Influence of the Special Relativity Theory on GPS

Gonçalo Filipe Andrade dos Santos

Dissertation for acquisition of the Master’s Degree in

Electrical and Computer Engineering

Supervisors: Professor Doutor Carlos Manuel dos Reis Paiva
Professora Doutora Filipa Isabel Rodrigues Prudêncio

Examination Committee

Chairperson: Professor Doutor José Eduardo Charters Ribeiro da Cunha Sanguino
Supervisor: Professor Doutor Carlos Manuel dos Reis Paiva
Members of the Committee: Professor Doutor Paulo Sérgio de Brito André

November 2018
“The GPS is a remarkable laboratory of the concepts of special and general relativity.”

Neil Ashby
Declaration

I hereby declare that this document is an original work of my authorship that fully compliant with the specifications of the Code of Conduct and Good Practices of Universidade de Lisboa.
Acknowledgements

This dissertation represents the end of a long walk, throughout the last 5 years, marked by a constant personal, scientific, and social development and characterized by hard work and perseverance. The realization of this goal – the conclusion of the master’s degree – was accomplished not only through my effort, but also by the people that were always by my side along this path. For that reason, it is important to acknowledge and be grateful to the people that have always supported me given that alone and by ourselves nothing can be accomplished. Without them this path would have been much more difficult.

This significant challenge would not be accomplished without the support, patience, and guidance of Professor Carlos Paiva, my supervisor and mentor. His wisdom, knowledge, and commitment to the highest standards inspired and motivated me. I would also like to thank Professor Filipa Prudêncio for accompanying me during the last months, continuously showing availability and patience.

A special thanks to my parents, they are the main pillars of my life and my reason to live. With no less importance a special thanks to my brother, João, for his effort and hours put into helping me in achieving the desired result. It is a privilege to have them has a family, I have no words to express my gratitude for what they have done for me.

Additionally, I would like to thank my grandmother as I would not be the person I am today without her values and teachings. A thank you to my girlfriend Inês, for helping me and for always having a friendly piece of advice at the right moment. Also, for being present every step of the way and for being a faithful companion, thanks to my dog Simba.

Lastly, I would like to thank my colleagues and friends, that watched me closely in this journey, contributing for my personal and professional growth.
Abstract

The Global Positioning System (GPS) provides both a global position and a time determination. The system utilizes stable and precise satellite atomic clocks. These clocks suffer from relativistic effects, mainly due to time dilation and gravitational frequency shifts, that cannot be neglected. Without considering these effects, GPS would not work properly because timing errors of $\Delta t = 1 \text{ ns}$ will lead to positioning errors of magnitude $\Delta x \approx 30 \text{ cm}$. In the end, the total correction per day, due to relativity (both special and general theories), is approximately, $39 \mu \text{s day}^{-1}$, which corresponds to an imprecision of $12 \text{ km day}^{-1}$.

The main objective of this dissertation is to discuss the conceptual basis, founded on special relativity, that ensure a proper operation of the GPS. All the consequences and effects stemming from this theory, are explained based on the geometric approach to the hyperbolic plane which is commonly known as Minkowski diagrams in the literature. To have a correct geometric intuition on this plane, equipped with a non-Euclidean metric, one has to make extensive use of two concepts: equilocs and equitemps. The main consequence of the universality of the speed of light (in vacuum) is that equitemps from different observers are not parallel, i.e. simultaneity is relative.

Quoting Neil Ashby: “The GPS is a remarkable laboratory of the concepts of special and general relativity.”

Keywords: GPS, Relativistic effects, Time dilation, Minkowski diagram, Hyperbolic plane, Bondi’s factor
Resumo

O GPS permite a determinação da posição e do tempo numa escala global, este sistema utiliza relógios atómicos incorporados em cada um dos satélites. Estes relógios sofrem dos efeitos relativistas, principalmente da dilatação do tempo e das mudanças de frequência gravitacional. Estes efeitos não podem ser desprezados, uma vez que sem os contabilizar o GPS não funcionaria corretamente, isto porque um atraso de tempo de $\Delta t = 1\, ns$ corresponde a um erro de posicionamento de $\Delta x \approx 30\, cm$. O valor da correção total devido à relatividade (restrita e geral), aplicado aos relógios dos satélites, ao fim de um dia sem correções é de $39\mu s\, dia^{-1}$. Este valor corresponde a um erro de posicionamento de $12\, km\, dia^{-1}$.

O objetivo desta dissertação é discutir os conceitos básicos da teoria da relatividade restrita, de forma a assegurar o funcionamento correto do GPS. Todas as consequências e efeitos provenientes desta teoria são explicados com base na abordagem geométrica ao plano hiperbólico, mais conhecido na literatura por diagramas de Minkowski. De forma a ter-se uma correta intuição geométrica sobre este plano, definido por uma métrica não euclidiana, tem de se recorrer aos conceitos fundamentais: equiloc e equitemp. A principal consequência da universalidade da velocidade da luz, no vazio, é que equitemps de diferentes observadores não são paralelas, i.e., simultaneidade é um conceito relativo.

Citando Neil Ashby: “The GPS is a remarkable laboratory of the concepts of special and general relativity.”

Palavras-chave: GPS, Efeitos relativistas, Dilatação do tempo, Diagramas de Minkowski, Plano hiperbólico, Factor de Bondi
# Table of Contents

Declaration ........................................................................................................................................ iii

Acknowledgements ........................................................................................................................... v

Abstract ....................................................................................................................................... vii

Resumo .......................................................................................................................................... ix

Table of Contents ............................................................................................................................ xi

List of Figures and Tables ................................................................................................................ xv

List of Acronyms ............................................................................................................................. xvii

List of Symbols .............................................................................................................................. xix

Chapter 1 – Introduction .................................................................................................................. 1
  1.1. Motivation ............................................................................................................................... 1
  1.2. Goals and Structure of the Report ......................................................................................... 2
  1.3. Historical Framework ........................................................................................................... 3
  1.4. Original Contributions ......................................................................................................... 11

Chapter 2 – GPS ............................................................................................................................ 13
  2.1. Historical Framework ........................................................................................................... 13
  2.2. GNSS Systems ..................................................................................................................... 16
  2.3. GPS Basic Idea ................................................................................................................... 17
  2.4. Satellite Communication .................................................................................................... 20

Chapter 3 – Special Relativity: Fundamental Aspects ................................................................. 21
  3.1. Introduction ......................................................................................................................... 21
  3.2. IFR – Inertial Frame of Reference ....................................................................................... 21
  3.3. Einstein’s Special Relativity Postulates .............................................................................. 22
  3.4. Galileu Transformation ....................................................................................................... 23
  3.5. Minkowski Diagrams .......................................................................................................... 25
  3.6. Some Definitions Derived from Einstein’s Second Postulate ........................................... 28
  3.7. Geometry of Relative Simultaneity ................................................................................... 30
List of Figures and Tables

Figure 1.1.: Mechanics field of applicability ................................................................................. 1
Figure 2.1.: Trilateration [28] ........................................................................................................ 18
Figure 3.1.: Galilean addition of velocities ...................................................................................... 23
Figure 3.2.: Two IFR’s in relative motion .......................................................................................... 24
Figure 3.3.: Galilean transformation graphical representation ......................................................... 25
Figure 3.4.: Spacetime Minkowski diagram ...................................................................................... 25
Figure 3.5.: Representation of electromagnetic signals in a Minkowski diagram ......................... 26
Figure 3.6.: Event interpretation from different observers in Minkowski diagrams [101] ............... 28
Figure 3.7.: Minkowski diagram of Alice’s point of view for the wagon example ......................... 31
Figure 3.8.: Example with two electromagnetic signals simultaneously from the wagon ............... 32
Figure 4.1.: Bondi’s radar method (1) ............................................................................................. 35
Figure 4.2.: Bondi’s radar method (2) ............................................................................................. 36
Figure 4.3.: Minkowski diagram example for Lorentz boost and interval invariance ..................... 40
Figure 4.4.: $\gamma$ factor in function of $\beta$ ..................................................................................... 42
Figure 4.5.: Geometrical representation of the set of events that are in the same spacetime interval ... 43
Figure 4.6.: Geometrical representation of the set of events that are in the same spacetime interval, defined by the two different metrics (Euclidean and Minkowski). .................................................................................. 44
Figure 4.7.: Representation of the unitary vectors with their respective different dimensions through Lorentz boost (active transformation) .................................................................................. 46
Figure 4.8.: Representation of the time dilation effect ..................................................................... 47
Figure 4.9.: Representation of the length contraction effect ............................................................. 49
Figure 4.10.: Representation of an example for the time reciprocity ............................................. 52
Figure 4.11.: Representation of an example for the space reciprocity ............................................. 53
Figure 5.1.: Relativistic Composition of Velocities ........................................................................ 56
Figure 5.2.: Representation of an electromagnetic signal that passes by three different observers ... 56
Figure 5.3.: Representation of an impossible example where the effect happens before the cause .... 58
Figure 5.4.: Representation of the collision example ..................................................................... 59
Figure 5.5.: Apparent superluminal speeds ....................................................................................... 60
Figure 5.6.: Representation of an observer that sends periodically signals to another observer .......... 61
Figure 5.7.: Circular movement .................................................................................................... 63
Figure 5.8.: Representation of both twins’ voyage ........................................................................... 65
Figure 5.9. a and b: Unitary vectors associated with each observer ......................................................... 66
Figure 5.10. a and b: Representation of all electromagnetic signals sent and received by the two twins... 68
Figure 6.1.: Sagnac effect illustration ........................................................................................................ 75
Figure B.1.: Different world lines in a spacetime Minkowski diagram ...................................................... 85
Figure B.2: The twin paradox with a parabolic section .............................................................................. 86
Figure C.1.: Introduction of the referee clock in order to synchronize both clocks .................................... 91
Figure C.2.: Time set for different clocks ................................................................................................ 92
Figure E.1.: Graphical representation of how to obtain the geometric mean of two lengths ................. 95
Figure F.1.: Relation between mass and energy of a particle ................................................................... 98
# List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>GNSS</td>
<td>Global Navigation Satellite System</td>
</tr>
<tr>
<td>IFR</td>
<td>Inertial Frame of Reference</td>
</tr>
<tr>
<td>AFR</td>
<td>Accelerated Frame of Reference</td>
</tr>
<tr>
<td>NIFR</td>
<td>Non-Inertial Frame of Reference</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>LOS</td>
<td>Line of Sight</td>
</tr>
<tr>
<td>NNSS</td>
<td>Navy Navigation Satellite System</td>
</tr>
<tr>
<td>DOD</td>
<td>Department of Defense</td>
</tr>
<tr>
<td>JPO</td>
<td>Joint Program Office</td>
</tr>
<tr>
<td>AS</td>
<td>Anti-Spoofing</td>
</tr>
<tr>
<td>SA</td>
<td>Selective Availability</td>
</tr>
<tr>
<td>IOC</td>
<td>Initial Operational Capability</td>
</tr>
<tr>
<td>CDMA</td>
<td>Code Division Multiple Access</td>
</tr>
<tr>
<td>SPS</td>
<td>Standard Positioning Service</td>
</tr>
<tr>
<td>PPS</td>
<td>Precise Positioning System</td>
</tr>
<tr>
<td>ECEF</td>
<td>Earth Centered Earth Fixed</td>
</tr>
<tr>
<td>ECI</td>
<td>Earth Centered System of Coordinates</td>
</tr>
<tr>
<td>SYPOR</td>
<td>Système de Positionnement Relativiste</td>
</tr>
<tr>
<td>GDOP</td>
<td>Geometric Dilution of Precision</td>
</tr>
</tbody>
</table>
List of Symbols

\( n \) Refraction index of the material
\( \eta \) Amount of dragged ether
\( c \) Speed of light in vacuum
\( v \) Relative velocity between two observers
\( m \) Inertial mass
\( a \) Acceleration
\( S \) Resting spacetime system of coordinates
\( S' \) Motion spacetime system of coordinates
\( x, y, z \) Space coordinates
\( t \) Time coordinates
\( \beta \) Normalized relative velocity
\( \mathbf{B} \) Magnetic field vector
\( \mathbf{H} \) Magnetic field intensity vector
\( \mathbf{D} \) Electric displacement field vector
\( \mathbf{E} \) Electric field vector
\( \mathbf{F} \) Lorentz force
\( q \) Electric charge
\( \mathbf{v} \) Instantaneous velocity
\( \mathbb{R}^4 \) Quadri-dimensional space
\( \mathbb{R}^{1,1} \) Minkowski spacetime with 1 time and space axis
\( \mathbb{R}^{3,1} \) Minkowski spacetime with 1 time axis and 3 space axes
\( \mathbf{E}_x \) Electric field x component
\( \mathbf{E}_y \) Electric field y component
\( \mathbf{E}_z \) Electric field z component
\( \mathbf{B}_x \) Magnetic field x component
\( \mathbf{B}_y \) Magnetic field y component
\( B_z \) Magnetic field z component
\( \varepsilon_0 \) Electric permittivity in free space
\( \mu_0 \) Magnetic permeability in free space
\( \nabla \) Gradient operator
\( J \) Current density
\( \rho \) Electric charge density
\( u_0, u_1, u_2 \) World Lines
\( \oplus \) Composition of velocities operator
\( L_0 \) Proper length
\( L \) Relative length
\( \theta \) Euclidean angle between different inertial frames of reference
\( \kappa \) Bondi’s factor
\( O, \mathcal{D}, \mathcal{Q}, \mathcal{E}, \mathcal{R} \) Equilocs from different observers in a spacetime Minkowski diagram
\( O_\perp, \mathcal{D}_\perp, \mathcal{Q}_\perp, \mathcal{E}_\perp, \mathcal{R}_\perp \) Equitemps from different observers in a spacetime Minkowski diagram
\( \tau_A \) Own time
\( s \) Slowing down factor
\( \gamma \) Lorentz transformation factor
\( \dot{\phi} \) Lorentz boost quickness
\( e_0, f_0, g_0 \) Time unitary vector of the Minkowski spacetime frame
\( e_1, f_1, g_1 \) Space unitary vector of the Minkowski spacetime frame
\( w, u \) Proper velocity of an event
\( f_r, f_l \) Frequencies
\( T, T' \) Period of time coordinates of the Minkowski spacetime frame
\( \varepsilon \) Error
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Phi$</td>
<td>Potential gravitational energy of a particle</td>
</tr>
<tr>
<td>$G$</td>
<td>Gravitational field</td>
</tr>
<tr>
<td>$M$</td>
<td>Earth's mass</td>
</tr>
<tr>
<td>$R$</td>
<td>Earth's radius</td>
</tr>
<tr>
<td>$\omega$</td>
<td>Angular velocity</td>
</tr>
<tr>
<td>$A$</td>
<td>Area covered by the vector</td>
</tr>
<tr>
<td>$e$</td>
<td>Eccentricity</td>
</tr>
</tbody>
</table>
Chapter 1 – INTRODUCTION

1.1. Motivation

The starting point for the development of this master’s dissertation was introduced in Photonics course, my choice being in part based on the contents lectured which increased my interest in learning more about the special relativity theory and relativistic effects in general. Besides the theoretical interest, relativistic effects are important practical engineering concerns related to daily life relevant situations. One of the first and most used applications that took into consideration all the relativistic effects was the GPS [1], this system deals with huge velocities. As the electromagnetic signals sent from the satellites travel at the speed of light and the system is subject to the gravitational force, the error associated with the relativistic effects cannot be neglected. The GPS operating system incorporates many important science and engineering areas such as: aerospace engineering for its satellite system; telecommunications engineering for the emission of radio wave signals; quantic mechanics for measuring its propagation time with atomic clocks; computers engineering and mathematics for the calculation of the real time position.

Albert Einstein, the author of special relativity theory, transformed classic theoretical physics and astronomy, introducing different concepts that include relativity of simultaneity, kinematic and gravitational time dilation, and length contraction.

As described in figure 1.1. Newton’s classical mechanics provided accurate results when referring to large objects, not extremely massive, and with velocities smaller than the speed of light. However, classical mechanics failed to explain some situations, for instance, if the object had the size of an atom, it was necessary to introduce quantum mechanics, if the velocities were comparable with the speed of light, special relativity had to be accounted for, and the study of objects extremely massive required general relativity.

![Figure 1.1.: Mechanics field of applicability [2].](image-url)
1.2. Goals and Structure of the Dissertation

The structure of the dissertation is divided in seven chapters, each one subdivided in several sections. In this introductory chapter it is initially reviewed the historical framework of relativity theories – with greater emphasis on special relativity. Subsequently, the motives that took me to the realization of this dissertation are presented, specifying objectives of the dissertation, its structure, organization, and planning. Lastly, the original contributions section showcases the new and innovative ways in which these subjects were addressed and graphically interpreted.

The second chapter begins by describing the evolution of the GPS over the years (historical framework). Afterwards, GNSS systems, and in particular the GPS and its constitutive parts, are presented. Finally, the basic idea behind this system and how it is able to work properly is addressed.

The third chapter starts by defining an IFR. This topic is followed by the Einstein’s postulates, which are the base of his article about the special relativity theory. Moreover, the Galileu transformation is presented along with Minkowski’s interpretation and representation of space and time diagrams. Consequently, some important conclusions drawn from the postulates are listed. The end of the chapter explains the concept of relative simultaneity and depicts the geometrical construction of an IFR.

The fourth chapter, entitled special relativity consequences, thoroughly describes the Bondi’s factor, time dilation, length contraction, Lorenz transformation, spacetime interval invariance and its conclusions, Euclidean geometry, the hyperbolic plane through the Minkowski diagrams, Minkowski quadratic space, and reciprocity of space and time. The Bondi’s factor is described to show the construction of a Minkowski diagram and to deduce and explain the time dilation through the Bondi’s radar methodology.

In the fifth chapter other important special relativity theory effects and paradoxes are introduced. This part begins with the description of the Einsteinian velocity composition and is followed by the causality effect. Additionally, other topics are described such as the apparent superluminal speeds, the longitudinal and transverse Doppler effects, and the twin paradox.

In the sixth chapter the main relativistic effects on the GPS are considered, particularly the time dilation effect, which originated with the special relativity theory. Furthermore, the impact of all the relevant relativistic effects are precisely calculated given that without accounting for them, the GPS would not work.

The seventh and last chapter of this dissertation provides the reader with the attained conclusions regarding all chapters and some suggestions for future work in this field, since the evolution of science requires a continuous effort.
The appendixes are composed by six important topics that need to be addressed in order to clarify and consolidate the main body of this dissertation.

1.3. Historical Framework

Physics is our attempt to conceptually justify the occurrences of the observable world. The ways in which individual particles interact with each other can be described by the fundamental interactions of physics [3], which are the forces of gravity and electromagnetism – present in everyday life – and the weak and strong interactions – introduced when discussing the nuclear phenomenon.

Until Galileo’s era (1564–1642), light was a mystery, something unknown. Galileo was the first person that tried to measure its speed, without success, having concluded that the speed of light was too big to be measured. Around the same period, Descartes (1569–1650), through his study of lunar eclipses, concluded that the speed of light was infinite [4]. Thus, the hypothesis that light was made of particles was excluded, assuming that no known particle has unlimited speed. This also meant that the particle was omnipresent because the propagation time, considering its velocity, had to be zero.

Influenced by Christiaan Huygens’ (1629–1695) Wave Front Theory [5], Sir Isaac Newton (1642–1727) defined a model, known as The Corpuscular Theory of Light [6], to explain the nature of light. This model represented light as the microscopic particles flux that is emitted from a light source at high speed. Such representation allowed the creation of a mechanic and deterministic model of moving material bodies with defined characteristics. Subsequently, according to Newton, the relativistic principle was applied to the light as it was to other materials. This principle stated that if a system of coordinates S is chosen, when physical laws are preserved, the same laws apply to any other system of coordinates, S', moving in uniform translation relatively to S.

The first reliable measurement of the speed of light which proved it to be finite, was done by Ole Rømer (1644–1710), in 1675. Studying the eclipses of Jupiter satellites, he learned that these phenomena happened at different times than expected when the Earth was moving away or getting closer to Jupiter [7]. This fact led to the conclusion that light took time to propagate at a certain distance, in this particular case, the orbital distance of Earth. In the end, Descartes theory that light was omnipresent, was refuted.

This conclusion was also supported by the Wave Front Theory mentioned before, which suggested that light could be constituted by waves. As thought at the time, any wave needs a material medium to propagate. Hence, light, or any other electromagnetic signal, was believed to propagate through a transparent medium that surrounded the entire space and that was designated ether [8]. Analogously to air for the sound, the “ether’s wind” should originate an anisotropy from the electromagnetic signals, that depend on the direction and the sense of propagation. Several experiences were developed to try and observe the ether’s influence on the
speed of light, but none of them could achieve a measurable speed for the so called “ether’s wind”. The lack of results of these experiments enlarged the mystery surrounding the relativity principle, and the theory of the luminiferous ether as the hypothetical medium for the propagation of light was widely debated.

Another important contribution was made by Thomas Young (1773–1829) [9]. Young’s interference experiment shown that there was interference in the context of light as a wave, proving, once again, that light behaves as a wave.

Until the beginning of the XIX century, people agreed on the Corpuscular Theory of Light. However, the mindset started to change after some important developments. Among these developments, the state of motion of Earth with respect to ether started to be discussed. More specifically, whether or not the ether is dragged by moving matter (partial ether dragging), theory defended by Augustin Fresnel (1788–1827).

François Arago (1786–1853), in 1810, claimed that the modification in the refractive index of a substance could provide a method to measure the light speed. With the different speed of the stars and the motion of Earth at different times, the expectation was to obtain a range of different angles of refraction, taking into account the refractive index depends on the speed of light. Instead, the results showed that there was no difference in the refraction index, this is because what he was truly observing was an astronomical phenomenon which produces an apparent motion of celestial objects around their true position, depending on the velocity of the observer [10]. In 1818, inspired by Arago’s results – even if light was transmitted as waves, the refractive index had to change –, Fresnel proposed that a transparent material body in motion could drag some of the ether with it [11]. In other words, the speed of light in the prism would need an adjustment, proportional to the amount of dragged ether ($\eta$), which only depended on the refractive index of the material ($n$) [12]

$$\eta = 1 - \frac{1}{n^2}.$$ (1.1)

The conclusion of Fresnel’s Theory of Wave Optics is that no relative motion exists between Earth and the ether and, consequently, light can pass through any object [8]. Young and Fresnel both believed that light propagated as a transversal wave within an elastical medium called ether.

