29.19)$$

It is important to observe that the radiation term, 2nd in Eq. (29.18), decreases as $1/r$. This must be so since Eq. (29.14) can also be written as

$$-\frac{1}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} + \nabla^2 \vec{E} = \frac{1}{\epsilon_0 c^2} a \delta^3(\vec{x}) + \nabla (\nabla \cdot \vec{E}). \quad (29.20)$$

The origin of the radiation $1/r$ component is in the term $\nabla^2 (a/r) = -4\pi a \delta^3(\vec{x})$.

According to Eq. (28.4) the outflow of radiated energy is governed by the divergence of the Poynting vector Eq. (28.2) $\vec{S}$. To compute $\vec{S}$ we also need to determine the magnetic field $\vec{B}$. The method we used to find $\vec{E}$ would yield $\vec{B} = 0$. However, the transformation from the laboratory frame to the instantaneous rest-frame of the particle is time dependent, and this time dependence drives the value of $\vec{B}$. A simpler path is to instead determine the value of the magnetic field in the instantaneous rest-frame from the requirement that the mass density of the field surrounding a charged particle as described by the invariant Eq. (28.13) cannot depend on the force we apply to the charged particle: this is the present day theoretical paradigm.

Inspecting Eq. (28.13) we realize that this means that $\vec{B}$ must be orthogonal to both parts of the electric field $\vec{E}$, Eq. (29.18), assuring $\vec{E} \cdot \vec{B} = 0$. At the same time $c^2 \vec{B}^2$ must compensate exactly in $\vec{E}^2$, Eq. (29.19), the second radiation term. Thus the magnitude of the magnetic field $c \vec{B}$ must be the same as that of the radiation term in the electric field, and to be orthogonal to both parts of the $\vec{E}$ we must have $\vec{B} \propto \dot{r} \times \dot{a}$. Thus $c \vec{B}$-field is determined except for the sign to read

$$c \vec{B} = \dot{r} \times \vec{E}_a = -\frac{e}{4\pi\epsilon_0 c^2} \frac{a}{r} (\hat{r} \times \hat{a}) \quad (29.21)$$

In the instantaneous rest-frame of the particle we find for the Poynting vector Eq. (28.2)

\[ \vec{S} = \frac{e^2}{4\pi\epsilon_0 c^3} \frac{2a^2}{4\pi r^2} (\hat{r} \times \hat{a})^2 \hat{r}, \]  

(29.22)

where we used

\[ ((\hat{r} \times \hat{a}) \times (\hat{r} \times \hat{a})) = \hat{r} (\hat{r} \times \hat{a})^2 - \hat{r} \times \hat{a} [(\hat{r} \times \hat{a}) \cdot \hat{r}] = \hat{r} (\hat{r} \times \hat{a})^2. \]  

(29.23)

The magnitude of the Poynting vector Eq. (29.22) has angular dependence. By choosing the z-axis to align with direction of \( \vec{a} \), we see that \( (\hat{r} \times \hat{a})^2 = \sin^2 \theta \); the maximum of emission is transverse to the direction of the uniform acceleration.

We now perform a surface integral around the charge, where we encounter the angular average \( \langle \sin^2 \theta \rangle = 2/3 \)

\[ \int d^2 A \cdot \vec{S} = \frac{e^2}{4\pi\epsilon_0 c^3} \frac{2a^2}{3}. \]  

(29.24)

Using Gauss’s theorem to rewrite the left hand side we have

\[ \int d^3 x \ \vec{\nabla} \cdot \vec{S} = \frac{e^2}{4\pi\epsilon_0 c^3} \frac{2a^2}{3}. \]  

(29.25)

This energy outflow through the surface is the energy loss due to radiation by an accelerated particle

\[ \frac{dE_{\text{rad}}}{dt} = \int d^3 x \ \vec{\nabla} \cdot \vec{S} = \frac{e^2}{4\pi\epsilon_0 c^3} \frac{2a^2}{3}. \]  

(29.26)

This is the well-known non-relativistic Larmor form for the EM radiation power loss. The sign in Eq. (29.21) was chosen to obtain the energy loss for the accelerated particle. To convert from SI to Gauss units we remember the third rule in Eq. (23.2).

The Larmor radiated power result Eq. (29.26) for energy loss is usually presented without signs, as an absolute value. In our study the sign was not determined but chosen, as commented below Eq. (29.21). We often see remarks that the Larmor radiated power Eq. (29.26) is valid only in the ‘radiation zone’, that is, far from the location of the charge. However, as obtained here, it is evident that Eq. (29.18) and Eq. (29.21) are exact; there is no need to distant from the particle – the amount of emitted radiation is the same near and far from the location of the charge of the particle as reported by an observer in the instantaneous rest-frame.
29.5 EM friction-force

We determine the equivalent radiation-reaction force \( \vec{F}_{\text{rad}} \) by evaluating the energy delivered by such a force. According to Eq. (29.26)

\[
\int \vec{v} \cdot \vec{F}_{\text{rad}} \, dt = -\int \frac{dE_{\text{rad}}}{dt} \cdot dt = -\frac{2}{3} \frac{e^2}{4\pi\epsilon_0c^3} \int \dddot{a}^2 \, dt \quad (29.27)
\]

We gain more insight by integrating by parts using \( \dddot{a} = \frac{d\dddot{v}}{dt} \)

\[
\int \vec{v} \cdot \vec{F}_{\text{rad}} \, dt = \frac{2}{3} \frac{e^2}{4\pi\epsilon_0c^3} \int \vec{v} \cdot \frac{d\dddot{v}}{dt} \, dt + \frac{2}{3} \frac{e^2}{4\pi\epsilon_0c^3} \vec{v} \cdot \left. \frac{d\dddot{v}}{dt} \right|_{t^-}^{t^+}. \quad (29.28)
\]

The last surface term vanishes for the initial and final integration boundaries outside of the acceleration domain, and for particles subject to periodic motion requiring that the device, such as a synchrotron, is equipped to compensate radiation losses. The second-to-last term in Eq. (29.28) characterizes the radiation-reaction force

\[
\vec{F}_{\text{rad}} = m\tau_0 \frac{d^2\vec{v}}{dt^2}. \quad \tau_0 = \frac{2}{3} \frac{e^2}{4\pi\epsilon_0mc^3} = \frac{2re}{3c} = 0.626 \times 10^{-23}, \text{s}. \quad (29.29)
\]

where the value of the time constant \( \tau_0 \) is stated for an electron.

