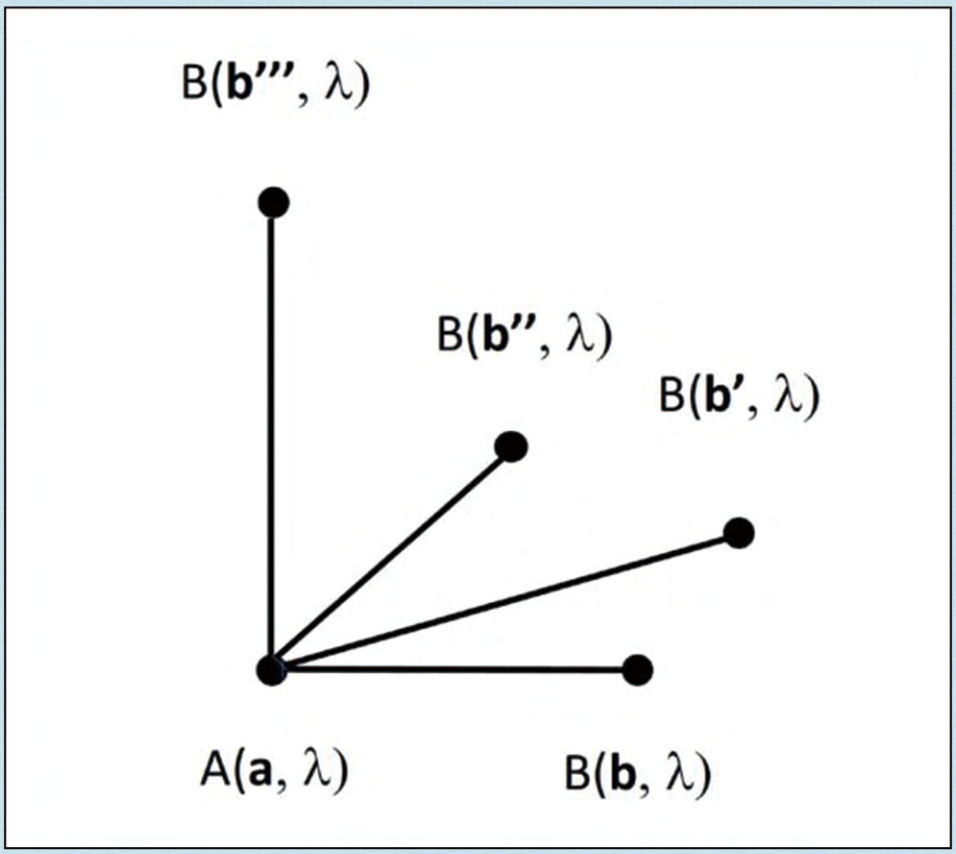
ISSN: 2153-1196 Volume 12, Number 9, July 2021

Scientific Research Publishing

Journal of Modern Physics



ISSN: 2153-1196

https://www.scirp.org/journal/jmp

Journal Editorial Board

ISSN: 2153-1196 (Print) ISSN: 2153-120X (Online)
https://www.scirp.org/journal/jmp

Editor-in-Chief

Editorial Board

Journal of Modern Physics, 2021, 12, 1183-1399
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Table of Contents

Volume 12 Number 9 July 2021

The figure on the front cover is from the article published in Journal of Modern Physics, 2021, Vol. 12, No. 9, pp. 1219-1236 by Karl Hess.

Journal of Modern Physics (JMP)
Journal Information

SUBSCRIPTIONS

The *Journal of Modern Physics* (Online at Scientific Research Publishing, https://www.scirp.org/) is published monthly by Scientific Research Publishing, Inc., USA.

Subscription rates:
Print: $89 per issue.
To subscribe, please contact Journals Subscriptions Department, E-mail: sub@scirp.org

SERVICES

Advertisements
Advertisement Sales Department, E-mail: service@scirp.org

Reprints (minimum quantity 100 copies)
Reprints Co-ordinator, Scientific Research Publishing, Inc., USA.
E-mail: sub@scirp.org

COPYRIGHT

PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:
E-mail: jmp@scirp.org

Journal of Modern Physics, 2021, 12, 1183-1189
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

The Delayed Quantum Eraser Experiment Explained Classically

Declan Traill

Independent Researcher, Melbourne, Australia
Email: declan@netspace.net.au

How to cite this paper: Traill, D. (2021) The Delayed Quantum Eraser Experiment Explained Classically. *Journal of Modern Physics*, **12**, 1183-1189.
https://doi.org/10.4236/jmp.2021.129072

Received: June 3, 2021
Accepted: July 2, 2021
Published: July 5, 2021

Abstract

This paper discusses the well-known delayed choice Quantum Eraser experiment performed by Kim *et al.* in 2000 and analyzes it from a Classical Physics perspective. I have included a diagram of the setup used in the experiment. I show that the result of the experiment can be explained by Classical Physics and does not require "Spooky action at a distance" due to entangled particles, as Einstein famously once put it, nor events modifying the past due to the delayed choice aspect of the experiment.

Keywords

Delayed, Choice, Quantum, Eraser, Photon, Laser, EPR, Entangled, Classical, Physics, Polarizing, Beam, Splitter, Coincidence, SPDC

1. Introduction

This paper is the second I have written on analyses of key experiments in Quantum Optics that claim to employ entangled photons in the experimental setup, resulting in correlations in photon detections which are claimed to demonstrate the Quantum nature of the behaviour of the light in the experiments. It is claimed by these experimental results that no Local, Real (Classical Physics) explanation for the results is possible, and the design of these experimental setups is done in such a way to supposedly close any loopholes that would allow for a Classical explanation.

In previous research I have conducted [1], I examined modern versions of the original Einstein-Podolsky-Rosen (EPR) thought experiment of Quantum Mechanics. The EPR experiment is thought to demonstrate that photons of light can become entangled when they are created by a process such as Spontaneous Parametric Down Conversion (SPDC) and then can travel on to different, spa-

tially separated, regions within the laboratory and then, upon measurement, display correlations in their respective polarizations despite their separation being such that a light signal cannot communicate between these two locations within the duration of each measurement.

In my examination and computer modelling of this experiment, I found that the correlation curve obtained in the experiment can be completely explained by a detection bias at certain detector polarization angles such that as the detector is rotated through 180 degrees, a cosine shaped correlation curve is obtained, rather than the linear curve which was expected from a Classical model. The Quantum Mechanical prediction, based on there being entanglement between the photons, is also a cosine curve thus giving confidence in the Quantum Mechanical interpretation. The experiment I examined used a Steering Inequality to assess the level of correlation between the photons in the experiment. Such a Steering Inequality is supposed to be a better method of assessing the correlation than the Bell or CHSH inequalities that are usually used and is claimed to close the detection-loophole (the detection bias that I modelled in my analysis). However, in my Classical model, I found not only did the model result in the same cosine curve that Quantum Mechanics predicts, but the Steering Inequality is comprehensively violated too. My model found a Steering Inequality of 1.622, yet a Classical model in only supposed to be able to give a Steering Inequality of 1 or less.

The experiment performed by Kim *et al.* [2] is a variation on this EPR type of experiment. Their experiment is called a Delayed Choice Quantum Eraser experiment and was devised to deliberately make one of the photon path lengths longer than the other such that one photon's polarization can be known before the other is measured. The rationale is that it is thought that by measuring the 2nd photon, the "which path" information of the photon becomes known and the wave-function of the entangled photon pair collapses, so by delaying its measurement until after the 1st photon has either interfered (or not) it should not be able to influence the 1st photon's measurement without signaling backward in time. Therefore, so the reasoning goes, if there is a correlation between the path taken by the 2nd photon and the pattern of the 1st photon (interfering or otherwise) then that indicates Quantum weirdness that cannot be explained with a Real, Local Physical model. Having successfully modelled the EPR experiment as a completely Classical process, I had my doubts as to whether there was really any Quantum entanglement required, or indeed happening at all; so, I thought it might be useful to turn my attention to the Delayed Choice Quantum Eraser experiment too and see if a rational Classical explanation might also exist.

2. Aim

The aim of this paper is to demonstrate that the usual assumptions made in the analysis of this experiment are incorrect and that a simple, real and local explanation exists that can fully explain the results of the experiment without the need to resort to faster-than-light nor backwards-in-time communication.

3. Method

The main features of the Kim *et al.* [2] setup shown in Figure 1 are:

- The pulsed input laser beam striking a double-slit barrier.
- The photon output from the double-slit enters a BBO crystal where SPDC occurs.
- The "entangled" pairs of photons from the SPDC process are split down two paths by a Glan-Thompson beam splitter.
- The two signal photons (one from each pair) are converged with a lens onto detector **D0** (which can be moved in the x-axis direction to observe the interference fringes)
- The idler photons (one from each pair) are further separated by a prism.
- Each idler photon passes through a polarizing beam splitter **BSa** or **BSb**.
- The idler photons then get recorded at either **D3** or **D4**, or pass through a further beam splitter **BSc** and then off mirrors **Ma** or **Mb** to detectors **D1** or **D2**.
- Detection results are recorded and correlated by a coincidence counter.

The main feature of the experimental setup is the initial photon from the pulsed laser passing through a double-slit, thus presenting an unknown path aspect to the situation, followed by the generation of entangled photons—one pair from each photon path through the double-slit. Then one photon from each pair (the signal photons) passes down a short path and are brought together and allowed to interfere, whilst the other photon from each pair (the idler photons) travels a longer distance and passes through several beam splitters before reaching 4 possible detectors (3 of the 4 are possible detectors for each idler photon). Due to the path length difference, the result of the signal photons' interference is

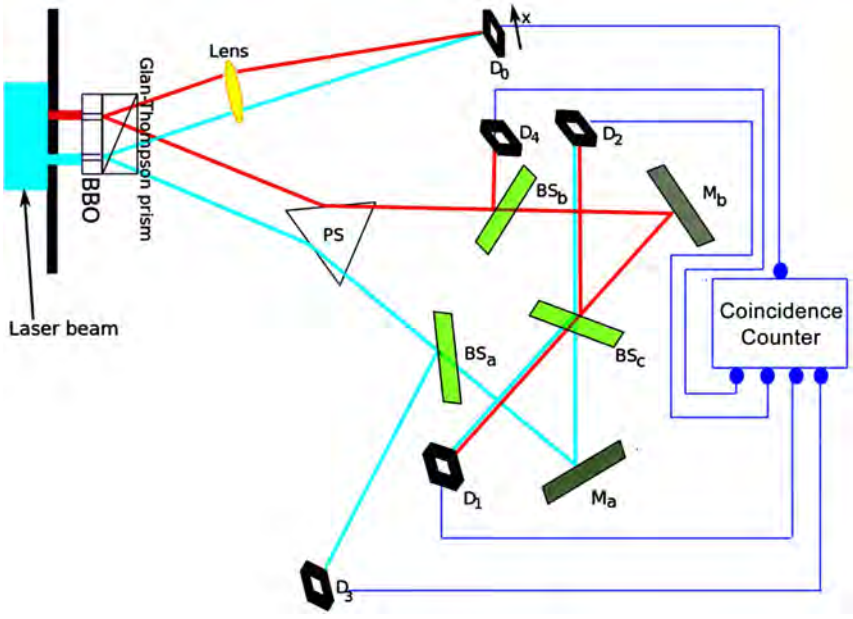


Figure 1. The experimental setup used by Kim *et al.* [2]. Attribution for this image is given in Ref. [3].

known before the path taken result (indicated by the detections of the idler photons) is known. The reasoning is that if knowing the path of the idler photons can collapse the Quantum Mechanical wave-function and affect the interference result of the signal photons, then by delaying the knowledge of the path taken until after the interference result is known any strange correlation in the results of all the photons in the experiment might be prevented. The Figures in the Kim *et al.* paper [2] do display a correlation that is thought to indicate entanglement between photons.

4. Analysis of Results

In order to analyze this experiment from a new, objective perspective I tried to imagine exactly what the light waves were doing at every point through the experimental apparatus. I aimed to keep Classical Physics front of mind and avoid having to resort to any non-real, non-local explanation for the observed results. One key feature of the experiment is that it used brief pulses of laser light such that single photons were traversing the experimental setup during the data collection. This is one of the features that make the results so confusing to understand, especially if one thinks of photons as discrete point particles.

The first thing one must realize in order to make sense of the experiment is that photons are not discrete particles but are actually continuous Electromagnetic waves. Sure, atoms emit and absorb Electromagnetic energy in fixed quantum amounts due to the changes in electron orbital energy levels, which gives rise to the wave packets of Electromagnetic energy that we refer to as Photons, but there is nothing binding this wave energy into discrete particles. Indeed, the wave packets that are referred to as photons may be meters long (depending on the frequency of the light) and are not point particles as is often assumed. Thus, when "photons" of light pass through a double-slit or a beam splitter the wave energy is free to be split up into different sub-quantum amounts down different paths.

We can see that this in fact must be the case when we observe single photons passing through Mach-Zehnder Interferometers [4] [5]. When single photons of light pass through such interferometers, interference patterns are still observed in the output when one of the path lengths in the Interferometer is altered to make it shorter or longer that the other path. This could only happen if Electromagnetic energy was passing down both paths simultaneously.

Thus, in the case of the experiment conducted by Kim *et al.* [2] (also see **Figure 1** above), the light photons will arrive at the double slit at the start of the experiment and, depending on the exact position and polarization of the photon, will either pass entirely (or almost entirely) through one of the two slits, or will diffract through both slits (with varying proportions of energy passing through each). When the latter occurs, sub-quantum amounts of Electromagnetic energy emerge from both slits A and B and proceed on to the BBO crystal. The process of SPDC which occurs in this crystal to convert 351.1 nm light into two beams of 702.2 nm light works based on the conservation of energy and momentum [6],

so the same laws will apply to the sub-quantum light waves that emerge from the double-slit and will result in sub-quantum 702.2 nm photons being generated in the BBO crystal.

After the BBO crystal, the sub-quantum 702.2 nm photons will continue through the apparatus to the beam splitters. The beam splitters are polarizing beam splitters that separate out orthogonal polarization components down their two output channels (causing either transmission or reflection). So, the photons will either transmit or reflect at these beam splitters depending on their polarization axes relative to the polarization axes of the beam splitters. It should be noted that the detection of light quanta by the detectors is a probability driven event that is based on the probability amplitude of the light signal. Thus, even though a sub-quantum amount of light is received at a detector, the probability of a detection event occurring at the detector increases with the amplitude of the received light signal. Thus, even sub-quantum amounts of light can cause detectors to trigger and generate "click" events.

There seems to be the mistaken assumption in the usual analysis of this experiment that 50:50 beam-splitters will separate each photon down each output channel with a 0.5 probability down either channel. This is NOT correct (unless non-polarizing beam splitters were used in the experiment). With unpolarized light beams (consisting of many photons with random polarization directions) there will be a 50:50 split in beam intensity down each channel, but for individual photons with specific polarizations the output channel is determined by the relative orientations of the polarization axes of the light and the beam-splitter. The choice of which of the two paths taken is not a 50:50 chance, but would range between 100:0 and 0:100 depending on the relative polarization of the photons with respect to the optical axes of the polarizing beam splitters.

As the SPDC process [6] used in the experiment produces photons with orthogonal optical axes (a type II SPDC process) the detections observed in the experiment are simply due to the two photon pairs produced having the same or different polarizations down the paths to the detectors.

So, if two orthogonal signal photons are produced down the path to detector **D0** then there will be no interference at that detector (as orthogonal light beams of photons do not interfere). The orientation of the optical axes of the other two photons, the 2nd photons (known as the idler photons) from the two pairs produced, is such that one photon will reflect and the other will transmit at the beam splitters **BSa** and **BSb** as these beam splitters are polarizing and will separate out different polarizations down each of their output paths.

Hence the observed detection curve R_{03} *or* R_{04} in **Figure 2**, showing no interference pattern, is obtained, (in the original Kim *et al.* paper [2] it is Figure 5 on page 4).

Then, if two photons (one from each pair produced), each having the same optical axis orientation, are sent down the path to detector **D0**, they will interfere at that detector. Assuming the optical axis of the other two photons (orthogonal to the first two) is aligned with the optical axis for transmission in **BSa** and **BSb**,

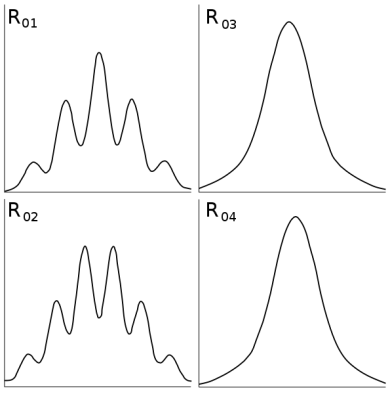


Figure 2. The experimental correlation results obtained by Kim *et al.* [2]. Attribution for this image is given in Ref [7].

they will always transmit through the beam splitters **BSa** and **BSb** and go on to be recorded at detectors **D1** or **D2**.

Thus, the detection results R_{01} *and* R_{02} in **Figure 2**, showing a correlation curve that has an interference pattern, are observed (as shown in Figure 3 & Figure 4 of the original Kim *et al.* paper [2]).

The raised center in the observed interference pattern of the joint-detection graphs of R_{01} *and* R_{02} in **Figure 2** is due to some of the "click" events at detectors **D1** and **D2** being due to the single idler photon from an orthogonal pair being transmitted at **BSa** or **BSb** (when the other two signal photons going to detector **D0** are also orthogonal to one another, so don't generate any interference pattern at detector **D0**). The result of such detections is evident in the two figures graphs R_{01} *and* R_{02} in **Figure 2**, as the interference fringe peaks vary in height (and from a different base level) following the shape of the curve seen in graph R_{03} in **Figure 2**. There would also be some contributions to the non-interfering photons detections (giving rise to this non-interfering component of the detection curve) from single photons that pass entirely through one of the slits at the start of the experiment and then go on to detectors **D1**, **D2**, **D3** or **D4**.

5. Conclusion

Therefore, we can see that once the nature of the SPDC process (generating orthogonally polarized photons) and the nature of how polarizing beam splitters work are taken into consideration, the results of the experiment are consistent with a Classical Physics (Local and Real) interpretation and there is no mystery requiring the notion of entangled pairs, nor faster than light communication. There is also a classical explanation for the original EPR experiment [1] that can explain the correlation in experimental results without requiring any notion of "entanglement" of the photons (or other particles such as electrons [8]).

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Traill, D.A. (2018) A Fundamental Misunderstanding (fqXi). https://www.researchgate.net/publication/327260788_A_Fundamental_Misunderstanding

[2] Kim, Y.-H., Yu, R., Kulik, S.P., Shih, Y.H. and Scully, M. (2000) *Physical Review Letters*, **84**, 1-5. https://doi.org/10.1103/PhysRevLett.84.1

[3] Noestmm, PNGHUT. https://pnghut.com/png/CW8jEALu4U/quantum-eraser-experiment-double-slit-delayed-choice-wave-particle-duality-wheelers-wave-transparent-png

[4] ScienceDirect, "Mach-Zehnder Interferometers" (2020) Elsevier B.V. https://www.sciencedirect.com/topics/physics-and-astronomy/mach-zehnder-interferometers

[5] Khokhlov, D.L. (2019) *Quantum Information Review*, **7**, 7-10. http://www.naturalspublishing.com/files/published/f721wno893296q.pdf

[6] Wikipedia, "Spontaneous Parametric Down-Conversion". https://en.wikipedia.org/wiki/Spontaneous_parametric_down-conversion

[7] Delayed Choice Quantum Eraser Graphs.svg, 2014, Wikimedia, *Stigmatella aurantiaca* at English Wikipedia, CC BY-SA 3.0. https://creativecommons.org/licenses/by-sa/3.0 https://commons.wikimedia.org/wiki/File:Delayed_Choice_Quantum_Eraser_Graphs.svg

[8] Jackson, P.A. (2021) The Measurement Problem, an Ontological Solution. https://www.researchgate.net/publication/352056822_The_Measurement_Problem_an_Ontological_Solution

Journal of Modern Physics, 2021, 12, 1190-1209
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Encoding Energy-Density as Geometry

Edwin Eugene Klingman

Cybernetic Micro Systems, Inc., San Gregorio, CA, USA
Email: klingman@geneman.com

How to cite this paper: Klingman, E.E. (2021) Encoding Energy-Density as Geometry. *Journal of Modern Physics*, **12**, 1190-1209.
https://doi.org/10.4236/jmp.2021.129073

Received: June 1, 2021
Accepted: July 5, 2021
Published: July 8, 2021

Abstract

Physicists possess an intuitive awareness of Euclidian space and time and Galilean transformation, and are then challenged with Minkowski space-time and Einstein's curved space-time. Relativistic experiments support the "time-dilation" interpretation and others support "curved space-time" interpretation. In this, and related work, we investigate the key issues in terms of the intuitive space-time frame. In particular, we provide alternative approaches to explain "time dilation" and to explain the energy density for gravity systems. We approach the latter problem from an information perspective.