On the opposite spectrum of the discussion, ether was considered to be a part of moving matter, being rigid for fast objects and fluid for slower objects. In other words, ether behaved as a viscous liquid, adherent to the surface of the bodies, being completely carried by Earth (complete ether dragging), hence, at rest in relation to it near the ground. The implication was that any optical experience near Earth was independent from Earth's movement [8], this theory was argued by George Stokes (1819–1903).

Despite there being a huge development that resulted from all these experiments and authors, there was still a struggle to measure the speed of Earth in relation to ether. Nonetheless, the
ether theory could not be discarded as it offered a qualitative explanation for electromagnetic forces and for light propagation in vacuum.

In 1851, Fizeau (1819–1896) performed an experiment that verified the theoretical prediction of Fresnel [13]. Taking into consideration it was already known that the speed of light with water at rest was $c / n$, the experiment consisted in measuring the speed of light within a tube filled with water moving inside with velocity $v$. The speed of light obtained was the sum of the speed of light, in relation to water, with a term corresponding to the water’s velocity – $\eta v + c / n$.

Fizeau’s experiment favored Fresnel’s partial ether dragging theory (stationary ether theory). Yet, a distinction between optical and electrodynamic phenomena must be highlighted and different models needed to be developed. After important work from Michael Faraday (1791–1867) and William Kelvin (1824–1907), James Clerk Maxwell (1831–1879), in 1864, created a precise electromagnetism theory, through a set of equations that brings together, electricity, magnetism, and inductance, based on the constitutive relations that exist in the vacuum

$$\begin{align*}
D &= \varepsilon_0 E \\
B &= \mu_0 H.
\end{align*}$$ (1.2)

The so-called Maxwell’s equations were written as follows

$$\begin{align*}
\nabla \times E &= -\frac{\partial B}{\partial t} . \\
\nabla \cdot B &= 0 \\
\nabla \times B &= \mu_0 J + \frac{1}{c} \frac{\partial E}{\partial t} . \\
\nabla \cdot E &= \frac{\rho}{\varepsilon_0} .
\end{align*}$$ (1.3)

Maxwell developed a complete mathematical model with an unsuitable theory attached, resulting in an inability to present a mechanical description of the ether. More specifically, the physicist supposed the phenomena were analyzed from a system of coordinates at rest in relation to the ether and that, possibly, new effects could appear when considering motion systems of coordinates. In the end, Maxwell conclude that, analogously to Faraday’s electromagnetic experiment (electromagnetic induction), in some cases only the motion of objects produced some effects. In 1831, Michael Faraday had shown that if there was movement between a magnet and a winding, when both were close to each other, a current would be produced in the conductor.

Nevertheless, only after Heinrich Hertz (1857–1894) shown the existence of electromagnetic waves, would Maxwell’s equations gain credibility.

August Foppl (1854–1924), based on Maxwell’s study, went even further and affirmed that the induction effect depended only on the relative motion between objects. In fact, if both objects moved together as a group, no effect would be produced [14].
In 1881, Albert Michelson (1852–1931) tried to measure the relative motion of Earth in relation to ether. Utilizing an interferometer, he could not determine any relative motion, interpreting the result as a validation of Stokes complete ether dragging theory. However, it was later proved that his calculations were wrong and that Stokes’ theory lead to contradictory consequences.

Years later, Albert Michelson (1852–1931) and Edward Morley (1838–1923) replicated the previous experiment [15] with higher measurement accuracy, achieving a negative result which could explain the complete ether dragging theory – this outcome directly contradicted Fresnel’s stationary ether theory. In other words, physicists were confronted with two apparently contradictory experiments made by Michelson.

Hendrik Lorentz (1853–1928) and FitzGerald (1851–1901) proposed, independently, a solution to Fresnel’s ether theory. FitzGerald, from observing Michelson-Morley’s experiment, concluded that the ether’s effect was being cancelled and speculated that the intermolecular forces were possibly of electrical origin in which, the motion of bodies through the ether would produce a contraction of their length [16]. This was in line with the work of Heaviside (1887) who had determined that the electrostatic fields in motion were deformed.

Another possible solution to the problem, contrary to FitzGerald’s, was shown by Woldemar Voigt (1850–1919). In 1887, the physicist published a work about the Doppler Effect on light waves, [17] where he analyzed the variations of frequency and wavelength when there was motion in relation to the ether. Let us consider the wave propagation equation and its luminous proprieties, in different systems of coordinates

\[
\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2 U}{\partial t^2} = 0. \tag{1.4}
\]

Voigt achieved his transformation relations by assuming that the previous equation is unchangeable for all different systems of coordinates (covariant), which would explain the negative result of the Michelson–Morley experiment, and by considering two systems of coordinates – \( S(x,t) \) and \( S'(x',t') \) – moving one in relation to the other with velocity \( v \) and direction of coordinate \( x \). This coordinate already corresponded to a relation with physical meaning known by classic mechanic

\[
x' = x - vt. \tag{1.5}
\]

If this equation, \( x' \) is the position of the object in relation to \( S' \), \( x \) is the position of the object in relation to \( S \) and \( t \) is the time elapsed since the instant in which the systems of coordinates intersect each other.
Considering \( y' = y, z' = z \) and \( t' = t \), the equation does not maintain the same form when changing the system of coordinates. In this case, Voigt transformation is presented as follows [18]

\[
\begin{align*}
    x' &= x - vt \\
    y' &= y \sqrt{1 - \frac{v^2}{c^2}} \\
    z' &= z \sqrt{1 - \frac{v^2}{c^2}} \\
    t' &= t - v \frac{x}{c^2}
\end{align*}
\]  

(1.6)

The Voigt transformation includes the Lorentz factor for the \( y \) and \( z \) coordinates and a new time variable, which was later called "local time". Despite its contributions, Voigt's work was completely ignored by his contemporaries.

Lorentz, in his model, assumed the existence of electrons separated from the ether, alongside the ideas defended by Stokes theory. Consequently, Lorentz assumed that the notion of the velocity of light was independent of the velocity of the source that emitted it. In essence, an abstract electromagnetic ether theory was created by trying to explain the mechanical processes through electromagnetic ones, without explaining neither their nature nor ether’s nature. In his theory, Lorentz calculated, like Oliver Heaviside (1850–1925), the contraction of the electrostatic fields.

In 1895, Lorentz introduced the Theorem of Corresponding States. This theorem states that, in relation to the ether, a moving observer or an observer at rest, from their own different systems of coordinates, make the same observations. At the time, the equations that were previously thought to be valid only in relation to resting systems of coordinates, became valid in relation to moving systems of coordinates as well. Consequently, if the same phenomenon produced different results for a moving system of coordinates and a resting one, that difference could determine Earth’s movement in relation to the ether.

Lorentz was able to prove that Maxwell’s equations were valid in any given system of coordinates through Lorentz transformation

\[
\begin{align*}
    x' &= x - vt \\
    y' &= y \\
    z' &= z \\
    t' &= t - v \frac{x}{c^2} \\
    E' &= E + v \times B \\
    B' &= B - v \times \frac{E}{c^2}
\end{align*}
\]  

(1.7)
As Voigt had done, Lorentz assumed the relation previously known for the coordinate $x$ and considered that the rest of the coordinates did not suffer any changes. An important part of this system was the definition of local time, despite it having no physical meaning [19]. Additionally, the transformation relations of the electric and magnetic fields were deduced from the force of Lorentz

$$ F = q(E + v \times B). \quad (1.8) $$

With these concepts, Lorentz could explain the Aberration of Light, the Doppler Effect, and the Fizeau experiment. The contraction length assumption made it possible to create a compatibility capable of explaining both Stokes’ and Fresnel’s theory.

In 1893, Joseph Larmor (1857–1942) published his work, in which he tried to justify, based on Maxwell’s equations, that light was an electromagnetic phenomenon and that it had not been possible until then to measure the speed of the ether in relation to Earth. His conclusion was that every electromagnetic phenomenon occurred in the exact same way, independently of the given system of coordinates, in relation to ether. Larmor went even further and was the first to develop an algebraically equivalent to the Lorentz transformation. Nonetheless, these equations maintain the form of Maxwell’s equations – to second order of $v / c$ –, hence they are not entirely correct.

In 1900, Larmor published a book where he correctly formulated the time and space transformations that maintain Maxwell’s equations invariant. The results obtained were the first correct version of Lorentz transformation [20]

$$
\begin{align*}
    x' &= \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}} \\
    y' &= y \\
    z' &= z \\
    t' &= t \sqrt{1 - \frac{v^2}{c^2}} - x' \frac{v}{c^2}
\end{align*}
$$

By substituting $x'$ on the fourth equation we obtain Lorentz transformation. This transformation is recognized as Lorentz instead of Lamor because these equations only make sense when combined with the transformation of the correct electromagnetic variants, combination only advanced by Lorentz.

Henri Poincaré (1854 – 1912), discussing Lamor’s ideas, affirmed that only the motion between material bodies could be measured, being impossible to measure either the absolute motion of matter or the relative motion of matter in relation to the ether. In 1899, Poincaré published an article entitled the “Relative Motion Principle” in which he stated that all the effects that depend on the movement of a system in relation to the ether should be cancelled, being impossible to detect the system’s movement. Moreover, Poincaré also said that the contraction of bodies, formulated by FrizGerald, was merely a justification for Michelson – Morley’s experiment. In 1900,
a new article was published, in which a physical explanation for Lorentz local time was given, proving that this time transformation represented the time measured when clocks were synchronized with the help of light signals [21].

Another article, published in 1904 and written by Lorentz, proposed an exact theory for the electromagnetism of bodies in motion. This article was composed of two different parts; in the first part, space and time transformations of coordinates passing by a resting system of coordinates \( S \) into a motion system of coordinates \( S' \) are presented by utilizing the classical physics transformations, also referred to as Galilean transformation

\[
\begin{align*}
    x' &= x - vt \\
    y' &= y \\
    z' &= z \\
    t' &= t
\end{align*}
\]  

(1.10)

In the second part, the article presents the following mathematical transformations, through which Maxwell laws become covariant

\[
\begin{align*}
    x'' &= \frac{x'}{\sqrt{1 - \frac{v^2}{c^2}}} \\
    y'' &= y' \\
    z'' &= z' \\
    t'' &= t' \sqrt{1 - \frac{v^2}{c^2}} - \frac{x' \frac{v}{c}}{\sqrt{1 - \frac{v^2}{c^2}}}
\end{align*}
\]  

(1.11)

By combining the previous sets of equations, we reach the sought for result

\[
\begin{align*}
    x'' &= \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}} \\
    y'' &= y \\
    z'' &= z \\
    t'' &= \frac{t - \frac{vx}{c}}{\sqrt{1 - \frac{v^2}{c^2}}}
\end{align*}
\]  

(1.12)
Lorentz also obtained the electromagnetic field transformations

\[
\begin{align*}
E_x' &= E_x \\
E_y' &= \frac{E_y - v \cdot B}{\sqrt{1 - \frac{v^2}{c^2}}} \\
E_z' &= \frac{E_z - v \cdot B}{\sqrt{1 - \frac{v^2}{c^2}}} \\
B_x' &= B_x \\
B_y' &= \frac{B_y + E_y \cdot \frac{v}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}} \\
B_z' &= \frac{B_z + E_z \cdot \frac{v}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}}
\end{align*}
\]

(1.13)

With this set of equations, Lorentz proved the impossibility of measuring Earth motion in relation to the ether, through either electromagnetic or optical media. Additionally, Lorentz also proved that the length contraction is not only a result from experiments but also a consequence of the change of forces associated with the particles that compose each body.

In 1905, Einstein (1879 – 1955) wrote two important articles, which played a significant role in the relativity theory [22]. The first one, entitled “On the electrodynamics of moving bodies” contains essentially what is now known as the special relativity theory. Special relativity applies to elementary particles and their interactions, explaining all their physical phenomena except for gravity. Whereas, general relativity describes the law of gravitation and its relation to the other forces of nature. In this article, Einstein obtains the same results obtained by many others before him, however, the main difference lies in the simple form trough which Einstein sheds light on intuitive ideas about measurements of space and time.

The second article, based on the first one, was entitled “Does the inertia of a body depend upon its energy-content?” where, for the first time, the equivalence of mass and energy, was proposed. It became henceforth clear that a body’s total energy or total mass (characteristics associated with the body in motion state) is greater than that same body rest energy or rest mass.
(characteristics associated with the body in stationary state), respectively. The following equivalence serves also to convert units of energy into units of mass through the formula

\[ E = mc^2. \tag{1.14} \]

Special relativity theory makes a wide revision of our common-sense conceptions. Before Einstein, the speed of light was interpreted in relation to the ether – if the ether could be used to characterize a system of coordinates in absolute rest, Maxwell’s equations were considered valid in that system. Special relativity, taking into account the several experiments that failed to detect its existence, denied the existence of the ether and, by extension, showed Maxwell’s equations were compatible with the principle of relativity – only theory to show this coherence. In fact, Einstein explained that the search for this medium was unnecessary as the only property that mattered was the ability to transmit electromagnetic waves, physical property the space had [23].

Einstein’s relativity theory is not complex or difficult to understand. The results, as presented by the physicist, were summarized in two postulates. These were clear from both a conceptual and a mathematical point of view, with a methodology transversal to every case. Thus, classic theoretical physics and astronomy were transformed, and new concepts were introduced, such as relativity of simultaneity, kinematic, gravitational time dilation, and length contraction.

In 1908, Herman Minkowski (1864 – 1909), in a lecture entitled “Space and Time”, attempted to explain to an audience without any mathematical background the special relativity theory, representing the movement of objects through diagrams (later called Minkowski Diagrams). Besides the movement of the object, Minkowski also illustrated, for each observer, the events that could happen in the future or in the past in relation to a certain reference. This lead him to make the distinction between time and space vectors \( r \in \mathbb{R}^{3,1} \).

### 1.4. Original Contributions

In order to truly understand the special relativity theory, the study of this subject should not start with the algebraic manipulations of the Lorentz transformation. Instead first it should be comprehended how the concept of simultaneity of two events, a relative concept, depends on the observer. This knowledge should be acquired in both a physical and a graphical manner. Afterwards, the main consequences of the special relativity theory should also be studied without resorting to Lorentz transformation.

The principal contribution of this dissertation lies on an innovative geometrical interpretation, which is a graphical and intuitive way of approaching this subject, without resorting to the more classical Lorentz transformation approach. This geometric approach to the hyperbolic plane through the Minkowski diagrams, and despite its fast development, has a slow acceptance within the scientific community. This slow acceptance is given, mainly, by the lack of knowledge, as a mathematic tool applicable in diverse subjects, particularly, in special relativity theory and electromagnetism.
A key feature of the relativity theory is that there is an intimate relation between physics and geometry. In special relativity the way to do physics is in a flat spacetime and the lesson this dissertation intends to convey is that the Euclidean metric is inappropriate for flat spacetime. Hence, it should be adopted, instead, the Minkowski metric.
Chapter 2 – GPS

2.1. Historical Framework

From an early age, mankind has looked to the sky with wonder, seeking intriguing signs. Men became a specialist in deciphering the mystery of the stars and developed rules for guiding their life based on the position of these bodies. For instance, the world’s first surveyors were priests that foretold the right time to plant the corps; the alignment of both the pyramids and the Stonehenge was due to celestial observations; Egyptian engineers used high distance control points to repair constructions destroyed by flooding; French surveyors utilized triangulations in large scale to measure the land extension of France; sea travelers confided on angular measurements of celestial bodies to pinpoint their location. The triangulation technique was posteriorly used for determining accurate coordinates over continental distances.

Man’s ambition to use science to help and develop society and the desire to master time and space took him to expand the chain of technical developments from the simple observer to the present satellite geodesists. The surveyor’s work that has remained the same over all these years, is to determine land borders, create maps of his environment and oversee the construction of projects [24].

In the beginning, the techniques utilized for surveying were the previously mentioned triangulation and trilateration. Trilateration, in geometry, is the process of determining absolute or relative locations of points by measuring distances, using the geometry of circles, spheres, or triangles. In surveying, triangulation involves only angle measurements, rather than measuring distances to the point directly, as in trilateration. As these techniques were limited by the Line of Sight (LOS), observers created survey towers to expand it. Utilizing these techniques and with the help of the stars position, it was possible to determine the position of any given point on the surface of the Earth. These astronomic positions could have a huge error associated hence, although each continent was well delimited, their interrelationship was inaccurate.

The world’s first artificial satellite – Sputnik I – was launched by the Soviet Union, on October 4, 1957. It had the size of a beach ball (58 cm of diameter), a weight of 83.6 kg, and was equipped with two battery-powered transmitters (20.005 MHz and 40.002 MHz) and four whip antennas. The satellite ended up reentering the atmosphere, in 1958. Sputnik II, with the dog Laika, was launched on November 3, 1957; on December 6, 1957, the United States failed to launch Vanguard TV3; on January 31, 1958, the US successfully launched Explorer I; NASA was created on October 1958. With these events, the US-USSR Space Race was started.

The invention of the artificial satellite, an optical stellar triangulation method, enabled the interrelation between continents to be accurately determined. One of the first worldwide satellite triangulation programs dates back to 1946 and was named BC-4. This system used a metric
camera alongside a chopping shutter and photographed reflective satellites against a star background. In the end, an image consisting of a series of dots – representing both a star’s and a satellite’s path – was created. The coordinates of specific dots were precisely measured using a photogrammetric comparator, and the associated spatial directions from the observing site to the satellite were then processed using an analytical photogrammetric model. If the same satellite was photographed, at the same time, from another location, a different set of spatial directions resulted. A pair of corresponding directions formed a plane that contained the observed points and the satellite. Additionally, the intersection of at least two planes gave us the spatial directions between the observing sites. These oriented directions were then used to create a global network, with the scale being derived from several terrestrial traverses. However, this technique had two main issues: it only worked with a clear sky and the equipment was expensive.

As time went by, the optical techniques were replaced by the electromagnetic ranging techniques. HIRAN, the first electronic ranging system, was utilized during World War II to position aircrafts. A technical breakthrough occurred when it was understood that the Doppler shift in the signal broadcasted by a satellite could be used as an observation to determine the exact time of closest approach of the satellite. By combining this with the capacity to predict satellite orbits according to Kepler’s laws, it became possible to determine any position in the world in an instantaneous and precise manner.

Navigation is defined as the science of getting from one place to another. At a basic level utilizing our eyes, common sense, and landmarks are necessary to navigate. When a more accurate knowledge of our position is needed, more aid is also required such as a clock to determine the velocity over a known distance or an odometer to keep track of the distance traveled. Around the 1920's, a more advanced technique called radionavigation emerged. Radionavigation transmits electronic signals, allowing navigators to locate the direction of shore-based transmitters when in range, and their own position to be known in turn. Afterwards, the users of the transmission process the radio signals and compute a fixed position. The receiver performs computations for the users to navigate to a specific location. There are two main types of radionavigation aids: ground-based and space-based. The accuracy of ground-based radionavigation techniques is proportional to their operating frequency. Highly accurate systems generally transmit at relatively short wavelengths and the user has to be within LOS, whereas systems with longer wavelengths are not limited to LOS but are less accurate.

The predecessors of today’s positioning system are the Navy Navigation Satellite System (NNSS), also known as TRANSIT system, and Timation, both having become operational in 1964. TRANSIT, composed of six satellites orbiting with circular polar orbits, provided a two-dimensional high-accuracy positioning service. This system was developed by the U.S. military with the objective of determining the coordinates of ships and airplanes. As time went by, the system was authorized for civilian use and became widely used both for navigation and surveying. Some experiments showed that accuracies of about one meter could be obtained by occupying a point
for several days and reducing the observations. Hence, the frequency of obtaining a position fix was dependent on the user’s latitude.

A TRANSIT user location in the equator could obtain a position fix once every 110 minutes, whereas at 80 degrees latitude the fix rate would improve to once every 30 minutes. The limitations are that each position fix requires 10 to 15 minutes of receiver processing and an estimate of the user’s position. These features were suitable for shipboard navigation because of the low speed, but for high speed or dynamic users the satellites were not able to provide continuous positioning.

Timation was a space-based navigation system program. By using experimental satellites, high stability clocks, time transfer, and two-dimensional navigation was improved, allowing the comprehension that atomic clocks had better frequency stability than quartz clocks. This knowledge significantly increased the accuracy of predictions of satellite orbits.

The modern positioning system, Global Positioning System (GPS), was invented in the early 1970’s. Considering the major drawbacks of the previous systems, several U.S. Government organizations, including the Department of Defense (DOD), were interested in developing satellite systems for three-dimensional position determination. The goal was to provide a continuous global positioning and timing capability, with global coverage under any weather conditions, an ability to serve high dynamic platforms with high accuracy, and a strategy to orbit a sufficient number of satellites to ensure that, at least, four were always electronically visible. This type of positioning system also needed to give users information about their velocity through a proper receiver. In 1969, the Defense Navigation Satellite System (DNSS) program was established combining the different research and development achievements so that a sole joint-use system was created. The NAVSTAR GPS program, directed by the GPS Joint Program Office (JPO), appeared initially due to military motivations however, in 1996, the change of paradigm resulted in sharing with civilians the Global Navigation Satellite System (GNSS). This program commanded the production of new satellites, and all the equipment associated with the system, including the U.S. military user’s receivers.

The ongoing development of the different GPS segments has been incremental over the years. This growth began with a concept validation phase and evolved into several production phases – a block of satellites is associated with each phase of development. The initial concept began with a series of 11 satellites which were called Block I. These were launched by the JPO, between 1978 and 1985, and had an experimental purpose; for instance, the inclination angle with respect to the equator was 63° at the time, characteristic modified in the following satellite generations.