The appearance of a higher derivative of acceleration in Eq. (29.29) has unpalatable consequences. Newton’s equation for a free particle now has two independent solutions

\[
m \frac{d\vec{v}}{dt} = \vec{F}_{\text{rad}} = m\tau_0 \frac{d^2\vec{v}}{dt^2}, \quad \rightarrow \quad \dddot{v} = \dddot{v}_0, \quad \& \quad \dddot{v}_{\text{run}}(t) = \dddot{v}_0 e^{t/\tau_0}. \quad (29.30)
\]

The removal of the remnants of the ‘runaway’ solution \( \dddot{v}_{\text{run}}(t) \) is a goal of anyone attempting to incorporate radiation-reaction into charged particle dynamics. This means that in order to determine a physical particle trajectory we need to prescribe additional boundary conditions capable of suitably restricting motion, and this turns out to be in general a procedure that requires causality violation: we must know the motion in the future to assure that no runaway occurs along the path.

We now generalize the radiation-reaction force Eq. (29.29) to the relativistic domain

\[
\mathcal{F}^\mu_{\text{rad}} \equiv m\tau_0 \left( g^{\mu\nu} - \frac{u^\mu u^\nu}{c^2} \right) \frac{d^2u_\nu}{d\tau^2}. \quad (29.31)
\]

The term in parentheses assures \( u \cdot \mathcal{F}_{\text{rad}} = 0 \), securing the \( u^2 = c^2 \) constraint. Other than that we have replaced the second derivative with respect to time by derivative with respect to proper times. This assures the non-relativistic limit.
This leads to the Lorentz-Abraham-Dirac (LAD) equation written in the following form by Dirac\textsuperscript{33}

\[
\frac{d(mu^\mu)}{d\tau} = (eF^\mu_{\nu\text{ext}} + eF^\mu_{\nu\text{rad}})u_\nu, \quad F^\mu_{\text{rad}} = \frac{m}{e} \frac{\tau_0}{c^2} (\ddot{u}^\mu u^\nu - u^\mu \ddot{u}^\nu),
\]

(29.32)

where \(F^\beta_{\alpha\text{ext}}\) is a prescribed external field. We use the notation \(\dddot{u} = \frac{d^2 u}{d\tau^2}\).

Equation (29.32) is a direct transcription of Eq. (29.31) using \(u^\nu u_\nu = c^2\). Without entering into detailed discussion we note that an integral form for the solution is known\textsuperscript{34}. The constraint required to deal with (generalized) runaway solution is in apparent violation of causality for time of order \(\tau_0\).

### 29.6 Landau-Lifshitz radiation force model

Another well known modification of dynamical equations was developed in the Landau-Lifshitz (LL) text in theoretical physics\textsuperscript{35}. There are three different presentations of the model, but it is hard to find a statement that this model differs from LAD radiation friction force. Since LAD is not a foundational model, one can view LL as a method of improving on the LAD approach solving the problem of runaway solutions and causality violation. Therefore a short derivation with a clarification of how and where LL differs from LAD will offer a useful complement to other presentations.

In the LL approach the problematic third proper time derivative of the position of a particle appearing in Eq. (29.32) is replaced by an expression obtained differentiating the covariant Lorentz force equation

\[
\dddot{u}^\mu = \frac{d}{d\tau} \left( \frac{e}{m} F^\mu_{\delta\chi} u^\delta \right) = \frac{e}{m} \ddot{u}^\gamma \partial_{\gamma} F^\mu_{\delta\chi} u^\delta + \frac{e}{m} F^\mu_{\delta\chi} u^\delta = \frac{e}{m} \ddot{u}^\gamma \partial_{\gamma} F^\mu_{\delta\chi} u^\delta + \left( \frac{e}{m} \right)^2 F^\mu_{\delta\chi} F^\chi_{\beta\gamma} u^\gamma u^\beta,
\]

(29.33)

where if and when \(F^\mu_{\nu\text{rad}} = F^\mu_{\nu\text{ext}} + F^\mu_{\nu\text{rad}}\) no approximation is being made. Using Eq. (29.33) in the definition of \(F^\mu_{\nu\text{rad}}\) in Eq. (29.32) we find for the radiation friction force

\[
K^\mu_{\text{rad}} = eF^\mu_{\nu\text{rad}} u_\nu = e\tau_0 \ddot{u}^\gamma \partial_{\gamma} F^\mu_{\delta\gamma} u_\delta + e\tau_0 \frac{e}{m} \left( F^\mu_{\delta\gamma} F^\chi_{\beta\gamma} u^\gamma u^\beta - u^\nu F^\nu_{\delta\gamma} u^\gamma u^\beta \right).
\]

(29.34)

Considering that \(F^\mu_{\nu\text{rad}} = F^\mu_{\nu\text{ext}} + F^\mu_{\nu\text{rad}}\), \(K^\mu_{\text{rad}}\) continues to implicitly contain \(\dddot{u}\) terms.