Keywords

Curved Space-Time, Gravitational Energy, Stress-Energy Tensor, Encoding Information in Geometry, Time Dilation, Equivalence Principle, Minkowski Relation, Schwarzschild Metric, Linearized General Relativity, Cosmology

1. Introduction

Every general relativity text contains some discussion of the fact that the energy of gravitating systems cannot be formulated in curved space-time. A recent Centennial paper [1] by Chen, *et al.* begins by stating:

"*How to give a meaningful description of energy-momentum for gravitating systems… has been an outstanding fundamental issue since Einstein began his search for gravity theory.*"

The problem of how best to describe the energy-momentum and angular momentum in gravitating system suffers from the fact that

"*It is known that these quantities cannot be given a local density.*"

A century-long failure to solve the fundamental problem is indicative of confusion; the goal of this work is to provide necessary clarification of the issue. The plan for this paper is as follows:

Section 2 introduces coordinate systems and their meaning, in the context of

invariance. We review the development of energy-time physics in absolute space and time, based on treating the Minkowski relation as an *invariance*, not as space-time.

Section 3 provides context for the geometric algebra approach applied to the problem of "clock slowdown" in moving systems. Two interpretations of the same *geometric algebra construct* yield two or more conceptions of physics based on corresponding ontological assumptions.

Section 4 reviews issues of coordinates in theories that formulate gravity as curved space-time. These issues are at the root of the problem of gravitational energy representation in curved space.

Section 5 presents the essence of the "never-solved" problem of general relativity; the fact that a century of effort has failed to formulate a meaningful description of energy-momentum for gravitating systems.

Section 6 introduces Einstein's equations and focuses on their inherent insolvability except in *One-body solutions* for essentially trivial cases.

Section 7 begins the central effort in this work: the encoding of gravitational energy density in geometry. It formulates the density for the primordial field and one body of mass M located in the field. The procedure for encoding the energy density in the geometry is prescribed.

Section 8 analyzes the encoding scheme of Section 7 and formulates the appropriate equations connecting flat-space coordinates to curved-space coordinates. The equations are shown to be equivalent to the Schwarzschild metric solution to Einstein's equation.

Section 9 derives the time metric (g_{00}) associated with curved 3-space and shows it to match the Schwarzschild metric.

Section 10 discusses the time metric as a function of position in the local field.

Section 11 discusses the conclusions we draw from this work.

2. Intro to Coordinate Systems and Their Meaning

One of the fundamental rules of physics is that coordinate systems have no effect on physical reality. When this is violated, physics becomes confusing. We begin by noting that Euclidean space is Pythagorean in that distances in space and time are unlimited.

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 \to \infty \tag{1}$$

(Pythagorean distance in Euclidean *space* and *time*)

This implies that Euclidian four-space can be mapped onto any and all events in the universe. That allows us to label every event and to relate any event to any other. These relations constitute our *physics* or *models of reality*. But what are the relations? They are intended to capture objective reality in some sense. Nozick, in "*Invariance: the structure of the objective world*" [2] observes that an objective fact is accessible from different angles, *i.e.*, "*an objective fact is invariant under various transformations*."

Consider the Galilean transformation describing a photon's position in 4-space

with time

$$\vec{x} = \vec{v}t \tag{2}$$

where velocity $\vec{v}$ can point in any direction and $|\pm\vec{v}| = c$. Thus $x = -ct$ represents the photon moving in the negative direction. Rearranging and taking a difference we have

$$\mathrm{d}(x - ct) = 0 \quad \text{or} \quad \mathrm{d}(x + ct) = 0 \tag{3}$$

Following Hestenes, [3] we define geometric algebra multi-vector $X = ct + \vec{x}$ and its complement $\tilde{X} = ct - \vec{x}$. It is simple to show that

$$\mathrm{d}X\,\mathrm{d}\tilde{X} = c^2\mathrm{d}t^2 - \mathrm{d}\vec{x}^2 = 0 \tag{4}$$

This is an invariance that holds for all distances $\mathrm{d}\vec{x}$ and corresponding duration $\mathrm{d}t$. It is, of course independent of coordinate systems. If we denote the value $\mathrm{d}X\,\mathrm{d}\tilde{X}$ by $\mathrm{d}s^2$ we obtain the invariance relation for the photon equation of motion (where we use $\mathrm{d}\vec{x}^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$):

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 = 0 \tag{5}$$

From the above discussion we interpret this as an *invariance relation* in Euclidian 4-space and we call it the *Minkowski invariance*. But Minkowski did not conceive of this as an invariance relation; he actually believed it was a description of "space-time", and famously stated: "*space and time by itself are doomed to fade away… only a kind of union of the two will preserve an independent reality*." In fact, applying the Lorentz transformation to this imagined space time he showed that space can be rotated into time and time rotated into space. Little wonder physicists who grew up with an intuitive grasp of Euclidian space and time found it difficult or impossible to grasp *Minkowski space-time*.

In "*Physics of clocks in absolute space and time*" [4] I choose Equation (5) as the fundamental photon-based invariance and use Hestenes' multi-vector to derive the classical Hamiltonian

$$H = E = \sqrt{m_0^2c^4 + c^2p^2}\ . \tag{6}$$

This is then used to show that if two clocks are built to tell identical time when side-by-side, and one clock is accelerated to velocity $\vec{v}$, the time interval read on the moving clock, $\mathrm{d}\tau$, will run more slowly than the time duration $\mathrm{d}t$ measured on the clock at rest according to

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} = \gamma = \left(\frac{1}{\sqrt{1 - \left(v^2/c^2\right)}}\right) \tag{7}$$

The physics used to derive this "time dilation" is based on Galilean transformations in absolute time and space and the "relativistic" inertia

$$m = \gamma m_0 \tag{8}$$

where m_0 is the rest mass. This result agrees with Lucas and Hodgson [5]:

"*If we… retain Newtonian dynamics, and the Newtonian definition of velocity and acceleration, then we… still obtain relativistically correct results if we… al-*

low the mass to depend on the velocity."

In special relativity rest mass is defined in all inertial reference frames in relative motion, and the Lorentz transformation is applied to "space" and "time", where space-time is imagined as defined by Minkowski. In the referenced *energy-time* derivation, the most solid experimental fact of special relativity, *time dilation*, is reproduced *exactly* in absolute space and time *without* paradoxes that mar special relativity. Energy-time theory does *not* yield *length contraction* nor Einstein's *law of velocity addition*; Rindler [6] says length contraction will probably *never* be tested; Einstein's *law of velocity addition* is known to be violated in LHC-type accelerators [7].

The physics of "clock slowing" in absolute time (*universal simultaneity*) and absolute space (with *preferred frame* defined by local gravity) follows from increased inertia of the clock mechanism and consequent decrease in acceleration of the restoring force common to all harmonic oscillators. Energy-time physics is compatible with all relativistic experiments [8], yet is based on Galilean transformation *in intuitive time and space*, in which *the Minkowski relation is an invariance relation*.

3. Remarks on Geometric Algebra Approach

Hestenes, as noted, formulates a geometric algebra multi-vector consisting of the scalar ct and vector $\vec{x}$: $X = ct + \vec{x}$. With its complement $\tilde{X} = ct - \vec{x}$ he derives the Minkowski invariance. Writing circa 1965, he was focused on showing that his mathematics is compatible with special relativity; he was not interested in reinterpreting relativity. He therefore viewed Minkowski as "space-time" and the Lorentz transformation as operating on space-time.

We, on the other hand, began in search of an invariant upon which to base physics in Euclidean space and time. Hestenes' formulation leads immediately to such invariance. After almost a century of proof of "time dilation" we know that clocks in motion run slower than clocks at rest. Therefore we extend the invariance to frames in relative motion by assigning each frame clock reading $\mathrm{d}t$ = duration in rest frame, $\mathrm{d}\tau$ = duration for moving clock. We find that invariance in relative frames leads to $\mathrm{d}\tau/\mathrm{d}t = \left(1 - v^2/c^2\right)^{-1/2}$. Hestenes' multi-vector formulation leads to the classical Hamiltonian in intuitive space and time, and is compatible with $m = \gamma m_0$.

Recognizing that all clocks count cycles and that cycles derive from simple harmonic oscillator mechanisms, it is easy to show that increased inertia of the material oscillator leads to resistance to the acceleration of the oscillator restoring force, and hence clocks in motion run slower than clocks at rest. Moreover they do so by *exactly* a factor of γ. In other words Hestenes' multi-vector formulation of time and space leads to special relativity if viewed as *Minkowski spacetime* operated on by Lorentz transformation, but leads to classical physics if treated as *Minkowski invariance*:

$X = ct + \vec{x}$ Minkowski space + Lorentz transformation $\Rightarrow$ 4D relativity

$X = ct + \vec{x}$ Minkowski invariance + Euclidean space + Galilean $\Rightarrow$ (3 + 1)D reality

Observe that *the mathematical construct* representing time and space is the same; the *ontological assumptions* determine what one does with it.

Amazingly, most treatments of special relativity openly discuss built-in paradoxes, where *paradox* is synonymous with *logical contradiction*, yet special relativity is held as sacred and our *intuition is dismissed as misleading*. The paradox-free theory summarized above is compatible with our intuition *and* compatible with all physics experiments; therein the Minkowski relation is not a description of *space-time*, but a photon-based *invariance relation* at the heart of physics.

4. Coordinates and Gravity as Curved Space-Time

In "*The Schwarzschild metric: it's the coordinates, stupid*", Will *et al.* [9] observe that

"*Every general relativity text emphasizes that coordinates have no physical meaning… Every student of general relativity is taught that coordinates are irrelevant to physics.*"

"The principle of general covariance, upon which general relativity is built, implies that coordinates are simply labels of space-time events that can be assigned completely arbitrarily. The only quantities that have physical meaning—the measurables—are those that are *invariant* under coordinate transformations." It is such invariance that we used to derive the physics of absolute space and time in reference 4. That approach led to physics of *energy-time* rather than the *space-time* focus of relativity. The success of this theory in matching the experimental data of the last century suggests that we apply the focus on *energy* to the general relativistic treatment of gravity.

Despite the catechism that coordinates can have no effect on physics, *the equivalence principle* upon which general relativity is typically founded yields a built-in contradiction: by transforming local coordinates the gravitational field can be banished. Weinberg [10] formulates the equivalence principle as

"*At every space-time point in an arbitrary gravitational field it is possible to choose a 'locally inertial coordinate system' such that* [*locally*] *the laws of nature take the same form as in unaccelerated Cartesian coordinate systems in the absence of gravitation.*"

In other words, despite grand statements to the effect that coordinate systems have no effect on physics, the proper choice of a locally inertial coordinate system effectively makes gravity vanish! Poisson and Will [11] state:

"*All local aspects of gravity can be turned off by doing physics in a freely moving* (*coordinate*) *frame of reference… Gravity is not present in these* (*coordinate*) *frames.*"

In other words, one can always mathematically *do away with* the local gravitational field and hence any associated energy. A consequence of this is ambiguous

treatment of gravitational energy in general relativity, a problem that has never been solved.

5. The Essence of the "Never-Solved" Problem

Chen *et al.* observe in "*Gravitational energy for GR…*", that in all of Einstein's published papers, "most include a significant consideration of the topics of gravitational energy."

Most physicists are aware of Emmy Noether's seminal work on symmetry and conservation. What most probably do not realize is that the reason Noether began investigations was "*to clarify the issue of gravitational energy*." In fact, in 1918 she proved that "*there is no covariant total energy-momentum density tensor for gravitating systems*." MTW [12] discuss this feature as "*a consequence of the equivalence principle…*"

A century of effort in this direction has produced many results; however *it has not solved the problem*. One might say physics has given up on the problem; Penrose in 1982 invented the idea that "*energy-momentum is quasi-local*: *i.e.*, *it is associated with a closed 2-surface…*". Finding an appropriate quasi-local notion of energy-momentum has been surprisingly difficult. In fact,

"*the state-of-the-art is typically postmodern: although there are several promising and useful suggestions*, *we not only have no ultimate*, *generally accepted expression for the energy-momentum and especially for the angular momentum*, *but there is not even a consensus in the relativity community on general questions… or on the list of the criteria of reasonableness of such expressions*."

The remainder of this paper will examine the nature of the problem that has frustrated general relativists for over a century.

6. Introduction to Einstein's Equations

A non-vanishing energy-density necessarily produces gravity (*i.e.*, the curvature of space-time). It is questionable whether most physicists even have an intuitive conception of *curved spacetime*. Can a re-interpretation of general relativity clarify "curved space-time" and bring it into congruity with our intuitive ideas of space and time? Recall that Einstein's vision of gravity is *pure geometry*. His basic equation

$$G_{\mu\nu} = \kappa T^{(m)}_{\mu\nu} \tag{6}$$

specifies that stress-energy tensor $T^{(m)}_{\mu\nu}$ defines the distribution of mass-energy while $G_{\mu\nu}$ defines the geometry of the "*curved space-time*" induced by the existence of the material stress-energy $T^{(m)}_{\mu\nu}$. The problem is to *derive the curved space geometry* from $T^{(m)}_{\mu\nu}$. Lorentz, Levi-Civita and Klein argued that the Einstein curvature tensor $G_{\mu\nu}$ was the only proper gravitational energy-momentum density; hence one should regard the Einstein equation in the form

$$-\kappa^{-1} G_{\mu\nu} + T^{(m)}_{\mu\nu} = 0 \tag{7}$$

as describing the vanishing sum of gravitation and material energy-momentum.

However, this is generally *self-contradictory* if the solution is in curved space-time and curvature is known only *after* the problem is solved; how could we possibly express the source distribution in the *not yet solved for* curvature geometry? We cannot solve for curvature without knowing the source; and *cannot* express source distribution $T_{\mu\nu}^{(m)}$ responsible for curvature in the curved space that is the object of solution. How is this problem handled in relativity? Chen *et al.* observe:

"*Specifically, in the case of GR, knowing the curvature gives, via the Einstein tensor, the symmetric Hilbert energy-momentum density.*"

This does *not* solve the problem. It states that, *if* the problem could be solved, *and* we knew the solution ("knowing the curvature") then we could represent the energy-momentum density. This is *not* entirely un-akin to the observation that, "*if we had some ham, we could have ham and eggs, if we had some eggs.*" It definitely does *not* solve the never-solved problem of gravitational energy density that induces curvature of space-time.

Feynman [13]: gravity theory suffers because

"*… one side of the equation is… geometric, and the other side is not* [*geometric*]*… even for very simple problems, we have no idea how to go about writing down a proper* $T^{\mu\nu}$."

Little wonder physicists generally have no intuitive conception of *curved space-time*. There are two possible solutions; a trivial and a non-trivial solution. The trivial solution is $G_{\mu\nu}=0$, where the stress-energy tensor is everywhere zero in all coordinate frames. This generally implies "flat space" but Vishwakarma has analyzed curvature in [14] in view of the fact that a proper energy-stress tensor of the gravitational field does not exist. He discusses the Kasner solution, which I have treated in [15], "*A Primordial Space-Time Metric*". As noted in his title, this trivial solution involves "*a new paradigm in GR*".

There is also a *nontrivial* way to avoid the paradox of solving an equation expressed in two nontrivially different coordinate systems. I observe that the only non-trivial way to avoid this paradox is to place the center of mass of a spherically symmetric body at *the only point* common to both flat space and curved space, *the origin*, and appeal to Birkhoff's *Shell theorem*. This inherently limits Einstein's general relativity to *One-body solutions* of the *N-body problem*.

Heaviside [16] extended Newtonian physics to include *field energy density* in flat space, and this is recognized as iteratively equivalent to Einstein's non-linear field equation. Feynman noted:

"*It is one of the peculiar aspects of the theory of gravitation, that it has both a field interpretation and a geometric interpretation…*"

It is not generally clear how a density-based field description is related to metric *curvature* solutions; our immediate goal is to clarify this.

7. Encoding Density in Geometry

We ask how one might understand the transition from *physical-field-based*

physics in flat space to *curved-space-time-geometry-based* physics. Despite the common perception that *Minkowski space-time* views space as *empty*, Einstein [17] came to realize that

"*There is no such thing as empty space, i.e., a space without a field. Space-time does not claim existence on its own, but only as a structural quality of the field.*"

Further, *physical fields are real and have energy*. Ohanian and Ruffini [18]

"*The gravitational field may be regarded as the material medium sought by Newton; the field is material because it possesses an energy density.*"

In other words, gravity is viewed not as *curved space-time*, but as a *field with energy density*. Yet special relativity is based on pretending that gravity does not exist at all; and general relativity on pretending that gravitational energy does not exist "locally"—it can always be transformed away. This despite universal agreement that *coordinate systems*, curved or flat, *can have no effect on physical reality*. The *logical contradiction* that a change in coordinate system can make a physically real field vanish is hidden in the *Equivalence Principle* used to derive our premier theory of gravity. The claim of non-local energy density yields "*gravity as space-time curvature*".

Despite recent focus [19] on "information" as a real physical entity, we examine information in its original *coding* perspective; we observe that there is absolutely *no energy density information encoded in flat space coordinates*. One can change the coordinate system with *absolutely no effect* on the field energy density; one can change the energy distribution with absolutely no effect on the coordinate system. All of the information about the physical field is contained in the *energy density distribution*; *none in the flat space coordinates*!

In **Table 1**, flat space information is in the energy density; no information (other than scale) is associated with units of the coordinate system. In curved space, normalized energy represents no information; the relevant information is encoded in the geometry specified by the metric.

What happens to the physics when we remove density information by claiming zero local energy density? *It must be replaced somehow!* If the only physical info we have is energy density and coordinate info, then when removing the density info we must replace it with coordinate info, and this is done by *replacing constant length flat space coordinates with variable length* (*metric*) *curved space coordinates*. No information has been lost; the real physical information encoded by the density distribution over flat space has been replaced by abstract information in which *physical energy density is encoded as* "*geometry*". According to Poisson and Will:

Table 1. Information matrix.

	Flat	Curved
energy density	1	0
metric length	0	1

"*the metric… achieves two purposes*: *it encodes geometrical information about coordinate system, and it encodes physical information about the gravitational field*."

I would say that *the metric encodes physical information about the gravitational field as geometrical information in a coordinate system*. This clearly recognizes that the physical field energy density is ontologically real, and is mathematically equivalent to the description of the real physical density encoded in the abstract formulation of curved space time!

One best distinguishes *physical density-based reality* from *curved space-time-based reality* via inertia. In *Energy-time theory* the special relativity γ-factor applies to *inertial mass*. In *curved spacetime* real inertial force is replaced by the abstraction of geodesics. This is at the root of Feynman's, Weinberg's, Padmanabhan's, and others insistence that

"*Curved space-time is not a necessary conception of gravity*."