Block II satellites were the first production satellites, while Block IIA were the upgraded production satellites, with an increase in the navigation message data storage. This way the satellites had the capacity of operating without ground support for longer periods. This series was composed of a total of 28 satellites, launched between 1989 and 1997, and the inclination angle was corrected to 55° with respect to the equator. For military use these satellites had some extra
security characteristics, in particular, they had two cryptographic features denoted as antispoofing (AS) and selective availability (SA). AS purpose was to stop deception jamming, which consisted on replicating ranging codes, navigation data, and carrier frequencies with the intention of deceiving the receiver. SA objective was to downgrade other users' accuracy by interfering with their clock and consequently corrupting measurements.

Meanwhile, the following generation of satellites was named Block IIR, known as the replenishment satellites, and was compatible with the previous generation, resulting in no differences for the conventional user. Composed of 21 satellites, launched between 1997 and 2005, this set of satellites introduced an autonomous navigation capability, mutual satellite ranging capabilities, and prediction of satellite orbits and clock data – information uploaded from the ground station.

As time went by, the Block IIF, composed of a total of 12 satellites, was created and launched between 2010 and 2016. These were the first satellites not to have selective availability hardware installed. Among others, new characteristics include better accuracy, resistance to jamming, and a reprogrammable processor that can receive software uploads.

Block IIIA and block IIIF are the third generation GPS satellites, with new signals and higher broadcasting power levels. This generation was created to reassess the entire GPS architecture, in order to achieve the appropriate system for the future. The two main goals of this program were to reduce government ownership costs and to provide an architectural flexibility to satisfy the requirements stipulated until 2030 [25].

2.2. GNSS Systems

The GNSS system is a system used to determine the position of an object. At its essence, the GNSS is formed by a satellite constellation, with global coverage, that sends signals to the receivers located on the surface of Earth, or near it, allowing the receivers location in space (longitude, latitude, and height) and time to be pinpointed. GPS is the most popular GNSS system in the world [26-27].

For this optimum positioning system to work it is necessary a particular scheme, known as the Initial Operational Capability (IOC), composed of a constellation of 24 satellites, placed in 6 practically circular 12-hour orbits with 4 satellites per plane, and with an inclination of 55° to the equatorial plane. Such conditions assure total coverage (often with more than the minimum number of satellites available with the minimum cost associated). As the time that takes a satellite to complete a full orbit is approximately half a day, a fixed observer on the ground will see a given satellite, on that orbit, at the same place, twice a day.

The GPS is divided in three main components: the user segment, the control segment, and the space segment. The space segment was just described in the previous paragraph. The system uses only two different carrier frequencies, one for the navigation data and the other for the
satellite broadcast ranging codes – hence, it is used a Code Division Multiple Access (CDMA), which allows the usage of different ranging codes by different satellites, for the same carrier. The navigation data allows the receiver to determine the location of the satellite at the time of signal transmission. The ranging codes grant the user’s receiver the means to obtain the propagation time of the signal and, through that, to determine the distance between the user’s receiver and the satellite. For the system to work properly and to measure the receiver’s location it is of great importance that each satellite possesses a highly accurate atomic clock onboard, so that the satellites can transmit synchronous timing signals. Additionally, an user’s receiver requires a crystal clock (to minimize costs, complexity and size). In essence, four measurements are needed to be known: latitude, longitude, height, and receiver’s clock offset. If the receiver’s clock was synchronized with the satellite clocks, 3 measurements would be sufficient.

The control segment is a ground network of tracking stations for control of the satellites’ status, the segment’s main purposes are, among others, to detect and predict satellites’ locations, system integrity, performance of the atomic watches, and weather conditions. Moreover, receiver’s navigation data is also uploaded into the satellite, through a master control station located at Colorado Springs where both satellites’ orbits and clock performance are computed. One of the biggest advantages of this system is the fact that the provided service can have unlimited users as long as they operate in the system as a receiver (passively), this is the reason why GPS is a one-way-ranging system.

When the signal of the GPS reaches the ground, its intensity is only about \(3 \times 10^{-14} \text{W m}^{-2}\); to measure such weak radio signals, GPS receivers must implement very specific techniques. The user segment provides separate services for civil and military users. The Standard Positioning Service (SPS) is available to all users worldwide, without any restrictions, in contrast to the Precise Positioning System (PPS), utilized by U.S. authorized military and selected government agency users. The access to PPS is controlled through cryptography.

2.3. GPS Basic Idea

As it has been stated before, each GPS satellite continuously transmits a microwave radio signal, which is divided in two carriers, two codes and a navigation message. The GPS satellites follow their respective orbits with high accuracy. When there is a receiver, it receives the satellites signals through its antenna. Once the signals are collected, the receiver processes them, using its software. The result of the signal processing are the distances to the satellites, derived by knowledge of the exact time at which the signal – composed of digital codes and a navigation message that contains the satellites coordinates – is sent.

To determine a tridimensional position of any given point on Earth, four satellites are necessary. One satellite allows us to calculate the position of an object placed on a sphere. The second satellite reduces the uncertainty and the position is then defined by the intersection of two
spheres, resulting in an intersection circle. The third satellite further reduces this circle into two single points. As one of these two points is an impossible answer located far away from Earth or moving at an impossible speed, by elimination, we can obtain a single precise location. The last satellite is there to synchronize satellite clocks and receivers. In the end, the GPS can locate, accurately, any given point. This calculation method is called trilateration.

Let us consider an instant \( t \) in which the receiver receives signals from 3 satellites. The signal from the first satellite indicates a set of coordinates \((x_1, y_1, z_1)\), that define its precise location, and the instant \( t_1 \) that defines the time at which the signal was sent. If \( c \) is the speed of light, the radio wave signal has travelled a distance given by \( c(t - t_1) \). As a result, the receiver knows its position \((x, y, z)\) located on the surface of a sphere with a radius of \( c(t - t_1) \) and center \((x_1, y_1, z_1)\). In other words we have the following equation

\[
\sqrt{(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2} = c(t - t_1).
\]

(2.1)

Accordingly, if the second satellite is in position \((x_2, y_2, z_2)\) and is sending its signal at the instant \( t_2 \), the receiver concludes that the satellite is somewhere on the surface of a sphere that corresponds to the equation

\[
(x - x_2)^2 + (y - y_2)^2 + (z - z_2)^2 = c^2(t - t_2)^2.
\]

(2.2)
As these two surfaces intersect in a circumference, by subtracting the previous equations, we get

\[ 2(x_1 - x_2)x + 2(y_1 - y_2)y + 2(z_1 - z_2)z = I_{12}, \]  

(2.3)

where,

\[ I_{12} = x_1^2 + y_1^2 + z_1^2 - c^2(t - t_1)^2 - x_2^2 - y_2^2 - z_2^2 + c^2(t - t_2)^2. \]

This equation is the equation of a plane that intersects the two spherical surfaces in a common circumference. Applying the same procedure to the third signal, we conclude that the three spherical surfaces intersect in two points: the intersection of the two planes results in a straight line that intersects the first spherical surfaces in two distinct points. Mathematically, the receiver solves the following system of three quadratic equations

\[
\begin{align*}
(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2 & = c^2(t - t_1)^2, \\
(x - x_2)^2 + (y - y_2)^2 + (z - z_2)^2 & = c^2(t - t_2)^2, \\
(x - x_3)^2 + (y - y_3)^2 + (z - z_3)^2 & = c^2(t - t_3)^2,
\end{align*}
\]

(2.4)

or the equivalent system with one quadratic equation and two linear equations

\[
\begin{align*}
2(x_1 - x_2)x + 2(y_1 - y_2)y + 2(z_1 - z_2)z & = I_{12}, \\
2(x_1 - x_3)x + 2(y_1 - y_3)y + 2(z_1 - z_3)z & = I_{13}.
\end{align*}
\]

(2.5)

As these equations are nonlinear and the solution is nontrivial, usually the receiver performs the linearization of the equations assuming an approximate position – either the center of the Earth or the last known position – obtaining, by iteration, the expected result as previously explained.

In the real world, the idea of how the receiver of the GPS works, is less straight forward. It was assumed that the clock at the receiver end was sufficiently precise, which in fact does not correspond to reality taking into consideration that the only precise clocks in this system are the satellite atomic clocks. In fact, there is a need to measure precisely the time that the radio wave signals take to travel from the satellites to the receiver. More specifically, these signals travel at approximately \(3 \times 10^8 \text{ km/s} \), 30 cm at each nanosecond. Hence, to have a precision of 5 meters, for instance, the measurement of the time interval must have a precision of 15 nanoseconds. Consequently, as mentioned before, we need a fourth satellite that handles the clocks’ offset. The end result is a system of four equations in order to the variants \((x, y, z, t)\)

\[
\begin{align*}
(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2 & = c^2(t - t_1)^2, \\
(x - x_2)^2 + (y - y_2)^2 + (z - z_2)^2 & = c^2(t - t_2)^2, \\
(x - x_3)^2 + (y - y_3)^2 + (z - z_3)^2 & = c^2(t - t_3)^2, \\
(x - x_4)^2 + (y - y_4)^2 + (z - z_4)^2 & = c^2(t - t_4)^2.
\end{align*}
\]

(2.6)
2.4. Satellite Communication

The navigation data and signal timing information sent by the satellite are continuously broadcasted through two different frequencies:

- L1: 1575.42 MHz
- L2: 1227.60 MHz

L1 is modulated with the civilian C/A code (1 Mbps) and military P(Y) code (10 Mbps). While L2 is modulated only with military P(Y) code. Military P(Y) code allows a more precise position calculation and can be encrypted. Both codes are based on pseudorandom noise sequence (PRN), utilized to measure multiple signal propagation delay simultaneously. Physically the signal is a digital code, a sequence of pulses.

Each code carries the navigation message with a 50 bps data rate, on a 1500 bits data frame. Every frame is composed by 6 subframes that contains the following data:

- Time of transmission
- Satellite ephemeris data
- Satellite clock offset information
- Ephemeris data from nearby satellites
- Signal propagation information
- System health status

The GPS receiver compares the time a signal was transmitted by a satellite to the time it was received. The time difference tells the GPS receiver how far away the satellite is.
Chapter 3 – Special Relativity: Fundamental Aspects

3.1. Introduction

Physics is concerned with the motion of objects: how planets and stars move, how electrons and protons move, and how the movement of the molecules results in emerging properties such as temperature. The role of relativity in physics is to study how that motion looks from different perspectives, while special relativity observes how motion looks from a limited or special set of perspectives.

For an observer, for example, looking at the Earth and the moon, it might look like the moon is moving around the Earth in a circle, back and forth in a straight line, or even that the Earth and moon together are tracing out a spiraling path through space, depending on where the observer is located and how he is moving. However, if the motion of the objects can be described in so many different ways, what does any of these different descriptions really tell us about the objects that they describe?

Special relativity poses questions such as how perspective affects the perception of reality, if there are some perspectives better than others, or if there is a preferred perspective to observe objects that closes in on the true description of reality.

In essence, relativity can be summed up through two basic objectives: on the one hand one has to figure out how objects and their motion look from different perspectives and, on the other hand, what properties of these objects remain invariant among different perspectives.
The first idea has been previously explained by the different ways the motion of Earth and moon can look from different perspectives. Representing the second idea is not as simple, in the Earth and moon case all different perspectives appear quite different but, regardless of the perspective, the maximum physical distance between the Earth and the moon appears to be the same hence, it is possible to affirm that there is something independent from the perspective (in this case the orbital radius). Therefore, relativity is so important in physics given it studies what changes or not in a physical system when the perspective changes. In the end, this scientific field leads to universal truths, facts that remain true from any perspective throughout the universe.

### 3.2. IFR – Inertial Frame of Reference

The solution presented by Einstein in the special relativity theory, was one of the most revolutionary ideas, changing the way we look at the traditional concepts of space and time. This solution was only possible due to Lorenz transformation which, more than the mathematical explanation of the theory, is a physics theory based on the revision of the concept of simultaneity. In order to present the conclusions, we first need to understand what an Inertial Frame of Reference (IFR) is. The before mentioned reference frame describes time and space homogeneously, isotropically, and in a time-independent manner, being also known as Galilean reference frame. According to Newton’s Force, there is a relative movement, uniform and in a straight line between two different frames of reference, considering one of them is fixed. In Newton’s mechanics, this reference frame respects the Newton’s first law in which all inertial frames are in a constant state (they are at rest or they move at a constant velocity along a straight line). In this frame of reference, bodies, with net force acting on them equal to zero, are not accelerated \[ a = 0 \] \[ F = m a \] (3.1)

Without acceleration \( a \), there will be no force \( F \) over a given particle with inertial mass \( m \).

The movement of this particle will be described by the following Newton equation

\[ a = 0 \Rightarrow v = v_i \Rightarrow r(t) = r_i + v_i t. \] (3.2)

And also

\[ a = \frac{dv}{dt} = \frac{\partial^2 r}{\partial t^2}, \] (3.3)

Where \( v \) is the velocity vector and \( r \) is the position vector. In a coordinate system \( r(t) = r_i + v_i t \) is the parametric equation of the straight line.

As shown, there is no global inertial system. The existence of a gravitational force implies the existence of acceleration which means that the gravitational force cannot be described by the
special relativity theory and that the frame of reference of a material particle with acceleration is not an IFR. Nevertheless, we can consider, approximately, a simulation of an IFR if locally speaking.

In the general relativity theory, the distinction between IFRs and Accelerated Frames of Reference (AFRs) ceases to exist as both frames of reference are described equally.

### 3.3. Einstein’s Special Relativity Postulates

There are two fundamental postulates in the special relativity theory:

**First postulate (principle of relativity):** The laws of physics are the same and applicable to all inertial frames of reference (observers).

**Second postulate (constancy of the speed of light):** Electromagnetic waves travel, in vacuum and regardless of the motion of their source, with constant speed $c$.

The conclusion of the second and most important postulate is that the speed of light is a cosmic limit, a finite value, independent of any observer and the speed of a given emitting source [30]. Additionally, any velocity that remains invariant in the different reference systems has to be equal to the speed of light [31]. This postulate shows that the space and time concepts are relative. Although velocity $c$ is constant to all observers, a velocity is obtained by the quotient of space and time, concepts which vary according to the observer and thus can have different values.

### 3.4. Galilean Transformation

Prior to the existence of relativistic physics, there was a fundamental incompatibility between the two postulates [32]. The Newton’s mechanics imply, wrongly, that there is no superior limit to the speed of a particle. Analogously, consider a train is moving away from a station with velocity $v$; inside the train an electromagnetic signal is emitted with the same sense as the train’s movement and another signal is emitted in the opposite sense, both with velocity $c$. For the first signal Newton’s mechanics declare that the velocity of the signal in relation to the station is $u = v + c > c$. For the second signal the velocity is $u = v - c < 0$, where $|u| = c - v$. The contradiction is visible, given the second postulate states that $u = c$. In the end, we have to understand that the addition of velocities, from the Galilean transformation, is in direct contradiction to the second postulate of the special relativity theory, becoming not applicable in the situation considered. As depicted in figure 3.1., Galilean addition of velocities rule is only applicable for small velocities, which are incomparable to the speed of light. The formula that calculates this rule is given by the algebraic sum of velocities.
There is a wrong belief, consequent of Newton’s mechanics, that time is absolute \( t = t' \). Moreover, Galilean addition of velocities cannot be verified in every case. For instance, consider two IRF, one being represented by the frame of reference \( S \rightarrow (x, y, z) \), and the other being represented by the frame of reference \( S' \rightarrow (x', y', z') \). Suppose one frame of reference is moving away from the other with velocity \( v = \beta c \) and that this relative movement only occurs at the \( x \) axis, as described in figure 3.2:

The origins of the two frames of reference are the same when \( t = t' = 0 \). Hence, \( x' = x - vt \). The movement of the frame of reference \( S' \), from the point of view of the frame of reference \( S \), is given by the equation \( x = vt \), since \( x' = 0 \). Thus, the trajectory of a particle from the
point of view of frame of reference \( S' \) is \( x' = u t' \), which corresponds, from the point of view of \( S \), to \( x = x' - v t' = (u + v) t' = w t \). If one accepts the absolute time (\( t = t' \)) assumption, the addition of velocities \( w = u + v \), from Galileu transformation, has to be accepted as well. However, a relativistic approach refutes this demonstration by stating that time is not absolute and that simultaneity does not have a universal meaning, independent from the frame of reference. As a result, we conclude that \( t \neq t' \).

In Galilean relativity \( c \) is just a constant. Thereby, the speed of light depends on the observer measuring it.

Figure 3.3. represents Galileu transformation graphically. Where, in frame of reference \( S \), the event \( A \) has coordinates \( (x_A, t_A) \) and, in frame of reference \( S' \), the same event has coordinates \( (x'_A, t'_A) \). According to Galileu transformation \( t'_A = t_A \) – time is considered absolute hence, \( x \) and \( x' \) axes represent time in the exact same way – and \( x'_A = x_A - v t_A \).

The \( t' \) axis characterized by the equation \( x' = 0 \) (in \( S' \)) and the equation \( x = v t \) (in \( S \)) intersects the \( t \) axis in the origin of the two coordinates systems.

Now that we have seen what is wrong with the Galileu transformation – the addition of velocities goes against the second postulate of the special relativity theory –, and assuming this postulate is based on unequivocally solid experiments, we have to question the addition of velocities.

### 3.5. Minkowski Diagrams
To better understand and visualize the questions related to the special relativity theory, we need to resort to the spacetime diagrams, also known in the literature as Minkowski diagrams [33]. With these diagrams’ conditions we reduce the four-dimensional continuous time-space to a bidimensional one, where only one spatial dimension exists as shown in figure 3.4.

\[ S(x, y, z, t) \mapsto S(x, t) \, . \]

Figure 3.4.: Spacetime Minkowski diagram.

In figure 3.4., the points \( A(x_a, t_a) \), \( B(x_b, t_b) \), and \( C(x_c, t_c) \) are called events (points of the spacetime diagram) [34], while the lines \( u_0, u_1, u_2 \), are designated as world lines. The line \( u_0 \) corresponds to the equation \( x = x_0 \), representing a particle at rest, with the position \( x_0 \). The line \( u_1 \) corresponds to the equation \( x - x_1 = vt \), representing an animated particle with an uniform movement and velocity \( v \), on the positive \( x \) axis sense. Both these lines represent an IFR observer, as they have a constant slope. Finally, the line \( u_2 \) represents an animated particle with an accelerated movement, since it is subject to a force that varies through time and the respective world line is curve, the slope of the line varies from point to point.

This geometrical representation includes not only the space but also the time in the same diagram [35]. Assuming that the time and the spatial axes do not have the same SI units, and imposing geometric (or natural) units, \( c = 1 \). This type of metric is chosen so that the equation \( x = ct \mapsto x = t \) may represent the electromagnetic signal by an \( 45^\circ \) angle, with both the \( x \) and the \( t \) axes. All the straight lines parallel or perpendicular to this signal are electromagnetic signals. Nonetheless, the latter propagate on the opposite sense of the former. This description of the electromagnetic signals is implicit in the second postulate which explains the concept of relative simultaneity, a new concept that appears alongside relativistic physics. Figure 3.5. is a representation of the electromagnetic signals.
There are two important concepts that we need to clarify [36]. Firstly, the temporal axis, \( t \), corresponds to the equation \( x = 0 \) – all the events that belong to this line occur in the same position \( x = 0 \). For this reason, the temporal axis is called an equiloc or, in special relativity, we can also call it a clock, because time can be measured through this line. We can define an equiloc as a set of events that happens at the same place (position), from the point of view of a given system of coordinates. Secondly, the spatial axis, \( x \), corresponds to the equation \( t = 0 \) – all the events on this line occur at the same time instant \( t = 0 \), they are simultaneous events. For this reason, the spatial axis is called an equitemp or an axis that is orthogonal to a clock. We can define an equitemp as a set of events that occurs at the same time (simultaneous) from the point of view of a given IFR. Finally, we can define an event as the intersection of an equitemp with an equiloc. As it was mentioned for the electromagnetic signals, when referring to a particular frame of reference (observer), all equitemps are parallel between each other, and all equilocs are also parallel between themselves. This does not necessarily mean that equilocs and equitemps are perpendicular, although they must be orthogonal. Moreover, equitemps or equilocs from different frames of reference are not parallel, due to time being a relative concept dependent on the observer, as a consequence of the second postulate.

An observer is identified by a given coordinates system, from a given IFR. The previous considerations are referent to an affine space – a space concept even more general than the vector space, a geometric structure that generalizes certain properties of parallel lines in Euclidean space. This affine space metric defines the concept of parallelism however, it does not define the concept of perpendicularity given the concept of orthogonality requires the existence of a different metric associated with the vector space, the Lorentzian metric (not yet defined).
Once again resorting to figure 3.4., we can now state that events \( B \) and \( C \) are simultaneous and events \( A \) and \( B \) occur in the same spatial position. In figure 3.6., we illustrate an example of a Minkowski diagram.

![Minkowski diagram](image)

Figure 3.6.: Event interpretation from different observers in Minkowski diagrams [37].

### 3.6. Some Definitions Derived from Einstein’s Second Postulate

The second postulate contains all the potential revolutionary ideas inherent to the special relativity theory. Hence, it is necessary to derive all the physics implications of this principle.

According to Maxwell’s equations

\[
c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}.
\]  

(3.4)
The first possible interpretation is that the light is an electromagnetic phenomenon. Such knowledge enables us to unify three of the four interactions that exist. The electricity (through $\varepsilon_0$), the magnetism (through $\mu_0$), and the optics (through $c$).

According to the first postulate of the special relativity theory, either Maxwell’s equations are valid in all IFR or the equations cannot be right. Maxwell’s equations are right – at the macroscopic level –, this is one of the most solid truths established in physics.

The definition of the meter is the distance traveled by light in vacuum, during a certain time interval $\Delta t = \frac{1}{299792458} \text{[s]}$. (3.5)

Because of that, we can affirm that the speed of light is a universal constant and that it has a constant value, in SI system, of $c = 299792458 \text{[m/s]}$. (3.6)

Besides this, we can also define the exact values of the magnetic permeability and electrical permittivity as being

$$\mu_0 = 4\pi \times 10^{-7} \text{[H/m]},$$

$$\varepsilon_0 = \frac{1}{c^2 \mu_0}.$$ (3.7)

Given the importance of the value of the speed of light in special relativity, and utilizing a different convention from the one presented above, it is possible to define that, through geometric (or natural) units [38]

$$c = 1.$$ (3.9)

By doing so, there are two possibilities to measure time and space. The first hypothesis is that time is measured in seconds in which case space is measured in light seconds. The second hypothesis is that space is measured in meters and time in light meters.