The resulting LL equation of motion with simplifying index modifications is

\[
m\ddot{u}^\mu = eF^\mu_{\nu\text{ext}} u_\nu + e\tau_0 \left\{ u^\gamma \partial_{\gamma} F^\mu_{\delta\gamma} u_\delta + \frac{e}{m} \left( g^\mu\gamma - \frac{u^\mu u^\gamma}{c^2} \right) F^\gamma_{\beta\delta} F^\beta_{\gamma\delta} u^\delta \right\}.
\]

(29.35)


\textsuperscript{34}See Section 21.11 in W.K.H. Panofski and M. Phillips, \textit{Classical Electricity and Magnetism}, Addison-Wesley (Reading MA, 1962)

As constructed by inserting Eq. (29.33) in Eq. (29.32) the radiative friction force satisfies $u_{\mu}K_{rad}^{\mu} = 0$, assuring that the modified Lorentz force equation will always produce solutions with $u^2 = c^2$.

It seems that as long as we use on the right hand side of Eq. (29.34) $F = F_{\text{ext}} + F_{\text{rad}}$ LL and LAD are equivalent. However, the LL equation as is in use never an exact rendition of LAD, which is not surprising, as LL faces no causality and runaway challenges. The reason is that there are in principle two ways to proceed and none renders LL equivalent to LAD:

1) We solve simultaneously the LL equation and the Maxwell equations for the fields $F$, particle velocity $u$ and position $x$. That means that we incorporate in the field $F$ all radiation effects for a self-consistently determined trajectory rather than adopting the simplifying format of the Larmor radiation formula formulation which leads to LAD. In this procedure, the solution we would obtain differs from the approximation inherent in the LAD form of response to emitted radiation. Moreover, this approach will encounter challenges related to the infinite EM energy, see section [28.2] and, one cannot be sure to find in this approach causal non-runaway solution.

2) To avoid solving Maxwell equations again, and in order to preserve the consistency with LAD, we insert into the LL force explicitly $F^{\mu\nu}_{\text{rad}} = F^{\mu\nu}_{\text{ext}} + F^{\mu\nu}_{\text{rad}}$. Since $F^{\mu\nu}_{\text{rad}}$ contains $\ddot{u}^{\mu}$, we need to repeat the procedure, using Eq. (29.33) to eliminate $\ddot{u}$. This iterative procedure generates a series with terms in powers of $\tau_0(d/d\tau)$ or/and equivalently $r_e(\partial/\partial x))$. However, in addition to classical series there are quantum effects to consider which scale with $\lambda_C(\partial/\partial x) = \alpha^{-1}r_e(\partial/\partial x))$. These are in general more relevant compared to the second classical step in the iterative approach to LL. Therefore, it is more appropriate to truncate the LL iteration scheme after the leading term. This means that on the right hand side in Eq. (29.34) we set $F \simeq F_{\text{ext}}$ adopting, the lowest order in $\tau_0$ friction force

$$K_{\text{rad}}^{\mu(1)} \equiv e\tau_0 u^\gamma \partial_\gamma F_{\text{ext}}^{\mu\delta} u_\delta + e\tau_0 \frac{e}{m} \left( g^{\mu\gamma} - \frac{u^\mu u^\gamma}{c^2} \right) F_{\text{ext}}^{\alpha\beta} F_{\text{ext}}^{\gamma\delta} u_\delta + \ldots.$$ (29.36)

We now read the LL equation of motion Eq. (29.35) as used in approximation $F \simeq F_{\text{ext}}$. This is how LL is in general used in study of radiation friction force.

What this means is that the radiation field does not impact the particle motion as described by the LL dynamical equation. LL can be seen as a LAD motivated radiation friction force approximation where fields instead of higher proper time derivatives are introduced. Neither LAD nor LL are exact renditions of the effect of radiation friction as, for example, neither contains back reaction of radiation field on particle motion.

We now describe the strong effect of radiation friction where an electron collides head-on with a laser plane wave light-front. The radiation friction is introduced using the LL force. For comparison we consider the solution for the case of
Figure 29-4: The longitudinal velocity $v_z$ and the transverse velocity $v_x$ (both in units of $c$) are shown as functions of time measured in fs ($10^{-15}$s) for the case of an electron with an initial energy $E_e=511$ MeV, i.e. $\gamma = 1000$ hitting a circular polarized (CP) light-front with amplitude $a_0 = 100$ and wavelength $\lambda = 942$ nm. The solid red line is for electron motion subject to the LF (Lorentz force) case see section 24.3, while the dashed (blue) line gives the deceleration according to the LL (Landau-Lifshitz force). Adopted from Ref. 31.

the Lorentz force (LF). The LF part coincides, with a change in initial conditions, with the LF solution we presented in section 24.3. For the case of LL friction, the solutions were obtained semi-analytically in Ref. 31. The results we adapt from this reference consider the case of an electron with an initial energy $E_e=511$ MeV i.e. $\gamma = 1000$, hitting a circular polarized (CP) plane wave light-front with an amplitude $a_0 = 100$ and wavelength $\lambda = 942$ nm.

In figure 29-4 we see the behavior of the particle velocity vector. The particle moves counter the positive sense of the propagation of the laser pulse which explains the negative initial value of $v_z$ in the left part of figure 29-4. On the right in figure 29-4 we see transverse velocity $v_x$ which begins with the particle impact on the plane wave front. For the LF case we see that the plane wave modulates the motion to a minor degree. When LL friction is introduced, $v_z$ rapidly decreases. After the particle moves across 10 wavelengths, all of the motion is either attenuated or/and transferred from longitudinal to transverse direction; the value of the Lorentz factor $\gamma$ (not shown) drops from 1000 to 100, ultimately the longitudinal motion reverses. If the calculation were to continue we would find the solution applicable to the particle surfing a plane wave as was described in section 24.3.