Consider the primordial gravitational field treated in "*The Primordial Principle of Self-interaction*" [20]. The Self-interaction equation $\vec{\nabla}\vec{\psi}=\vec{\psi}\vec{\psi}$ is used to derive the Heaviside equations known to be equivalent under iterations to Einstein's field equations. The geometrical algebra formulation is the multi-vector $\vec{\psi}=\vec{G}+i\vec{C}$ where $\vec{G}$ is the gravitational field vector and $\vec{C}$ is the gravitomagnetic bivector. The duality operator i converts $\vec{C}$ to $\vec{C}$. In the absence of circulation $\vec{C}$ the solution to $\vec{\nabla}\vec{\psi}=\vec{\psi}\vec{\psi}\Rightarrow\vec{\nabla}\vec{G}=\vec{G}\vec{G}$ is $\vec{G}=1/\vec{r}\left(\equiv\vec{r}/r^2\right)$. Hence $\vec{G}\vec{G}=1/r^2$ is the energy density of the primordial field. As shown in **Figure 1**, the mass of the field inside a sphere of radius r is (modulo $4\pi/3$)

$$m\cong r^3(\rho)=r^3\left(\vec{G}\vec{G}\right)=r^3\left(\frac{1}{r^2}\right)=r\quad\Rightarrow\quad\frac{m}{r}\cong 1 \tag{8}$$

Thus potential $\phi=-(m/r\cong 1)$ where Newton's gravitational constant $g=1$ and speed of light $c=1$. Let us assume that this energy is distributed with unit density. Consider a volume defined by $\mathrm{d}x\mathrm{d}y\mathrm{d}z$ where $\mathrm{d}x=1,\mathrm{d}y=1,\mathrm{d}z=1$ as shown in **Figure 2**. Ignoring the sign of ϕ we define unit energy density of the primordial gravitational field with no material objects in the field.

Recall that the difference in potential energy of two points is the work done in moving mass m between two points.

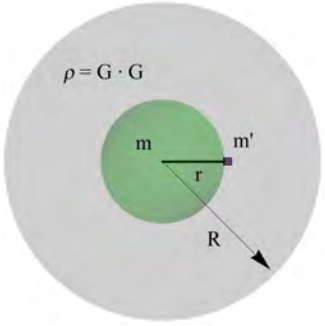


Figure 1. Mass m at $\vec{r}$.

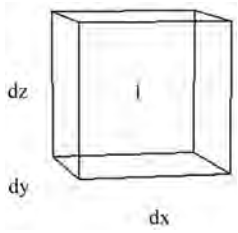


Figure 2. The energy density of the primordial gravitational field is chosen to be unity in a unit differential volume.

$$U(r)=W_{\infty r}=\int_{\infty}^{r}\vec{F}\cdot\mathrm{d}\vec{r}=\int\frac{gMm}{r^2}\mathrm{d}r=gMm\left[-\frac{1}{r}\right]_{\infty}^{r}=-\frac{gMm}{r}=m\phi \tag{9}$$

The "gravitational energy per unit mass", ϕ describes the potential energy at the point in question. This is the interaction energy $m\phi$ that would exist if m units of mass were to be introduced at that point. In the literature, this "test mass" m is presumed small enough that its own gravitational potential need not be taken into account.

Next we add a mass M to space. As noted, the only point common to both flat space and curved space is the origin $(0,0,0)$, assumed coincident. Although the massive body M does not exist at a point we assume it is symmetrical and thus we apply Birkhoff's *Shell theorem* to reduce M to an equivalent point mass. At distance r from the origin the gravitational energy (per unit mass) is $\phi=-M/r$. Figure 3 represents the energy of the primordial field and mass M.

The total gravitational energy in a given region is the sum of the primordial gravitational field plus the potential due to mass M located at the origin

$$\frac{m}{r}+\frac{M}{r}\quad\Rightarrow\quad 1+\phi\,. \tag{10}$$

We've chosen the unit cube $\mathrm{d}x\mathrm{d}y\mathrm{d}z$ where $\mathrm{d}x=1,\mathrm{d}y=1,\mathrm{d}z=1$ as the region containing energy density $1+\phi$. We have suppressed the negative sign of the potential(s) to simplify the treatment, but we will find [in Equation (16)] that the term of interest is always positive, whether we use $(1+\phi)$ or $-(1+\phi)$. Our goal is to normalize this density $1+\phi\to 1$ by choosing new variable coordinates such that $\mathrm{d}x'\neq 1,\mathrm{d}y'\neq 1,\mathrm{d}z'\neq 1$. Based on the equivalence principle one might assume that $1+\phi\to 0$ since "local density" cannot be defined. This is not the same as saying that local energy density of the field disappears; it merely ceases to provide physical information. We normalize the energy in every relevant volume, such that $\mathrm{d}x'\mathrm{d}y'\mathrm{d}z'$ always contains unit energy density as shown on the right side of Figure 4.

In this way the information has been transformed from the energy density of the field ϕ, in constant (informationless) coordinates of flat space, to the information of the (variable) coordinates containing a constant (informationless) energy density in curved space-time. This is consistent with the fundamental requirement that coordinate systems cannot affect physical reality. They bring no information to reality; they only label physical reality such that we all agree

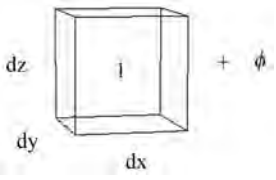


Figure 3. Unit primordial field plus local M field.

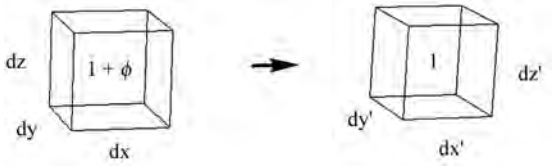


Figure 4. The primordial gravitational field alone has $\phi_0 = m/r = r/r = 1$ shown in **Figure 2** in flat space, $\mathrm{d}x, \mathrm{d}y, \mathrm{d}z = 1$. For mass M at the origin, we add potential energy $\phi = M/r$. The goal is to transform to a unit density (no density information) in curved space-time as shown in the cube at right, $\mathrm{d}x', \mathrm{d}y', \mathrm{d}z' \neq 1$.

upon the points under discussion. Of course, depending upon the physical configuration involved, one coordinate system, say spherical, may be more convenient (easier calculations) than another (say cylindrical) but neither coordinate system contains any more information than the other.

One might ask why the value 1 is used for the apparent energy density in curved space instead of the value *zero* that seems to be implied by the equivalence principle. We do not require that the energy density vanish, only that *information about the energy density must vanish*, in accordance with Einstein's statement that "*gravitational energy is not localizable*". Thus we are not working to get rid of the physical energy density; we are working to get rid of the energy density *information*. This is accomplished by normalizing the energy density such that every local region in curved space contains exactly one unit of energy density in a region bounded by $\mathrm{d}x', \mathrm{d}y', \mathrm{d}z' \neq 1$, as shown at right in **Figure 4**. By defining $\mathrm{d}x_i'$ to accomplish this goal we transfer the physical density information in flat space to the curvature (metric) information of curved space. In flat space every differential has unit length and there is no information contained in the normalized unit of the coordinate system. In curved space, the metric intervals are defined to normalize the energy density of the field. This removes the information from the energy density and transfers it to the curved space metric.

8. Analyzing the Encoding Scheme

We ensure that "gravitational energy is not localizable" by requiring that, in curved space-time, every local region contains a unit of energy indistinguishable from any other local region, whereas the local metric at any point is distinguishable from the metric at any neighboring point. This is captured by the expression of the generalized Pythagorean for curved space-time:

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^{\mu}\mathrm{d}x^{\nu}. \tag{11}$$

The goal, of course, is to calculate the metric $g_{\mu\nu}$ necessary to transform the energy density information contained in ϕ into the local coordinate information contained in $g_{\mu\nu}$. Whereas $\mathrm{d}x\mathrm{d}y\mathrm{d}z$ coordinate differentials are independent of ϕ, the transformed coordinates are totally dependent on ϕ, *i.e.*,

$$\mathrm{d}x'\mathrm{d}y'\mathrm{d}z' \equiv \mathrm{d}x'(\phi)\mathrm{d}y'(\phi)\mathrm{d}z'(\phi). \tag{12}$$

For the most general solution we force each metric-based component to satisfy the relations

$$\begin{aligned}(1+\phi)\mathrm{d}x_1'(\phi) &= \mathrm{d}x_1\\ (1+\phi)\mathrm{d}x_2'(\phi) &= \mathrm{d}x_2\\ (1+\phi)\mathrm{d}x_3'(\phi) &= \mathrm{d}x_3\end{aligned} \tag{13}$$

In this case the key coordinate variable relation becomes

$$\mathrm{d}x_i'(\phi) = \frac{\mathrm{d}x_i}{1+\phi} \tag{14}$$

The consequent volume element becomes

$$\mathrm{d}x_1'(\phi)\mathrm{d}x_2'(\phi)\mathrm{d}x_3'(\phi) \equiv \frac{\mathrm{d}x_1}{1+\phi}\frac{\mathrm{d}x_2}{1+\phi}\frac{\mathrm{d}x_3}{1+\phi}. \tag{15}$$

And for any two of these dimensions we obtain the (always positive) product:

$$\mathrm{d}x_i'(\phi)\mathrm{d}x_j'(\phi) \equiv \frac{\mathrm{d}x_i}{1+\phi}\frac{\mathrm{d}x_j}{1+\phi}. \tag{16}$$

The above analysis, restricted to 3-space, is effectively a "time-slice" of Euclidian reality, $\mathrm{d}t \equiv 0$; we use the convention $i, j \in \{1,2,3\}$, $\mu,\nu \in \{0,1,2,3\}$. The subtlety of 4-dimensional space-time certainly contributes to the difficulty of reconciling Einstein's theory with our intuitive ideas of time and space. As we've only considered 3-space energy density we begin with the Pythagorean relation $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu$. From Equations (11) and (16) we see that

$$g_{ii} = \frac{1}{(1+\phi)^2} = (1+\phi)^{-2}. \tag{17}$$

In order to make sense of this we quote John Wheeler's remarks about a white dwarf star: "*It is small, but not terribly small; dense, but not terribly dense. Space-time is 'flat' within it…*" The term "flat" means ϕ is small, approaching zero, therefore we use the binomial expansion

$$(1+\phi)^{-2} \cong (1-2\phi). \tag{18}$$

Yilmaz, [21] in his "*Field theory of Gravity*" derives a metric proportional to $\mathrm{e}^{-2\phi}$; for consistency with consensus GR we use the approximation

$$g_{ij} = 1-2\phi \tag{19}$$

This approximation is consistent with the linearized equation

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{20}$$

where $\eta_{\mu\nu}$ is the Minkowski metric with signature $(1,-1,-1,-1)$ and

$$h_{\mu\nu} = -2\phi \tag{21}$$

Limiting ourselves to 3-space ($\mathrm{d}t \equiv 0$) we obtain

$$\mathrm{d}s^2 = \frac{-1}{1+2\phi}\left(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right). \tag{22}$$

since the binomial expansion also yields

$$\frac{1}{\left(1+\phi\right)^2} = \frac{1}{1+2\phi} \tag{23}$$

Equation (22) represents diagonals term $\mathrm{d}x_i\mathrm{d}x_i$ with factor h_{ii}. In "*A primordial space-time metric*" the representation of the gravitomagnetic field of the dense region of the field moving in the z-direction satisfies $h_{xy} - h_{yx} = 0$. This angular momentum relation causes $\mathrm{d}x\mathrm{d}y$ and $\mathrm{d}y\mathrm{d}x$ terms to cancel, yielding Equation (22).

The moving density with $h_{xy} \neq 0, h_{yx} = -h_{xy}$ is seen to be compatible with Equation (22). The exact solution we've been considering however is based on the mass M located at the origin. The spherical symmetry of this case implies $h_{xy} = h_{yx} = 0$ (and cyclical iterations); also compatible with Equation (22). The alternative form of Equation (22) based on Equation (19) yields

$$\mathrm{d}s^2 = -\left(1-2\phi\right)\left(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right). \tag{24}$$

In order to touch base with general relativity, we use Equation (9) to observe that

$$\phi = -\frac{gM}{r} \tag{25}$$

Rewrite Equations (24) as

$$\mathrm{d}s^2 = -\left(1+\frac{2gM}{r}\right)\left(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right) \tag{26}$$

Compare this to Ohanian and Ruffini's Equation (12) p.179 from the approximate metric tensor:

$$g_{\mu\nu} = \begin{bmatrix} 1-2gm/r & 0 & 0 & 0 \\ 0 & -\left(1+2gm/r\right) & 0 & 0 \\ 0 & 0 & -\left(1+2gm/r\right) & 0 \\ 0 & 0 & 0 & -\left(1+2gm/r\right) \end{bmatrix} \tag{27}$$

Then compare to their space-time interval (Equation (13) p. 179):

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu = \left(1-\frac{2gm}{r}\right)\mathrm{d}t^2 - \left(1+\frac{2gm}{r}\right)\left(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right) \tag{28a}$$

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu = \left(1+2\phi\right)\mathrm{d}t^2 - \left(1-2\phi\right)\left(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right) \tag{28b}$$

Let us use the alternative Minkowski metric $\left(-1,1,1,1\right)$ and let Newton's constant $g = 1$ and speed of light $c = 1$. In this case $\phi = -\left(M/r\right)$ and we write

$$\mathrm{d}s^2 = -\left(1+\frac{2M}{r}\right)\left(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\right) + \left(1-\frac{2M}{r}\right)\mathrm{d}t^2 \tag{29a}$$

$$ds^2 = -(1-2\phi)\left(dx^2 + dy^2 + dz^2\right) + (1+2\phi)dt^2 \tag{29b}$$

This is exactly the form of Tolman's [22] expression (Equation (82.15)) for the Schwarzschild line element in approximate form (page 205). Similarly, Weinberg derives the Schwarzschild metric on page 180, Equation (8.2.12) using signature $(1,-1,-1,-1)$

$$d\tau^2 = \left[1-\frac{2Mg}{r}\right]dt^2 - \left[1-\frac{2Mg}{r}\right]^{-1} dr^2 - \left[r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2\right] \tag{30a}$$

$$d\tau^2 = \left[1+2\phi\right]dt^2 - \left[1+2\phi\right]^{-1} dr^2 - \left[r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2\right] \tag{30b}$$

Many other general relativity books can be consulted for derivation of the Schwarzschild metric surrounding mass M at the origin. Here we have used the encoding process to derive *exactly* the three-space metric representing the gravitational energy density equivalent information. For completeness and comparison I have included the *time-time* metric, g_{00}, but have not yet derived it; I do so in the next section. We conclude this section by observing that our approach of encoding energy density as geometry has yielded the Schwarzschild metric of curved space, the first exact solution to Einstein's equation. Rindler remarks that it is important to understand the spatial geometry of the Schwarzschild metric in order to apply to physical problems. Specifically:

"*One way to visualizing the curved 3-space like that of the Schwarzschild lattice, whose metric is given by*

$$dl^2 = \frac{dr^2}{1+2\phi} + \left[r^2\left(d\theta^2 + \sin^2\theta d\varphi^2\right)\right] \tag{31}$$

is to pretend that it is really flat, but that it's rulers behave strangely…"

In Equation (31), we consider a *radial ruler*, [in which case $d\theta = d\varphi = 0$] and note that the rulers shrink by factor $(1+2\phi)^{1/2} \approx 1+\phi$. This is *exactly* the result for dx, dy and dz we achieved in Equation (14) by transforming from actual physical energy density in flat space (constant rulers) to the equivalent information encoded as geometry of curved space (variable metric). QED.

The computation necessary to derive the Schwarzschild metric in general relativity is far from beautiful; it is rather obtuse. Rindler, at the top of his preface quotes Bernhard Riemann

"*The ideal is to reach proofs by comprehension rather than by computation.*"

By comprehending that we are *encoding energy-density as geometry* and drawing a few simple diagrams we have transformed from energy density in flat space (physical reality) to curved space with almost no computation and obtained the standard Schwarzschild solutions.

9. The Time Metric Associated with Curved 3-Space

Our focus in Section 2 was on "time dilation" in special relativity from the perspective of energy-time theory in absolute space and time. Sections 4 - 7 were focused on curved space and the associated "*never solved*" problem of gravitation-

al energy in curved space. This led to the conclusion that gravitational energy density in flat space is banished by the equivalence principle, and yet the physical information associated with the density of the field cannot be simply discarded—it must be transformed into the geometry associated with variable metric coordinate systems. This conclusion, and associated perspective, led almost immediately to the appropriate metric for the non-trivial solution represented by the Schwarzschild metric. The assumption of absolute space and time underlying the physics of inertial clocks is based on defining "absolute time" as *universal simultaneity*, and hence our curved space solution holds at any time (slice) t with $\mathrm{d}t \equiv 0$. This caused the time dimension to drop out of the Schwarzschild metric solution and reduce the problem to that of curved 3-space as derived in Section 8.

Recall that our goal was to encode the energy density in a unit cube in flat space $\mathrm{d}x\mathrm{d}y\mathrm{d}z = 1$ in the variable metric $\mathrm{d}x', \mathrm{d}y', \mathrm{d}z' \neq 1$ and we did so via Equation (14):

$$\mathrm{d}x_i'(\phi) = \frac{\mathrm{d}x_i}{1+\phi}.$$

In other words we derived the variable metric $\mathrm{d}x_i'$ from the flat space metric $\mathrm{d}x_i$ with its constant scale. This formulation is in Euclidean space and time derived by Equation (1). We now review the analysis context of Minkowski invariant defined by Equation (5)

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 = 0$$

It is called "invariant" because it does not vary with coordinate systems, hence

$$\mathrm{d}s'^2 = c^2\mathrm{d}t'^2 - \mathrm{d}x'^2 - \mathrm{d}y'^2 - \mathrm{d}z'^2 = 0 \tag{32}$$

Let us rewrite this as

$$\mathrm{d}s^2 = 0 = \mathrm{d}t^2 - \mathrm{d}r^2 = \mathrm{d}t'^2 - \mathrm{d}r'^2 \tag{33}$$

From our Equation (13) we express the coordinate differentials as $\mathrm{d}x_i$ and $\mathrm{d}x_i'$ therefore we use

$$\mathrm{d}s^2 = g^{\mu\nu}\mathrm{d}x_\mu\mathrm{d}x_\nu \tag{34}$$

Hence, when I obtain (35) from Equation (33) I do so in the form

$$\mathrm{d}t'^2 = \mathrm{d}r'^2 = \frac{\mathrm{d}r^2}{1+2\phi} = g^{00}\mathrm{d}x_0\mathrm{d}x_0 = g^{00}\mathrm{d}t^2 \tag{35}$$

Minkowski-invariant-based energy-time theory derives from the equation of motion for photons

$$\mathrm{d}r^2 - c^2\mathrm{d}t^2 = 0 \quad \Rightarrow \quad \frac{\mathrm{d}r^2}{\mathrm{d}t^2} = c^2. \tag{36}$$

Therefore, from Equation (35) we find

$$\frac{\mathrm{d}r^2}{\mathrm{d}t^2} = g^{00}(1+2\phi) = c^2. \tag{37}$$

The inverse metric $g_{00}=1/g^{00}$ and when $c^2=1$ we obtain

$$g^{00}g_{00}=1 \quad \text{and} \quad g^{00}(1+2\phi)=1. \tag{38}$$

Therefore

$$g_{00}=(1+2\phi) \tag{39}$$

is compared with Equations (29b) and (30b), and $g_{00}=(1+2\phi)$ is found to be in *exact agreement* with those expressions of the Schwarzschild time-time metric. Therefore, our flat space distribution of field energy density, $\phi(\vec{r})$ encodes energy density information as geo(metric) information of curved space and the corresponding time (the conjugate of energy).

10. Positional Dependence in the Local Field

Energy-time analysis is based not on space-time symmetry, but on inertial mass $m=m_0\gamma(v)$; inertial factor $\gamma(v)$ relates to inertial mass, not to space and time. We extend this relation to general relativity [23] as

$$m=m_0\gamma(\dot{\vec{r}})\mathrm{e}^{\phi(\vec{r})} \tag{40}$$

Here the mass of the particle depends upon the kinetic energy in a local absolute gravitational field as well as its position-dependent energy within a gravitational field with potential ϕ. This inertial relation explains clock-slowing, as known and practiced in GPS.