One light second is the distance travelled by light during a time interval of one second. One light meter is the time that the light takes to travel one meter. Therefore,

$$1 \text{ light second} = 299792458 \text{ m},$$

$$1 \text{ light meter} \approx 3.335640951981520 \text{ ns},$$ (3.10)

with $c = 1$, $\beta$ becomes a normalized variable ($-1 \leq \beta \leq 1$) and dimensionless speed will be equal to
According to these geometric units, when \( v = c = 1 \), an electromagnetic signal always has \( \pm 45^\circ \) angle hence, \( \tan(\theta) = \beta \) and \( \theta = \pm 45^\circ \). Additionally, all electromagnetic signals have to be parallel to each other, which is also stated by the second postulate. As a consequence, the universal concepts of simultaneity and time collapse. We can go even further and state that the equitemps from a given IFR cannot be parallel to the equitemps from a different IFR. Only when the equitemps and the equilocs are parallel to each other can we claim that time is absolute, and that the Galileu transformation is right and applicable. Nonetheless, most of the cases require a new coordinate transformation, which is called Lorentz transformation, and spacetime diagrams called Minkowski diagrams [39].

There is one last thing that is important to refer: in classic electrodynamics, a material medium is a set of charges and currents that, macroscopically speaking, alters the constitutive relationship in vacuum. A simple material medium (homogeneous, linear, isotropic and time independent), can be described with a certain refraction index \( n \). Moreover, if there is no dispersion, this index is a constant, independent of the frequency (theoretically speaking). Thus, the speed of light will be described as

\[
v = \frac{c}{n}.
\]  

The vacuum or any invariant medium, in a Lorentz transformation, is the one that has \( n = 1 \) and, consequently, \( v = c \).

Despite this, the question of why light is so relevant to the special relativity might persist. The main characteristic that sets apart light from other signals is that the photons – light’s elementary particles – have no mass [40]. The reason behind the cosmic limit of velocities being identified with the value \( c \) is not only the main characteristic previously mentioned but also the fact that this value is the maximum achievable by the speed propagation of a fundamental particle. In the end, the cosmic limit of velocities is the same as the speed of light.

**3.7. Geometry of Relative Simultaneity**

As it was stated before, the belief that time is absolute is incompatible with Einstein’s second postulate. With the next examples the concept of simultaneity, based on geometry simultaneity, shall be clarified and reviewed.

As a first example, there are two IRF, one represented by the frame of reference \( S \to (x,t) \), which describes Alice who is seated (at rest) at the train station, and the other represented by the frame of reference \( S' \to (x',t') \), which describes Bob who is at the middle point of a train
wagon, \( M(x_0', t_0') \), with length \( L \) and moving away from the station. Bob, from his point of view, emits simultaneously two electromagnetic signals with opposite senses; when one of the signals reaches the extreme left of the wagon, event \( A \) has occurred. Alternatively, when the other signal reaches the extreme right of the wagon, it is deemed event \( B \). From Bob’s point of view, events \( A \) and \( B \) are simultaneous because the distance between Bob’s position and the two extreme points of the wagon is the same: \( \Delta x' = L / 2 \). In this case, the velocity is the same at any given point of this frame of reference \( (c = 1) \). As a result, \( t_A' = t_B' \iff \Delta t' = t_B' - t_A' = 0 \).

In contrast, according to Alice’s point of view, the same events, \( A \) and \( B \), are not simultaneous given the train is moving from left to right. Hence, \( t_A < t_B \iff \Delta t = t_B - t_A > 0 \), as shown in figure 3.7. – event \( A \) occurs before the event \( B \).

![Minkowski diagram of Alice’s point of view for the wagon example.](image)

In essence, the two events, \( A \) and \( B \), are simultaneous for \( S' \), while occurring at different times for \( S \). In other words, the simultaneity of the events is a relative concept that depends on the IFR chosen. This implies that time is neither absolute, nor the same for different IRFs. Furthermore, the equitemps from different IRFs do not have to be parallel to each other – to deal coherently with time – as a consequence of time not being absolute.

### 3.8. Geometrical Construction of a Frame of Reference

The example below, similar to figure 3.7. but in which the axes of the IFR \( S \) are represented, is devised in order to determine the axes of the frame of reference \( S' \), which has velocity \( \beta \) in relation to \( S \). In figure 3.8., it is considered that Bob, inside the train, sends two electromagnetic signals simultaneously at instant \( t' = 0 \), with each one of the signals being sent at a different extreme of the wagon. These signals will meet at event \( M' \), located at the middle point of the train.
wagon.

Figure 3.8.: Example with two electromagnetic signals simultaneously from the wagon.

The events $A$ and $B$ of figure 3.8. represent the moment in which the electromagnetic signals are emitted from Bob's wagon. From his point of view ($S'$ frame of reference) these two events are simultaneous hence, the straight line that connects them is an equitemp for $S'$. Regarding the equiloc from this frame of reference, there is the necessity of discovering the coordinates from event $M$ through the intersection of the two electromagnetic signals: $t = x$ and $t = -x + b$, with the middle line of the wagon being $x = \beta t + L / 2$. As a result

$$x_M = \frac{L}{2(1 - \beta)},$$

$$t_M = \frac{L}{2(1 - \beta)}.$$

With these coordinates, getting the $b$ value from the equation $t = -x + b$ – for the electromagnetic signal – is now possible

$$t = -x + \frac{L}{1 - \beta}.$$  

(3.14)
After calculating event $B$ coordinates, through the intersection of $t = -x + L / (1 - \beta)$ with $x = \beta t + L$, and with the equitemp (from frame of reference $S'$) that unites events $A$ and $B$

\[ t - t_A = m(x - x_A). \]  

(3.15)

And, consequently

\[ x_B = \frac{L}{(1 - \beta)(1 + \beta)}, \]

\[ t_B = \frac{\beta L}{(1 - \beta)(1 + \beta)}. \]  

(3.16)

Finally, with these coordinates from events $A$ and $B$, it is now possible to write the equation that describes the equiloc $x'$ from $S$ point of view; the slope of $x'$ is

\[ m = \frac{t_B - t_A}{x_B - x_A} = \beta. \]  

(3.17)

With this result, it is possible to conclude that the equiloc is $x' = 0$. In other words, the $t'$ axis is described by the equation $t = \beta x$. Moreover, the $t'$ axis is parallel to the two extremes of the train wagon and is described by the equation $x' = 0$, which is the equiloc of IFR $S'$. IFR $S$, is described by equation $x = \beta t$. In the end, the way in which the axes of IFR $S'(x', t')$ are constructed is determined. Additionally, one can also derive that the equitemps of $S$ are not parallel to the equitemps of $S'$.

As the slopes of both these axes are reciprocal (the angle between $t$ and $t'$ is the same as the angle between $x$ and $x'$), the angle limited by them is denominated $\theta$. Considering an Euclidean point of view, the angle $\theta$ depends on the relative velocity $\beta$ in other words, it depends on the relative movement speed between the different frames of reference,

\[ \tan(\theta) = \beta. \]  

(3.18)

This angle is Euclidian and, therefore, it does not represent an angle in the Minkowski plane. As we will see afterwards, relativistic physics implies a non-Euclidian geometry. In fact, Minkowski diagrams introduce a new type of geometry where the Euclidean metric (defined through definite positive quadratic forms) is replaced by a Lorentzian metric (defined through indefinite quadratic forms).

In general, considering two observers, $A$ and $B$, each one with his own frame of reference, when the observer $A$ is moving away from observer $B$ with velocity $\beta$, the observer $B$ is also moving away from observer $A$ with velocity $-\beta$. These aspects are summarized in the following relations:
- $x$ axis ($S$ equitemp) $\rightarrow t = 0 \mapsto t' = -\beta x$.
- $t$ axis ($S$ equiloc) $\rightarrow x = 0 \mapsto x' = -\beta t'$.
- $x'$ axis ($S'$ equitemp) $\rightarrow t' = 0 \mapsto t = \beta x$.
- $t'$ axis ($S'$ equiloc) $\rightarrow x' = 0 \mapsto x = \beta t$.

In the end, the belief that time is absolute, brought by Newton’s mechanics, is disproved. Time depends on the frame of reference considered hence, it is, indeed, relative. Two observers moving away from each other have equitemps that are not parallel and only when equitemps from different frames of reference are parallel is time absolute. However, different observers moving away from each other do not have parallel equitemps and in that case the conclusion is that time is relative [41].
4.1. Bondi’s Factor and Time Dilation

Another important way of deducing the representation of the equitemp and the equiloc from a given IFR forces us to introduce the Bondi’s factor $\kappa$.

Let us consider that to Alice corresponds the equiloc $\xi_0$ characterized by the straight line $x = 0$.

In the following example, the purpose is to get the coordinates $(x_A, \tau_A)$ from event $A$, according to Alice’s point of view. Event $A$ represents the time instance when Bob receives Alice’s signal, with Bob being inside a train moving away with relative normalized velocity $\beta$. In order to obtain these coordinates, we use Bondi’s radar method [42].

Assuming that Alice sends an electromagnetic signal at time instant $t^A$, with the direction of event $A$, by the time Bob receives this signal his watch marks time instant $\tau_A$. Once this signal is received by Bob, he sends back another electromagnetic signal which is perpendicular to the one sent by Alice because they have opposite senses. When Alice receives this signal, the time instant is $t^A$. Moreover, let us also consider an event $B$ that represents an event simultaneous with event $A$, from Alice’s point of view, as we can see in figure 4.1.
In this example, we assume that the light speed is \( c = 1 \), in geometric units. Event \( B \) must occur with equal distance to \( t_- \) and to \( t_+ \), this is a consequence of light spending the same time from \( t_- \) to event \( A \) and from event \( A \) to \( t_+ \). In other words, event \( A \) is at a fixed distance from Alice. If we assume that event \( B \) occurs at time instant \( t_A \), we can define \( t_A \) as

\[
t_A = \frac{1}{2}(t_+ + t_-). \tag{4.1}
\]

Additionally, as event \( A \) is at a fixed distance from observer \( \mathcal{O}_A \), then

\[
x_A = \frac{1}{2}(t_+ - t_-). \tag{4.2}
\]

Figure 4.1. shows how it is possible to determine the equitemp \( \mathcal{O}_{\perp} \) of Alice – represented by the straight line that unites events \( A \) and \( B \).

By algebraically summing and subtracting the two previous equations, we can calculate \( t_+ \) and \( t_- \)

\[
\begin{align*}
t_+ &= t_A + x_A, \\
t_- &= t_A - x_A.
\end{align*} \tag{4.3}
\]

When adding a new observer (Bob) \( \mathcal{O}' \rightarrow (x', t') \), defined by the equiloc \( \mathcal{O}' \) to the last figure, as we can see in the figure 4.2. below, we have a new equiloc containing event \( A \).
As previously mentioned, Bob and Alice are moving away with relative velocity $\beta$ hence, the origin event $O(0,0)$ is the only event that belongs to both observers. Moreover, $x_A$ is linearly increasing with value $\beta$

$$x_A = \beta t_A \Leftrightarrow \beta = \frac{x_A}{t_A} = \frac{t_+ - t_-}{t_+ + t_-}. \tag{4.4}$$

Taking into account that simultaneity is a relative concept, time is not absolute and event $A$ has coordinates $A(x_A, t_A)$ from Alice’s point of view, and coordinates $A(x'_A, t'_A) = A(0, \tau_A)$ from Bob’s point of view, where $\tau_A$ represents $A$’s own time. With the coordinates of event $A$ well defined for each observer, we shall introduce Bondi’s factor $\kappa$. According to Bondi, we know that $\tau_A$, the time in which Bob receives the signal, is equal to a constant $\kappa$ times $t_-$

$$\tau_A = \kappa t_- \tag{4.5}$$

There is reciprocity: if Alice sends a signal and Bob receives it, the same behavior should occur when Bob sends a signal at instant $\tau_A$ and Alice receives it at instant $t_+$:

$$t_+ = \kappa \tau_A \tag{4.6}$$

According to Bondi’s factor $\kappa$, $t_+, t_- \in \mathbb{O}$ and $\tau_A \in \mathbb{O}$ hence, if we accept the previous equations

$$\tau_A^2 = \tau_A \tau_A = (\kappa t_-) \left( \frac{t_+}{\kappa} \right) = t_+ t_- \tag{4.7}$$

With the previous result we can deduce the Minkowski theorem
\[ \tau_A = \sqrt{t_+ t_-}. \] (4.8)

This means that \( \tau_A \) is the geometric mean between \( t_+ \) and \( t_- \), which is verifiable through Bondi’s notations – \( t_A \) is the arithmetic mean between \( t_+ \) and \( t_- \). Since the arithmetic mean is at least equal to the geometric mean, as proved through appendix E, then

\[ \tau_A \leq t_A. \] (4.9)

And, considering the result from equations 4.3. and 4.4.

\[
\begin{align*}
  t_+ &= t_A + x_A \\
  t_- &= t_A - x_A & \Leftrightarrow & \begin{cases} 
  t_+ = t_A + \beta t_A \\
  t_- = t_A - \beta t_A \\
  t_+ = t_A (1 + \beta) \\
  t_- = t_A (1 - \beta).
\end{cases}
\end{align*}
\] (4.10)

In the end,

\[ \tau_A = \sqrt{t_+ t_-} \Leftrightarrow \tau_A = t_A \sqrt{1 - \beta^2}. \] (4.11)

with the term \( \sqrt{1 - \beta^2} \) being the slowing down factor – because it is smaller or equal to one –, according to the special relativity theory

\[ \tau_A = t_A s, \text{ with } s \leq 1. \] (4.12)

This effect can be ignored in daily situations when the velocity \( v \) is far lower than \( c \), case in which \( t_A \) and \( \tau_A \) are equal (\( \beta = 0 \)). On the contrary, in situations that involve velocities comparable to the speed of light, for instance the particle accelerators or the precision needed to measure the time intervals in the GPS, this time dilation has to be taken into account.

When, from Alice’s point of view, event A has time \( t_A \), event B has a smaller time as a consequence of the slowing down effect. Alice conclusion is that as Bob is moving away from her, his time is slowing down. The more general approach is that moving clocks slowdown, which proves the existence of time dilation. Since there is reciprocity, the corresponding conclusion can be taken from Bob’s point of view. In other words, given Bob is inside the moving train, for him, Alice is the one moving away from his IFR.

Finally, we can deduce Bondi’s factor \( \kappa \)

\[
\tau_A = \kappa t_\perp \Leftrightarrow \kappa = \frac{\tau_A}{t_\perp} = \frac{t_A s}{t_A - x_A} = \frac{t_A \sqrt{1 - \beta^2}}{t_A (1 + \beta)} = \frac{\sqrt{1 + \beta} \cdot \sqrt{1 - \beta}}{1 - \beta} = \frac{1 + \beta}{\sqrt{1 - \beta}}. \] (4.13)

In case there is no movement, \( \beta = 0 \), and \( \kappa = 1 \), while if \( \beta \rightarrow 1 \), \( \kappa \rightarrow \infty \).
4.2. Length Contraction

Another important effect that can be derived from Bondi’s factor $\kappa$ is the length contraction. In fact, it is not just time that is relative but also length (space). Mathematics allows the demonstration that, by approaching the speed of light, space and time change. This is because, although they both perceive the same relative velocity between them, different observers register different times and distances for a certain event. More specifically, the faster the velocity of an observer is, the smaller the objects are perceived and the shorter the distance travelled seems to be. This phenomenon is called length contraction.

In the following example, in which Alice represents a stationary IFR and Bob represents a moving IFR with constant speed $\beta$, from Alice’s point of view, an object travelling in Bob’s IFR has a length denominated $L_0$ – referred to as proper length – and a longer time interval denominated $\Delta t$. While, from Bob’s point of view and as a consequence of the time dilation effect, the interval of time will be different – $\Delta t_0$. This time interval is referred to as proper time interval which is the time interval as measured by the IFR where the two events (extreme points of the object) occur. The contracted length, from Bob’s point of view, is labelled $L$, which is given by the proper length multiplied by the shrinking factor $s = \sqrt{1 - \beta^2}$.

If something moves while an observer is measuring it, that measurement does not represent its length. As such, to correctly measure the length of the object from Alice’s point of view, the measure of the extreme points of the object has to be made at the same time – simultaneously, on the same equitemp according to Alice’s perspective. Hence, length contraction is given by

$$L = L_0 \sqrt{1 - \beta^2}$$

(4.14)

As a result, the length of the object is shorter in Bob’s IFR, from Alice’s perspective. In other words, the measured length of moving objects is shorter than when those objects are measured while not moving. The closer we are to light speed, the more the relative perception of length becomes distorted. Thus, the proper time and length may not always be measured by the same observer. It is important to keep in mind that time dilation and length contraction are not the same concept.

4.3. Lorentz Transformation and Spacetime Interval Invariance

As it was explained in section 3.4. Galileu transformation is in direct contradiction to the second postulate of the special relativity theory. More specifically, this transformation is only applicable for small velocities – unlike the speed of light, for instance.
In special relativity the coordinates in a given IFR can be transposed to another IFR through a transformation called Lorentz transformation. Lorentz transformation, in what concerns the interaction between algebra and geometry, represents the recognition of everything that apparently goes against our intuition – Newtonian and Euclidean – due to the representation of the new Lorentzian metric, which diverges from the Euclidean metric – our geometric intuition of distance.

It is important to recall that the Lorentz transformation, while mathematically correct, had an incorrect physical interpretation. Let us consider Alice, represented by the equiloc $O$, Bob represented by the equiloc $\mathcal{D}$, and an event $A$ that does not belong to any of these equilocs. For Alice, event $A$ has coordinates $A \mapsto (x, t)$ while for Bob it has coordinates $A \mapsto (x', t')$. Lorentz transformation represents the linear transformation that relates the coordinates of Alice to Bob's. Both Alice (at instant $t_-$) and Bob (at instant $t'_-$) send electromagnetic signals in direction to event $A$, with different frequencies, and as soon as both signals reach this event, they are reflected back to their sources reaching Alice (at instant $t_+$) and Bob (at instant $t'_+$), as illustrated in figure 4.3.

![Figure 4.3.: Minkowski diagram example for Lorentz boost and interval invariance.](image)

In order to determine Bob's coordinates of event $A \mapsto (x', t')$, we use the coordinates $A \mapsto (x, t)$ and Bondi's approach (derivation), through which we can define the relation between $t'_-$ and $t_-$ and the relation between $t_+$ and $t'_+$. Additionally, considering the Bondi's factor $\kappa$, we reach
\begin{align}
\begin{cases}
t'_- = \kappa t_-
\vspace{0.5em}
t'_+ = \kappa t'_+
\end{cases}
\quad (4.15)
\end{align}

Which also defines
\begin{align}
\begin{cases}
t_- = t - x
\vspace{0.5em}
t_+ = t + x
\end{cases}
\quad (4.16)
\end{align}

In the end, the transformation results in Bob’s coordinates, as intended
\begin{align}
\begin{cases}
t'_- - x' = \kappa(t - x)
\vspace{0.5em}
t'_+ + x' = \kappa(t' + x')
\end{cases}
\Rightarrow \begin{cases}
t'_+ + x' = \frac{1}{\kappa}(t + x)
\vspace{0.5em}
t'_- - x' = \kappa(t - x)
\end{cases}.
\quad (4.17)
\end{align}

One can draw two important conclusions from the result. The first conclusion is obtained by the ordered multiplication of both equations
\begin{align}
(t'_+ + x')(t'_- - x') = (t + x)(t - x) \Rightarrow (t')^2 - (x')^2 = t^2 - x^2.
\quad (4.18)
\end{align}

The final equation proves the spacetime interval invariance which, consequently, establishes the Lorentzian metric. The spacetime interval represents a Lorentz invariant. It is proved that not everything is relative in relativity theory [43]. In fact, time and space are relative, but a new absolute arises associated with the spacetime concept: the interval is not relative, it is always the same value independently of the IFR considered [44]. Based on this result it is possible to affirm that the geometry of Minkowski spacetime is not Euclidian and that the plane \((x, t)\) has a hyperbolic geometry. Thus, the concept of invariance of the distance is reviewed in special relativity theory. Whereas in an Euclidian plane the geometric place of all the points that are at a fixed distance from the same point (center) is given by a circumference, in an hyperbolic plane the geometric place of the events that are at a fixed spacetime interval of a certain event is given by an hyperbole.

The second conclusion is obtained by the ordered addition and subtraction of both equations
\begin{align}
\begin{cases}
t'_- - x' = \kappa(t - x)
\vspace{0.5em}
t'_+ + x' = \kappa(t' + x')
\end{cases}
\Rightarrow \begin{cases}
2t' = \left(\kappa + \frac{1}{\kappa}\right)t - \left(\kappa - \frac{1}{\kappa}\right)x
\vspace{0.5em}
2x' = \left(\kappa + \frac{1}{\kappa}\right)x - \left(\kappa - \frac{1}{\kappa}\right)t
\end{cases}.
\quad (4.19)
\end{align}

With algebraic manipulations
\[ \kappa + \frac{1}{\kappa} = \sqrt{\frac{1 + \beta}{1 - \beta}} + \sqrt{\frac{1 - \beta}{1 + \beta}} = \frac{1 + \beta}{\sqrt{1 - \beta^2}} + \frac{1 - \beta}{\sqrt{1 - \beta^2}} = \frac{2}{\kappa} = 2\gamma. \quad (4.20) \]

\[ \gamma = \frac{1}{s} = \frac{1}{\sqrt{1 - \beta^2}}. \quad (4.21) \]

\[ \kappa - \frac{1}{\kappa} = \frac{1 + \beta}{\sqrt{1 - \beta^2}} - \frac{1 - \beta}{\sqrt{1 - \beta^2}} = \frac{2\beta}{s} = 2\gamma\beta. \quad (4.22) \]

If we consider geometric units, the second postulate imposes that \(-1 \leq \beta \leq 1\) and, given that the slowing down factor \(s\) is always smaller or equal to one, \(\gamma \geq 1\). Figure 4.4 represents the \(\gamma\) factor in function of the normalized relative velocity \(\beta\).