The question why there is such a significant difference in the LL response compared to LF relates to the condition of critical acceleration that has been exceeded. In figure 29-5 the invariant acceleration $a = \sqrt{-a_\mu a^\mu}$ that is achieved in collision of light-front with electron beam in the LL case is seen to exceed unity. We see that once the acceleration drops below critical value, the radiation response diminishes, and it fades out when $a < 0.2$. Note that without radiation friction in the LF case the particle also experiences a large acceleration, but in
Figure 29-5: Electron invariant acceleration $a = \sqrt{-a_\mu a^\mu}$ in natural units of critical acceleration, see Eq. (22.5), is shown as a function of time measured in fs ($10^{-15}$s) for the case of an electron with an initial energy $E_e=511$ MeV, i.e. $\gamma = 1000$ hitting a circular polarized laser plane wave with amplitude $a_0 = 100$ and wavelength $\lambda = 942$ nm. The solid red line is for electron motion subject to the LF (Lorentz force) case, while the dashed (blue) line gives the deceleration according to the LL (Landau-Lifshitz force). Adopted from Ref. 31.

the absence of radiative energy loss the charged particle motion remains, on the time scale considered, unaffected.

The strong response when radiation friction force is included is due to radiation power depending on acceleration (squared). When friction is large, (in this case) deceleration is large, which feeds back into the radiative loss, and the radiative loss runs away in producing a gigantic energy loss effect. We should keep in mind that the figure time resolution time scale is $10^{-15}$ s, while the expansion parameter for the radiation response is $\tau_0 = 0.63 \times 10^{-23}$ s. Thus what we see in figure [29-5] is the outcome of many much smaller reaction processes.

29.7 Open questions: radiation reaction

The radiation-reaction force is obtained under the assumption of radiation emission by a locally accelerated particle. Since the radiation-reaction force incorporates a higher derivative, one can argue that in a more consistent non-perturbative study a greater non-locality could emerge. A proposal was made by Caldirola\textsuperscript{[36]} to generalize the Newton part of the Lorentz-force to a nonlocal form.

Table 29-1: Models of radiation-reaction extensions of the Lorentz-force.

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maxwell-Lorentz</td>
<td>$\dot{u}^\mu = e F^{\mu\nu} u_\nu$</td>
</tr>
<tr>
<td>LAD$^{33}$</td>
<td>$\dot{u}^\mu = e F^{\mu\nu} u_\nu + m \tau_0 \left[ g^{\mu\nu} - \frac{u^\mu u^\nu}{c^2} \right] \ddot{u}_\nu, \quad \tau_0 = \frac{2}{3} \frac{e^2}{4\pi\epsilon_0 mc^3}$</td>
</tr>
<tr>
<td>Landau-Lifshitz$^{35}$</td>
<td>$\dot{u}^\mu = e F^{\mu\nu} u_\nu + e \tau_0 \left{ u^\gamma \partial_\gamma F^{\mu\delta} u_\delta + \frac{e}{m} \left( g^{\mu\gamma} - \frac{u^\mu u^\gamma}{c^2} \right) F_{\gamma\beta} F^{\beta\delta} u_\delta \right}$</td>
</tr>
<tr>
<td>Caldirola$^{36}$</td>
<td>$0 = e F^{\mu\nu}(\tau) u_\nu(\tau) + \frac{m}{2\tau_0} \left[ g^{\mu\nu} - \frac{u^{\mu(\tau)} u^{\nu(\tau)}}{c^2} \right] u_\nu(\tau - 2\tau_0)$</td>
</tr>
</tbody>
</table>

The Cardirola model keeps the conservative form of the Lorentz force in terms of fields. Instead a nonlocal inertial term is introduced

$$ K^\mu = m \frac{d u^\mu(\tau)}{d\tau} \rightarrow K^\mu_C \equiv \left( g^{\mu\nu} - \frac{u^\mu u^\nu}{c^2} \right) m \frac{u_\nu(\tau - 2\tau_0)}{-2\tau_0}, \quad (29.37) $$

which is an ad hoc modification of the equation of motion needing further exploration. For example it seems of some importance to know in what way particle motion is modified, which requires repeating the chain of solutions we have presented in Part IX, a step beyond the scope of this book.

Multiplying Eq. (29.37) from left by $u_\mu(\tau)$, we see that Caldirola solutions are constrained to $u^2 = c^2$. Expanding in powers of $\tau_0$, we find at lowest order $(\tau_0)^0$, i.e. not dependent on $\tau_0$, that the inertial term is the usual inertial force. At first order in $\tau_0$, LAD is recovered. This nonlocal modification of the inertial term avoids the LAD solution difficulty of acausal and/or runaway solutions: the past value of the 4-velocity of the particle determines its present 4-velocity, yet the effect of radiation-reaction arises from the inertia nonlocality.

To close this review of efforts to improve the Lorentz-force as summarized in table [29-1], we found that by accounting for the radiation-friction effect we are led to the LAD modification of the Lorentz-force. This effective dynamical equation contains spurious solutions which can be eliminated at the cost of violating causality. Two ad hoc improvements (among a few others) of LAD have been described that are opposite to each other in that each of these considers a change in one of the ‘sides’ of the Lorentz force.

1) The LL model of radiation friction amounts to a modification of the field-
Afterword: Acceleration

particle force (Lorentz force). However, an appropriate fundamental action that addresses the proposed modification has not been found. This is so since the LL model invokes particle motion that is not conservative, containing a friction effect: the transfer of energy from the particle to the emitted radiation that is not tracked. These LL non-conservative particle dynamics can therefore be studied only at the level of the proposed equation of motion.