Our "gravity free" definition of inertial mass, $m=\gamma m_0$ was found to include the kinetic energy in addition to rest mass: $m=m_0(1-v^2/c^2)^{-1/2}$. In a gravitational field the energy of a mass depends upon the Newtonian potential $\phi=-gM/r$ so let us naïvely add the gravitational term $2\phi/c^2$ to the kinetic energy term as follows

$$m=m_0\left[1-\left(\frac{v^2}{c^2}+\frac{2\phi}{c^2}\right)\right]^{-1/2} \quad \Rightarrow \quad m_0\left[1+\frac{1}{2}\frac{v^2}{c^2}+\frac{\phi}{c^2}\right] \tag{41}$$

yielding inertial mass and associated energies:

$$\begin{array}{c} mc^2=m_0c^2+m_0v^2/2+m_0\phi \\ / \qquad | \qquad\quad | \qquad\quad \backslash \\ \text{total energy} \quad \text{rest} \quad \text{kinetic} \quad \text{gravitational} \end{array} \tag{42}$$

This formulation demonstrates the dependence of inertial mass on velocity $\dot{\vec{r}}$ *through* the local gravitational field and on position $\vec{r}$ *in* the local gravitational field. Our simple addition of the gravitational potential $\phi=-gM/r$ is only intuitively justified, however Poisson and Will derive the same result in orthodox fashion based on the action

$$S=-m_0c^2\int \mathrm{d}\tau=\int L\mathrm{d}t \tag{43}$$

where $\mathrm{d}\tau$ is the proper time after clock slowdown (speedup) is considered, and the Lagrangian

$$L = -mc\sqrt{-\eta_{\alpha\beta}\frac{\mathrm{d}r^{\alpha}}{\mathrm{d}t}\frac{\mathrm{d}r^{\beta}}{\mathrm{d}t}} \tag{44}$$

is evaluated as

$$L = -mc\sqrt{1-\frac{2\phi}{c^2}-\frac{v^2}{c^2}} \tag{45}$$

in which case

$$L = -m_0c^2 + m_0v^2/2 + m_0\phi\,. \tag{46}$$

As explained in Section 5, our energy-time Hamiltonian is completely compatible with the classical Lagrangian approach, therefore the Lagrangian in Equation (45) yields physics in agreement with Equation (42). We again see that the time-dependence of clocks follows from energy-based analysis rather than geometry.

Dimensional analysis shows that $\phi = gm/r$ has dimensions of velocity squared, v^2. Therefore to achieve the dimensionless factor we divide ϕ by $c^2 = 1$ and obtain $\frac{\phi}{c^2} \sim \frac{v^2}{c^2}$. Although we have obtained this via dimensional analysis, Weinberg (page 212) observes that it is a familiar result of Newtonian mechanics that the typical kinetic energy $Mv^2/2$ will be roughly of the same order of magnitude as the typical potential energy gM^2/r hence

$$Mv^2/2 \approx gM^2/r \quad \Rightarrow \quad v^2 \sim \frac{gM}{r}\,. \tag{47}$$

Moreover, a test particle in circular orbit of radius r about a central mass M will have velocity v given by the exact formula $v_g^2 = gM/r$.

In Equation (41), the special relativity term $\left[1+\frac{1}{2}\frac{v^2}{c^2}\right]$ represents positive energy, while $\phi = -gM/r$ is negative energy. Therefore

$$m = m_0\left[1+\left[\frac{1}{2}\frac{v^2}{c^2}-\frac{v_g^2}{c^2}\right]\right]. \tag{48}$$

11. Conclusions

As noted, physicists have never solved the problem of including the energy-momentum of the gravitational field in Einstein's field equations. Based upon our analysis of this problem, we concur with Feynman, Weinberg, and others: the concept of curved space-time is *not* necessary for gravity. In contrast to this, we view gravity as a field whose energy density determines the physics in Euclidean 3-space, with absolute time spanning 3-space as universal simultaneity. Ontological arguments imply that both theories are not equally valid.

The equivalence principle is generally interpreted to imply that the field energy can be made to vanish locally despite universally agreed-upon rules that coordinate systems have no effect on physical reality. This "disappearance" of the gravitational energy of the field when curved coordinate frames appear has

confused general relativists since Einstein proposed his geometric approach. In *energy-time theory*, the specification of field energy density in flat space determines the physics. If this information is removed, it must be replaced by some other source of information, and the curved coordinates supply this information. The actual gravitational field does *not* vanish; its energy density is normalized. That is, every corresponding differential volume in curved space coordinates, $\mathrm{d}x'\mathrm{d}y'\mathrm{d}z'$, contains exactly *one* unit of energy. This approach is formulated herein for the primordial gravitational field and the field of mass M at the origin. The normalization procedure determines the metric and that metric is *exactly* equal to the Schwarzschild solution. The energy of the gravitational field does *not* vanish in this approach; the normalization transfers the gravitational field information to the curved space coordinates.

One might wonder why physicists continue to work with "curved space-time" despite problems interpreting gravitational energy. Many physicists appear to believe in physical reality of curved space-time, as evidenced by a very recent paper in *Phys Rev Letters* [24]: "*Traversable worm holes in Einstein-Dirac-Maxwell theory*" constructs a specific example of a class of traversable worm holes in four space-time dimensions. The Heaviside equations, best suited for handling energy density, are generally *believed* to be useful only in "weak field" situations. However, Will and Poisson have remarked that this *flat space* approach is actually useful for very strong fields, but they offer no reason for this fact. In "*The primordial principle of self-interaction*", I derive the Heaviside equations with no "weak field" assumptions. This implies that the Heaviside equations work for *all* field strengths and thus present a complete theory of gravity. This is radically different from the prevailing view that only curved space-time equations are complete.

Dozens of theories of gravity exist; most are formulated in the 4D "curved space-time" of Riemann-Cartan. The current standard theory, ADM [25], splits $4\mathrm{D} \to (3+1)\mathrm{D}$ and the space-time metric is replaced by the *spatial-metric* plus *lapse* and *shift* time steps. This 3 + 1 split, compatible to a degree with absolute space and time, is required for the dynamical Hamiltonian formulation of ADM. Like ADM, Chen *et al.* base their theories on the Hamiltonian, which separates the equation into two sets (shown for electromagnetic field):

The initial value constraints: $\vec{\nabla}\cdot\vec{E} = \rho/\varepsilon_0$, $\vec{\nabla}\cdot\vec{B} = 0$

The dynamic time equations: $\dot{\vec{B}} + \vec{\nabla}\times\vec{E} = 0$, $\vec{\nabla}\times\vec{B} = \mu_0\vec{J} + \dot{\vec{G}}$

The treatment of time is *absolute*, as in cosmological time—the universe-spanning dimension that establishes the NOW, missing from Einstein's theories. To this ADM-model, Chen *et al.* add a (Penrose) quasi-local 2D boundary. I interpret this as averaging over various-size cells, each of which contains a normalized *unit-of-energy*. Depending upon the 2D-quasi-local boundary chosen, this may help "undo" the normalization procedure and restore the variable energy-density upon which the curved-space formulation was based. The (3 + 1)-split is invoked for Hamiltonian-based formulations which essentially require that the time derivative be singled out from the spatial derivatives, and thus

cannot be truly *four-covariant*.

Why, if the non-solvable nature of Einstein-type theories of *gravity as geometry* is endemic, do physicists invariably believe the Einstein nonlinear field equations? I suspect that the key reason is their belief that linearized equations are appropriate only in the "weak field" approximation. In other words, no matter how preferable it might be to work gravity problems in flat space (instead of curved *and* flat space) the consensus belief is that the linear equations are *incomplete*. That is, curved space-time is physical reality. This belief is challenged by the recent development of the theory of gravity in *The primordial principle of self-interaction*, wherein the issue of "weak field" never arises; the development is good for all field strengths. Because the field interacts with its own mass, the nonlinear aspect of gravity appears in *iterative dynamics*. Feynman, Weinberg, and many others discuss the *exact equivalence* of the iterated linear equations to the nonlinear curved-space formalism, but the truth is effectively overridden by the "weak field" misnomer.

Physically, dynamical energy density physics can evolve over time, in contrast to the Schwarzschild metric, which is static and is solved for "all at once". The metric does not evolve to a solution; it is *a priori*. The implications of *the primordial principle of self-interaction* and of *encoding energy-density in geometry* are that it is actually the curved-space formalisms that are incomplete, and the gravitational field theory in Euclidean space represents physical reality. This ontological shift to a density-based formalism has significant consequences, some of which are being explored.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Chen, C.-N., *et al.* (2015) *International Journal of Modern Physics D*, **24**, Article ID: 1530026.

[2] Nozick, R. (2001) Invariance: The Structure of the Objective World. Belknap-Harvard, Cambridge, MA.

[3] Hestenes, D. (1986) New Foundations for Classical Mechanics. 2nd Edition, Kluwer Academic Publishers, Dordrecht. https://doi.org/10.1007/978-94-009-4802-0

[4] Klingman, E. (2020) *Journal of Modern Physics*, **11**, 1950-1968. https://doi.org/10.4236/jmp.2020.1112123

[5] Lucas and Hodgson (1990) Space-Time and Electromagnetics. Oxford University Press, Oxford.

[6] Rindler, W. (1991) Introduction to Special Relativity. 2nd Edition, Oxford Science Pub., Oxford.

[7] Cannoni, M. (2016) Lorentz Invariant Relative Velocity and Relativistic Binary Collisions.

[8] Klingman, E. (2018) Everything's Relative, Or Is It.

[9] Fromholz, P., Will, C. and Poisson, E. (2013) The Schwarzschild Metric: It's the Coordinates, Stupid.

[10] Weinberg, S. (1972) Gravitation and Cosmology. John Wiley & Sons, New York.

[11] Poisson, E. and Will, C. (2014) Gravity: Newtonian, Post-Newtonian, Relativistic. Cambridge University Press, Cambridge.

[12] Misner, C.W., Thorne, K.S. and Wheeler, J.A. (1973) Gravitation. W.H. Freeman and Company, New York.

[13] Feynman, R. (1995) Feynman Lectures on Gravitation. Westview Press, Boulder.

[14] Vishwakarma, R. (2013) Gravity of $R_{\mu\nu} = 0$: A New Paradigm in GR.

[15] Klingman, E. (2019) *Prespacetime Journal*, **10**, 671-680.

[16] Heaviside, O. (1893) *The Electrician*, **31**, 281-282.

[17] Einstein, A. (1952) Relativity: The Special and General Theory. Crown Publishers Inc., New York.

[18] Ohanian, H. and Ruffini, R. (2013) Gravitation and Spacetime. 3rd Edition, Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139003391

[19] Verlinde, E. (2010) On the Origin of Gravity and the Laws of Newton.

[20] Klingman, E. (2020) *Journal of Modern Physics*, **12**, 65-81. https://doi.org/10.4236/jmp.2021.122007

[21] Yilmaz, H. (1992) *Il Nuovo Cimento*, **107B**, 941-960. https://doi.org/10.1007/BF02899296

[22] Tolman, R. (1962) Relativity, Thermodynamics and Cosmology. University Press, Oxford.

[23] Robertson, S. (2019) *Prespacetime Journal*, **10**, 69-74.

[24] Blazques-Salcedo, J. (*2021*) *Physical Review Letters*, **126**, Article ID: 101102. https://doi.org/10.1103/PhysRevLett.126.101102

[25] Arnowitt, R., Deser, S. and Misner, C. (2004) The Dynamics of General Relativity.

Journal of Modern Physics, 2021, 12, 1210-1218
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Resonant Energy Transfer, with Creation of Hyper-Excited Atoms, and Molecular Auto-Ionization in a Cold Rydberg Gas

Marwan Rasamny, Alan Martinez, Lianxin Xin, Jun Ren, Essaid Zerrad

Division of Physics, Engineering, Mathematics, and Computer Science, Delaware State University, Dover, DE, USA
Email: ezerrad@desu.edu

How to cite this paper: Rasamny, M., Martinez, A., Xin, L.X., Ren, J. and Zerrad, E. (2021) Resonant Energy Transfer, with Creation of Hyper-Excited Atoms, and Molecular Auto-Ionization in a Cold Rydberg Gas. *Journal of Modern Physics*, **12**, 1210-1218.
https://doi.org/10.4236/jmp.2021.129074

Received: May 23, 2021
Accepted: July 6, 2021
Published: July 9, 2021

Abstract

A cold Rydberg gas, with its atoms prepared initially all in the excited state $|n_0\rangle$, with $n_0 \gg 1$, contains an excessive amount of energy, and presumably is to relax by the Penning-type *molecular auto-ionization* (*MAI*), in which a portion of excess energy of one atom is given to another near-by atom and ionizing it. Its complementary process, the *resonant energy transfer* (*RET*), is discussed, in which the excess energy of one atom is used on another to form a hyper-excited atomic state $|n_a\rangle$ with $n_a \gg n_0$. This process is always present, provided certain resonance energy conditions are satisfied. In this report, the n_0 and density dependences of the RET rates are studied in detail, employing a simple model: 1) at low densities, the RET is mediated by the dipole-dipole coupling V_{dd} and its rates are generally much smaller than that of MAI, especially for small n_0. But 2) as the density increases, our model shows that the rates become of comparable magnitude or even larger than the MAI rates. The V_{dd} is no longer adequate. We, then construct a semi-empirical potential to describe the RET process. 3) At high densities, we show that the atomic orbital of $|n_a\rangle$ overlaps with that of neighboring atoms, and the electron-electron potential becomes prominent, resulting in much higher rates.

Keywords

Resonant Energy Transfer (RET), Molecular Auto-Ionization (MAI), Cold Rydberg Gas

1. Introduction

The relaxation of a cold Rydberg gas produced initially by exciting the atoms to

Rydberg states $|n_0\rangle$, with the principal quantum number $n_0 \gg 1$, has been the subject of much recent studies, both theoretically [1]-[7] and experimentally [8]. At low densities, the Rydberg atoms (RyA's) in the gas interact with each other by the dipole-dipole coupling V_{dd}. One of the prime breakdown mechanisms of the gas, mediated by V_{dd}, is the Penning-type molecular auto-ionization (MAI), in which the de-excitation energy of one RyA is transferred to the second one and *ionizing* it. However, at the density of $N_A^{exp} \approx 10^9$ cm^{-3}, the decay rates were observed [8] to be too large, by as much as a factor of two orders of magnitude or more of the MAI rates. Besides, a large number of low energy free electrons as well as readily field-ionizable excited state atoms were detected.

In this paper, we discuss the *resonant energy transfer* (RET), a Forster-type [9] [10] [11] process, in which a pair of RyA's shares its internal energies and creates new *bound* states, one higher and the other lower than the original state $|n_0\rangle$. This RET process is complementary to MAI and is always present if certain energy resonance conditions are satisfied. However, the RET process has been omitted in the past, because its rates are known to be very small compared to that of MAI, especially when n_0 is small and the density of the gas N_A is low. By a simple model, we examine in detail the n_0 and density dependences of the RET rates as well as that of MAI. Especially for the RET, as the density increases, the dipole-dipole interaction V_{dd} between the RyA's may no longer be valid and must be modified.

The density of a cold gas N_A (cm^{-3}) defines the average separation R_N between a pair of atoms, given by $R_N(a_B) \approx 1.9\times 10^8 \left[N_A(\text{cm}^{-3})\right]^{-1/3}$. E.g. $N_A = 10^{11}$ cm^{-3} gives $R_N \approx 4 \times 10^4 a_B$, where a_B is the Bohr radius which we set equal to one. Furthermore, the radius of the RyA is defined *simply* as $r_0 = n_0^2$ and $r_a = n_a^2$, neglecting the factor 3/2 and angular momentum dependence. In discussing the RET, the $r_{aT} = 2r_a$ is the important parameter [1] [3], and we simply define (a) the low density as $R_N > 2r_{aT}$, (b) the moderate density as $2r_{aT} > R_N > r_{aT}$, and (c) the high density as $R_N < r_{aT}$.

2. The MAI at Low Density

For a pair of RyA's A and B, each in state $|n_0, l_0\rangle$, and neglecting the l_0 part, the two-body MAI involves the transition $\{i \to f\}$, where $[i] = \left[(n_0)_A, (n_0)_B\right]$ and $[f] = \left[(n_b)_A, (c)_B\right]$, and where $n_b \ll n_0$ and $(c)_B$ denote the ionized electron from B. At low densities and low temperature of interest here, the motion of the pair and van der Waal's attraction may be neglected. States are specified by [., .], while transitions are specified by {[..] → [..]}.

1) The MAI amplitude. The system Hamiltonian is defined as

$$H = \left(K_{I,II} + K_{1,I} + K_{2,II} + V_{1,I} + V_{2,II}\right) + \left(V_{1,II} + V_{2,1} + V_{12} + V_{I,II}\right) \equiv H_0 + V \quad (2.1)$$

where *I, II* denote the ions and 1, 2 denote the ionized electrons. While being attracted to each other by the van der Waal's potential (which is quadratic in V_{dd} and adiabatic), the MAI proceeds as

$$\{i \to f\} = \left\{(n_0 \to n_b < n_0)_A \left(n_0 \to c_b = continuum\right)_B\right\} \quad \text{(MAI)} \quad (2.2)$$

where simply $[i] = [n_0, n_0]$ and $[f] = [n_b, c_b]$; the transition (2.2) produces a free electron, an ion $II = B^+$, and a bound atom A in state $|n_b\rangle$ with $n_b \ll n_0$. The energy conservation for (2.2) is

$$E_i\left(=2e_0\right) = E_f\left(=e_{nb}+e_{cb}\right) \tag{2.3}$$

where $e_0 = -1/n_0^2$ Ry, $e_{nb} = -1/n_b^2$ Ry, etc, and $e_{cb} \geq 0$ for ionization to take place. The *highest allowed value* n_b^x of n_b is obtained by setting $e_{cb} = 0$ in (2.3), as

$$n_b^x = n_0/\sqrt{2}, \text{ (or the closest integer below it)} \tag{2.4}$$

For $n_0 = 50$, we have $n_b \leq 35 \equiv n_b^x$.

The MAI amplitude (superscript ma) is defined, in the "prior" form, as

$$M_{fi}^{ma} = \left(\Psi_f, V_i\Phi_i\right) \tag{2.5}$$

where $\left(H-E\right)\Psi_f = 0$, and $\left(H_i-E\right)\Phi_i = 0$ with $H_i = H_0$ and $V_i = V$. In the Born approximation and at low densities, we let $\Psi_f \to \Phi_f$ with $\left(H_0-E\right)\Phi_f = 0$, while $V_i \to V_{dd}$, as

$$M_{fi}^{ma} \approx \left(\Phi_f, V_{dd}\Phi_i\right) \tag{2.6}$$

where

$$V_{dd} = 2\left[\left(\boldsymbol{r}_1\cdot\boldsymbol{r}_2\right) - 3\left(\boldsymbol{r}_1\cdot\boldsymbol{\varepsilon}\right)\left(\boldsymbol{r}_2\cdot\boldsymbol{\varepsilon}\right)\right]/R^3, \boldsymbol{\varepsilon} = \boldsymbol{R}/R \tag{2.7}$$

which makes the Born amplitude integral separable in $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$.

2) The total MAI probabilities. The MAI transition probability per unit time is given by

$$P_{fi}^{ma} = 2\pi\left|M_{fi}^{ma}\right|^2 \tag{2.8}$$

The R^{-3} dependence of V_{dd}, causes the P_{fi}^{ma} to decrease as R^{-6}.