Finally, it is possible to obtain the Lorentz boost

\[
\begin{align*}
2t' &= 2\gamma t - 2\gamma \beta x \\
2x' &= 2\gamma x - 2\gamma \beta t
\end{align*} \Rightarrow \begin{cases}
t' = \gamma(t - \beta x) \\
x' = \gamma(x - \beta t)
\end{cases}. \quad (4.23)
\]

The Lorentz boost corresponds to the transformation of coordinates from one IFR to another and it can also be written in matrixial form

\[
\begin{bmatrix} t' \\ x' \end{bmatrix} = \gamma \begin{bmatrix} 1 & -\beta \\ -\beta & 1 \end{bmatrix} \begin{bmatrix} t \\ x \end{bmatrix} \Leftrightarrow \begin{bmatrix} t \\ x \end{bmatrix} = \gamma \begin{bmatrix} 1 & \beta \\ -\beta & 1 \end{bmatrix} \begin{bmatrix} t' \\ x' \end{bmatrix}. \quad (4.24)
\]

The explanation for Lorentz transformation is not that the moving bodies are contracting, but instead that simultaneity is a relative concept. In other words, different observers have different ways of interpreting what simultaneity is given two observers with relative motion between them have equitemps that are not parallel. In fact, only when the equitemps from different frames of reference are parallel, is time absolute and can the Galilean transformation be applied. However, different observers moving away from each other do not have parallel equitemps – equitemps
meet at some point – hence, time is relative. The spacetime interval invariance, as represented through equation 4.18., leads to effects such as time dilation and space contraction.

### 4.4. The Euclidian Geometry vs Hyperbolic Geometry

As it was discussed before in section 3.8. Geometrical Construction of a Frame of Reference), from an Euclidian point of view, the angle designated by $\theta$ with $\tan(\theta) = \beta$ is determined exclusively by the relative speed between the two IFR. Additionally, as it is an Euclidian angle, it cannot truly represent an angle on the Minkowski plane. By replacing the $\theta$ angle by an hyperbolic angle designated by $\phi$ – Lorentz boost quickness – which, by association with Bondi’s factor $\kappa$, leads to

$$\kappa = e^\phi = \frac{1 + \beta}{\sqrt{1 - \beta^2}} \Rightarrow \phi = \ln(\kappa) = \frac{1}{2} \ln\left(\frac{1 + \beta}{1 - \beta}\right). \quad (4.25)$$

With the introduction of this new term it should be noted that

$$
geq \begin{align*}
\gamma &= \cosh(\phi) = \frac{1}{2} (\kappa + \kappa^{-1}) \\
\gamma/\beta &= \sinh(\phi) = \frac{1}{2} (\kappa - \kappa^{-1}) \\
\beta &= \tanh(\phi) = \frac{\kappa^2 - 1}{\kappa^2 + 1}.
\end{align*} \quad (4.26)
$$

The Euclidian plane $(x,y) \in \mathbb{R}^2$, defined by two spatial axes and with $x^2 + y^2 = r^2 \geq 0$, equation which represents the circumference of radius $r$, has to be replaced by the hyperbolic plane, $(x,t) \in \mathbb{R}^{1,1}$, defined by a spatial and a temporal axes and with $x^2 - t^2 \in \mathbb{R}^{1,1}$. Let us consider two different coordinates systems, $S \mapsto (x,t)$ and $S' \mapsto (x',t')$, and an event $A$ that can be represented by the set of coordinates $(x_A, t_A) \in S$ or by the set of coordinates $(x'_A, t'_A) \in S'$. The set of events $(x,t) \in S$ are in the same spacetime interval, in relation to event origin (here and now) $O(0,0)$. They must belong to one of the three different possibilities, as shown by the spacetime interval invariance:

- Hyperbole time equation: $t^2 - x^2 = \tau^2$, if $t_A^2 - x_A^2 = \tau^2 > 0$;
- Straight line equations (electromagnetic signals): $x^2 = \pm t$, if $x_A^2 - t_A^2 = 0$;
- Hyperbole space equation: $x^2 - t^2 = \varepsilon^2$, if $x_A^2 - t_A^2 = \varepsilon^2 > 0$;

The geometrical representation is presented in figure 4.5.
In the previous example, the system of coordinates $S \mapsto (x,t)$ represents the observer $O$, the $t$ axis is designated by $O$, and the $x$ axis is designated by $O_\perp$, while the system of coordinates $S' \mapsto (x',t')$ represents the observer $P'$, the $t'$ axis is designated by $P'$, and the $x'$ axis is designated by $P'_\perp$. The hyperbola of equation $t^2 - x^2 = \tau^2$ (an hyperbola corresponding to the temporal axis) corresponds to the geometric place of the event $A(x_A,t_A)$ where the interval between events $A$ and $O(0,0)$ is $(t_A^2 - (x_A)^2) = t^2 - x^2 = \tau^2$.

The hyperbola of equation $x'^2 - t'^2 = \xi^2$ (an hyperbola corresponding to the spatial axis) corresponds to the geometric place of the event $B(x_B,t_B)$ where the interval between events $B$ and $O(0,0)$ is $(t_B^2 - (x_B)^2) = t'^2 - x'^2 = -\xi^2$.

Finally, a circumference corresponds to the geometric place of the point $P(x_P,t_P)$ where the Euclidian distance in relation to $O(0,0)$ is $O-P$. The events where the interval in relation to the origin is $(t^2 - (x)^2) = t^2 - x^2 = 0$ are represented by the straight line equation $t = x$.

Figure 4.6. displays the comparison between the representation of the two metrics in a Minkowski diagram. In this graph, the Euclidian metric is represented by the circumference with an Euclidean distance to the origin, while the Lorentzian metric is defined by one of the three equations of the hyperbolic plane previously described.
Let us consider two events: $A_1(x_1, t_1)$ and $A_2(x_2, t_2)$. The interval between them is defined as $\mathcal{I}(A_1, A_2) = (t_2 - t_1)^2 - (x_2 - x_1)^2$, which in turn enables the definition of the interval measurement as

$$\mu(A_1, A_2) = \sqrt{\mathcal{I}(A_1, A_2)} = \sqrt{(t_2 - t_1)^2 - (x_2 - x_1)^2}.$$ (4.27)

In this case, the measurement that corresponds to the hyperbole equation $t^2 - x^2 = \tau^2$ is $\mu = \tau$, while the measurement corresponding to the hyperbole equation $x^2 - t^2 = \xi^2$ is $\mu = \xi$. The concept of measurement is, in this case, a concept of the hyperbolic plane analogous to the concept of distance in the Euclidean plane. Moreover, the Euclidean distance is given by

$$\rho = \sqrt{t_A^2 + x_A^2}.$$ (4.28)

Additionally, $A \mapsto (x_A, t_A) = (\beta, 1)t_A$, considering $\mathcal{O}$'s point of view, and $B \mapsto (x_B, t_B) = (1, \beta)x_B$, which results in, respectively,

$$\rho = \sqrt{t_A^2 + x_A^2} = t_A\sqrt{1 + \beta^2} = \rho,$$

$$\rho = \sqrt{t_A^2 + x_A^2} = x_B\sqrt{1 + \beta^2} = \rho.$$ (4.29)

As the $\mathcal{P}$ equiloc, from $\mathcal{O}$'s point of view, is represented by the equation $x = \beta t$ and $\mathcal{P}_1$ equitemp, from the same point of view is represented by the equation $t = \beta x$ we have
Finally, for the general case, the measure of the interval (Lorentzian metric) and the distance (Euclidian metric) can be related through the equation

\[ \mu = \tau = \sqrt{t_A^2 + x_A^2} = t_A \sqrt{1 - \beta^2} = \frac{t_A}{\gamma} = \frac{1 - \beta^2}{1 + \beta^2}. \]  

(4.30)

\[ \mu = \xi = \sqrt{x_B^2 - t_B^2} = x_B \sqrt{1 - \beta^2} = \frac{x_B}{\gamma} = \frac{1 - \beta^2}{1 + \beta^2}. \]

When \( \beta = 0 \), case in which observers \( O \) and \( O' \) coincide, the Euclidian distance is the same as the interval measurement \( (\mu = \tau = \xi = \rho) \). Alternatively, when \( \beta \to 1 \), a degenerated hyperbole is obtained with \( \mu \to 0 \).

4.5. The Minkowski Quadratic Space

The Minkowski quadratic space is a vectoral space with a Lorentzian metric added. By using a spacetime simplification we reduce the quadratic space to a bi-dimensional one. In this simplification a single spatial dimension is considered, enabling the identification of the spacetime model through the hyperbolic plane \( \mathbb{R}^{1,1} \). Conversely, when considering the spacetime model without the simplification, a vectoral space defined in \( \mathbb{R}^{3,1} \) must be accounted for. Nonetheless, many relevant examples in the Minkowski quadratic space are solved through the simplification to a bi-dimensional quadratic space (hyperbolic plane). Hence, whenever there is no necessity to use the complete Minkowski space, the simplification will be used. Appendix A details the analysis and definitions of the vectoral space of the hyperbolic plane.

Figure 4.7 depicts a Minkowski diagram that represents the non-Euclidean nature with the four unitary vectors \( \{e_0, e_1, f_0, f_1\} \), which have the same interval invariance

\[ \mu(e_0) = \mu(e_1) = \mu(f_0) = \mu(f_1) = 1. \]  

(4.32)

The unitary vector of time type \( (e_0 \text{ or } f_0) \) defines a set of equilocs, while the unitary vector of space type \( (e_1 \text{ or } f_1) \) defines a set of equitemps. By using the Lorentz boost (active transformation) the unitary vectors \( \{e_0, e_1\} \) are transformed into \( \{f_0, f_1\} \). Additionally, the angle \( \phi \) represents an angle in the hyperbolic plane and the hyperbole \( t^2 - x^2 = 1 \) is the geometric place of the affixes of the vectors of space type. This hyperbole is centered in the origin and has an unitary Lorentzian measurement. In contrast, the hyperbole \( x^2 - t^2 = 1 \) is the geometric place of the
affixes from the vectors of time type with center in the origin and an unitary Lorentzian measurement. Any vector located in the straight line $t = x$ has a null Lorentzian measurement.

Figure 4.7.: Representation of the unitary vectors with their respective different dimensions through Lorentz boost (active transformation).

4.6. The Main Consequences of the Interval Invariance

According to the example of the previous section, from a physical point of view, there are two effects that result from the interval invariance: time dilation and length contraction. These are also the most recognizable consequences of the special relativity theory, as proven before in Bondi’s factor $\kappa$ section.

4.6.1. Time Dilation

If from $O$’s point of view the clock measures the time interval $\tau$ then, from $O'$’s point of view, the corresponding time interval measured is $t_A = \gamma \tau > \tau$ (special relativity theory’s time dilation effect).

The example below, depicted in figure 4.8., has two observers: Alice $\rightarrow (x,t)$ and Bob $\rightarrow (x',t')$. The former is characterized by the equiloc $O$ and the equitemp $O_{\perp}$ while the latter is characterized by the equiloc $O'$ and the equitemp $O'_{\perp}$. Bob is moving away from Alice with velocity $\beta$ with the only point common to both equilocs being the origin $O(0,0)$, where $t_0 = t'_0 = 0$. Event $A$ is defined as an event that belongs to Bob’s equiloc. Additionally, Bob is observing his clock which registers a proper time – a time measured when the clock is at rest – associated with this event and denominated $t'_{A} = T_0$. At the same time, Alice registers her clock
time – \( t_A = T \) – defined by an event \( A' \) simultaneous with event \( A \), from Alice’s point of view – these two events belong to the same equitemp \( O \). In the end, the question that needs to be solved is if \( T \) and \( T_0 \) represent the same value given that, as proven before, they represent different values.

![Diagram](image)

**Figure 4.8.: Representation of the time dilation effect.**

By making use of the analysis and definitions of the vectoral space of the hyperbolic plane present in appendix A, and according to figure 4.8., it is possible to define

\[
\overrightarrow{OA} = \overrightarrow{O A'} + \overrightarrow{A A'}.
\]

\[
\overrightarrow{OA} = t_A \mathbf{f}_0 + x_A \mathbf{f}_1 = t_A \mathbf{f}_0 = T \mathbf{f}_0 = T_0 \mathbf{f}_0.
\]

\[
\overrightarrow{OA'} = t_A \mathbf{e}_0 + x_A \mathbf{e}_1 = t_A \mathbf{e}_0 = T \mathbf{e}_0.
\]

\[
\overrightarrow{A A'} = x_A \mathbf{e}_1.
\]

Hence, by manipulation of the previous equations

\[
\overrightarrow{OA} = \overrightarrow{O A'} + \overrightarrow{A A'} \Rightarrow T_0 \mathbf{f}_0 = T \mathbf{e}_0 + x_A \mathbf{e}_1.
\]

The internal product of both members by the vector \( \mathbf{e}_0 \) leads to the elimination of the last term \( x_A \mathbf{e}_1 \), since \( \mathbf{e}_0 \cdot \mathbf{e}_1 = 0 \) and \( \mathbf{e}_0 \cdot \mathbf{e}_0 = \mathbf{e}_0^2 = 1 \), these changes result in

\[
T_0 (\mathbf{f}_0 \cdot \mathbf{e}_0) = T (\mathbf{e}_0 \cdot \mathbf{e}_0) \iff T_0 (\mathbf{f}_0 \cdot \mathbf{e}_0) = T
\]

From this it can be derive that

\[
f_0 = \gamma (\mathbf{e}_0 + \beta \mathbf{e}_1) \iff f_0 \cdot \mathbf{e}_0 = \gamma \left[ (\mathbf{e}_0 \cdot \mathbf{e}_0) + \beta (\mathbf{e}_1 \cdot \mathbf{e}_0) \right] \iff f_0 \cdot \mathbf{e}_0 = \gamma.
\]

Furthermore, it is possible to deduce the time dilation effect through
\[ f_0, e_0 = \gamma = \frac{1}{s} = \frac{1}{\sqrt{1 - \beta^2}} \geq 1. \] (4.37)

Which, in the end, is summarized as

\[ T = \gamma T_0 \geq T_0. \]

\[ T = e_0, \overrightarrow{OA} = T_0 (f_0, e_0) = \gamma T_0 \] (4.38)

### 4.6.1. Length Contraction

If an object of length \( x_B \), at rest in relation to observer \( O_i \) is measured through the equitemp \( \mathcal{P}_\perp \) by the observer \( \mathcal{P} \), the measurement is given by \( \xi = \frac{x_B}{\gamma} \approx x_B \) (this corresponds to the contraction length special relativistic effect).

The definitions of the previous example apply to the next one, shown in figure 4.9. Although the examples are similar, this time there are two equilocs from Bob’s IFR – \( \mathcal{P}_1 \) and \( \mathcal{P}_2 \). In Bob’s IFR an object is represented by its proper length – \( L_0 \) –, given it is at rest from the observer point of view. The two extreme points of the object are given by the introduced equilocs \( \mathcal{P}_1 \) (left extreme point) and \( \mathcal{P}_2 \) (right extreme point). The measurement of the object length is measured through equitemp \( \mathcal{P}_\perp \) since it allows the measurement of the length of events occurring at the same time instant – simultaneously. Conversely, Alice observes the object on Bob’s IFR moving away with velocity \( \beta \) and measures the length of the object as \( L \). Event \( A \) is an event that belongs to Bob’s equitemp, while event \( A' \) represents the event that belongs to Alice’s equitemp and occurs at the same instant in space as event \( A \) – both events belong to the same equiloc \( \mathcal{P} \). Analogously to the previous example, the question that needs to be solved is whether \( L_0 \) and \( L \) represent the same value given that, as proven before, they represent different values.
In order to correctly measure the extreme points of the object from Alice’s point of view, the measurements have to be simultaneous, over the same equitemp \( O' \). Resorting once again to the analysis and definitions of the vectoral space of the hyperbolic plane, present in appendix A, and according to figure 4.9.

\[
\overrightarrow{OA} = \overrightarrow{OA'} + \overrightarrow{A'A}.
\]

\[
\overrightarrow{OA} = t_{x', f_0} + x_{x', f_1} = x_{x', f_1} = L_0 f_1.
\]

\[
\overrightarrow{OA'} = t_{x, e_0} + x_{x, e_1} = x_{x, e_1} = L_e.
\]

The internal product of both members by the vector \( f_1 \), leads to the elimination of the last term \( t_{x', f_0} \), since \( f_0 \cdot f_1 = 0 \) and \( f_1 \cdot f_1 = f_1^2 = -1 \), these changes lead to

\[
L_0 (f_1, f_1) = L(e_1, f_1) \iff -L_0 = L(e_1, f_1)
\]

And, from this it is known that

\[
f_i = \gamma(\mathbf{e}_i + \beta \mathbf{e}_0) \iff \mathbf{f}_i \cdot \mathbf{e}_i = \gamma \left[ (\mathbf{e}_i \cdot \mathbf{e}_i) + \beta (\mathbf{e}_0 \cdot \mathbf{e}_i) \right] \iff \mathbf{f}_i \cdot \mathbf{e}_i = -\gamma.
\]

Finally, it is possible to deduce the time dilation effect through
\[ \mathbf{f}_0 \cdot \mathbf{e}_i = -\gamma = -\frac{1}{s} = -\frac{1}{\sqrt{1 - \beta^2}}. \]  
(4.43)

\[ \gamma L = -L_0 \Leftrightarrow \left[ L = \frac{L_0}{\gamma} \leq L_0 \right] \]

Which, in the end, is summarized as

\[ L_0 = -\mathbf{f}_0 \cdot \overline{OA} = -L (\mathbf{f}_0 \cdot \mathbf{e}_i) = \gamma L \]  
(4.44)

### 4.6.3. Conclusion

In essence, time dilation of moving objects is the direct consequence of Lorentz transformation consecutively stretching different time coordinates. In contrast, length contraction of moving objects is a combination of a stretching effect of Lorentz transformation on spatial distances (distance dilation) with the changing of a length from different IFRs to a single one (this allows the length to be measured simultaneously).

Time dilation compares the time of the same event for different observers pairing with distance dilation which compares the positions of the same event for different observers. In contrast, length contraction compares positions at the same time according to different observers.

### 4.7. Reciprocity

Both time dilation and length contraction are real and reciprocal. Additionally, in these effects the different points of view of different observers are equally correct. In a three-dimensional space plane, the shortest distance between two points is the straight line that unites them. However, in a bi-dimensional spacetime plane the straight world line (line with a clock attached) between the two events does not represent the shortest path between them, instead it represents the longest interval between the two [45,46].

Both observers are correct as in both are making correct measurements. Time dilation and length contraction do manifest themselves, both being actual physical consequences of the lack of absolute simultaneity [47]. In fact, the effects are manifestations of the fabric of spacetime. Comoving observers do not measure time dilation or length contraction, while non-comoving observers do measure clocks slowing down and objects shrinking.

51
4.7.1 Time Reciprocity

Two observers with relative speed to one another, perceive each other’s perception of time as running slow according to the same factor. In fact, the relative motion causes our perception of time duration between events to seem longer or dilated.

The reason why it is possible to conceive that two observers perceive each other’s time as running slow is that the respective world lines of each observer, which correspond to the observer’s time line axis, rotate in relation to each other hence, each one attributes a projection of the other’s world lines length, as represented through time and space.

The factors by which the intervals are dilated depend on how fast the observers are moving in relation to each other. In order to illustrate the reciprocity effect, it is necessary to make use of figure 4.10., where there are two different observers – $\mathcal{O}$ and $\mathcal{O'}$, which are represented by their respective equiloc and equitemp – moving away from each other with velocity $\beta$. Moreover, there is also a representation of two hyperboles of time type, which the following equations describe

$$t^2 - x^2 = T^2$$
$$t'^2 - x'^2 = T'_0^2$$

The time interval between events $A$ and $O$ has the value $T$ ($\overline{OA} = T$), as represented by the first hyperbole equation, while the time interval between events $B$ and $O$ has the value $T'_0$ ($\overline{OB} = T'_0$). Both these events belong to equiloc $\mathcal{O}$. From $\mathcal{O}$’s point of view, the event belonging to equiloc $\mathcal{O'}$, which is simultaneous with event $A$, is event $B'$. However, event $B'$ belongs to the second hyperbole equation as well, as a result $\mathcal{O'}$’s clock marks a time $T'_0 < T$ in $B'$. In the end, this shows the time dilation effect from $\mathcal{O}$’s point of view. If the analysis is done from the other observer perspective – $\mathcal{O'}$ –, the time interval between events $A'$ and $O$, which belong to the equiloc $\mathcal{O'}$, is given by $T$, as they are also a part of the first hyperbolic equation defined. From $\mathcal{O'}$’s point of view, the event that belongs to the equiloc $\mathcal{O}$ and is simultaneous with event $A'$ is event $B$ which is also part of the second hyperbolic equation. Additionally, observer $\mathcal{O}$’s clock registers a time $T'_0 < T$, in $B$. In the end, the conclusion confirms the expectations: a moving clock runs slower than a clock at rest. As a matter of fact, the result could not have been different because the relativity principle derived from the first postulate determines that if time dilation is observed in one IFR, then it has to be observed in the second IFR in relation to the first one – given the motion between two IFRs is relative, time dilation is reciprocal.
4.7.2 Space Reciprocity

The length contraction effect is reciprocal as well. To show the effect’s reciprocity it is necessary to analyze figure 4.11., where there are two different observers – \( O \) and \( P' \), which are represented by their respective equiloc and equitemp \( \perp \), moving away from each other with velocity \( \beta \). Moreover, there is also a representation of two hyperboles of space type, which the following equations describe

\[
\begin{align*}
  x^2 - t^2 &= L^2 \\
  x^2 - t^2 &= L_0^2
\end{align*}
\]  

(4.46)

From \( O \)'s point of view, an object at rest with proper length \( L_0 \) has its extreme points located at the equilocs \( O \mapsto x = 0 \) and \( O \mapsto x = L_0 \) – measurement performed on equitemp \( \perp \). The space interval (length) between events \( B \) and \( O \) has the value \( L_0 \left( \overline{OB} = L_0 \right) \). By contrast, from \( P' \)'s point of view, the same object is moving hence, it has the measurement \( L \) on the equitemp \( \perp \). This measure corresponds to the length between \( A' \) and \( O \left( \overline{OA'} = L \right) \). The two hyperbole equations above are a clear evidence that there is length contraction because \( L < L_0 \).