2) For the Cardirola proposed ad hoc modification of the inertial term an action principle from which the dynamics could be derived has not been obtained. This approach and these shortcomings imply numerous theoretical and phenomenological consequences that are not understood in full.

The framework we have developed in this book comprises the present paradigm of two distinct interacting dynamical equations characterizing particle dynamics on one hand, and the EM-field dynamics on the other. These are two independent physics concepts with some unresolved details as we have seen exploring the inertial part of the particle action in section 25.2, and section 25.3, and the EM field energy in section 28.2. We do not yet have a theoretical framework extending SR to critical acceleration domain.

One could argue that the radiation-reaction difficulties we are discussing relate to the absence of the proper understanding of acceleration. This conjecture is supported by an explicit computation showing that at condition of critical acceleration the radiation-reaction correction exceeds in magnitude the Lorentz-force, see Ref. 31 and section 29.6. What happens when acceleration approaches and exceeds the critical value is a research frontier. We do not actually know how particles and fields behave in the context of strong acceleration. Without any strong acceleration experimental guidance involving particle and field dynamics it is difficult to make progress.

Discussion XI-1 – The acceleration frontier

**Topic:** Are there novel research opportunities at the radiation-reaction and acceleration physics frontier?

**Student:** Everyone I spoke with who has studied electromagnetism in depth has misgivings about EM-theory because of the radiation-reaction problem.

**Simplicius:** My friends say we must remember quantum physics. Electrons in atoms do not radiate. My friends have suggested that the radiation-reaction problem is a classical problem resolved in quantum theory.

**Professor:** Please ask your friends about cosmic binaries emitting gravitational radiation. We cannot claim today that all classical physics phenomena originate in the quantum physics realm. In many cases such claims are without merit.

**Simplicius:** Can I tell if radiation is emitted by a quantum, or by a classical electron? **Professor:** A quantum transition often produces discrete energy when an excited quasi-
stable state transits emitting a photon. Such a quantum jump has little to do with classical continuous radiation.

Student: Could a quantum transition that bridges a highly excited state with many, many $\hbar$, leading to a much lower level, be similar to classical radiation?

Professor: Performing a quantum calculation, you find that radiative transitions across many $\hbar$ are extremely rare.

Simplicius: Meaning that there is no classical radiation limit for an electron sitting in a quantum orbit? So how do we get synchrotron radiation out of electrons such as circulating in a macroscopic storage ring?

Professor: That synchrotron electron has such a crazy large number of $\hbar$ that the idea of a quantum state model is unthinkable. We are no longer in quantum jump condition.

Student: Seeking online I have not found a convincing source of wisdom on how to connect classical and quantum radiation schemes.

Professor: We know how to increase classical radiated power. It grows with acceleration (squared). Furthermore there is no limit to this radiation-friction. The stronger the friction, the more friction we see as the particle radiates more strongly.

Simplicius: Are you saying the particle comes to rest?

Professor: There is no absolute rest in SR; what the particle does is decelerate towards some inertial motion condition. Since we measure the radiation power in some frame of reference, in that frame of reference the radiation-friction diminishes as the reservoir of energy and power emitted diminish.

Simplicius: However, you said that an ultra-relativistic particle can decelerate in an ultra-short distance because the radiation-reaction helps!

Professor: Yes, one sees this when solving the radiation-reaction improved Lorentz-force equations (loc.cit. Ref. 31). There is runaway in the radiation-friction effect; the more the particle decelerates by radiation, the more energy is radiated, the more the particle is decelerated, and so on. I hope more physics is waiting to be discovered that relates to radiation friction. The old idea has been that EM-force unifies with gravity. Many people have looked for this, Weyl, Kaluza, Klein, Infeld, Einstein, to name a few. Yet no scheme has gained widespread acceptance.

Simplicius: Browsing the web to see how people think about particle motion and gravity phenomena I found that collisions at the CERN large hadron collider (LHC) produce black holes. I found a Wiki entry describing how particle physicists had to respond to concerns that high energy particle collisions could end the world.

Student: I would think that the formation of black holes is a pretty radical gravitational effect. I do not know anyone expecting black holes to accompany strong acceleration

---


38 Safety of high-energy particle collision experiments, https://en.wikipedia.org/wiki/Safety_of_high-energy_particle_collision_experiments
phenomena. But speaking of experiments, are there other manifestations of strong EM acceleration phenomena offering connection to gravity?

Professor: We believe that in quantum field theory context strong fields=strong acceleration imply instability, fields decaying into particle pairs. Such instability could imply the need for a modification of the theoretical context. Since strong acceleration looks like a hidden Planck scale, finding a way to explore inertia response to very extreme acceleration conditions could be a challenging but promising objective. Moreover, the equivalence principle on which GR relies has been tested in the presence of extremely weak acceleration. Put this together and it could be that a study of EM radiation friction could help uncover a relation between GR and EM which many have been seeking for more than a century. The parameter here that turns from small to large is acceleration=strong field. Thus to paraphrase those who work at energy and intensity frontier I must say that in the foreseeable future we will be performing experiments at the acceleration frontier where current understanding clearly is at its limit.
Index