The total MAI probability is obtained by *summing* the P_{fi}^{da} *over the restricted set of final states* [2], as

$$\Gamma^{ma} = \sum_{nb=1}^{n_{bx}} P_{fi}^{ma},\ f = \left(n_b, c_b\right) \tag{2.9}$$

In (2.9), *the sum* $\sum_{nb}$ *is over the range* $1 \leq n_b \leq n_b^x$. (For type-setting, we used $n_b^x \equiv n_{bx}$, $c_b \equiv c_{nb}$, $nb \equiv n_b$). It has been established that the transition involving $n_b = n_b^x$ *is the most dominant* [1], as the *both* dipole matrix elements associated with V_{dd} decrease as n_b decreases from n_b^x: the *threshold dominance*. **Figure 1** illustrates two cases, $n_b = n_b^x$ and $n_b = n_b^x - 1$.

It is important to point out that *the states with* $n_b' > n_b^x$ *are omitted* in the sum (Equation (2.9)). They represent the RET, and thus *the* n_b^x *is the dividing point of the spectrum between the MAI and RET*.

3) A Model Calculation. For the actual calculation of the P_{fi}^{ma}, we define the model parameters, as $n_b = 10$, $n_0 = 20$, and $n_a = 50 = 1/k$ for the continuum function. (These parameters are not quite the actual MAI in specific cases, but represent the physics reasonably well, both for MAI and RET.) We let the orbital functions $\varphi = \left(u/r\right)\cdot Y_{lm}$. Since the V_{dd} is separable and proportional to R^{-3}, the amplitude P_{fi}^{ma} can be written in the form

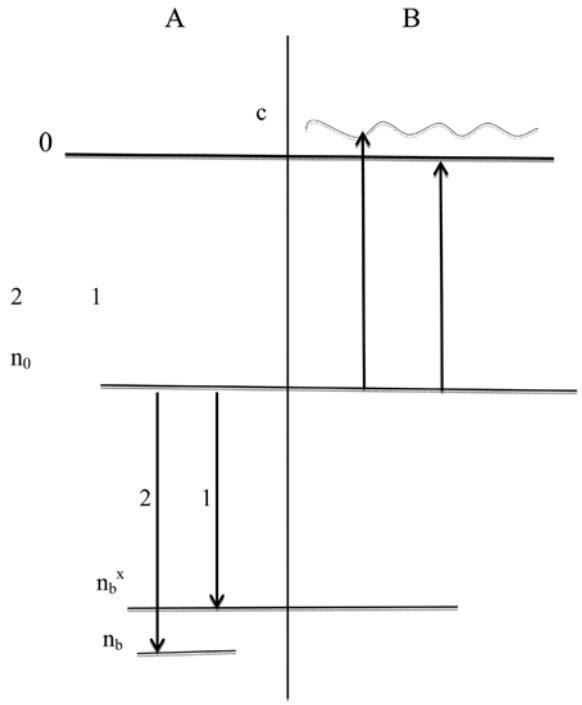


Figure 1. Illustration of levels and transitions for the MAI for two case: Case1: $n_b = n_b^x$, $e_{cb} = 0$; Case2: $n_b = n_b^x - 1$, $e_{cb} > 0$.

$$P^{ma}\left(nl, nl \to n_b l', c_b l'\right) = C_p I_J \cdot J \Big/ R_N^6 \tag{2.10}$$

where $C_p = 8\pi A_y^2$, with $A_y = \left|\int \mathrm{d}\Omega Y_{lm} Y_{lm+m'} Y_{l'm'}^*\right|^2$ for $l' = l \pm 1$. We have A_y = 1/(4π) for $l = 0$ and $A_y = 1/(5\pi)$ for $l = 1$, $m = 0$. Explicitly, $C_{p0} = 1/(2\pi) \approx 0.16$ for $l = 0$, e.g. and $C_{p1} = 8/(25\pi) \approx 0.11$ for $l = 1$. We obtain

$$I_J\left(nl, n_b l'\right) = \int r_1 u_{nl} u_{nbl'} \mathrm{d}r_1 = 3.19 \tag{2.11a}$$

$$J\left(nl, c_b l'\right) = \int r_2 u_{nl} u_{cbl'} \mathrm{d}r_2 = 7.47 \times 10^4 \tag{2.11b}$$

Thus, with $R_N = r_{aT} = 2r_a = 5.0 \times 10^3$ and $R_N^6 = 1.6 \times 10^{22}$,

$$P^{ma} \approx C_{p0} \times 1.5 \times 10^{-17} \tag{2.12}$$

3. The RET at Low Density

In RET, a pair of RyA's, A and B, can exchange some of their internal energy, such that both atoms remain in the *bound states of different n*'s; the process must satisfy the strict *resonance energy condition*. This can be satisfied either by an accidental matching of relevant energies, by a *Stark shift*, or by imposing an external electric field. An estimate of the Stark shift for example shows that, at a given density, high enough n_a can satisfy the resonance condition. In the following, we simply assume that such conditions are met, and focus on the rates.

1) The RET probabilities (superscript *re*). The RET transition involves $[i] = \left[\left(n_0\right)_A \left(n_0\right)_B\right] \equiv \left[n_0, n_0\right]$ and $[j] = \left[\left(n_b'\right)_A, \left(n_a\right)_B\right]$, with $n_b' \ll n_0$ and $n_a \gg n_0$. Thus, we have

$$\{i \to j\} = \left\{\left(n_0 \to n_b' = n_b^x + m'\right)_A \left(n_0 \to n_a \gg n_0\right)_B\right\} \tag{3.1}$$

with $m' = 1, 2, \cdots, n_0 - n_b^x$, and the *resonance energy conditions*

$$E_i = 2e_{n0} = e_{nb'} + e_{na} = E_j, \ n_b' > n_b^x \tag{3.2}$$

The RET probability P_{ji}^{re} for $\{i \to j\}$ is given in the Born approximation as

$$P_{ji}^{re} = 2\pi \left|M_{ji}^{re}\right|^2, \text{ with } M_{ji}^{re} \approx \left(\Phi_j, V_{dd}\Phi_i\right) \tag{3.3}$$

A complete list of allowed n_a for $n_0 = 50$ and $n_b^x = 35$ is obtained from (Equation (3.2)) for the final state of RET $[j] = (n_b', n_a)$: (36, 188), (37, 120), (38, 96), (39, 84), $\cdots$, (48, 52), (49, 51), provided (Equation (3.2)) is satisfied. For the maximum $n_a = 188$, we have $r_a \approx n_a^2 a_B \approx 16 r_0 = 3.5 \times 10^4 a_B$; in this case, R_N must be larger than $8R_0$ to have the V_{dd} applicable. Incidentally, we note that the RET for near-by states transitions, such as (48, 52) and (49, 51), have often been considered [11], but not for maximum n_a. (The RET rates for small n_0 and large n_a are usually very small, but not for large $n_0 < n_a$.)

Generally, the maximum n_a depends sensitively on the lowest n_b', and varies widely, roughly between 150 to 300 and more. For the general discussion, we simply take the typical value $n_a = 200$. Since at $R_N \approx r_{aT} + r_{b'T} \approx r_{aT}$, the "tails" of two orbital functions for $|n_a\rangle$ and $|n_b'\rangle$ start to overlap each other [1] [3], and where the transition probabilities start to increase exponentially, the r_{aT} is the very parameter that can be used to define the regions of different densities, as given in Section 1.

2) The total RET probability. The total probability that includes *the on-shell part of the sum in* $j = (n_b', n_a)$ is given by

$$\Gamma^{re} = \sum_{nb'=nbx+1}^{n0-1} P_{ji}^{re}, \ j = (n_b', n_a), \tag{3.4}$$

where the *sum over* n_b' *is* $n_b^x + 1 \le n_b' \le n_0 - 1$, *provided the resonance condition is satisfied*. (For typesetting we used $nb' = n_b'$, $nbx = n_b^x$, and $n0 - 1 = n_0 - 1$.) This is to be compared with the total MAI probability given by (2.4), $\Gamma^{da} = \sum_{nb=1}^{nbx} P_{fi}^{da}$, where $f' = (n_b, c_b)$. Evidently, Γ^{re} *complements* Γ^{da} as the combined set $[n_b]_f + [n_b']_j$ is complete; the sum includes all values up to n_0 and the contributions from values above n_0 are already included as the states for A are exchanged to that of B.

For the special case with hyper RyA, we can estimate the RET probability P_{ji}^{re} for a two-body system by comparing with the P_{fi}^{ma} of MAI and adopting the quantum defect theory [12] [13]. To have a smooth analytic extrapolation over the threshold $e_c = 0$, the continuum function in the P_{fi}^{ma} is assumed energy-normalized [14] [15] [16]. For densities $R_N > 2r_{aT}$, define

$$Q \equiv P_{ji}^{re}\left(n_a, n_0, R_N\right) \Big/ P_{fi}^{ma}\left(k_{nb} = \frac{1}{n_a}, n_0, R_N\right). \tag{3.5}$$

As expected, the quantum defect theory gives $Q = n_a^{-3}$, for $n_0 < 10$, but slowly increases with n_0, and becomes large.

3) Our model calculation. Consider again the model with the parameters $n_{b'}$ = 10, n_0 = 20, and n_a = 50. This corresponds roughly to one-half of the third set $[j]=(n_b', n_a)=(38,96)$. (Here, or simplicity, we take the same $n_{b'}$ as n_b of MAI.)

We write P_{fi}^{re} *as*

$$P_{ji}^{re}\left(n_0 l, n_0 l \to n_b' l', n_a l'\right) = C_p I_I \cdot I / R_N^6 , \tag{3.6}$$

where C_p is defined earlier, and $l' = l \pm 1$. Taking $I = 0$, we obtain

$$I_I\left(n_0 l, n_b' l'\right) = \int r_1 u_{n_0 l} u_{nb'l'} \mathrm{d}r_1 = 3.19 . \tag{3.7a}$$

$$I\left(n_0 l, n_a l'\right) = \int r_2 u_{n_0 l} u_{n_a l'} \mathrm{d}r_2 = 1.67 . \tag{3.7b}$$

Thus, with $R_N = r_{at} = 5.0\times10^3$ we have

$$P_{ji}^{re} \approx C_{p0} \times 3.3\times10^{-22} . \tag{3.8}$$

This is to be compared with Equation (2.12)

$$P_{fi}^{ma}\left(n_0 l, n_0 l \to n_b l', k = n_a^{-1} l'\right) = C_{p0} \times 1.5\times10^{-17} .$$

In general, $I_I > I_J$ and are nearly of same magnitudes for $n_b = n_b^x$ and $n_b' = n_b^x + 1$.

The main difference between the RET and MAI comes from I and J. They are conveniently compared by

$$Q = P^{re} / P^{ma} = \left(I_I \cdot I\right) / \left(I_J \cdot J\right) \approx I / J , \tag{3.9}$$

especially when $n_b \le n_b^x$ and $n_b' \ge n_b^x + 1$ are close to each other and $n_b^x \gg 1$.

Our model calculation, with V_{dd} and the parameters n_0 = 20, n_b = 10, and n_a = 50, shows that Q is very small for $n_0 < n_a/2$. Evidently, the RET is negligible for low n_0, and this may be the principal reason for neglecting it. More generally, because of the unusual dependence of $I(n_0, n_a)$ on n_0 for a fixed n_a (**Figure 2**), Q behaves approximately as

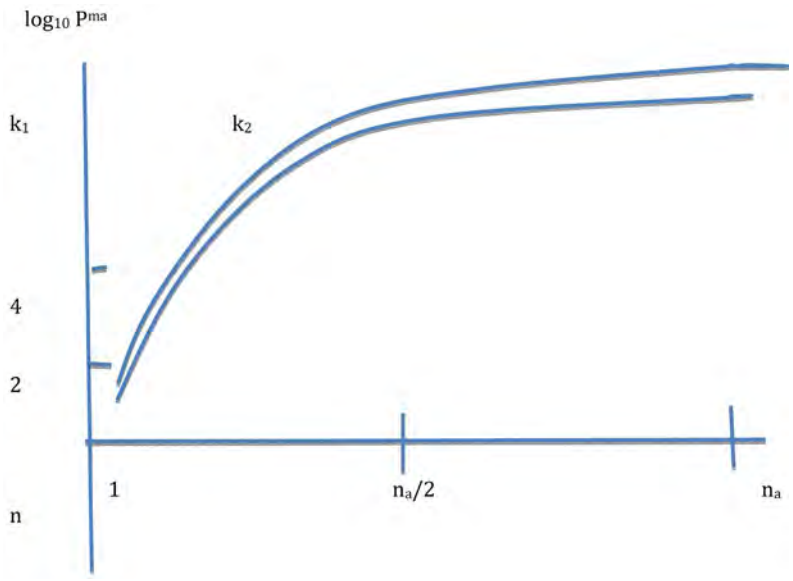


Figure 2. P^{ma}: n_0 and k dependences where $k = 1/n_a$.

$$n_a^{-3} \le Q < n_a^{-2}, \text{ for } 1 \le n_0 < n_a/4 \quad \text{(QDT)}, \tag{3.10a}$$

$$n_a^{-2} < Q < 1, \text{ for } n_a/4 < n_0 < 3n_a/4, \tag{3.10b}$$

$$1 < Q < n_a^2, \text{ for } 3n_a/4 < n_0 < n_a. \tag{3.10c}$$

For applications, (3.10b) is the most relevant. The above trend continues to hold for larger n_a.

4. The RET at High Densities

The RET probabilities discussed in the previous section are for low density, with $R > 2r_{aT}$, where the dipole-dipole coupling V_{dd} is effective. However, as the density increases and R approaches r_{aT}, the V_{dd} is no longer adequate in so far as treating the RET is concerned. By contrast, for the MAI, the V_{dd} should be valid for $R_N \ge r_{aT}/4 \ge R_0 = 4r_0$. Thus, the result of Section 2 is unchanged; in fact, the MAI does not involve hyper-Rydberg states.

1) The RET amplitude (superscript h for high density) via V_{12}. We write the RET amplitude in the "post" form as

$$M_{ji}^{re-h} = \left(\Phi_j^M, V_j \Psi_i\right), \tag{4.1}$$

where Φ_j^M now satisfies $\left(H_j - E\right)\Phi_j^M = 0$, with $H_j = H - V_{12}$ and thus $V_j = V_{ee} = V_{12}$. Note that the Φ_j^M is a two-center molecular orbitals. In the Born approximation for the $\Psi_i \to \Phi_i$, we have (superscript h for high density) for the post form

$$M_{ji}^{re-h} = \left(\Phi_j^M, V_{12}\Phi_i\right), \tag{4.2}$$

and the RET probability is

$$P_{ji}^{re-h} = 2\pi \left| M_{ji}^{reM} \right|^2. \tag{4.3}$$

2) Evaluation of the amplitude integrals—approximations. The integrals involved in the M_{ji}^{re} are rather complex, and in the following, we make some simple approximations to obtain an estimate of the amplitude. Firstly, the molecular Φ_j^M is replaced by its LCAO (Linear Combination of Atomic Orbitals) form, Φ_J which is a solution of H_0; this will give only a crude estimate of the amplitude. Secondly, the variables involved in the $V_{12} = 1/r_{12}$ are complicated, as $r_{12} = \left|\boldsymbol{r}_{2I} - \boldsymbol{r}_1\right|$ with $\boldsymbol{r}_{2I} = \left|\boldsymbol{R} + \boldsymbol{r}_2\right| \equiv s_2$, or $r_{12} = \left|\boldsymbol{r}_2 - \boldsymbol{r}_{1,II}\right|$ with $\boldsymbol{r}_{1,II} = \left|\boldsymbol{r}_1 - \boldsymbol{R}\right| \equiv s_1$. We first set the dipole case, as $V_{12} \to Y_{1m}(1)Y_{1m}^*(2)r_<'/r_>'^2$, where $r' = r_2$ or s_1. Second, we let all the orbital functions be in the form $\varphi = (u/r)\cdot Y_{lm}$. To further simplify the task, e.g., let s_1 to be strictly a function of r_1 and R, but not its angular dependence. Then, the angular parts become trivial, and we have

$$P_{ji}^{re-h} = C_p \times (1/4) \times \left|T^{re}\right|^2,$$

where the C_p was defined earlier, and the amplitude integral is given by

$$T^{re} = \int_0^\infty \mathrm{d}r_2 u_{nl} u_{n_a l'} \left\{ \left(1/r_2^2\right) \int_0^{r_2} \mathrm{d}r_1 s_1 u_{nl} u_{nb'l'} + r_2 \int_{r_2}^\infty \mathrm{d}r_2 u_{nl} u_{nb'l'} / s_1^2 \right\} \equiv T_1 + T_2$$

Various simplifying approximations are considered: (si) $s_1 = r_1$, (sii) $s_1 = 2n_0^2$,

(siii) $s_1 = 2n_a^2$. All three cases make the angular integrals simple.

We also define for convenience, $W_1 = \int_0^{r_2} \mathrm{d}r_1 s_1 u_{nl} u_{n_a l'}$ and $W_2 = \int_{r_2}^{\infty} \mathrm{d}r_1 u_{nl} u_{nb'l'} / s_1^2$ Then, e.g. for (si) $W_1 \sim \int_0^{r_2} \mathrm{d}r_1 r_1 u_{nl} u_{n_a l'}$ and $W_2 \sim \int_{r_2}^{\infty} \mathrm{d}r_1 u_{nl} u_{nb'l'} / r_1^2$, etc.

The results for the model are:

si: $T_1 = 7.11 \times 10^{-7}$, $T_2 = -1.68 \times 10^{-7}$; $T^2 = 2.9 \times 10^{-13}$, Molecular

sii: $T_1 = 2.54 \times 10^{-5}$, $T_2 = 5.73 \times 10^{-9}$; $T^2 = 6.45 \times 10^{-10}$, n_0 Overlapping

siii: $T_1 = 1.59 \times 10^{-4}$, $T_2 = 1.47 \times 10^{-10}$; $T^2 = 2.53 \times 10^{-8}$, n_a touching.

5. Summary and Conclusion

We have presented the RET as the potentially important process that affects the relaxation of cold Rydberg gas. This process has been neglected in all the previous studies of the decay of the gas, presumably because of its extreme low rates at small n_0. Our present study has shown the importance of the RET, especially because of the large physical size of the hyper-Rydberg state; it basically changes the effective density by many folds.

The MAI is shown to be the dominant process at low density, $R_N > 2r_{aT}$, where the V_{dd} is effective, while the RET is always present, but at very low probabilities. The n_0 dependence of the P^{re} and P^{ia} is studied in detail; for the initial excited state $|n_0\rangle$ not too high, $1 \le n_0 < 3n_a/4$. The ratio $Q = P^{re}/P^{ia}$ is found to be very small, of the order of n_a^{-3} to n_a^{-1}. The lower limit of Q follows from the quantum defect theory, which is approximately valid for $n_0 < 10$, but starts to break down for higher n_0. As the n_0 approaches n_a, Q increases to one, and grows rapidly, to as much as n_a^2.

For the cold Rydberg gas at moderate density, $r_{aT} < R_N < 2r_{aT}$, the wave function φ_a starts to overlap with the neighboring RyA's, and the V_{dd} is no longer applicable. It is replaced by the electron-electron interaction V_{ee}. The modified RET probabilities, P_{ji}^{re-M}, are estimated in several approximations, all of which indicate the resulting Q^M to be much larger than the Q, of the order to n_a. As expected, the P_{ji}^{re-h} at high density is much larger, with the overlap of orbital wave functions.

The RET can create a hyper RyA of large size, which in turn immediately forms a giant auto-ionizing clusters, *enveloping many near-by atoms*, and contains a huge amount of excess internal energies, making it highly unstable. Multiple production of clusters in the gas leads to a cascade decay of the gas. This problem requires a careful analysis with rate equations.

The dominance of states near the ionization threshold both in the MAI and RET processes yields multitude of low energy free electrons, via MAI, and many weakly bound electrons, via RET. These results are consistent with the experimental observation [8].