From a different perspective – in this case, \( P' \)'s point of view –, the object is at rest, having proper length \( L_0 \) and extreme points located at the equilocs \( P' \mapsto x' = 0 \) and \( P' \mapsto x' = L_0 \) – measurement done on equitemp \( \perp \). The length between events \( B' \) and \( O \) has the value \( L_0 \left( \overline{OB'} = L_0 \right) \). However, from \( O \)'s point of view, the same object is moving, with a length of \( L \).
on equitemp $Q_1$, which results in a length between $A$ and $O$ with value $L\left(\overline{OA'} = L\right)$. Once again, the two hyperbola equations above are a clear evidence that there is length contraction because $L < L_0$ – as the motion between two IFRs is relative, length contraction is reciprocal.

Figure 4.11.: Representation of an example for the space reciprocity.
Chapter 5 – Special Relativity: Other Effects and Paradoxes

5.1. Introduction

Two of the most known consequences of the special relativity theory are time dilation and length contraction. Both these effects occur due to the concept of relative simultaneity, which, in turn, is a consequence of the fact that light is a cosmic limit and a finite value. Nonetheless, there are more important effects derived from the special relativity theory. This section aims to describe them: the Einsteinian velocity composition, the causality, the apparent superluminal speeds, the Doppler effect and the twin paradox.
5.2. Einsteinian Velocity Composition

As previously addressed in section Galilean transformation, in order to describe Einsteinian composition of velocities, a velocity \( w \) does not equal the algebraic sum of any other velocities \( u \) and \( v \). This is a direct consequence of the fact that if we consider either \( u \) or \( v \) as a velocity equal to the speed of light \( c \), the result would be a physical impossibility – a velocity greater than the speed of light \( w = c + v > c \). Additionally, this situation contradicts Einstein’s second postulate. This composition of velocities is a consequence, as previously explained, of simultaneity being a relative concept dependent on the observer (IFR) instead of an absolute concept, as considered in Galilean transformation. When there are only small velocities, the Galilean transformation is applicable because the error associated can be neglected. As a result, \( w = u \oplus v = u + v \).

Einsteinian velocity composition is described by the formula \( w = u \oplus v \neq u + v \). This formula implies that for any given particle, when considering high velocities, the error associated with the Galilean addition of velocities cannot be neglected. Consequently, considering \( u = c \), the relativistic composition of velocities is given by \( w = c \oplus v = c \). In the end, this proves that the velocity of any particle with no mass (e.g.: photons) is always \( c \) and that the second postulate is verified.

![Figure 5.1.: Relativistic Composition of Velocities.](image)

In order to deduce the new Einsteinian velocity composition, it is necessary to define the velocity between the different IFRs through two different, although correct, conventions.

As figure 5.1. depicts, the first convention is described by geometric units. In fact, it is possible to define that the IFR \( S"(x",t") \) is moving with normalized velocity \( \beta_2 \) in relation to the IFR \( S'(x',t') \), the IFR \( S'(x',t') \) is moving with velocity \( \beta_1 \) in relation to the IFR \( S(x,t) \), and the IFR \( S"(x",t") \) is moving with velocity \( \beta \) in relation to the IFR \( S(x,t) \). Einsteinian velocity composition’s result is \( \beta = \beta_1 \oplus \beta_2 \neq \beta_1 + \beta_2 \). The example in figure 5.2., represents the
previously mentioned IFRs through three different observers and their clocks 
\( \mathcal{O} \mapsto S(x,t) \), \( \mathcal{O}' \mapsto S'(x',t') \), and \( \mathcal{Q} \mapsto S''(x'',t'') \), respectively.

![Diagram of electromagnetic signal passage](image)

**Figure 5.2.:** Representation of an electromagnetic signal that passes by three different observers.

In the figure an electromagnetic signal is sent, at time instant \( t_0 \). This signal is sent from observer \( \mathcal{O} \), passing through observer \( \mathcal{O}' \), at instant \( t'_0 \), and reaching observer \( \mathcal{Q} \), at instant \( t''_0 \).

Resorting to the Bondi’s factor \( \kappa \) radar method, it is possible to define

\[
\begin{align*}
    t'_0 &= \kappa_1 t_0, \\
    t''_0 &= \kappa_2 t'_0, \\
    t''_0 &= \kappa t_0, \\
    \kappa &= \frac{1 + \beta}{\sqrt{1 - \beta^2}}. \\
\end{align*}
\]

Through algebraic manipulation

\[
    t''_0 = \kappa t_0 = \kappa_2 t'_0 = \kappa_2 \kappa_1 t_0 \Rightarrow \kappa = \kappa_2 \kappa_1. 
\]

By substituting \( \kappa \) we reach the expected result for the new velocity composition

\[
    \kappa = \kappa_2 \kappa_1 \iff \sqrt{1 + \beta} = \sqrt{1 + \beta_1} \sqrt{1 + \beta_2} \iff \beta = \frac{\beta_1 + \beta_2}{1 + \beta_1 \beta_2}. 
\]

Utilizing SI units, namely \( u = \beta_1 c \), \( v = \beta_2 c \), and \( w = \beta c \), the same velocity composition can be written as
\[ w = \frac{u + v}{\frac{c}{1 + \frac{uv}{c^2}}} \Leftrightarrow \frac{w = u + v}{1 + \frac{uv}{c^2}} \] 

Hence, considering \( u = c \), the relativistic composition of velocities is given by

\[ w = c \oplus v = \frac{c + v}{1 + \frac{cv}{c^2}} = \frac{c + v}{c} = c. \]

In the end, as initially mentioned, there is proof that the new rule of velocity composition does not contradict Einstein’s second postulate and that \( c \) is a universal constant.

Einsteinian velocity composition is radically different from the prior relativistic Galilean addition of velocities. According to the former, the relativistic composition of two velocities with values below the speed of light is also a velocity with a lower value. From a theoretical point of view, the result of a composition of two velocities, in which one of them has a higher value than light’s, is a velocity that can be, although not necessarily, a velocity superior to light’s.

Nonetheless, in the real world, there is no signal that propagates with a velocity superior to the speed of light, theme which will be addressed in the following section. Thus, the existence of such a signal would invalidate the principle of causality.

5.3. Causality

The principle of causality states that a given event, \( A \), originates another event, \( B \). In this case, \( A \) is the cause; as a result, it has to always occur before event \( B \) – which is the effect –, considering all the different points of view of the different observers. The objective of this section is to prove that electromagnetic signals with velocities bigger than the constant speed of light violate the principle of causality, scenario which implies the physical impossibility, that an effect happen before its cause.

Figure 5.3.: Representation of an impossible example where the effect happens before the cause.
Figure 5.3. depicts an example where two observers are represented, \( \mathcal{O} \mapsto S(x,t) \) and \( \mathcal{D} \mapsto S'(x',t') \). An electromagnetic signal described by the equation \( x = \beta_1(t - t_A) \), with \( \beta_1 > 1 \), is sent from the observer \( \mathcal{O} \) with a velocity that has a higher value than the constant speed of light at event \( A(0,t_A) \). This signal reaches the observer \( \mathcal{D} \) at event \( B(x_B,t_B) \). Instantaneously, upon receiving the signal, the observer \( \mathcal{D} \) sends another electromagnetic signal with a velocity above the constant speed of light. In this case, the equation \( x = \beta_2(t - t_C) \), with \( \beta_2 > 1 \), describes the signal that reaches observer \( \mathcal{O} \) at event \( C(0,t_C) \). This event, from both points of view, occurs before the event \( A \) — hence, \( t_C < t_A \) —, which means that the second electromagnetic signal reaches the observer \( \mathcal{O} \) in the past when measured against the sending of the first signal.

Following a mathematical approach, this example is correct however, from a logical point of view, it is impossible for an observer that sends an electromagnetic signal to receive, as a response, another signal of the same type which has an arrival time prior to the sending time of the first signal considered. In the end, the conclusion that can be drawn from this example is that, as explained before, all physical actions propagate with velocities that are, at most, equal to the speed of light.

Furthermore, it is also important to mention that, in special relativity, relative simultaneity implies that a set of events can be interpreted as having different sequences of events from different IFRs.

In another example, which figure 5.4. represents, two photons are created at event \( C \), as a result of the collision between an electron \( (e^-) \) — created at event \( A \) — and a positron \( (e^+) \) — created at event \( B \). The two photons are then absorbed, one at event \( D \) and the other at event \( E \).

![Figure 5.4.: Representation of the collision example.](image)
The set of events in figure 5.4. are described by two different observers characterized by the equilocs \( \{ Q, Q' \} \). The temporal sequence of the events depends on the observer analyzing, according to the way in which different equitemps, from a given observer, intersect chronologically the events. The resulting sequence of events given by the different observers is

\[
\begin{align*}
\text{observer with clock} \ O: & \quad A \rightarrow B \rightarrow C \rightarrow D \rightarrow E \\
\text{observer with clock} \ O': & \quad B \rightarrow A \rightarrow C \rightarrow E \rightarrow D
\end{align*}
\]

(5.6)

Confirming the statement that different observers associate a different order to the events. This is a direct consequence of the concept of relative simultaneity.

### 5.4. Apparent Superluminal Speeds

Although neither light nor any conceivable object propagate with bigger velocities than the speed of light, a distant object can appear to travel faster than the speed of light across our line of sight. If a superluminal velocity appears, causality faces serious logical problems.

Considering the following example, in figure 5.5., at instant \( t = 0 \), an observer \( O \) watches two light pulses emitted from \( x = 0 \), with opposite directions. The light pulse propagating to the left is described by equation \( x_1 = -ct \) while the other light pulse, which is propagating to the right, is
described by equation \( x_2 = ct \). Both signals, from \( \mathcal{O}' \)'s point of view, move away with closing velocity

\[
\frac{dw}{dt} = \frac{d}{dt}(x_2 - x_1) = \frac{d}{dt}(2ct) = 2c > c
\]

(5.7)

\[
\beta = \frac{w}{c} = 2 > 1
\]

Which is an apparent superluminal speed. The different observers are represented by their respective IFRs: left pulse \( \mathbb{S}(x, t) \to \mathcal{O} \to \mathbb{S}'(x', t') \) and right pulse \( \mathbb{S}''(x'', t'') \). If \( \delta > 0 \) or \( \delta \ll 1 \), it is possible to define the relative velocity between the different IFRs, as explored in the section relative to composition of velocities, that the IFR \( \mathbb{S}'(x', t') \) is moving with normalized velocity \( \beta_2 \) in relation to the IFR \( \mathbb{S}'(x', t') \), the IFR \( \mathbb{S}'(x', t') \) is moving with velocity \( \beta_1 \) in relation to the IFR \( \mathbb{S}(x, t) \), and the IFR \( \mathbb{S}''(x'', t'') \) is moving with velocity \( \beta \) in relation to the IFR \( \mathbb{S}(x, t) \).

If we define

\[
\beta_1 = \delta - 1 \\
\beta_2 = 1
\]

(5.8)

And apply the composition of velocities formula

\[
\beta = \beta_1 \oplus \beta_2 = \frac{\beta_1 + \beta_2}{1 + \beta_1 \beta_2} = \frac{\delta}{\delta} \to 1, \text{ when } \delta \to 0.
\]

(5.9)

The conclusion is that, although the closing velocity – the rate of change of radial distance between two objects – is \( \beta = 2 > 1 \), the relative velocity is actually \( \beta = 1 \). From a mathematical point of view, the closing velocity is a valid concept however, this velocity does not represent the propagation velocity of any physical entity. Moreover, this concept must not be confused with the relative speed which indicates the speed of a body in the reference frame of another [44].

### 5.5. Longitudinal Doppler Effect

Another important effect that the special relativity encapsulates is the Doppler effect. More specifically, according to the special relativity theory, one of the consequences of the Doppler effect is that, for a certain observer, the relative perception of the frequency of a given emitting source depends on the relative movement between them [48]. This is different from the frequency of the emission source, with relative motion in respect to the observer.
In this section the longitudinal Doppler effect will be deduced by utilizing the Bondi’s factor radar method. Taking into consideration that the movement in study only has one spatial direction, it is not possible to deduce the transverse Doppler effect.

![Diagram](image)

Figure 5.6.: Representation of an observer that sends periodically signals to another observer.

Figure 5.6. represents an example in which there are two different observers, $\mathcal{O} \mapsto (x,t)$ and $\mathcal{R} \mapsto (x',t')$ – represented by their respective equilocs. The source $\mathcal{O}$ is moving away from the receptor $\mathcal{R}$ with relative velocity $\beta$, in relation to the latter. An electromagnetic signal is sent periodically, at instant $t_x$, reaching its goal at instant $t_x'$. As a result, the time interval between two electromagnetic signals is $T_e = t_{x+1} - t_x$ according to $\mathcal{O}$’s point of view while being $T_r = t_{x+1}' - t_x'$ from $\mathcal{R}$’s point of view. Through Bondi’s factor radar method, it is possible to establish that

$$t_x' = \kappa t_x \Rightarrow \begin{cases} t_1' = \kappa t_1 \\ t_2' = \kappa t_2 \end{cases}.$$  

$$\kappa = \frac{1+\beta}{\sqrt{1-\beta}} = e^{\phi}.$$  

(5.10)

Which, by usage of algebra manipulation, results in

$$T_r = t_{x+1}' - t_x' = \kappa (t_2 - t_1) = \kappa T_e.$$  

(5.11)

As the emission frequency is the inverse of the emission period and the reception frequency is the inverse of the reception period

$$\begin{cases} f_e = \frac{1}{T_e} \\ f_r = \frac{1}{T_r} \end{cases}.$$  

(5.12)
Since the relation between $T_e$ and $T_r$ is already known, it is also possible to determine the relation between $f_e$ and $f_r$, resorting to equations 5.11. and 5.12.

$$\frac{1}{f_r} = \kappa \frac{1}{f_e} \iff f_r = \frac{f_e}{\kappa} \iff f_r = \sqrt{\frac{1-\beta}{1+\beta}} f_e.$$  \hspace{1cm} (5.13)

This equation shows the relativistic Doppler effect when the observers are moving away from each other, with relative velocity of $\beta > 0$. In the end, it is possible to draw the conclusion that $f_r < f_e$. In other words, there is a red shift of the frequency – a deviation to lower frequencies. Through this it is possible to determine that the universe is in expansion: by analyzing the spectral emissions from different stars – signals present a red shift of their frequency, it is observable that the distance between the source (star) and the receptor (Earth) is getting larger.

A different scenario occurs when the observers are getting closer to each other. In that case the relative velocity is $\beta < 0$ consequently, $f_r > f_e$. There is a blue shift of the frequency – a deviation to higher frequencies. In this second scenario, the equation changes to

$$f_r = \kappa f_e \iff f_r = \sqrt{\frac{1+\beta}{1-\beta}} f_e.$$  \hspace{1cm} (5.14)

Furthermore, there is also a classic Doppler effect that occurs, for instance, in sound propagation. In fact, when a moving object emitting a sound is getting closer, the sound becomes more treble – with higher frequency – and as the same object is getting away, the sound becomes more bass – with lower frequency.

### 5.6. Transverse Doppler Effect

Let us consider a circular movement, defined in a single plane. The study of this type of movement has at least two spatial dimensions, in order to describe the circumference, one temporal dimension, and the distance between the source and the receptor is constant and equal to $r$. This type of movement is, for instance, the same a satellite performs around the Earth. Looking at an object moving transversely across the line of sight, the frequency is not the same as the frequency emitted by the source. Nowadays, with the relativistic physics reality being considered, a pure manifestation of the time dilation effect occurs, this is the reason why this effect does not exist in classical physics.

As depicted in figure 5.7., the source (origin) only requires a clock associated with a single IFR hence, it is at rest fixed at the same position. Alternatively, the receptor has a clock that is not associated with a single IFR hence, it is necessary to make use of a different clock to describe the circular movement, with each instant being characterized by an IFR composed by three
different axes. The first axis is described by its tangential velocity – this velocity might have a constant magnitude, but it is not constant due to the centripetal acceleration that leads to the change of direction of the velocity. The second axis represents the centripetal acceleration with constant magnitude. Lastly, there is a third axis which is perpendicular to both axes mentioned before.

Figure 5.7.: Circular movement.

The relative motion between two observers, one being at the center and the other describing a circular movement around the first one, is always the same; what changes for the observer describing a circular movement is the local IFR, which is different at every instant. The general relativistic Doppler effect, as seen before, is given by

\[ f_r = \frac{\sqrt{1 - \beta^2}}{1 + \beta \cos(\theta_0)} f_e \]  

(5.15)

Where \( \theta_0 \) represents the angle defined by the direction of the emitting source and the direction of the observer. In the specific case of the circular movement, where \( \theta_0 = \pi / 2 \), the result obtained is the transverse red shifted frequency relative to the observer. This is considered a pure relativistic effect

\[ f_r = \frac{f_e}{\gamma} = \sqrt{1 - \beta^2} f_e \]  

(5.16)

Thus, according to the special relativity theory, the moving object’s emitted frequency and the received frequency are both reduced by the Lorentz factor.
5.7. The Twin Paradox

A famous paradox from literature is The Twin Paradox which is deduced through the Doppler effect. In essence, this paradox is represented by the example in which there are two different observers, in this case twin brothers. One of them stays on Earth while the other embarks on a spacial travel at a high speed – near the speed of light. Years later, when the travelling twin brother returns back to Earth, the brother that remained on Earth is much older than the brother that went away. This phenomenon is a direct consequence of the time dilation effect.

However, as there is a time reciprocity, it is also proper to affirm that the opposite situation might be true [49]. Through the relativity principle, it can be considered that the traveling brother is represented by the IFR at rest while his twin brother that stayed on Earth is the one with an IFR moving at high speed. Considering this situation, the twin brother that remained on Earth would be the one that aged slowly and hence the younger of the two. Taking into consideration that both situations cannot be simultaneously true, there is, indeed, a paradox. In order to solve this problem, it is necessary to make use of the numerical example below.

There are two observers, who are twin brothers according to the paradox, Bob described by \( \mathcal{O} \rightarrow S(x,t) \) and Alice described by \( \mathcal{O}' \rightarrow S'(x',t') \). In this case, Alice is the travelling astronaut who does a round trip composed by two different stages: the departure and the arrival. In both stages the relative velocity between the observers is uniform with magnitude \( \beta = \frac{3}{5} \).

Using geometric units – time is measured in years and distance is measured in light-years – Alice is moving away from Earth until she reaches a distance to Earth of \( L = 3 \) light-years. Once this distance is achieved Alice reverses her motion, instantaneously, going back to Earth. The only instance during the trip in which it can be considered there is acceleration is on the turning point, where Alice’s velocity passes from \( +\beta \) to \( -\beta \) – which is considered an instantaneous infinite acceleration. This moment is the most relevant given it is the point where the demystification of the paradox occurs since it breaks the reciprocity of both twins due to the impossibility of describing Alice’s round trip through a single IFR.

Figure 5.8. contains the numerical example through a Minkowski diagram of the situation just described.
In figure 5.8., from O’s point of view, event A(L, T/2) belongs to the equiloc ☐ and event B(0, T) belongs to equiloc ☐, where L represents the distance traveled by Alice in a single direction and T represents the time of one round trip done by Alice. Bob’s world line corresponds to the sequence of events O→B while Alice’s corresponds to O→A→B.

From O’s point of view, and given Alice’s trip lasted T years for the distance $L = 3$ light-years

$$L = \beta \frac{T}{2} \Leftrightarrow T = \frac{2L}{\beta} = 10 \text{ years} \quad (5.17)$$

However, from ☐’s point of view, Alice’s trip took $T' = 8$ years to be completed

$$T' = \frac{T}{\gamma} = 8 \text{ years, with } \gamma = \frac{1}{\sqrt{1 - \beta^2}} = \frac{5}{4} \quad (5.18)$$

Consequently, considering Alice and Bob were 20 years old at the beginning of the trip, they would be, respectively, 28 and 30 years old when they met again at the end of the voyage.

Bob’s IFR, $O \mapsto S(x, t)$, is described by two constant unitary vectors $(\mathbf{e}_0, \mathbf{e}_1)$. However, Alice’s IFR, $\mathcal{D} \mapsto S'(x', t')$, as a consequence of the turning point (instantaneous infinite acceleration) of the trip, can be described by two different IFR, with the infinite acceleration point (turning point) being neglected. In fact, the first IFR, corresponding to the departure trip, is $\mathcal{D}_1 \mapsto S'_1(x', t')$ – described by two constant unitary vectors $(\mathbf{f}_0, \mathbf{f}_1)$ – while the second IFR, corresponding to the arrival trip, is $\mathcal{D}_2 \mapsto S'_2(x', t')$ – described by two constant unitary vectors $(\mathbf{g}_0, \mathbf{g}_1)$. As a result, it is known that
\[
\begin{align*}
\mathbf{f}_0 &= \gamma (\mathbf{e}_0 + \beta \mathbf{e}_1) \\
\mathbf{f}_1 &= \gamma (\mathbf{e}_1 + \beta \mathbf{e}_0) \\
\mathbf{g}_0 &= \gamma (\mathbf{e}_0 - \beta \mathbf{e}_1) \\
\mathbf{g}_1 &= \gamma (\mathbf{e}_1 - \beta \mathbf{e}_0)
\end{align*}
\] (5.19)

Figure 5.9. represents the unitary vectors associated with each of the IFRs.