æther, xiv, 14
    drag, 28, 101
Einstein, 2
Larmor, 27
luminiferous, 2
Maxwell, 13
quantum vacuum, 18, 441
wind, 24, 28

aberration
    light, 10, 98, 100
    ultrarelativistic, 103
absolute space, Newton, 440
acceleration, 305, 439
    critical, 304, 444
    frontier, 460
    instantaneous rest-frame, 306, 449
    orthogonal to 4-velocity, 307
    particle rapidity, 308
    plane wave EM field, 356, 359
    small, 304
    spacelike, 307
action, 402
    covariant Hamiltonian, 375
    EM-field, 401
particle
    covariant, 367
    relativistic, 368
antimatter, see also positron, 213
annihilation, 195
antiproton production, 255
apsidal precession, 343
atomic clock, 7
atoms, 55
axial vector, 381
bac-cab rule, 330, 331
Bell
    body contraction, 32, 34
    inequality, 161
    John S., iv, xv, 136
    Larmor time dilation, 27
    rockets, 135
Big Bang, 153
black hole, 459
body contraction, see
    Lorentz-FitzGerald body contraction
Bohr radius, 342
Boltzmann constant, 205
boost, 81, 280
Born, Max, 435
Born-Infeld EM theory, 435
    bound on electric field, 437
Bradley, James, 9
canonical
    equations of motion, 336
    momentum, 333, 335, 375
causality, [151]
    particle entanglement, 159
    principle of, 6
    violation when \( v > c \), 154
center of momentum, 233
    frame, 234
CERN
    LHC, 253, 459
QGP discovery, 448
SPS, 259
charge quantization, 415
CM
(acronym) center of momentum, vi
-energy $\sqrt{s}$, 235 252 296
-frame
finding the reference frame, 235
two body collisions, 236
CMB, 25 35 191
(acronym) cosmic microwave background, vi
reference frame, 26, 442
relative motion, 20
CMF (acronym) CM-frame, vi
collision
deep inelastic, 299
elastic, 260
inelastic, 248, 258
light-pulse
electron beam, 455
moving relativistic wall, 266
nonrelativistic moving wall, 270
Compton
Arthur H., 263
collision, 264, 299
electron wavelength, 161
scattering, 262
wavelength, 149 431
wavelength shift, 264
contraction, see Lorentz-FitzGerald
body contraction
coordinate transformation
active and passive, 75
Galilean, 22 75
Lorentz, 38, 80
metric, 286
cosmic frame of reference, 26
cosmological red-shift, 25, 175
Coulomb
atomic orbit, 55
Charles-Augustin de, 406
electron orbit, 342
field, 406
energy atomic nuclei, 430
mass density, 428
force, 338
gauge, 329
potential, 338
covariant and contravariant, 277, 364
CPT symmetry, 189
critical field, 352, 444
current
conservation, 404
density, 402
point structure, 417
cyclotron frequency, 322
d’Alambert equation, 409
dark
energy, 156, 175, 443
matter, 52, 156
deuterium, 211
dipole moment, magnetic, 417
Dirac
magnetic monopole, 415
Paul A. M., 415
Doppler
shift, 128
Christian A., 173
CMB radiation, 26
relativity test, 184
shift, 173, 175, 178, 196
line of sight, 179
dual field, 383
Earth energy balance, 213
EeV, 204
Einstein
æther, xv, 2
Albert, 4, 42
mass energy equivalence, 195
proper time, 5
relativity, xiv
summation convention, 278
electron, 55
  classical radius, 265, 320
  Coulomb orbits, 342
  motion subject to electric field, 324
  photon scattering, 265
  plane wave surfing, 354
EM (acronym) electromagnetic, vi
EM-field
  action, 401, 404
  dual tensor, 383
  energy momentum content, 422
  GR and æther, 15
  invariants, 381
  mass-energy, 429, 434, 436
    proton, 433
  plane wave mass density, 428
  radiation, 450
  tensor, 379, 389
  zero field frame, 386
energy
  body at rest, 198
  conservation, 233
  Earth balance, 213
  invariant, 298
  invariant in two body reactions, 252
  kinetic, 198
  nuclear, 212
  origin, 203, 213
  reaction threshold, 248
  renewable, 214
  rest, 195
energy-momentum
  four-vector, 294
  particle distribution, 428
  tensor, 422
  equal velocity reference frame, 258
  equivalence principle, 439
  Euler-Lagrange variational equation, 405
eV, 204
Fermi paradox, 171
<table>
<thead>
<tr>
<th>Index</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPS, 47, 190</td>
<td></td>
</tr>
<tr>
<td>GR (acronym) General Relativity, vi</td>
<td></td>
</tr>
<tr>
<td>gravity</td>
<td></td>
</tr>
<tr>
<td>acceleration, 439</td>
<td></td>
</tr>
<tr>
<td>frequency shift, 175</td>
<td></td>
</tr>
<tr>
<td>mass defect, 208</td>
<td></td>
</tr>
<tr>
<td>radiation, 34, 442</td>
<td></td>
</tr>
<tr>
<td>waves, 442</td>
<td></td>
</tr>
<tr>
<td>GT (acronym) Galilean transformation, vi</td>
<td></td>
</tr>
<tr>
<td>Hamiltonian</td>
<td></td>
</tr>
<tr>
<td>covariant action principle, 375</td>
<td></td>
</tr>
<tr>
<td>relativistic, 337</td>
<td></td>
</tr>
<tr>
<td>heavy ion collision, 32, 213</td>
<td></td>
</tr>
<tr>
<td>acceleration, 447</td>