Acknowledgements

This research work was supported by the National Science Foundation (NSF). The award number is 1901397.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Hahn, Y. (2000) *Journal of Physics B*, **33**, L655. https://doi.org/10.1088/0953-4075/33/20/101

[2] Amthor, T., *et al.* (2009) *The European Physical Journal D*, **53**, 329. https://doi.org/10.1140/epjd/e2009-00119-4

[3] Kiffner, M., *et al.* (2016) *Journal of Physics B*, **49**, Article ID: 204004. https://doi.org/10.1088/0953-4075/49/20/204004

[4] Robicheaux, F. (2005) *Journal of Physics B*, **38**, S333. https://doi.org/10.1088/0953-4075/38/2/024

[5] Amthor, T., *et al.* (2007) *Physical Review A*, **76**, Article ID: 054702.

[6] Amthor, T., *et al.* (2007) *Physical Review Letters*, **98**, Article ID: 023004.

[7] Robicheaux, F., *et al.* (2014) *Physical Review A*, **90**, Article ID: 022712.

[8] Tanner, J.T., *et al.* (2008) *Physical Review Letters*, **100**, Article ID: 043002.

[9] Forster, Th. (1948) *Annalen der Physik*, **437**, 35. https://doi.org/10.1002/andp.19484370105

[10] Maineult, W., *et al.* (2016) *Journal of Physics B*, **49**, Article ID: 214001. https://doi.org/10.1088/0953-4075/49/21/214001

[11] Comparat, D. and Pillet, P. (2010) *Journal of the Optical Society of America B*, **27**, A208. https://doi.org/10.1364/JOSAB.27.00A208

[12] Hahn, Y. (2002) *Physics Letters A*, **293**, 266. https://doi.org/10.1016/S0375-9601(01)00854-4

[13] Bethe, H.A. and Salpeter, E.E. (1957) Quantum Mechanics of One- and Two-Electron Atoms. Springer-Verlag, Berlin, 264. https://doi.org/10.1007/978-3-662-12869-5

[14] Hahn, Y. (1977) *Physical Review Letters*, **39**, 82. https://doi.org/10.1103/PhysRevLett.39.82

[15] Hahn, Y. (1985) *Advances in Atomic and Molecular Physics*, **21**, 123. https://doi.org/10.1016/S0065-2199(08)60142-6

[16] Seaton, M.J. (1955) *Comptes Rendu*, **240**, 1317.

Journal of Modern Physics, 2021, 12, 1219-1236
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

What Do Bell-Tests Prove? A Detailed Critique of Clauser-Horne-Shimony-Holt Including Counterexamples

Karl Hess

Center for Advanced Study, University of Illinois, Urbana, Illinois, USA
Email: karlfhess@gmail.com

How to cite this paper: Hess, K. (2021) What Do Bell-Tests Prove? A Detailed Critique of Clauser-Horne-Shimony-Holt Including Counterexamples. *Journal of Modern Physics*, **12**, 1219-1236.
https://doi.org/10.4236/jmp.2021.129075

Received: June 8, 2021
Accepted: July 6, 2021
Published: July 9, 2021

Abstract

Many future directions of scientific endeavors depend on quantum theory and the precise interpretation and significance of the entanglement of quantum-particles. This interpretation depends in turn on the physical meaning of so called Bell-tests that are mostly performed using entangled photons and randomly switched polarizers to measure their polarization at distant locations. This paper presents a detailed critique of the well known theory of Bell tests given by Clauser-Horne-Shimony-Holt (CHSH). It is demonstrated that several important steps of the CHSH derivations contain serious inaccuracies of the underlying physics and probability theory and even a calculus error. As a consequence, the Bell-CHSH theory cannot be used to demonstrate extreme and opposite interpretations of entanglement such as super-luminal influences or alternatively super-determinism that cast aspersions on Einstein's concepts of locality and separability.

Keywords

Bell Theorem, Clauser-Horne-Shimomy-Holt, Einstein-Podolsky-Rosen

1. Introduction

If one asks for a list of significant problems in branches of current science, one is bound to find pointers toward developments of quantum mechanics that include the concept of quantum entanglement. Quantum computing, quantum teleportation and even topics in biology are using quantum entanglement for their basic considerations. The concept of quantum entanglement was defined and described by Schroedinger and has been shown to represent a main feature of quantum theory itself.

Quantum entanglement of isolated and distant pairs has been experimentally demonstrated by Kocher and Commins [1] and has also been discussed by Nordén [2], with a perspective on questions that arose around Bell's theorem.

The most intriguing questions related to the Bell theorem emerged from measurements by Aspect [3] and Zeilinger and coworkers [4]. They have suggested links of quantum entanglement to superluminal influences based on their work; a daring suggestion that also has given rise to the astounding proposition of quantum teleportation.

The basic question of the possible existence of superluminal influences is actually contained in many a scientific theory, for example that of Newton, and has seen many discussions and opposing views (of Leibnitz, in Newtons case). Since the event of Einstein's relativity, ideas of superluminal influences became unthinkable, but have been revived in the last decades based on Bell's Theorem and the experimental Bell-tests. Recently a sensational Bell-test has been performed by Zeilinger and related groups [4] involving photons from billions of years past as well as the "free will" of fellow researchers [5].

To be sure, no direct proof has ever been provided by any experiments. We do not have any measurement equipment available that can claim to transmit information faster than the speed of light in vacuum and thus we cannot measure superluminal influences in one single shot, as is well known from Einstein's special theory of relativity. The recent ideas of superluminal influences are based on the statistical results underlying the Aspect-Zeilinger type of experiments. These researchers claim that the absolute randomness and ultrafast switching between their distant measurements provides the definite proof for their assertions. This fact has even been discussed in popular TV-shows such as "Einstein's Quantum Riddle"—NOVA, which describes the Einstein-Podolsky-Rosen-type (EPR) experiments that are the basis for the work of Aspect [3], Zeilinger [4] [5] and related groups.

I show in this paper, that among other factors there are certain aspects of the random switching that render Bell-tests questionable. In fact, such switching was not included in any direct way into the theories that describe Bell tests, but has only been recommended to the experimenters in order to support Einstein locality. To demonstrate these facts, I discuss in great detail and in fairly elementary mathematical terms, the important theoretical framework of Clauser, Horne, Shimony and Holt (CHSH) [6].

Each of the following sections starts with verbatim-quotes or at least extremely faithful descriptions of the most relevant definitions, assumptions, inequalities and corollaries given in the original CHSH work. The criticisms and comments follow.

It is demonstrated that the main results of CHSH are based on assumptions lacking generality and accuracy and that their mathematical and physical steps contain several serious problems and even a calculus error. Their results are further invalidated by their lack of use of modern probability theory and particularly the work of Vorob'ev [7], which demonstrates that their inequality (1b)

does not follow from their inequality (1a) without extensive conditions that CHSH were not aware of. CHSH inequality (1b) provides the important connection of the CHSH work to the results of quantum mechanics and to the experimental results.

My criticism (of the CHSH theorem that has been scrutinized for more than 50 years) may appear too harsh and even suspect to some, because physically speaking, the theorem is based on innocuous and plausible assumptions. Furthermore, the theorem is undoubtedly correct within certain mathematical frameworks. The physical consequences of the theorem, however, are such that only the extremes of super-luminal effects and/or super-determinism appear as possible explanations. This fact opens the question whether the physical assumptions to derive the theorem are really that innocuous and whether the mathematical premises correspond to the physics of actual experiments (see also [8] [9] and references therein).

The seriousness of this question is demonstrated below by explicit counter-examples to the CHSH theory, including one that invalidates the reasoning in the TV-show "Einstein's Quantum Riddle"—NOVA.

2. Connection of CHSH to Bell's Theorem and EPR Experiments

2.1. CHSH Quotation

"—Consider an ensemble of correlated pairs of particles moving so that one enters apparatus I_a and the other apparatus II_b, where a and b are adjustable apparatus parameters. In each apparatus a particle must select one of two channels labeled +1 and −1. Let the result of these selections be represented by $A(a)$ and $B(b)$, each of which equals ±1 according as the first or second channel is selected."

2.2. Criticism

To work within the aims of Bell-CHSH, we are bound to use Einstein-type space-time (or, in approximation, space and time) physics. Thus, Bell-CHSH have replaced the operators and eigenvalues of quantum theory by functions $A(a,...), B(b,...)$ with the "..." being not yet fully specified, but later related to the information λ from a source (see below).

The choice of two integer numbers to define two channels and the codomain of the corresponding functions $A, B = \pm 1$ is innocuous and contradiction-free if indeed only one pair of apparatus parameters (fixed a, b) is in discussion. However, the identical choice of +1 or −1 for all possible equipment (apparatus) settings and in particular the 4 different equipment settings used by CHSH (see below) represents an inaccuracy that leads to inconsistencies, because channels pointing into different directions are denoted by the same integer. The application of integer algebra leads, in fact, to direct contradictions, as is shown in expression (5) below.

The inaccuracy of CHSH (and Bell) to deal with only 2 equipment settings, while the bulk of their paper deals with 2 settings in each station, is further aggravated by the random switching between the settings in each station: a, a' in I_a and b, b' in II_b. (Please note that we are using throughout the symbols a, a' and b, b' for the polarizer or magnet settings, as is usual in more recent publications. CHSH have used different symbols in their original paper.) The random switching had been suggested to the experimenters already by Bell and was realized by Aspect and coworkers [3] and in a sensational way, involving photons from billions years past, by Rauch and coworkers [4].

However, the possibility of random switching has not been included into the mathematical formalism of the theories of Bell and CHSH. From the viewpoint of probability theory, they needed to introduce random variables j, j' with the possible outcomes $j = a, a'$ and $j' = b, b'$ in the respective stations. Thus, it is nontrivial to write the functions A, B in terms of these random variables and to equate all possible outcomes to ± 1: $A(j,...) = \pm 1$ and $B(j',...) = \pm 1$, because the +1 or −1 may then indeed describe channels that point in different directions, depending on the actual setting outcome. That inconsistency is hidden in the Bell-CHSH-type theories, because they only deal with the actual outcomes $a, a'; b, b'$ and leave the switching reserved for the experimenters.

We, therefore, ask why different geometric equipment-arrangements need only be described by two integer numbers and follow their algebra? In the authors' opinion, this oversimplification has been accepted by almost everyone, because of the illusion that we deal with something analog to the eigenvalues in quantum mechanics, while the Bell and CHSH formalism must not be quantum mechanical as emphasized by both Bell and CHSH (see next section below).

Experiments with polarized photons are usually characterized by descriptions such as [*horizontal*, *a*, …] instead of just +1 and [*vertical*, *a*, …] instead of −1 for a given setting *a* and analogous notations for the other settings. Thus, the experimenters imply a connection between the definition of *horizontal* or *vertical* to the equipment setting directions *a* (or *b* etc.). A more precise theoretical notation would, therefore, use *horizontal*^{a} and *vertical*^{a} and similar notations (such as *horizontal*^{b}) for the other settings, if one wishes to recognize the fact that the polarizer direction co-determines what is meant by *horizontal* and *vertical*.

Wigner [10] did notice and correct part of this problem. He and later d'Espagnat [11] have made use of group theory and a more general codomain for the functions A, B. They counted the *equal* and *not-equal* outcomes in the two stations and produced an inequality involving the counted numbers of *equal* vs *not-equal* results for a variety of different equipment setting pairs. The Wigner-d'Espagnat procedure is more general than that of Bell and CHSH. However, they did not account for all consequences of the random switching: As explained, one needs to have a consistent understanding what the expressions *horizontal* and *vertical* mean for the entire experimental system and for all pair-measurements. The Wigner-d'Espagnat inequality deals with 3 pair-observations. Such an experiment has not been performed yet and their approach leads to contradictions as

discussed in reference [12].

Strictly speaking, one may only use the judgements "*equal*" and "*not-equal*" for parallel polarizers or Stern-Gerlach magnets in both stations. It is still possible to introduce different definitions such as *horizontal*^{a} and *vertical*^{b} in the respective separate stations. However, a precise notation would then also use superscripts such as *equal*^{a,b} or *not-equal*^{a,b} and similar superscripts for other settings. The use of superscripts from both stations is completely acceptable, because the notions of *equal* and *not-equal* can only be introduced after merging the data from both stations. Thus, one cannot exclude probability measures for the results *equal* and *not-equal* that are functions of the polarizer-settings of both stations and even of the angle between those settings. The standard objection that experimenters Alice and Bob in the respective stations know nothing about each other does simply not apply to the theoretician who deals with the merged data from both EPRB stations and compares them.

Similarities with relativity theory do surface here. If we have two spaceships with pilots Alice and Bob, respectively, who know absolutely nothing of each other, no correlation of any physical processes in the spaceships can ever be found. The times displayed by their respective clocks are unknown until investigated, for example, by light-beams tracking the spaceships; alternatively the clocks may be compared when the spaceships are brought together. Only in these ways may their clocks be correlated and found to depend on the difference of their velocity-histories (see also [12]), which represent, of course, a nonlocal piece of information.

As done with the velocities in relativity, one may also fix the polarizer in one station and vary the angle between the polarizers settings (see **Appendix** and references [13] as well as [14]).

Summarizing, one can state that the Bell-CHSH choice of $A, B = \pm 1$, implying that the function values can be handled as two integer numbers for all possible equipment settings, oversimplifies the actual physical facts of the Einstein-Podolsky-Rosen (EPR) type experiments. The extensive repair work of Wigner and d'Espagnat represents a major improvement, but did not demonstrate a general validity of the CHSH work, particularly not for pair measurements with multiple random polarizer settings on both sides (as shown in detail in sections below).

3. Information from the Source

3.1. CHSH Quotation

"—Suppose now that the statistical correlation of $A(a)$ and $B(b)$ is due to information carried by and localized within each particle, and that at some time in the past the particles constituting one pair were in contact and communication regarding this information. The information, which emphatically is not quantum mechanical, is part of the content of a set of hidden variables, denoted collectively by λ. The results of the selections are

then to be deterministic functions $A(a,\lambda)$ and $B(b,\lambda)$. Locality reasonably requires $A(a,\lambda)$ to be independent of the parameter b and $B(b,\lambda)$ to be likewise independent of a, since (sic) the two selections may occur at an arbitrary great distance from each other."

3.2. Criticism

We only note a minor inaccuracy: λ is defined as a "set of hidden variables" and certainly may be used as such. However, in most of what follows in the CHSH paper, the same symbol λ is used for the possible values of that hidden variable, which represent Einstein's elements of physical reality. This twofold meaning of λ has led to some confusions in the literature. We attempt to present all the following explanations in a way that avoids confusion. However, we still continue to use λ with the dualistic meaning that it has in the CHSH paper (variable as well as value of variable), in order not to deviate too much from their notation.

4. Involvement of Probability

4.1. CHSH Quotation

"—Finally, since the pair of particles is generally emitted by a source in a manner physically independent of the adjustable parameters a and b, we assume that the normalized probability distribution $\rho(\lambda)$ characterizing the ensemble is independent of a and b."

4.2. Criticism

The possible equipment settings a,b are still regarded as "adjustable parameters" that somehow also encompass the random switching. However a,b as well as all other equipment settings are used in the equations below as given. As mentioned, the introduction of random variables $j=a,a'$; $j'=b,b'$ is a necessity if random switching ought to be included. Furthermore we need to introduce a probabilistic tool that accounts for the physical fact that measurements with different setting pairs in any given station cannot occur simultaneously. The kinematics of random switching and any dynamics of the many body physics involved in the interactions between incoming particles and measurement equipment (see also reference [15]), can generally not be described by the functions and probability density as introduced by Bell-CHSH. For example, the possible values of λ may be all different and, therefore each value λ interacts exclusively with a single pair of polarizer settings.

The only way to cover such a situation consistently with a classical probability theory, such as the framework of Kolmogorov, is by use of stochastic processes [16] [17]. Thus, a general treatment of the random switching of setting pairs requires the introduction of a time-like variable t and a two dimensional vector stochastic process of the kind $\left[A(a,\lambda,t),B(b,\lambda,t)\right]$. The probabilistic Bell-CHSH approach is not general enough, because it cannot include such stochastic

processes [17].

This latter fact becomes visible in the immediately following mathematical steps of CHSH. They use in all their algebraic expressions the identical value of λ, but now for different pair-functions that correspond to different pair measurements as well as different polarizer setting-pairs. This procedure has been justified in two ways. First, by counterfactual reasoning of the kind "had we measured with a different setting pair, we would have encountered the same λ. We discuss this way in our criticism following the mathematical steps and the second justification in more detail afterwards.

5. Mathematical Steps of the CHSH Derivation; All with Identical λ

5.1. CHSH Quotation

Defining the correlation function

$$P(a,b)=\int_{\Gamma}A(a,\lambda)B(b,\lambda)\rho(\lambda)\mathrm{d}\lambda, \tag{1}$$

where Γ is the total λ space, we have

$$\begin{aligned}&\left|P(a,b)-P(a,b')\right|\\ &\le\int_{\Gamma}\left|A(a,\lambda)B(b,\lambda)-A(a,\lambda)B(b',\lambda)\right|\rho(\lambda)\mathrm{d}\lambda &(2)\\ &=\int_{\Gamma}\left|A(a,\lambda)B(b,\lambda)\right|\left[1-B(b,\lambda)B(b',\lambda)\right]\rho(\lambda)\mathrm{d}\lambda &(3)\\ &=\int_{\Gamma}\left[1-B(b,\lambda)B(b',\lambda)\right]\rho(\lambda)\mathrm{d}\lambda &(4)\end{aligned}$$

5.2. Critizism

5.2.1. Failure of Counterfactual Reasoning

The steps from Equation (2) to Equation (4) use the axioms of integer numbers as well as the concept of the absolute value that have no obvious validity for the more general outcomes of *horizontal* and *vertical* and "products" of them. This fact does not matter for Equation (1), because we may adopt a convention to just subtract the numbers of *equal* and *not-equal* products as Wigner did. However, the algebraic steps from Equations (2) to (4) lead to the product:

$$B(b,\lambda)B(b',\lambda) \tag{5}$$

that now exhibits identical λs for two elements of reality that must be, generally and physically speaking, different, because both functions B symbolize measurements in the same station for different settings. Here we encounter a spectacular failure of the counterfactual reasoning and the assumption that the same λ may be used in the mathematical expressions above. Expression (5) is definitely a red flag for the basic assumptions that govern the domain and co-domain of the Bell-CHSH functions and their products.

5.2.2. Reordering the Possible Outcomes

Bell, CHSH and their followers have attempted to circumvent some criticism of

the counterfactual arguments in the following way, which agrees also with the main reasoning in references [3] [4] [5] as well as many textbooks:

Because λ represents the possible outcome-value of measurements, the claim is made that one can reorder these possible outcomes in such a way that about all the actual outcomes may be arranged in quadruples, each with the same value λ_n to obtain the following inequality:

$$\begin{aligned} -2 \le A(a,\lambda_n)B(b,\lambda_n) - A(a,\lambda_n)B(b',\lambda_n) \\ + A(a',\lambda_n)B(b,\lambda_n) + A(a',\lambda_n)B(b',\lambda_n) \le +2. \end{aligned} \tag{6}$$

where $n = 1,2,3,\cdots,N$, with N being a large number. Inserting all possible values of ± 1 for the functions A, B supplies immediate verification. CHSH correctly have deduced from the random switching that Γ, the set of all possible λ_n, must be independent of the equipment-settings, because of Einstein locality. They maintain, as the majority of experts do, that this fact justifies the reordering of data into a large number of quadruples each with the same outcome-value λ_n. We will see below that this type of reordering indeed leads to an inequality that is only slightly different from the one derived by CHSH. As and aside, the possibility of reordering is actually mathematically only guaranteed for countable values of λ_n. If the variable λ represents a continuum of some form, reordering may not be a valid procedure [16]. However, for our current purpose this problem may be ignored, because of additional reasons that supersede these finer points.

The additional reasons will be discussed in detail when connecting the CHSH-Bell-type inequalities to quantum theory and the actual experiments after inequality (11) below. These reasons follow from the findings of the mathematician Vorob'ev and are also explained in the **Appendix**. We first proceed, however, with the derivation of the CHSH inequalities.