As in Bob's IFR, Alice's departure trip is described by the equation \( x = \beta t \), equiloc \( \mathcal{P}_1 \), and the equitemp described by the equation \( t = \beta x + t_a \). Yet, from the same point of view, Alice's arrival trip is described by the equation \( x = \beta (T - t) \), equiloc \( \mathcal{P}_2 \), and the equitemp described by the equation \( t = -\beta x + t_b \). Additionally, \( T / 2 = L / \beta \).
\[ t_a = (1 - \beta^2) \left( \frac{T}{2} \right) = \frac{T}{2} - \beta L = 3.2 \text{ years} \]
\[ t_b = (1 + \beta^2) \left( \frac{T}{2} \right) = \frac{T}{2} + \beta L = 6.8 \text{ years} \]

(5.20)

On the turning point of the trip, at time instant \( T / 2 \), there is a change of direction in Alice’s equitemps which pass, instantaneously, from time instant \( t_a = 3.2 \text{ years} \) to time instant \( t_b = 6.8 \text{ years} \). In other words, by turning around instantaneously, Alice loses a period of time which Bob lives [50]. This period, unknown for Alice, is given by the temporal difference between both time instants

\[ \Delta t = t_b - t_a = \beta^2 T = 3.6 \text{ years.} \]

(5.21)

This is the reason why the turning point is so relevant: it breaks the reciprocity between the twins [51]. Another instantaneous change for Alice is the IFR changing from \( \mathcal{O} \mapsto S_1' \leftrightarrow (x', t') \mapsto (f_0', e_0') \) to \( \mathcal{O} \mapsto S_2' \leftrightarrow (x', t') \mapsto (g_0, e_0) \). Regardless of these changes, Bob does not lose any time period because his world line is described by a unique IFR \( \mathcal{O} \mapsto S(x, t) \mapsto (e_0, e_1) \).

In the moment immediately before event A (turning point), Alice sends an electromagnetic signal described by the equation \( t = -x + t_i \). As a result, Bob, through the signals received from Alice, only perceives that Alice turned around at instant \( t = t_i = (1 + \beta) (T / 2) = T / 2 + L = 8 \) years. Consequently, from Bob’s point of view, Alice’s arrival trip should take \( t_a = T - t_i = (1 - \beta) (T / 2) = T / 2 - L = 2 \) years. In the end, from Bob’s point of view Alice’s voyage has a duration of \( T = t_i + t_a = 10 \) years while from Alice’s point of view the same trip has the duration of \( T' = t_1' + t_2' = 8 \) years, where \( t_1' = (T / 2) / \gamma = 4 \) years.

With the purpose of comparing the different points of view, both twins send an electromagnetic signal to each other with the same frequency – frequency measured in their own IFR, from where the signals are emitted. However, as proved by the Doppler effect, the frequency received differs from the frequency emitted. In Alice’s departure trip the frequency suffers a red shift \( f \mapsto f' = f / \kappa < f \), as both observers are moving away, while in the arrival trip the frequency suffers a blue shift \( f \mapsto f'' = \kappa f > f \), given both observers are getting closer. Figure 5.10. presents a comparative analysis of the emitted and received signals in relation to both points of view.
In this figure it is considered that each twin emits a signal once a year \( f = 1 \text{year}^{-1} \). Additionally, \( \beta = 3/5 \), \( \gamma = 5/4 \), \( \kappa = 2 \) and due to the Doppler effect, \( f' = f/\kappa = 0.5 \text{year}^{-1} \) and \( f'' = \kappa f = 2 \text{years}^{-1} \). Throughout the whole trip, the total number of signals sent from Bob was \( N = fT = 10 \) and the total number of signals sent from Alice was \( N' = f'T' = 8 \). Although the total number of signals received by Bob and Alice, respectively, on the departure trip were \( M_1 = f_1't_1 = 4 \) and \( M'_2 = f'_2't'_1 = 8 \), for the arrival trip the signals received by Bob and Alice were, respectively, \( M_2 = f_2't_2 = 4 \) and \( M'_2 = f'_2't'_1 = 8 \). Hence, Bob received a total of \( M = M_1 + M_2 = 8 \) signals and Alice received a total of \( M' = M'_1 + M'_2 = 10 \) signals. In the end, both Bob and Alice are in agreement given \( M' = N' \) and Bob correctly predicts that his sister aged 8 years while Alice correctly predicts that her twin aged 10 years.

This analysis has a problem that remains to be solved. In fact, neglecting the turning point (simplification), a point where Alice has an instantaneous infinite acceleration (an impossible situation that would lead to her death) requires us to revisit the twin paradox, in appendix B, considering the world line of Alice as a parabolic section.
Chapter 6 – Relativistic Effects and GPS

6.1. Introduction

Both the general relativity theory and the special relativity theory, conceived by Albert Einstein, are essential to the measurement of distances and to the phase carrier. In fact, without them the GPS would not work properly [52]. The satellites are affected by relativity in three different ways: movement equations, signal propagation, and satellites’ clock rhythm. As a result, the important effects that require correction are the gravitational deviation of the frequency of the clock satellite, time dilation, the Sagnac effect and the eccentricity effect [53-56]. The principle of the constancy of \( c \) finds application as the fundamental concept on which the GPS is based. Timing errors of \( \Delta t = 1 \text{ ns} \) will lead to positioning errors of magnitude \( \Delta x \approx 30 \text{ cm} \).

When considering navigation purposes, the effects presented are enough to achieve a precise location on the surface of the Earth. However, in some specific applications, it might be necessary higher accuracy. For these particular cases with accuracy levels needed of a few centimeters or even millimeters, the delays cannot be larger than a few picoseconds. Researchers are modelling systematic effects down to this described level to be applied in future and more complex applications. Among these effects are, one must especially consider the signal propagation delay, the effect on geodetic distance, the phase wrap-up effect, and the effect of other solar system bodies.

GPS is based on the Newton’s model, utilizing an NIFR (Non-Inertial Frame of Reference) with a rotational movement alongside Earth’s and with origin at the center of Earth – ECEF (Earth Centered Earth Fixed), for clocks on the Earth surface at rest. A system of inertial time reference centered on Earth – ECI (Earth Centered System of Coordinates), for clocks orbiting Earth, is also utilized.

A correction is necessary given the special relativity theory predicts that a moving clock (in an IFR) is slower than a stationary one. As a result, as a satellite orbits Earth its clock will slow down in relation to the clock at the receiver down on Earth. This happens because it is considered that the signals from the satellite and from the receiver are moving relatively to an IFR.

Additionally, the general relativity theory predicts that clocks move faster in a higher gravitational field. Since clocks’ rhythms do not depend only on their velocity, there is also a deviation of the frequency of the clock satellite due to the higher gravitational potential. The difference in the gravitational potential between the satellite and the receiver at the ground, translate in the receiver’s clock moving slower than the clock on the satellite that is further apart from the gravitating mass considered – Earth.
Although the two effects are acting in opposite directions, their magnitude is not equal, thus they do not cancel each other out. Moreover, these errors are cumulative, which implies that, in case of them not being corrected, one would not compensate for the difference in clock speeds, resulting in a wrong distance measurement and in a position estimation, potentially, hundreds of meters off. These failures would render the GPS virtually useless.

The third effect arises given the GPS’s time scale is defined through a single IFR in contrast to Earth’s NIFR associated with the planet’s rotation. This difference between the characteristics of each frame of reference leads to the emergence of the Sagnac effect. In essence, this effect is a consequence of the difference in location when the electromagnetic signal of the satellite is sent when compared to when it is received. More specifically when the receiver on Earth suffers a finite rotation due to Earth’s rotation motion in relation to the system of inertial time reference centered on Earth – ECI.

### 6.2. Relativistic Effects Correction

The satellite of a navigation system, when orbiting, is accelerated with the direction of the center of Earth due to gravitational effects. The variable  \( v_s \) (satellite velocity) is constant –  \( v_s = \|v\| \) – and can be obtained through the following equation

\[
v_s = \sqrt{\frac{GM}{r_s}},
\]

(6.1)

Considering the vector variable \( \mathbf{r} \) is

\[
\|\mathbf{r}\| = r_s = 26562 \text{km}.
\]

(6.2)

And considering the equatorial radius of Earth

\[
r_e = 6378 \text{km},
\]

(6.3)

and the orbital height of the satellite, which for the GPS satellites is of 20184 km the Gravitational parameter of Earth is defined by

\[
GM = \mu = 3.986005 \times 10^{14} \text{ m}^3\text{s}^{-2},
\]

(6.4)

With \( G \) being the universal gravitational constant and \( M \) representing the Earth’s mass.

Finally, we can conclude that  \( v_s = 3874 \text{ m s}^{-1} \) for a GPS.
6.3. Effect of Time Dilation

One of the most important relativistic effects on the GPS is the slowing down factor associated with the clocks that are in relative motion in relation to Earth.

As it was already shown, the time dilation is given by the equation

$$\tau_A = t_A \sqrt{1 - \beta^2}. \tag{6.5}$$

In order to calculate precisely this effect during a day, we need to know how many seconds a day has, from the moving system of coordinates point of view (Bob)

$$\tau_A = 24 \times 3600 = 86400 \text{ s.} \tag{6.6}$$

As explained before, the GPS satellite velocity is $v_s = 3874 \text{ m s}^{-1}$, in relation to ECI, due to Earth’s rotation. In geometric units, this velocity would be $\beta = 1.2922 \times 10^{-5}$. An observer at rest in the ECI will notice the satellite clock is affected by the slowing down factor

$$s = \sqrt{1 - \beta^2}. \tag{6.7}$$

This factor can be approximately expanded using the binomial theorem,

$$\sqrt{1 - \beta^2} \approx 1 - \frac{1}{2} \beta^2. \tag{6.8}$$

Taking into account the GPS satellite velocities, the error per day due to time dilation effect is

$$\varepsilon = -\frac{1}{2} \beta^2 \approx -\frac{1}{2} \frac{v_s^2}{c^2} \approx -8.3485 \times 10^{-11}. \tag{6.9}$$

Converting this result to seconds

$$\delta t = \tau_A - t_A = \varepsilon \tau_A = -7.2131 \mu\text{s.} \tag{6.10}$$

A referential clock in the equatorial plane has the same movement as an ECI due to Earth’s rotation. However, its velocity is inferior to the satellites’ by

$$\omega \cdot r_e \approx 465 \text{ m s}^{-1}. \tag{6.11}$$

This represents the velocity of Earth’s rotation, with an average angular Earth velocity of $\omega = 7.292115 \times 10^{-5} \text{ rad s}^{-1}$. To obtain the difference between a GPS satellite clock (ECI) and a reference clock in the equator (ECEF), first we need to calculate the order of magnitude of the time dilation effect

$$-\frac{1}{2} \beta^2 \approx -\frac{1}{2} \frac{(\omega r_e)^2}{c^2} \approx -1.2013 \times 10^{-12} \approx -103.792 \text{ ns.} \tag{6.12}$$
Afterwards, it is necessary to compute the fractional frequency difference

\[
\Delta f = -\frac{1}{2} \frac{v^2}{c^2} - \left( -\frac{1}{2} \frac{\omega r}{c^2} \right) = -8.2284 \times 10^{-11} \approx -7.1093 \mu s.
\] (6.13)

Finally, we can conclude that if this navigational error was not accounted for, it would increase resulting in a distance error, at the end of a day, of

\[
\delta x = c \left| \delta t \right| = 2.1328 \text{ km d}^{-1}.
\] (6.14)

### 6.4. Newtonian Gravity

The Newtonian gravitational field is described by the potential gravitational function \( \Phi : \mathbb{R}^3 \rightarrow \mathbb{R} \), through which the gravitational field \( G : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \) is described according to the formula \( G = -\nabla \Phi \). The trajectory \( r : \mathbb{R} \rightarrow \mathbb{R}^3 \) of a particle with mass \( m \) in free fall on the gravitational field is described through the differential equation, that obeys Newton’s second Law

\[
m \frac{d^2 r}{dt^2} = m G
\] (6.15)

This means that the acceleration suffered by a particle while free falling is \( G \), independently of the particle observed. Consequently, the total energy conserved is

\[
E = \frac{1}{2} m \frac{dr}{dt} + m \Phi.
\]

\[
\frac{dE}{dt} = 0.
\] (6.16)

The term \( m \Phi \) represents the potential gravitational energy of a particle. The field is considered weak if the characteristic velocity has a value much lower than the speed of light and the equations above can be applied. However, if the velocities are almost equal to the speed of light, it is necessary to take into consideration the time dilation effect and make use of the formulas that describe the theory of general relativity.
6.5. Effect of Gravitational Deviation of the Frequency of the Clock Satellite

If we consider two clocks, a reference clock that stays at rest on Earth’s equator with radius $a_i$ and an atomic clock orbiting the Earth with radius $r_s$ – satellite clock. Remembering that, the gravitational potential of a satellite created by a mass $M$ (Earth’s mass) spherically symmetrical is given by

$$\Phi_s = -\frac{GM}{r_s} \approx -14.984979 \times 10^6. \quad (6.17)$$

Considering that $GM = 3.986004415 \times 10^{14} \text{ m}^3/\text{s}^2$ and $r_s = 26600 \text{ km}$. However, the gravitational potential of a specific point on Earth’s surface is affected given Earth is not a perfect sphere hence, the real gravitational potential is

$$\Phi(r) = -\frac{GM}{r(\theta)} \left( 1 - J_2 \frac{a_i^2}{r^2(\theta)} \left( \frac{1}{2} (3 \sin^2 \theta - 1) \right) - \frac{1}{2} \left( wr(\theta) \cos \theta \right)^2 \right). \quad (6.18)$$

With the Earth’s quadrupole moment coefficient being $J_2 = 1.0863 \times 10^{-3}$, the angle $\theta$ being measured between the equator and the location of the receiver on Earth’s surface, and $r(\theta)$ being the radius of the Earth depending on the value of the location $\theta$. The gravitational potential is divided in two parts: the potential caused by the mass of the Earth ($\Phi_{\text{static}}$) and the potential from the centripetal forces consequent of Earth’s rotation – this last term of equation 5.4. compensates the time dilation effect that originates from the motion of the resting clocks on Earth (in a ECI). By considering a clock on the equator ($\theta = 0$), the equivalent gravitational potential is simplified

$$\Phi(r) = -\frac{GM}{a_i} \left( 1 + \frac{J_2}{2} \right) - \frac{1}{2} \left( wa_i \right) \approx -7 \frac{GM}{a_i} \left( 1 + \frac{J_2}{2} \right) \approx -59.8798 \times 10^6. \quad (6.19)$$

Taking into account that the time dilation effect ($\Delta t / t$) for an observer at rest with potential $\Phi(r)$ in relation to a clock with potential $\Phi_s$ is given by

$$\frac{\Delta t}{t} = \frac{\Phi_s - \Phi(r)}{c^2}. \quad (6.20)$$
Furthermore, we can neglect the quadrupole moment, consider that the satellites are in free fall, hence neglecting the centripetal potential as well, and it is known that Earth’s radius is $6371 \text{ km}$.

By combining the previous equations it is possible to calculate the order of magnitude of the gravitational frequency shift of GPS satellite clocks

$$
\varepsilon = \frac{\Delta t}{t} = \frac{\Phi_s - \Phi(r)}{c^2} = \left( -\frac{GM}{r_s} \right) - \left( -\frac{GM}{a_i} \left( 1 + \frac{J_2}{2} \right) \right) \approx 5.288 \times 10^{-10}. \quad (6.21)
$$

When multiplying this result by the number of seconds in a day, it is possible to derive the error per day associated with the gravitational frequency shift of GPS satellite clocks

$$
\delta t = 5.288 \times 10^{-10} \times 86400 = 45688.32 \text{ ns d}^{-1}. \quad (6.22)
$$

Finally, we can conclude that if this navigational error was not accounted for, it would increase, resulting in a distance error, at the end of a day, of

$$
\delta x = c |\delta t| = 13.706 \text{ km d}^{-1}. \quad (6.23)
$$

### 6.6. Sagnac Effect

The Sagnac effect accounts for the receiver’s movement while taking into consideration Earth’s rotation – considering a resting receiver on the surface of the Earth, the error introduced by the Sagnac effect derives from the difference between the moment the signal is sent from the satellite and the moment the signal is received by the receiver. This effect is represented in figure 6.1.

![Sagnac Effect Illustration](image)

**Figure 6.1.:** Sagnac effect illustration.
Considering Earth's rotational frame, the navigation equations can be described by

\[ t = t_j + \frac{|r(t) - r_j|}{c} = \frac{R}{c} + \frac{v \times (t - t_j)}{c}. \]  

(6.24)

With \( r(t) = 6378 \text{ km} \) representing the receiver's position at instant \( t \), \( v \) being the velocity of the receiver according to the satellite's signal transmission time – which is small compared to the speed of light and for that reason can be neglected –, and \( R = 26562 \text{ km} \) standing for the distance between sender and receiver at the time that the signal was sent. By simplification, the signal's arrival time for a fixed receptor in an IFR would be

\[ t = \left| \frac{r(t_j) - r_j}{c} \right| = t_j + \frac{R}{c}. \]  

(6.25)

By applying this result, we obtain a time correction of

\[ t = t_j + \sqrt{\frac{R^2}{c^2} + 2 \frac{R \cdot v (t - t_j)}{c^2}} \approx t_j + \frac{R}{c} + \frac{R v}{c^2} \approx \frac{R \cdot v}{c^2}. \]  

(6.26)

Accounting for the fact that Earth's rotation applies a velocity to the receiver of order

\[ v = w \times r(t_j). \]  

(6.27)

Where \( w = 7.292115 \times 10^{-5} \text{ rads}^{-1} \) represents the mean angular velocity of Earth. As a result, the correction of the formula error associated with the Sagnac effect is given by

\[ \delta t_{\text{Sagnac}} = \frac{R \cdot v}{c^2} = \frac{2w \cdot A}{c^2} = 137.454 \text{ ns d}^{-1}. \]

(6.28)

\[ A = \frac{1}{2} r(t_j) \times R = 8.4706 \times 10^{13} \text{ m}^2. \]

\[ \delta x_{\text{Sagnac}} = c |\delta t| = 41.20767 \text{ m d}^{-1}. \]

With the variable \( A \) corresponding to the area covered by the vector – from the rotation axis, where the signal is originally sent, until the position where the signal is received.

### 6.7. Eccentricity Effect

In all the computations until this point it was considered that the satellite orbits were circular. However, this is not true as orbits are not perfectly circular, presenting instead an eccentricity of \( e < 0.02 \). On these type of orbits, and given the radius \( r_j \) is not constant, the velocity \( v_j \) and the gravitational potential \( \Phi_j \) change their value periodically. This eccentricity correction is possible
to apply because of the combination of effects on the satellite clock: the gravitational frequency shifts and the Doppler effect, which varies according to the orbital radius. The formula that describes the correction is given by

\[
\Delta t_r = \frac{2r \cdot v}{c^2}
\]

(6.29)

With \( r \) and \( v \) being, respectively, the position and the velocity at the time the satellite’s signal is sent, considering an ECI. This equation is derived by resorting to the expressions for the Keplerian orbits of the satellites and is, by approximation, equal to

\[
\Delta t_r \approx 2 \cdot \frac{\sqrt{GMa}}{c^2} \cdot e \cdot \sin(E) = 75\text{ns}
\]

(6.30)

Where the variables \( \sqrt{a} \) and \( E \) represent, respectively, the semi-major axis of the orbit and the eccentricity anomaly. At the beginning, the computing power of the satellites was limited and the correction of the anomaly was applied to the transmitted time from the ground station. However, nowadays, the satellite has capabilities to make this correction before the signal is transmitted. This alteration of procedures ensures a diminishing of the error associated with the signal sent.
Chapter 7 – Conclusion and Final Remarks

7.1. Conclusion

The number of applications that use the GPS is almost uncountable. This system is worthy of a more detailed attention from relativity researchers and specialists. Therefore, this may be considered a source of pedagogical examples.

The influence of special and general relativity on the GPS is considered in this work, through time dilation (special relativity) and Einstein gravitational shift (general relativity). In the end, it is shown that these effects are too considerable to be disregarded [57]. The total correction per day, due to relativity (both special and general theories), is approximately $39 \mu s \text{ day}^{-1}$, which corresponds to an imprecision of $12 \text{ km day}^{-1}$.

The understanding of these effects is fundamental and can be correctly predicted. The learning process is possible thought experiments, based on principles, that although may seem complex, are simple and easily explained with the help of the non-Euclidean concept associated with the algebraic example of the hyperbolic plane used in the Minkowski diagrams.

The coordinated relativistic time is an inherent part of the GPS. In fact, the GPS receivers make use of a software that applies the required relativistic corrections. GPS clocks in orbit have been modified to realize coordinate time with more accuracy.

Accordingly, the GPS user, even if unknowingly, has become dependent of the revolutionary concepts of Spacetime introduced by Einstein and graphically interpreted by Minkowski.

7.2. Future Work Perspectives

In the special relativity theory most of the results are possible to obtain by utilizing a hyperbolic plane with a single time dimension alongside a single space dimension. However, in reality, the hyperbolic plane is defined by four dimensions with three space dimensions plus one time dimension. This dissertation leads to future detailed works where the simplification of the spacetime diagrams cannot be applied such as:

- Thomas rotation which allows to calculate the velocity composition of non-colinear frames of reference.
- The utilization of geometric algebra for the study of motion media for non-isotropic media.
- The general relativity theory where the existence of energy or matter curves the spacetime and this spacetime perturbations propagate through gravitational waves.
Additionally, the actual system is based on the Newtonian model which requires, as previously shown, important relativistic corrections. However, a paradigmatic shift is needed in how the GPS is operated. A new relativistic and totally autonomous system that does not need constant correcting should be developed.

Masterminded by Bartolomé Coll, the SYPOR (Système de Positionnement Relativiste) is a theoretical system where relativistic effects coexist [58]. This new idea was conceived with the purpose of improving the precision and the efficiency of the existing navigation systems. The innovation of this project relies on the connection of two different segments of the GPS (control and spatial segments) in a single and unique one. This new configuration will enable the determination of the receiver’s coordinates in relation to the satellite constellation, instead of resorting to the receiver’s coordinates on Earth. This will only be possible if there is exchange of the correspondent time of each satellite clock with their satellite constellation and with the receivers that passively use the system [59]. Finally, the satellite system could also be used for measuring the gravitational field acting on the constellation.
References


[13] Fizeau, H. "The Hypotheses Relating to the Luminous Aether, and an Experiment which Appears to Demonstrate that the Motion of Bodies Alters the Velocity with which Light Propagates itself in their Interior". Philosophical Magazine. 1851, 2: 568–573.