<td></td>
</tr>
<tr>
<td>helium, 205</td>
<td></td>
</tr>
<tr>
<td>Hertz, Heinrich R., 14</td>
<td></td>
</tr>
<tr>
<td>Higgs field, 16, 18, 53, 428</td>
<td></td>
</tr>
<tr>
<td>horizon problem, 153</td>
<td></td>
</tr>
<tr>
<td>hyperfine splitting, 418</td>
<td></td>
</tr>
<tr>
<td>inertial</td>
<td></td>
</tr>
<tr>
<td>observers, 22</td>
<td></td>
</tr>
<tr>
<td>resistance, 311</td>
<td></td>
</tr>
<tr>
<td>Infeld, Leopold, 435</td>
<td></td>
</tr>
<tr>
<td>interaction</td>
<td></td>
</tr>
<tr>
<td>field-matter, 402</td>
<td></td>
</tr>
<tr>
<td>interactions</td>
<td></td>
</tr>
<tr>
<td>weak, 55</td>
<td></td>
</tr>
<tr>
<td>invariant</td>
<td></td>
</tr>
<tr>
<td>proper time, 149</td>
<td></td>
</tr>
<tr>
<td>Ives and Stilwell experiment, 184</td>
<td></td>
</tr>
<tr>
<td>kinetic energy</td>
<td></td>
</tr>
<tr>
<td>relativistic, 334</td>
<td></td>
</tr>
<tr>
<td>Klein-Nishina cross section, 265</td>
<td></td>
</tr>
<tr>
<td>Kronecker-delta, 278</td>
<td></td>
</tr>
<tr>
<td>Lagrange function, 333</td>
<td></td>
</tr>
<tr>
<td>Larmor</td>
<td></td>
</tr>
<tr>
<td>Lorentz transformation, 81</td>
<td></td>
</tr>
<tr>
<td>radiation power, 451</td>
<td></td>
</tr>
<tr>
<td>Sir Joseph, 27, 41</td>
<td></td>
</tr>
<tr>
<td>time dilation, 27</td>
<td></td>
</tr>
<tr>
<td>laser</td>
<td></td>
</tr>
<tr>
<td>intensity and power, 353</td>
<td></td>
</tr>
<tr>
<td>particle acceleration, 356, 395</td>
<td></td>
</tr>
<tr>
<td>pulse, 345</td>
<td></td>
</tr>
<tr>
<td>electron collision, 447</td>
<td></td>
</tr>
<tr>
<td>mass density, 428</td>
<td></td>
</tr>
<tr>
<td>relativistic optics, 354</td>
<td></td>
</tr>
<tr>
<td>Laue, Max von, 42, 96</td>
<td></td>
</tr>
<tr>
<td>Lavoisier, Antoine-Laurent de, 203</td>
<td></td>
</tr>
<tr>
<td>LI (acronym) Lorentz invariant, vi</td>
<td></td>
</tr>
<tr>
<td>Liénard-Wichert</td>
<td></td>
</tr>
<tr>
<td>EM fields, 450</td>
<td></td>
</tr>
<tr>
<td>potential, 412</td>
<td></td>
</tr>
<tr>
<td>light</td>
<td></td>
</tr>
<tr>
<td>aberration, 10, 101</td>
<td></td>
</tr>
<tr>
<td>ultrarelativistic, 103</td>
<td></td>
</tr>
<tr>
<td>clock, 49</td>
<td></td>
</tr>
<tr>
<td>interferometer, 19</td>
<td></td>
</tr>
<tr>
<td>speed, 7, 9, 10, 13</td>
<td></td>
</tr>
<tr>
<td>universal speed, 81</td>
<td></td>
</tr>
<tr>
<td>wave, 13, 345</td>
<td></td>
</tr>
<tr>
<td>year, 7</td>
<td></td>
</tr>
<tr>
<td>light cone, 78, 120, 148</td>
<td></td>
</tr>
<tr>
<td>coordinates, 150</td>
<td></td>
</tr>
<tr>
<td>Lorentz</td>
<td></td>
</tr>
<tr>
<td>-FitzGerald body contraction, xiv, 28, 30, 63</td>
<td></td>
</tr>
<tr>
<td>Bell’s experiment, 32</td>
<td></td>
</tr>
<tr>
<td>equal time, 126</td>
<td></td>
</tr>
<tr>
<td>measurement, 139</td>
<td></td>
</tr>
<tr>
<td>photograph, 127</td>
<td></td>
</tr>
<tr>
<td>quasi-rotation, 130</td>
<td></td>
</tr>
<tr>
<td>two frame comparison, 134</td>
<td></td>
</tr>
<tr>
<td>-FitzGerald contraction</td>
<td></td>
</tr>
<tr>
<td>spatial separation, 157</td>
<td></td>
</tr>
<tr>
<td>-force</td>
<td></td>
</tr>
<tr>
<td>covariant fields, 389</td>
<td></td>
</tr>
<tr>
<td>density format, 429</td>
<td></td>
</tr>
<tr>
<td>matrix format, 392</td>
<td></td>
</tr>
<tr>
<td>transformation</td>
<td></td>
</tr>
<tr>
<td>proper, 149</td>
<td></td>
</tr>
<tr>
<td>boost, 76, 81</td>
<td></td>
</tr>
<tr>
<td>boosts, 281</td>
<td></td>
</tr>
<tr>
<td>factor, 197</td>
<td></td>
</tr>
</tbody>
</table>
force, 319
force with potentials, 332
Hendrik A., 30, 41
invariance, 217
metric, 286
invariant, 105, 148, 275
mass, 295
transformation, 38
angle, 222
EM fields, 380
explicit form, 81
graphic representation, 120
history, 74
inverse, 87
naming, 74
non-commuting, 109, 284
nonrelativistic limit, 85
principle of relativity, 77
speed of light, 79
two in sequence, 107, 283
two orthogonal, 108
Lorentz
gauge, 411
gauge constraint, 409
LT (acronym) Lorentz
transformation, vi
Mach
Ernst, 41, 440
principle, 440
magnetar, 445
magnetic
dipole, 416, 417
electron $g$-factor, 419
field source, 415
force helical motion, 323
monopole, 328
Mandelstam variables, 295
mass
-potential particle motion, 373
and energy, 195
defect
atomic, 208
chemical, 210
nuclear, 211, 241
effective, 200
energy equivalence, 198
Lorentz invariance, 217, 295
origin, 16, 441
origin in the Universe, 213
potential equivalence, 206
rotational motion, 209
Maxwell
equations
covariant, 403
source-less, 327, 415
waves, 406
J. Clerk, xiii, 13, 14, 41
Mercury apsidal precession, 344
meter, 7
MeV, 204
Michelson
-Morley Experiment, 19
present day result, 187
Albert, 20, 41