6. The CHSH Inequalities

6.1. Remark on a Calculus Error of CHSH

CHSH use the following assumptions and equations to derive their original inequality from the inequality derived in Equations (1) - (4).

CHSH assume $P(a',b) = 1-\delta$ with $0 \le \delta \le 1$ and deduce

$$\int_{\Gamma_-} \rho(\lambda)\mathrm{d}\lambda = \frac{\delta}{2} \tag{7}$$

with Γ_- representing the set of all λ for which $A(a',\lambda) = -B(b,\lambda)$. In addition CHSH imply

$$\delta = -2\int_{\Gamma_-} A(a',\lambda)B(b',\lambda)\rho(\lambda)\mathrm{d}\lambda, \tag{8}$$

Criticism

Equation (8) is only valid if also $P(a',b') = 1-\delta$. CHSH assume Γ_- to be independent of the setting-pairs, which is incorrect. In this way they derive the

following inequality.

6.2. The Original CHSH Inequality

"And therefore

$$\left|P(a,b)-P(a,b')\right| \le 2-P(a',b)-P(a',b')" \tag{9}$$

6.3. Comment

Although a calculus-error was made to derive this original CHSH inequality (9), the error is largely inconsequential, because it is removed by the additional precondition that $P(a',b')=1-\delta$, which is just more restrictive. This original CHSH inequality does, however, still suffer from the problems with the product (5) given above.

The nowadays mostly used variation of the CHSH inequality is not encumbered by the latter problem and is only somewhat "weaker" than (9):

$$\left|P(a,b)-P(a,b')+P(a',b)+P(a',b')\right| \le 2. \tag{10}$$

This inequality follows immediately from inequality (6) by summation over all n with $1 \le n \le N$. It does involve the reordering discussed above.

7. Connection to the Result of Quantum Theory and Experiment

7.1. CHSH Quotation

CHSH made the connection of their inequality to quantum theory by noting that $P(a,b)$ is a function of $b-a$. This function depends on the actual entanglement (pair-correlations) and is typically given by $-\cos(b-a)$ and similar functions for the pairs a,b', a',b and a',b', which thus are functions of the angle Δ between the polarizer settings; a result that is obtained by quantum theory and also corroborated experimentally. CHSH further note that Equation (9) can be written by using only three differences resulting in three numbers that commonly are called Bell- or CHSH-angles. This fact follows from the cyclical arrangement of the CHSH polarizer settings, which means that three setting pairs fully determine the fourth. CHSH arbitrarily chose $\alpha = b-a$, $\beta = b'-b$ and $\gamma = b-a'$ to obtain CHSH Equation (1b):

$$\left|P(\alpha)-P(\alpha+\beta)\right| \le 2-P(\gamma)-P(\beta+\gamma) \tag{11}$$

7.2. Criticism

Seen from the viewpoint of rotational invariance, the angles between the polarizer settings indeed appear as the important physical variables, while the setting directions for themselves do not [14]. Thus, the introduction of the Bell-CHSH angles (or more generally scalar vector products) is indeed a necessity in order to properly compare Bell-CHSH-type theories with quantum theory and with actual measurements and experiments. This fact has, however, far reaching con-

sequences that invalidate inequality (11).

7.3. Vorob'ev's Necessary Condition

Vorob'ev's work [7] represents not just another way to prove CHSH-Bell-type inequalities for functions on a common probability space. Its main corollary is that no constraint of the CHSH-Bell-type can be derived or proven without the existence of what Vorob'ev calls a combinatorial topological cyclicity such as the one just described above in terms of the cyclical arrangement of the polarizer settings.

The term "cyclicity" expresses the precise appearance and recurrence of the equipment settings in the different terms of inequalities such as (10), which according to Vorob'ev are a sine qua non for the constraints that can be deduced in form of inequalities or otherwise. As a consequence of the cyclicity only three of the Ps in inequality (10) (e.g. $P(a,b), P(a,b'), P(a',b)$) may be chosen freely within their given codomain $-1 \le P \le +1$, the fourth, $P(a',b')$, cannot be freely chosen, which is the exact reason for the constraints that are expressed by the CHSH inequality.

CHSH were obviously not aware of the work of Vorob'ev [7], who proved in general terms that the validity of inequality (11) is conditional to the cyclicity of the equipment settings that label the expectation values (correlation functions (1)) in the inequalities (9) and (10). The specialization of Vorob'ev's more general work to Bell-CHSH inequalities has been discussed in in detail in references [17] [18]. A short review of the significance of Vorob'evs findings for the validity (or lack of validity) of inequality (11) is given in **Appendix**.

7.4. Removal of the Vorob'ev Cyclicity

This cyclicity, necessary for any constraint on the expectation values, is usually not mentioned in discussions of Bell-type and CHSH-type inequalities, because most of these discussions stop at inequality (10) and do not continue to describe the link with quantum theory and experiment in any detail. Inequality (10) contains the cyclicity automatically. When the step to use Bell- and CHSH-angles is taken, this is no longer the case and the inequality (11) is then invalid for the following reasons.

The choice of the three CHSH angles α, β, γ to obtain inequality (11) is arbitrary and a multitude of other choices could have been made to obtain inequalities different from (11). Their choice also does not guarantee the cyclicity of the original equipment settings. In fact, going back from inequality (11) to inequality (10), one can choose for each term an infinity (cardinality of the real numbers) of non-cyclical setting arrangements $\left[a'',b'';a''',b''';a^{iv},b^{iv};a^{v},b^{v}\right]$ that express the same four angles:

$$b''-a''=b-a;\ b'''-a'''=b'-a;\ b^{iv}-a^{iv}=b-a' \text{ and } b^{v}-a^{v}=b'-a'. \quad (12)$$

For these modified equipment settings, a CHSH inequality does not exist, be-

cause of the lacking cyclicity.

8. Completely Random Polarizer Settings and Counterexamples to Inequality (11)

The claim of using completely random polarizer settings in references [4] [5] and other works of Aspect, Zeilinger, Giustina and coworkers, is misleading. This fact can be seen with particular clarity in the TV-show "Einstein's Quantum Riddle"—NOVA, which creates the illusion that two absolutely random polarizer settings are used on two respective islands and the strong correlation of the measurement outcomes have, therefore, only one explanation: instantaneous influences from one island to the other, just the kind of influences that Einstein called "spooky". In an attempt to appear entirely convincing, the randomness of the setting-choices is derived from photons of billion years past [4] and from the "free will" of collaborating researchers [5]. This illusion of randomness, however, can only be created for a small number of polarizer settings, two in the published cases. These two settings are, however, carefully pre-chosen and pre-determined such that the polarizers on the two islands always exhibit one of the four desired Bell-CHSH angles; independent of the random switching. The relative position of the polarizers to each other is, thus, not random at all.

Based on the findings in the section "Removal of the Vorob'ev cyclicity", it will be shown immediately that measurements with large numbers of random polarizer settings leave only negligibly few setting combinations for which inequality (11) remains indeed valid and we will discuss in the remarks at the end of this section how even these few exceptions may be dealt with to avoid any vestige of hints toward instantaneous influences between islands.

Note that the following counterexamples to the validity of inequality (11) do not contest the independence of the set Λ of all λs from the polarizer settings. The counterexamples are merely based on the use of a much larger number of random settings and in some cases on the fact that the correlation functions P are by law of nature invariant to rotations of the polarizer-pairs involved in any specific pair-measurement, while the Vorob'ev cyclicity represents a mathematical abstraction that is not subject to laws of nature.

8.1. Counterexample 1

First, use a large number of completely random polarizer settings in both measurement stations (wings). In this case we encounter, with only negligible probability, measurement-pairs that exhibit CHSH angles. We may select these measurements-pairs in order to form a CHSH-quadrupel. However, even among these selected quadruples, it is only a negligible number that can be arranged into a Vorob'ev cyclicity. Thus the necessary cyclicity that validates (11) is rarely encountered (with probability close to 0). All of this follows immediately from the algebra discussed in the subsection "Removal of the Vorob'ev cyclicity". We further discuss in the remarks below, how even this negligible set of measurements that form a Vorob'ev cyclicity may be dealt with.

8.2. Counterexample 2

Second, use again a large number (instead of just two as the Aspect- and Zeilinger-groups do) of completely random polarizer settings in one of the EPRB wings and use corresponding settings in the other wing that conserve (with equal frequency of occurrence) the 4 CHSH angles between the two polarizer directions and thus the 4 pair correlations P. The overwhelming majority of the corresponding CHSH-quadruples will again not exhibit any Vorob'ev cyclicity as explained above and may thus violate inequality (11). Large numbers of such measurements may indeed be performed by using the techniques of Giustina *et al.* [19] that include electro optical modulators. One may then ask the question why such measurements that exhibit a greater randomness than those of Aspect-Zeilinger-Giustina (who have only two different but also correlated settings in each wing), are not constrained by the CHSH or Bell-type inequalities and do, therefore, not require any instantaneous influences for their explanation, while the reported Aspect-Zeilinger-Giustina-measurements supposedly do. We also see that the Aspect-Zeilinger-Giustina claim of using completely random measurements in both wings is incorrect and certainly misleading.

As mentioned, two polarizer settings are indeed switched randomly on each island in the NOVA—TV-show and in references [4] [5]. However, the polarizer settings on the two islands are correlated by predetermined choice of the CHSH angles. The experimenters would have discovered the importance of this fact, had they attempted to use many random settings in each station. This "randomness" limited by the chosen CHSH angles still guarantees the independence of Λ but not the validity of inequality (11).

It is informative in this context, to consider how the equipment settings for the measurements in ordinary space are entering the operators that act on the Hilbert-state-vectors of quantum theory. For EPRB experiments with spin $\frac{1}{2}$ particles, we require a tensor product of two Pauli matrices acting on the tensor product of spinors. One Pauli matrix (related to the measurements performed with apparatus I) may be chosen corresponding to an arbitrarily defined coordinate system that describes an arbitrary magnet-setting in ordinary space. The second Pauli matrix (related to apparatus II) is not entirely arbitrary but must account for the experimentally given Bell-CHSH angle. These facts demonstrate the Achilles heel of the work of Bell and CHSH: The validity of their inequality requires a particular cyclicity of the equipment settings in ordinary space. The choice of the quantum operators in relation to these equipment settings is, on the other hand, to a large extent arbitrary and requires only the conservation of certain vector products that result in the same Bell-CHSH angles but do not necessarily correspond to cyclical equipment settings. This fact opens the possibility of violations of the Bell-CHSH type inequalities.

8.3. Counterexample 3

As another counterexample that violates (11), consider the use of a fixed setting

in one measurement station (wing) and arbitrary random settings in the other, as described in more detail in **Appendix**. This experimental arrangement permits us to obtain setting pairs that have the same Bell-CHSH angles but no cyclicity that imposes any restrictions whatsoever [14] [17] [20]. An explicit example of this kind may also be implemented on two computers (one representing the fixed setting and the other and arbitrary polarizer setting that may also be switched rapidly) and has been presented by the author [12]. As described in the **Appendix**, such measurements were indeed performed by Kwiat and coworkers and violate (11), which is no surprise, however, because no Vorob'ev cyclicity is involved.

8.4. Counterexample 4

It is important to note that the quantum results for the pair correlations P as well as the corresponding experimental results are invariant under rotations in ordinary three-dimensional space. The specific Vorob'ev cyclicity used by CHSH for the four measurement pairs is not invariant under rotations of the separate setting pairs. In addition, the four-pair inequality (6) represents, as mentioned above, a mathematical construct that includes reordering of the λs and does not follow any natural symmetry-law. This latter fact permits the removal of the Vorob'ev cyclicity by the following plausible counterexample.

Link the frame of reference to the distant stars, and consider the rotation of the polarizer settings with the rotation of the earth. This rotation will also remove the cyclical arrangement of the terms in (6) and, therefore, any CHSH constraint, while it will leave the pair correlations P unchanged. The NOVA—TV-show does not discuss the rotation of the earth relative to the stars.

8.5. What If the Vorob'ev Cyclicity Exists?

We may ask ourselves what it means when the quadruple-cyclicity indeed exists for some subset of the data or even for about all of them?

Counterexample 3 relates to the most precise EPRB-measurements ever performed to the authors' knowledge. However, due to the fixing of the polarizer setting on one side, the cyclicity is completely removed and inequality (11) is, therefore, not valid.

Counterexample 4 applies directly to the Aspect-Zeilinger-type of experiments. Rotation of the polarizer settings does certainly remove some of the cyclicity. It does also leave the possibility for some remaining cyclical or almost cyclical rearrangements of CHSH quadruples. However, a typical experimental run lasts for about one hour, which leads to significant angular changes in a system rotating with Earth. That rotation (or any other relevant rotation) together with the need of reordering the λs and the need to use consistent definitions of *horizontal* and *vertical* (as emphasized in Subsection 2.2) may lead to considerable violations of the cyclicity. Inequalities (11) and (9) have, thus, lost their theorem-character and become rather a multivariate Monte Carlo approximation problem. This fact means, in turn, that the experiments and measurements need

not only be investigated with respect to how much they deviate from the originally fixed upper limit 2, but need also be investigated with respect to their, often significant, deviation from the results of quantum theory that have usually been ignored.

There are important (with respect to clarification and explanation) experiments that have not yet been performed. Assume, for example, that we have 4 Aspect-Zeilinger-type experiments with four setting pairs $a,b;a,b';a',b$ and a',b' and also four sources of entangled pairs that may all be run simultaneously. Select from the runs of the 4 pair-experiments the subset C of data for which indeed the simultaneous quadruple measurement-outcomes are equal for all measurements with equal polarizer settings. This subset C fulfills then inequalities (9) as well as (11). The quantum result for this subset must obviously also fulfill the inequalities because otherwise it would contradict algebra and the measurements (see [21]).

Last but not least, I would like to comment about the fact that for the experiments of the Aspect-Zeilinger-type, the "closure of the cyclical arrangement is not guaranteed by the simultaneity of the measurements of CHSH quadruples, but only involves sequential measurements of pairs collected with synchronized clock-pair-readings. For such sequential pair-measurements, many variations of possible violations have been suggested in the past. For example, the two-dimensional vector stochastic processes from Subsection 4.2 remove, at least in principle, any cyclicity, because the pair measurements are performed at different times and any dependency of the relative pair-outcomes on measurement times will remove the cyclicity. Violations are particularly easy to construct, when only small subsets of the totality of measurements can be arranged in cyclical form, as is the case for all of the above counterexamples.

The possibility of geometric phases that also prevent the closure of the cyclicity (as shown in **Figure A1** of the **Appendix**) was recently added as an additional option to the previously existing extensive list of possible violations [14] [20].

In summary, the reported violations of CHSH inequality (1b) (identical to (11) above), are actually no violations at all for all the experiments and measurements that do not exhibit a Vorob'ev cyclicity as clearly shown in counterexample 3. The measurements by Kwiat and coworkers (see **Appendix**) represent the most accurate Bell-test yet performed. However, these measurements exhibit no cyclicity and, therefore, are not subject to any constraint.

Furthermore, the CHSH constraints are, generally speaking, diminished as demonstrated in counterexamples 1, 2, 4, because of the limited presence of cyclicities that are a necessary condition for CHSH inequality 1b. This fact adds many more possibilities to the already previously published deficiencies of CHSH-Bell-type inequalities when applied to actual experiments (see also the work of Kupczynski [22]).

9. Conclusion

The well known CHSH results [6] are expressed by their inequalities (1a) and

(1b) and by the somewhat modified inequalities (10) and (11). It has been shown that none of these inequalities is beyond reproach and that particularly inequality (11) that connects the CHSH framework to actual experiments is violated by straightforward counterexamples. I conclude that the derivations of the CHSH inequalities and all similar inequalities, including that of Wigner, as well as their application to actual EPRB experiments, are highly questionable from both a mathematical and physical point of view and must certainly not be used to deny Einstein's views of physical locality and separability.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Kocher, C.A. and Commins, E.D. (1967) Polarization Correlation of Photons Emitted in an Atomic Cascade. *Physical Review Letters*, **18**, 575-577. https://doi.org/10.1103/PhysRevLett.18.575

[2] Nordén, B. (2016) Quantum Entanglement: Facts and Fiction—How Wrong Was Einstein after All. *QRB Discovery*, **49**, 1-12. https://doi.org/10.1017/S0033583516000111

[3] Aspect, A., Dalibard, J. and Roger, J. (1982) Experimental Test of Bell's Inequalities Using Time-Varying Analyzers. *Physical Review Letters*, **49**, 1804-1807. https://doi.org/10.1103/PhysRevLett.49.1804

[4] Rauch, D., *et al.* (2018) Cosmic Bell Test Using Random Measurement Settings from High-Redshift Quasars. *Physical Review Letters*, **121**, Article ID: 080403. https://doi.org/10.1103/PhysRevLett.121.080403

[5] The Big Bell Test Collaboration (2018) Challenging Local Realism with Human Choices. *Nature*, **557**, 212-216. https://doi.org/10.1038/s41586-018-0085-3

[6] Clauser, J.F., Horne, M.A., Shimony, A. and Holt, R.A. (1969) Proposed Experiment to Test Local Hidden-Variable Theories. *Physical Review Letters*, **23**, 880-884. https://doi.org/10.1103/PhysRevLett.23.880

[7] Vorob'ev, N.N. (1962) Consistent Families of Measures and Their Extension. *Theory of Probability and Its Applications*, **7**, 147-163. https://doi.org/10.1137/1107014

[8] Khrennikov, A. and Alodjants, A. (2019) Classical (Local and Contextual) Probability Model for Bohm-Bell Type Experimants: No-Signaling as Independence of Random Variables. *Entropy*, **21**, 157. https://doi.org/10.3390/e21020157

[9] Khrennikov, A. (2020) Has the CHSH-Inequality Any Relation to the EPR-Argument? *Quantum Bio-Informatics*, **5**, 87-92. https://doi.org/10.1142/9789811217838_0008

[10] Wigner, E.P. (1970) On Hidden Variables and Quantum Mechanical Probabilities. *American Journal of Physics*, **38**, 1005-1009. https://doi.org/10.1119/1.1976526

[11] d'Espagnat, B. (1979) The Quantum Theory and Reality. *Scientific American*, **241**, 158-181. https://doi.org/10.1038/scientificamerican1179-158

[12] Hess, K. (2018) Bell's Theorem and Instantaneous Influences at a Distance. *Journal of Modern Physics*, **9**, 1573-1590. https://doi.org/10.4236/jmp.2018.98099

[13] Hess, K. (2020) Kolmogorov's Probability Spaces for "Entangled" Data-Subsets of EPRB Experiments: No Violation of Einstein's Separation Principle. *Journal of Modern Physics*, **11**, 683-702. https://doi.org/10.4236/jmp.2020.115044

[14] Oaknin, D.H. (2020) Bell's Theorem Revisited: Geometric Phases in Gauge Theories. *Frontiers in Physics*, **8**, 142-162. https://doi.org/10.3389/fphy.2020.00142

[15] Schatten, K.H. (2021) A Physical Origin for Quantum Entanglement and Probabilistic Behaviors. *Journal of Modern Physics*, **12**, 50-58. https://doi.org/10.4236/jmp.2021.121005

[16] Hess, K. and Philipp, W. (2004) Breakdown of Bell's Theorem for Certain Objective Local Parameter Spaces. *PNAS*, **101**, 1799-1805. https://doi.org/10.1073/pnas.0307479100

[17] Hess, K., Philipp, W. and Aschwanden, M. (2006) What Is Quantum Information? *International Journal of Quantum Information*, **4**, 585-625. https://doi.org/10.1142/S0219749906002080

[18] Hess, K. and Philipp, W. (2005) The Bell Theorem as a Special Case of a Theorem of Bass. *Foundations of Physics*, **35**, 1749-1767. https://doi.org/10.1007/s10701-005-6520-y

[19] Giustina, M., *et al.* (2015) Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons. *Physical Review Letters*, **115**, Article ID: 250401.