[37] Source of the image:


Appendix A – Vectoral Space of the Hyperbolic Plane

From a vectoral space point of view \( \mathbb{R}^2 \), an event in the hyperbolic plane or a set of coordinates in an Euclidian plane is defined as a vector

\[
\mathbf{r} = t \mathbf{e}_0 + x \mathbf{e}_1. \tag{8.1}
\]

This vector specifies the event \((x, t)\) which is defined by the linear independent canonic base of unitary vectors

\[
\mathcal{B} = \{ \mathbf{e}_0, \mathbf{e}_1 \} \tag{8.2}
\]

Where, by definition,

\[
\mathbf{e}_0 = (t = 1, x = 0), \quad \mathbf{e}_1 = (t = 0, x = 1). \tag{8.3}
\]

With the purpose of defining a relation between both vectors or their length it is also necessary to define a metric. Only with the definition of this metric is it possible to address the notion of quadratic space. The metric is defined through the matrix

\[
\mathcal{G} = \begin{bmatrix} g_{00} & g_{01} \\ g_{10} & g_{11} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_0 \cdot \mathbf{e}_0 & \mathbf{e}_0 \cdot \mathbf{e}_1 \\ \mathbf{e}_1 \cdot \mathbf{e}_0 & \mathbf{e}_1 \cdot \mathbf{e}_1 \end{bmatrix}. \tag{8.4}
\]

The notion of orthogonality (and the notion of length as well) only has meaning when the determination of the internal product between vectors is feasible. Thus, the Euclidian metric is defined through the identity matrix

\[
\mathcal{G} = \begin{bmatrix} \mathbf{e}_0 \cdot \mathbf{e}_0 & \mathbf{e}_0 \cdot \mathbf{e}_1 \\ \mathbf{e}_1 \cdot \mathbf{e}_0 & \mathbf{e}_1 \cdot \mathbf{e}_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{8.5}
\]

However, this defined metric is physically incompatible with the relativity theory hence, it follows that the invariance is defined by

\[
\mathbf{r}^2 = \mathbf{r} \cdot \mathbf{r} = (t \mathbf{e}_0 + x \mathbf{e}_1) \cdot (t \mathbf{e}_0 + x \mathbf{e}_1) = t^2 (\mathbf{e}_0 \cdot \mathbf{e}_0) + tx(\mathbf{e}_0 \cdot \mathbf{e}_1 + \mathbf{e}_1 \cdot \mathbf{e}_0) + x^2 (\mathbf{e}_1 \cdot \mathbf{e}_1). \tag{8.6}
\]

Nevertheless, as the Euclidean metric space is defined as

\[
\begin{bmatrix} \mathbf{e}_0 \cdot \mathbf{e}_1 = 0 \\ \mathbf{e}_0^2 = \mathbf{e}_0 \cdot \mathbf{e}_0 = 1. \\ \mathbf{e}_1^2 = \mathbf{e}_1 \cdot \mathbf{e}_1 = 1 \end{bmatrix} \tag{8.7}
\]
The internal product is symmetric, and the invariance is
\[
\begin{align*}
e_0 \cdot e_i &= e_i \cdot e_0 = 0 \\
r^2 &= t^2(e_0 \cdot e_0) + 2tx(e_0 \cdot e_i) + x^2(e_i \cdot e_i) \\
&= t^2e_0^2 + 2tx(e_0 \cdot e_i) + x^2e_i^2
\end{align*}
\]  
(8.8)

Which implies that the quadratic invariance form is
\[
r^2 = t^2 + x^2 = (t')^2 - (x')^2.
\]  
(8.9)

The Lorentzian metric is different from the Euclidean metric. More specifically, the Lorentzian metric is defined as
\[
\mathcal{G} = \begin{bmatrix} g_{00} & g_{01} \\ g_{10} & g_{11} \end{bmatrix} = \begin{bmatrix} e_0 \cdot e_0 & e_0 \cdot e_i \\ e_i \cdot e_0 & e_i \cdot e_i \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.
\]  
(8.10)

Therefore
\[
\begin{align*}
e_0 \cdot e_i &= 0 \\
e_0^2 &= e_0 \cdot e_0 = 1 \\
e_i^2 &= e_i \cdot e_i = -1
\end{align*}
\]  
(8.11)

Which implies that the quadratic invariance form is
\[
r^2 = t^2 + x^2 = (t')^2 - (x')^2 \iff \text{special Lorentzian hyperbolic relativity metric plane).
\]  
(8.12)

As previously seen, with this result it is proven that the interval invariance is correct. The identity matrix – Euclidean metric – does not have a physical existence in the spacetime context. The hyperbolic plane corresponds to the relativistic physics. In other words, it physically defines the Lorentzian metric.
Appendix B – Twin Paradox Revisited

A clock is a mechanism responsible for measuring time intervals while a measuring rod is a device responsible for measuring space intervals. In order to measure the distance \( \tau \) of type time between events \( A \) and \( B \), through the world line which is also called own time, it is necessary to resort to the expression

\[
\tau_{AB} = \int_{\tau_A}^{\tau_B} \sqrt{1 - \left(\frac{dx}{dt}\right)^2} \, dt.
\]  

(9.1)

This integral not only depends on events \( A \) and \( B \) but also on the specific world line that unites the two points. Figure B.1. depicts different world lines.

Unlike the absolute time from Newtonian mechanics, the relativistic time is given by the distance measured over the world line that unites both events. If a different world line connecting the same two events was considered, the own time measured would be different than the previously obtained.

Considering again the example given on figure B.1., the own time of Bob’s world line is given by

\[
T = \int_{\tau_0}^{\tau_B} dt = \frac{T}{2} + \frac{T}{2} = T.
\]  

(9.2)

While the own time from Alice’s world line is given by
\[ T' = \int_{t_0}^{t_f} \sqrt{1 - \beta^2} \, dt + \int_{t_0}^{t_f} \sqrt{1 - \beta'^2} \, dt = \frac{1}{\gamma} \left( \int_{t_0}^{t_f} \, dt + \int_{t_0}^{t_f} \, dt \right) = \frac{1}{\gamma} \left( \frac{T + T'}{2} \right) = \frac{T}{\gamma}. \] (9.3)

The conclusion, which is very important for the twin paradox, is that

\[ 0 < |\beta| < 1 \Rightarrow \gamma = \frac{1}{\sqrt{1 - \beta^2}} > 1 \Rightarrow T' < T. \] (9.4)

In order to avoid the abrupt acceleration change considered previously, we will consider Alice's world line as a parabolic section, which is represented in figure B.2.

In this figure the world line of Alice, previously represented by the set of events \( O \rightarrow A \rightarrow B \), is represented by the new set of events \( O \rightarrow A_1 \rightarrow A_2 \rightarrow B \), in which section \( A_1 \rightarrow A_2 \) represents a finite interval, different from zero. Furthermore, the acceleration affecting Alice has a finite value.

Bob's world line continues to be described by the vertical line \( O \rightarrow B \), to which corresponds a total own time \( T \) represented by a unique IFR \( S(x, t) \) – while Alice's IFR \( S'(\chi, \tau) \) – is not described by a single one or two IFRs, as in the twin paradox section. Alice's world line is, by definition, \( \chi = 0 \). This world line is constituted by three different sections of which two are similar, being described by the set of events \( O \rightarrow A_1 \) and \( A_2 \rightarrow B \), respectively, and each being represented by a single IFR. The third section is described by \( A_1 \rightarrow A_2 \) and an infinity of IFRs, one for each position, with each one representing one of the positions that coincide with \( S'(x', t') \).
This section represents, in the hyperbolic plane \( \mathbb{H}^{1,1} \), a parabola arc and, considering \( t_0 = \frac{T}{2} \), Alice’s world line is defined in this section as

\[
t_1 \leq t \leq t_2 \implies x(t) = -\frac{1}{2} \alpha (t - t_0)^2 + x_0.
\]

(9.5)

Where \( x_0 = x(t_0) = x(T/2) \) and \( t \) represents Bob’s own time. The velocity in this section is

\[
\frac{dx}{dt} = -\alpha (t - t_0),
\]

(9.6)

and the acceleration \( \alpha > 0 \) – constant – is represented by

\[
\frac{d^2x}{dt^2} = -\alpha.
\]

(9.7)

The instantaneous velocity of Alice’s world line in the different sections is given by

\[
\begin{align*}
\frac{dx}{dt} \bigg|_{t=t_1} &= -\alpha(t_1 - t_0) = \beta \\
\frac{dx}{dt} \bigg|_{t=t_0} &= 0 \\
\frac{dx}{dt} \bigg|_{t=t_2} &= -\alpha(t_2 - t_0) = -\beta
\end{align*}
\]

(9.8)

Consequently, considering \( A_1(x_1, t_1) \), \( A_2(x_2, t_2) \) and \( x_1 = x_2 = L - \delta \)

\[
\begin{align*}
t_0 &= \frac{L}{\beta} \\
t_1 &= -\frac{\beta}{2} + \frac{T}{2} = \frac{L - \delta}{\beta} \\
t_2 &= \frac{L + \delta}{\beta}
\end{align*}
\]

(9.9)

Additionally,

\[
x_1 = L - \delta = -\frac{1}{2} \alpha (t_1 - t_0)^2 + x_0 \implies x_0 = L - \frac{\delta}{2}.
\]

(9.10)

Where \( x_0 \) represents the position where Alice turns around and starts to move in the opposite direction – position correspondent to half of the total trip duration.
Consequently, Alice’s world line from Bob’s point of view is described by equations

\[ x(t) = \begin{cases} 
\beta t, & 0 \leq t \leq t_1 \\
\left( L - \frac{\delta}{2} \right) - \frac{1}{2} \frac{\beta^2}{\delta} \left( t - \frac{L}{\beta} \right)^2, & t_1 \leq t \leq t_2 \\
-\beta(t-T), & t_2 \leq t \leq T
\end{cases} \]  \hspace{1cm} (9.11)

In order to infer Alice’s own time \( \tau = \tau(t) \), in her own referential where her world line is characterized by the equation \( \chi = 0 \), it is necessary to utilize the previous expression

\[ d\tau = dt \sqrt{1 - \left( \frac{dx}{dt} \right)^2} = dt \sqrt{1 - \alpha^2 \left( t - \frac{T}{2} \right)^2}, \quad t_1 \leq t \leq t_2 \]  \hspace{1cm} (9.12)

Since

\[ \int \sqrt{1 - \alpha^2 \left( t - \frac{T}{2} \right)^2} \, dt = \frac{1}{2} \left( t - \frac{T}{2} \right) \sqrt{1 - \alpha^2 \left( t - \frac{T}{2} \right)^2} + \frac{1}{2\alpha} \sin^{-1} \left[ \alpha \left( t - \frac{T}{2} \right) \right], \]  \hspace{1cm} (9.13)

It is true that

\[ \tau(t) = \tau_0 + \frac{1}{2} \left( t - \frac{T}{2} \right) \sqrt{1 - \alpha^2 \left( t - \frac{T}{2} \right)^2} + \frac{1}{2\alpha} \sin^{-1} \left[ \alpha \left( t - \frac{T}{2} \right) \right], \quad t_1 \leq t \leq t_2 \]  \hspace{1cm} (9.14)

Where \( \tau_0 = \tau(t_0) = \tau(T/2) \) and

\[ \tau(t_i) = \frac{L - \delta}{\gamma \beta} \]
\[ = \tau_0 + \frac{1}{2} \left( t_i - \frac{T}{2} \right) \sqrt{1 - \alpha^2 \left( t_i - \frac{T}{2} \right)^2} + \frac{1}{2\alpha} \sin^{-1} \left[ \alpha \left( t_i - \frac{T}{2} \right) \right] \]  \hspace{1cm} (9.15)
\[ = \tau_0 - \frac{\delta}{2\beta} \sqrt{1 - \beta^2 - \frac{\delta}{2\beta^2} \sin^{-1}(\beta)} \]

As a result,

\[ \tau_0 = \frac{L - \delta}{\gamma \beta} + \frac{\delta}{2\beta^2} \left[ \frac{\beta}{\gamma} + \sin^{-1}(\beta) \right] = \frac{T'}{2} + \frac{\delta}{2\beta^2} \left[ \sin^{-1}(\beta) - \frac{\beta}{\gamma} \right] \]  \hspace{1cm} (9.16)

And, in the end, Alice’s own time is given by

\[ \tau(t) = \begin{cases} 
\sqrt{1 - \beta^2}, & 0 \leq t \leq t_1 \\
\tau_0 + \frac{1}{2} \left( t - \frac{T}{2} \right) \sqrt{1 - \alpha^2 \left( t - \frac{T}{2} \right)^2} + \frac{1}{2\alpha} \sin^{-1} \left[ \alpha \left( t - \frac{T}{2} \right) \right], & t_1 \leq t \leq t_2 \\
\tau(T) + (t-T)\sqrt{1 - \beta^2}, & t_2 \leq t \leq T
\end{cases} \]  \hspace{1cm} (9.17)
With $\tau(T)$ representing the trip’s total time from Alice’s point of view. Thus,

$$
\begin{align*}
\tau_1 &= \tau(t_1) = \frac{t_1}{\gamma} = t_1 \sqrt{1 - \beta^2} = \frac{T'}{2} - \frac{\delta}{\beta} \sqrt{1 - \beta^2}, \\
\tau_2 &= \tau(t_2) = \frac{T'}{2} + \frac{\delta}{\beta^2} \sin^{-1}(\beta)
\end{align*}
$$

(9.18)

Which in turn implies that

$$\Delta\tau = \tau_2 - \tau_1 = \frac{2\delta}{\beta^2} \sin^{-1}(\beta)$$

(9.19)

According to this expression, when $\delta = 0$, the result is $\Delta\tau = 0$. However, the own time interval $\Delta\tau$ is equal to $\tau(T)$ when $\delta = L$. Moreover, it is also important to notice that

$$\Delta\tau = \tau_2 - \tau_1 = \frac{2\delta}{\beta^2} \sin^{-1}(\beta) = \frac{\Delta t}{\gamma} = \frac{t_2 - t_1}{\gamma} = \frac{2\delta}{\beta} \sqrt{1 - \beta^2}.$$ 

(9.20)

Hence,

$$\tau(T) + (t_2 - T)\sqrt{1 - \beta^2} = \tau_0 + \frac{1}{2}(t_2 - T)^2 \sqrt{1 - \left[ \frac{\alpha(t_2 - T)}{\alpha} \right]^2} + \frac{1}{2\alpha} \sin^{-1} \left[ \alpha \left( t_2 - \frac{T}{2} \right) \right]$$

(9.21)

Which results in

$$\tau(T) = 2\tau_0 = T' + \frac{\delta}{\beta^2} \left[ \sin^{-1}(\beta) - \frac{\beta}{\gamma} \right].$$

(9.22)

In this case, when $\delta = 0$ the result is, as previously obtained, $\tau(T) = T' = (2L) / (\gamma \beta)$. The $\delta$ parameter characterizes the parabola family, where $0 \leq \delta \leq L$. When $\delta \neq 0$, a parabolic movement is obtained $\tau(T) > T'$, which represents a larger own time interval for Alice's world line compared to the simplified case presented previously presented.
Appendix C – Clock Synchronization

In order to compare two clocks of two different (inertial) observers in relative uniform motion, it is necessary to make use of a third observer – the referee –, who is always in the middle of those two clocks. In figure C.1, the referee clock \( \mathcal{R} \) is placed between clocks \( \mathcal{C} \) and \( \mathcal{D} \).

![Figure C.1. Introduction of the referee clock in order to synchronize both clocks.](image)

It can be established that the two initial clocks with identical mechanisms, clock \( \mathcal{C} \), which sets the time \( \tau' \) at event \( A' \), and the clock \( \mathcal{D} \), which sets the time \( \tau \) at the same event, are the same: \( \tau' = \tau \). As a result, two electromagnetic signals emitted at event \( S \in \mathcal{R} \) – one in \( \mathcal{C} \) ‘s direction and reflected back to \( \mathcal{R} \) in event \( A' \) and another in \( \mathcal{D} \) ‘s direction and reflected back to \( \mathcal{R} \) in event \( A \) – are simultaneously received by the referee clock, at the same event \( r \in \mathcal{R} \). In other words, as the speed of light is an universal constant \( c = 1 \), a new metric is defined through which two lengths from two distinct observers are defined as equal, by the referee’s clock

\[
|OA| = |OA'| = \tau_A
\]  

(10.1)

From the referee’s point of view, events \( A \) and \( A' \) are at the same spacetime interval. Moreover, the equilocs \( \mathcal{C} \) and \( \mathcal{D} \) are moving away from \( \mathcal{R} \) with the same velocity and opposite directions.

In the Euclidean plane, this equality would not be valid hence, the circumference centered in \( O \) with radius \( OA \) does not pass on event \( A' \).

As explained before, time is not absolute but relative. Therefore, Newtonian time is an illegitimate abstraction – an illusion. In fact, time is a route-dependent quantity and it must be treated as a...
coordinate. In order to coherently handle time, a correct synchronization procedure must be applied.

According to Alice, Bob receives the signal at $t = t_0$. When Bob’s clock reads off $t' = \tau$, Alice’s clock reads off $t = t_0 = \gamma \tau > \tau$.

Meanwhile, from Bob’s point of view, he sends, at $t' = t_-, an electromagnetic signal to Alice, who receives it at $t = \tau$ and immediately sends it back to Bob, and Bob receives the signal back from Alice, at $t' = t_+, according to Bob, Alice receives his signal at $t' = t_0$. When Alice’s clock reads off $t = \tau$ Bob’s clock reads off $t' = t_0 = \gamma \tau > \tau$.

Also, in this figure, a calibration hyperbole, which passes by the time coordinates $\tau$ of both IFRs, is introduced. Another hyperbole could also be drawn connecting both $t_+$ or connecting both $t_-$. This hyperbolic plane is defined through a Lorentzian metric which comes from the new mathematical result that expresses the special relativity physics.
Appendix D – Non-Relativistic Effects

There are other corrections that are required to precisely correct the calculation for the position of the GPS receiver. Part of these correction derive from the fact that the satellites signals propagate through the atmosphere and so their velocity is not exactly the speed of the light, in vacuum \( c \). Besides, in function of the ionization and humidity of the different atmospheric layers, this velocity varies with space and time. Other corrections become necessary for the elimination of the electromagnetic signals’ reflection effects originated by the GPS signal, mainly in populated areas with high skyscrapers. The precision of the receiver depends on a complex interaction of several factors:

- Clock error: it is required to know precisely the dates of transmission and reception, in that case satellites can transmit their location and the GPS can receive it, in a synchronized time.

- Satellite orbit error: this error is caused by the imperfect modulation of the dynamics of the satellites, in other words, this error results from the divergence between the real position value of the satellite orbit and the value achieved through the modulation.

- Satellites geometry error: known as GDOP (Geometric Dilution of Precision) describes the importance of the precision of the measurements obtained through the satellites constellation configuration. A high value associated with the GDOP implies that the visible satellites are close between each other, which corresponds to a weak geometry.

There are several methods for implementing the desired corrections, as using two different frequencies for the GPS signal to react to the ionization in a different way, allowing to determine the delay due to this phenomenon.
Appendix E – Arithmetic and Geometric Mean.

The purpose of this annex is to prove that the arithmetic mean is at least equal to the geometric mean, as shown in section 4.1. Bondi’s Factor and Time Dilation.

Resorting to the previous equations of Bondi’s factor section

\[
\begin{align*}
t_A &= \frac{1}{2}(t_+ + t_-) \\
\tau_A &= \sqrt{t_+ t_-} = t_A \tau = t_A \sqrt{1 - \beta^2}.
\end{align*}
\]  

(12.1)

Geometrically, it is possible to draw a semi-circumference, as depicted in figure E.1.

Figure E.1.: Graphical representation of how to obtain the geometric mean of two lengths.

In this figure, it is possible to define

\[
\begin{align*}
t_+ &= t_A + x_A \\
t_- &= t_A - x_A.
\end{align*}
\]  

(12.2)
And also

$$\beta = \frac{x_A}{t_A}.$$  \quad (12.3)

Applying the Pythagoras’ Theorem to the marked triangle in the figure, the result is

$$t_A^2 = t_A^2 + x_A^2 \Leftrightarrow t_A = \sqrt{t_A^2 - x_A^2} = t_A s = t_A \sqrt{1 - \beta^2}.$$  \quad (12.4)

Only if $\beta = 0,$ $\tau_A = t_A,$ while $\tau_A < t_A$ if $0 < \beta < 1.$
Appendix F – The Equivalence Between Mass and Energy

As it was previously explained, the speed of light is a cosmic limit – a finite value. The objective of this appendix is to prove that photons are the massless elementary particles of light.

Generally speaking any particle with a mass \( m \), that transports energy, is associated with an intrinsic (or proper) energy – \( E_0 \). This intrinsic energy can be defined as

\[
E_0 = mc^2. \tag{13.1}
\]

It can also be defined the total energy of a particle, \( E \), such that

\[
E = \gamma E_0, \quad \text{with} \quad \gamma = \frac{1}{\sqrt{1 - \beta^2}}. \tag{13.2}
\]

The relation between mass and energy of a particle is described through the relativistic formula

\[
E^2 = E_0^2 + c^2 p^2, \quad p = \gamma m v, \tag{13.3}
\]

where \( p \) is the linear momentum. The velocity of the particle, for a given frame of reference, in both classical and relativistic physics, is always given by

\[
v = \frac{dE}{dp}. \tag{13.4}
\]

Consequently, deriving both members of the equation 13.3, one obtains that

\[
2E \frac{dE}{dp} = 2c^2 p \iff 2E v = 2c^2 p, \tag{13.5}
\]

then

\[
v = \frac{c^2 p}{E}. \tag{13.6}
\]

In general, one has

\[
\beta = \frac{v}{c} \quad \Rightarrow \quad \beta = \frac{c p}{E} = \cos(\theta). \tag{13.7}
\]
Therefore, on the one hand if $\cos(\theta) \to 0$, the linear momentum vanishes and $\gamma = \gamma_0$, whereas, on the other hand, when $\cos(\theta) \to 1$, (the massless case) where $\theta \to \frac{\pi}{2}$ and $\gamma_0 \to 0$, thereby leading to $m \to 0$. Accordingly, for a photon, where $v = c$, one gets (figure F.1. depicts this example)

$$c = \frac{\gamma^2 p}{\gamma} \Rightarrow \gamma = \gamma_0 \Rightarrow m = 0$$

(13.8)

Figure F.1.: Relation between mass and energy of a particle.