Michelson-Morley Experiment, 19
Minkowski
Hermann, 5
space-time, 5, 277
Minkowski, Hermann, 42
MM (acronym) Michelson-Morley, vi
Molyneux, Samuel, 9
momentum, see also
four-momentum
conservation, 233
transfer, 299
Morley, Edward W., 19, 41
motion, relativistic constraints, 314
muon, 55
covariant motion, 301
decay
in rest-frame, 302
in the lab, 301
mean lifespan, 55
range, 301
time dilation, 84	ravel range, 56, 107
natural constants, variation of, 26
neutrino, 55
neutron, 55
magnetic g-factor, 419
star, 445
Noether
Amalie E., 233
theorem, 233
nuclear
energy in water, 213
fusion, 213
nucleon, 55
Occam’s Razor, 17
optics, relativistic regime, 354
pair production, 445
parallax, 9
parity, 381, 401
particle
beams, 224
collision CM-rapidity, 237
Data Group, 20
decay
Λ → π− + p−, 240
energy balance, 241
in CM-frame, 238
in flight, 242
into two photons, 243
product at rest, 244
entanglement, 159
momentum, 215
plane wave motion, 350
production, 258
relativistic focusing, 94
rapidity, 219
rapidity addition theorem, 220
spin entangled, 161
travel distance, 302
PeV, 204
photon, 201
conversion into a pair, 247
decay(?), 245
electron scattering, 265
Thompson limit, 266
Planck
constant h, 149, 264
energy, 188
length, 188 444
Max, 40
quantum of action h, 195
spacecraft collaboration, 191
plane wave
electromagnetic, 345
electron surfing, 354
polarization, 346, 356
power transfer to particle, 348
units, 351
Poincaré, Henri J., 39, 42
point-current, circulating, 417
positron
-i-um, 106
annihilation in flight, 251
production in heavy ion collisions, 447
potential
4-vector, 363
electromagnetic, 327
mass equivalence, 206, 207
retarded in time, 412
Poynting
force, 395
John H., 347
vector, 426
radiation, 451
vector motion, 396
principle of equivalence, 210
principle of relativity, 23
accelerated observers, 33
Lorentz transformation, 77
proper body energy, 195
proper time, 5, 27, 45, 50, 149
evolution parameter, 368
Lorentz invariance, 105, 276
proton, 55
magnetic g-factor, 419
to electron mass ratio, 26
pseudo-scalar, 383
QED, 432
quantum physics
  nonlocality, 159
quark
  -gluon plasma, 17, 32, 213, 445
  formation, 448
  confinement, 16, 428, 441
  particles, 55
radiation
  EM-field, 450
  friction, 459
    Caldirola, 457
    gravitational, 442
    Landau-Lifshitz (LL), 453
    Lorentz-Abraham-Dirac (LAD), 453
gauge, 329
power, 451
reaction force, 429
solar, 213
synchrotron, 448
rapidity, 109
  4-velocity, 292
  addition, 112, 113
  CM-frame, 237
movies, 114
particle, 219, 220
particle acceleration, 308
rocket, 227
rotation, 281
spaceship, 168
relativistic atomic orbits, 337
relativity, double special, 188
RHI collisions, 258
rocket
  nonrelativistic, 225
  relativistic, 227
Rutherford, Ernest, 265
scattering
  angle, 299
  Compton, 265
  Rutherford, 265
Thompson, 266
Schwarzschild radius, solar, 344
Schwinger critical field, 352, 445
second, 7
SI-units, 7, 204, 320
eV complement, 204
simultaneity of events, 121
space-time
  metric
tensor, 279
space
  isotropy, 40
  isotropy and homogeneity, 76
  rocket, 225
  travel, 167, 170, 225
    aging, 106
space-time
  homogeneity, 40
  metric, 277
    Lorentz invariance, 286
Minkowski, 5
spacelike, 152
speed of light, 7
  aberration effect, 10
  Fizeau-Foucault terrestrial measurement, 10
  in astronomy, 9
  terrestrial measurement, 11
spin covariant dynamics, 420
SR (acronym) special relativity, vi
SR test, 184
  balancing Doppler shift, 184
  coordinate transformation, 188
discrete symmetry, 189
GPS signal, 47, 186
MM interferometer, 186
proper time, 185
time and gravity, 190
time dilation, 49, 184
cancel gravity, 49
star light aberration, 98, 100
summation convention, 278
superposition principle, 401
synchronization, 152
Index

synchrotron radiation, 448

tachyons, 52, 53, 155
tensor, second rank, 279
test of SR, see SR test
TeV, 204
Thompson cross section, 266
time, 6
cosmological, 191
dilation, 27, 46, 123, 164
atomic clocks, 49
clock on a plane, 46
gravitational, 190
Larmor, 27, 80
muon, 56, 84
rocket, 171
satellite, 47
space travel, 169
test SR, 184
twins, 32
imaginary, 276
laboratory, 6
measurement, 59, 188
proper, 5, 50, 149
travel, 6
unit, 7

universality of measurement, 59
time dilation, 123
timelike, 151
transverse gauge
condition, 329
potential, 330

units
eV complement to SI, 204
Gauss, 320
natural, 8
SI, 7, 204, 320
Universe, expansion, 25, 213

vacuum, quantum, 441
variation
least proper time, 371
principle, 333, 367
vector, see four-vector
velocity, see four-velocity
Vlasov equation, 403

work-energy theorem, 215, 313
world line, 3, 119, 291

ZeV, 204