[20] Oaknin, D.H. and Hess, K. (2020) On the Role of Vorob'ev Cyclicities and Berry's Phase in the EPR Paradox and Bell Tests.

[21] Khrennikov, A. (2014) CHSH Inequality: Quantum Probabilities as Classical Conditional Probabilities. *Foundations of Physics*, **45**, 711-725. https://doi.org/10.1007/s10701-014-9851-8

[22] Kupczynski, M. (2020) Is the Moon There If Nobody Looks: Bell Inequalities and Physical Reality. *Frontiers in Physics*, **8**, 273. https://doi.org/10.3389/fphy.2020.00273

[23] Kwiat, P.G., Waks, E., White, A.G., Appelbaum, I. and Eberhard, P.H. (1999) Ultrabright Source of Polarization-Entangled Photons. *Physical Review A*, **60**, 773-776. https://doi.org/10.1103/PhysRevA.60.R773

Appendix

The function products in all Bell-CHSH-type inequalities contain a cyclicity that the mathematician Vorob'ev [7] identified as an absolutely necessary condition to obtain the constraint imposed by these inequalities given a common probability space for all random variables. This cyclicity is illustrated in **Figure A1** that shows a quadrangle who's vertices represent the functions (random variables) in the CHSH inequality. The reader may imagine arbitrary distorted forms of the quadrangle and relate, symbolically, joint pair probability-distributions to the length of the lines between the functions. Vorobe'v noted (in a much more general way) that the arbitrary prescription of joint pair probability-distributions for three pairs does not permit complete freedom to choose the joint distribution of the last fourth pair. This fact is the direct reason for the constraints that Bell-CHSH-inequalities introduce, as shown in detail in [18]. No constraints exists without the Vorob'ev cyclicity.

It turns out that the cyclicity may rather easily be removed in view of Equation (11), because it is not the single polarizer settings that are important for the statistics of the outcomes but only the angle between them.

The most straightforward removal of the Vorob'ev cyclicity is accomplished by the experimental choice of a fixed polarizer setting in one wing as shown in **Figure A2**. For purposes of illustration, imagine again that the different lengths of the lines between the functions relate to the joint pair probability-distributions. Because they may now all be different and because of the lack of a cyclicity, there exists no constraint in form of a CHSH inequality (see also [20]).

The fact that such a rather straightforward removal of the cyclicity removes also all restrictions that the CHSH inequality and all other related inequalities impose, has received little attention by the majority of researchers in this field. Instead they have concentrated on other conditions to derive the inequalities,

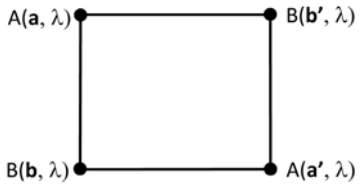


Figure A1. Symbolized Vorob'ev cyclicity for CHSH-type inequalities.

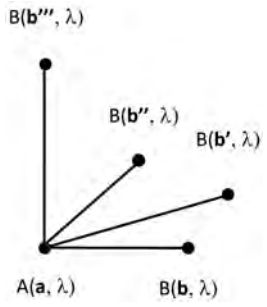


Figure A2. Removed Vorob'ev cyclicity for CHSH-type inequalities.

conditions such as "Bell locality" or the "complete randomness" and the "free will" related to setting choices, conditions that are neither necessary nor sufficient for validations or violations of the inequalities.

Experiments corresponding to this precise situation, experiments that violate the CHSH inequality with extremely high precision (probably the highest yet achieved), have been performed by Kwiat and coworkers [23] (see their **Figure A2(a)**). These experimenters have indeed fixed the polarizer-setting in one wing and varied that in the other, which amounts to varying the angle Δ between the polarizer directions. The graph of their data can be covered with great precision by a $\cos(2\Delta)$ graph which represents the quantum result for entangled photons.

The agreement with the quantum result, corresponding to the very high visibility of $V = 99.6 \pm 0.3\%$ is, however, not surprising. There exists, in this case, no valid CHSH inequality that would put any constraints on the measurement results. The explicit proof of this fact is straightforward: just insert the settings of **Figure A2** into inequalities (10) or (11), while conserving the CHSH angles and notice that the sum of all function pairs may now violate the CHSH inequalities, because the cyclicity has vanished.

It is important to realize that the wing with variable polarizer setting may also be subjected to rapid switching in this particular experiment. Rapid switching in both wings from positive to negative polarizer direction may be performed in addition, all without changing any expectation values (the correlations P).

Rapid switching of the Kwiat and coworker experiment would be useful to test Einstein's separation principle. It also would guarantee, now in a consistent way, the independence of the sets Λ from the polarizer-setting-pairs. Therefore, conspiracy-theory loopholes would be excluded as they have been excluded in the work that uses human and cosmic random number generators [4] [5]. Now, however, no constraints exist, because no Vorob'ev cyclicity exists in the measurement arrangements and no loophole needs to be closed to start with and ordinary random number generators as well as no switching at all are expected to give the same results.

Journal of Modern Physics, 2021, 12, 1237-1247
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

A Universal Binding Mechanism in Molecular Covalent Bonding and Nucleon-Nucleon Interaction

Nicolae Bogdan Mandache

National Institute for Laser, Plasma and Radiation Physics, Magurele, Romania
Email: mandache@infim.ro

How to cite this paper: Mandache, N.B. (2021) A Universal Binding Mechanism in Molecular Covalent Bonding and Nucleon-Nucleon Interaction. *Journal of Modern Physics*, **12**, 1237-1247.
https://doi.org/10.4236/jmp.2021.129076

Received: June 18, 2021
Accepted: July 9, 2021
Published: July 12, 2021

Abstract

In the hydrogen molecular ion, the kinetic energy lowering of the electron is associated with its delocalization due to electron exchange between the two protons of the molecule. This decrease in the kinetic energy of the exchanged electron in the hydrogen molecular ion and the decrease in the dynamical mass of the two exchanged pions in the nucleon-nucleon interaction are at the origin of the attraction mechanism in the molecular covalent bonding and in the nuclear interaction. Based on this unitary approach of the attraction mechanism, the formulas of molecular potential and central nucleon-nucleon potential were derived. The decrease in the mass of the exchanged pions in the nucleon-nucleon bound state, actually means the decrease in the mass of the nucleons. This nucleon mass decrease could be a manifestation of the partial chiral symmetry restoration in nuclear matter.

Keywords

Electron and Pion Exchange, Dynamical Mass, Molecular and Central Nucleon-Nucleon Potentials, Nucleon Mass Decrease, Partial Chiral Symmetry Restoration

1. Introduction

The mechanism of molecular covalent bonding by electron exchange and the mechanism of nucleon-nucleon attraction by pion exchange are important in our understanding of bound states.

In the hydrogen molecular ion H_2^+, the kinetic energy lowering of the electron associated with its delocalization due to electron exchange between the two protons of the molecule, plays a fundamental role in the mechanism of covalent

bonding [1] [2] [3]. The maximum molecular attraction is realized in the diatomic molecular bond by exchange of two electrons [1] [3]. This is the case of the hydrogen molecule H_2 in which the two electrons are exchanged between the two protons.

Heisenberg was the first who presented the attractive force between proton and neutron in analogy to that in the hydrogen molecular ion H_2^+ [4]. Yukawa was the first who proposed the pion as the particle exchanged in the nucleon-nucleon interaction. The basic mechanism of nuclear attraction at intermediate range is still under debate. The two pion exchange seems to be the most favored candidate [4] [5] [6] [7] [8].

Feynman made a simple and unitary analysis of the mechanism of both covalent bonding and nucleon-nucleon interaction [3]. The probability amplitude for a particle to get from one place to another a distance R is the key ingredient used in [3] to describe the behavior of the exchanged particle, the electron in the covalent bonding and the pion in the nucleon-nucleon interaction. The interaction energy is proportional to this probability amplitude [3].

Starting from Feynman analysis, a quantitative approach to the mechanism of nucleon-nucleon attraction was presented in [9] [10]. By analogy with the decrease in the kinetic energy of the exchanged electron in the molecule (a decrease of dynamical mass from a relativistic point of view), it has been proposed that the decrease in the dynamical mass of exchanged pions in the nucleon-nucleon interaction is the main mechanism responsible for nucleon-nucleon attraction. The delocalization of exchanged pions is strongly limited by the probability of pions to tunnel from one nucleon to another. A formula for the central nucleon-nucleon potential, which does not contain any parameter, was derived [10]. The probability concept was used instead of the probability amplitude.

In the present paper, using the concept of probability amplitude as in [3], both the potential of the hydrogen molecular ion H_2^+ (Section 2) and the central potential of the nucleon-nucleon interaction due to two pion exchange (Section 3) are derived. The two formulas of the potentials do not contain any unknown parameter. A universal binding mechanism in the molecular covalent bonding and in the central nucleon-nucleon interaction is demonstrated.

A possible manifestation of partial chiral symmetry restoration in nuclear matter is analyzed.

2. Molecular Covalent Bonding

In the hydrogen molecular ion H_2^+, since there are two protons, there is more space where the electron can have a low potential energy than in the case of hydrogen atom. The exchanged electron spreads out lowering its kinetic energy, in accord with uncertainty relation. This kinetic energy decrease is at the origin of the molecular attraction in covalent bond, in particular in the H_2^+ ion [1] [2] [3].

A simple analysis of this binding mechanism of hydrogen molecular ion H_2^+ is presented by Feynman in [3]. For large distances between the two protons of the H_2^+ ion the electrostatic potential energy of the exchanged electron is nearly zero over most of the space between the protons and the electron moves nearly like a free particle in empty space but with "negative" kinetic energy [3]:

$$\frac{p^2}{2m} = -W_H . \tag{1}$$

where W_H is the binding energy (13.6 eV) of the hydrogen atom. This means that p is an imaginary number:

$$p = i\sqrt{2m_e W_H} . \tag{2}$$

The probability amplitude A for a particle of definite energy to get from one place to another a distance R away is proportional to [3]:

$$A \sim \frac{e^{(i/\hbar)pR}}{R} . \tag{3}$$

If the particle goes in one direction the probability amplitude is [3]:

$$A \sim e^{(i/\hbar)pR} . \tag{4}$$

Replacing p (Formula (2)) one obtains that the amplitude of jumping of electron from one proton to the other proton of the molecule, will vary as [3]:

$$A \sim e^{-\frac{1}{\hbar}\left(\sqrt{2m_e W_H}\right)R} = e^{-R/a_0} . \tag{5}$$

where $a_0 = \hbar/\sqrt{2m_e W_H}$ is the Bohr radius. In other words A is the probability amplitude that the bound electron would penetrate the barrier in the space between the two protons. One can note that this exponential is just the wave function (excepting a normalization constant) of the fundamental level of the hydrogen atom. The interaction energy due to electron exchange is proportional to A [3].

Starting from this approach presented in [3] let's make an estimation of the kinetic energy lowering of the exchange electron in the H_2^+ ion. When a hydrogen atom approaches a proton, they exchange the electron which tunnels back and forth between the two protons. This is equivalent with a slight delocalization of the electron from a region of characteristic dimension a_0 associated with H atom, to a region of characteristic dimension $a_0 + \Delta(R)$ associated with the molecular ion H_2^+, where R is the distance between the two protons. This means that the mean kinetic energy of the exchanged electron in H_2^+ ion is lower than its mean kinetic energy in the H atom.

The slight delocalization $\Delta(R)$, is direct proportional to the distance R between the two protons and is strongly limited by the probability amplitude (5) of the electron to penetrate the potential barrier between the two protons. From a physical point of view one expects that this probability amplitude of transmission is 1 for a barrier width $R \to 0$. Therefore, $\Delta(R)$, which is proportional both to the distance R between the protons and to the probability amplitude of transmission of the exchanged electron, can be written as:

$$\Delta(R) = Re^{-R/a_0}. \tag{6}$$

The kinetic energy of the delocalized electron in H_2^+ ion is:

$$E_{kin,H_2^+} = \frac{p^2}{2m_e} = \frac{\hbar^2}{2m_e\left(a_0 + \Delta(R)\right)^2} = \frac{\hbar^2}{2m_e\left(a_0 + Re^{-R/a_0}\right)^2}. \tag{7}$$

where $p = \hbar/\left[a_0 + \Delta(R)\right]$ from uncertainty relation, and $\Delta(R)$ is given by relation (6).

The mean kinetic energy of electron in the H atom $E_{kin,H}$ is:

$$E_{kin,H} = \frac{\hbar^2}{2m_e a_0^2}. \tag{8}$$

The decrease of the kinetic energy of the electron bound into the H_2^+ ion (7) compared to the kinetic energy of the electron bound into the H atom (8) is directly related to the formation of the molecular bound state [1] [2] [3]. The negative quantity $T(R)$:

$$\begin{aligned} T(R) &= \frac{\hbar^2}{2m_e\left(a_0 + Re^{-R/a_0}\right)^2} - \frac{\hbar^2}{2m_e a_0^2} \\ &= -\frac{\hbar^2}{2m_e a_0^2}\left(\frac{R}{a_0}\exp\left(-R/a_0\right)\right)\frac{2 + \frac{R}{a_0}\exp\left(-R/a_0\right)}{\left(1 + \frac{R}{a_0}\exp\left(-R/a_0\right)\right)^2} \end{aligned} \tag{9}$$

represents the main contribution to the attraction mechanism in the molecular H_2^+ ion.

The curve $T(R)$ is shown in Figure 1. The distance R between the two protons is given in a_0 units. This curve is similar to those obtained by detailed quantum mechanics calculus [1] [2].

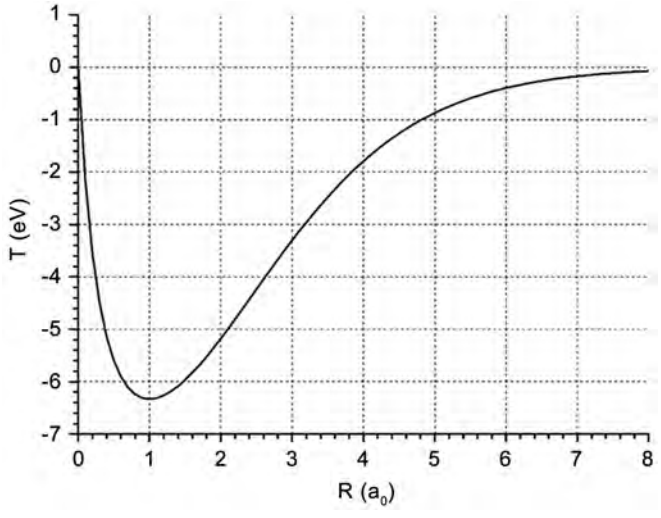


Figure 1. The decrease of the kinetic energy $T(R)$ of the exchanged electron in H_2^+ ion relative to its kinetic energy in the H atom. The distance R between the two protons is in a_0 units.

To obtain the total energy $E(R)$ of the H_2^+ ion one must add to $T(R)$ the contribution of the electrostatic interaction energy. The electron jumps from one proton to another, so it can be approximated that in the H_2^+ ion there is always a H atom (proton surrounded by the electron cloud) that interacts electrostatically with the other proton. Therefore the electrostatic repulsion between the two protons of the H_2^+ ion is partially reduced by the electron cloud. An estimation of this interaction can be obtained starting from the time-averaged potential of the (neutral) H atom [11]:

$$\phi_H(R) = \frac{e}{a_0}\frac{\exp(-2R/a_0)}{R/a_0}\left(1+\frac{R}{a_0}\right) \tag{10}$$

The energy due to electrostatic interaction between the H atom and the other proton is $e\cdot\phi_H(R)$. The total energy of the H_2^+ ion is:

$$E(R) = T(R) + e\cdot\phi_H(R). \tag{11}$$

and is shown in **Figure 2**. The distance R between the two protons is given in a_0 units.

In **Figure 2**, the value of the minimum energy is $B = -4.4$ eV and the position of the minimum is at $R = 2a_0$. The experimental values are $B_{exp} = -2.8$ eV and $R_{exp} = 1.06$ Å $\cong 2a_0$.

In a similar way one can analyses the covalent bond of H_2 molecule in which two electrons are exchange. The probability amplitude (5) of exchange of an electron will be replaced by the probability amplitude of simultaneous exchange of the two electrons between the H atoms [3];

$$A_{2e} \sim e^{-R/a_0}\cdot e^{-R/a_0} = e^{-2R/\lambda_\pi a_0}. \tag{12}$$

The lowering of the kinetic energies of both electrons gives the main contribution to the attraction mechanism in the H_2 molecule.

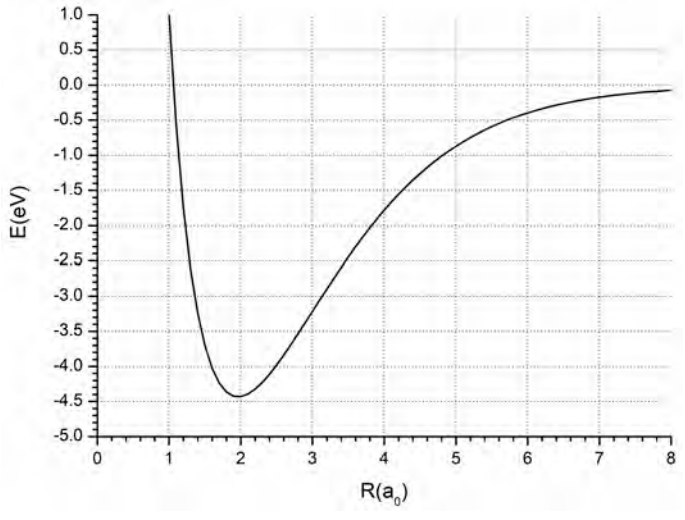


Figure 2. The total energy of the H_2^+ ion $E(R) = T(R) + e\cdot\phi_H(R)$ relative to the energy of a free H atom and a free proton. The distance R between the two protons is in a_0 units.

3. The Central Nucleon-Nucleon Potential Due to Two Pion Exchange

The effective degrees of freedom in the nuclear interaction at low energy, in particular in the nuclear bound state, are the nucleons and the pions [4] [5] [6] [7] [12] [13] [14] [15]. The intermediate range attraction is due to two pion exchange.

The nuclear interaction which takes place between a neutron and a proton by pion exchange is described by Feynman with similar arguments he used for the covalent bound of H_2^+ ion. The energy of the H atom is less than that of the proton by W_H (calculating nonrelativistically, and omitting the rest energy $m_e c^2$ of electron), so the electron which penetrates the barrier in the space between the two protons of the molecule has a "negative" kinetic energy. Since in the nuclear process the proton and the neutron have almost equal masses, an exchanged pion will have zero total energy [3]. For a pion of mass m_π and momentum p the total energy E is:

$$E^2 = p^2c^2 + m_\pi^2 c^4 . \tag{13}$$

For zero total energy p is again an imaginary number [3]:

$$p = im_\pi c . \tag{14}$$

Taking into account the Formula (4) and the Formula (14) the probability amplitude for the pion to jump from one nucleon to another is [3]:

$$A_\pi \sim \mathrm{e}^{-(m_\pi c/\hbar)R} = \mathrm{e}^{-R/\lambda\!\!\!^{-}_\pi} . \tag{15}$$

where $\lambda\!\!\!^{-}_\pi$ is the Compton wavelength of pion. The exponential function, which is typical for a Yukawa potential or exponential potential, limits drastically the probability amplitude for a pion to penetrate the barrier in the space between the two nucleons for large R [3].

Heisenberg was the first who presented the attractive force between proton and neutron in analogy to that in the hydrogen molecular ion H_2^+. The exchange of electron between the two protons is a real process, the electron is put in common by the two protons. Due to this delocalization the kinetic energy of the electron decreases in the bound system, as shown in Section 2.

We also treat the exchange of pions between the two nucleons like a real process. In consequence, if a pion leaves from nucleon 1 to tunnel to nucleon 2, a pion from nucleon 2 must tunnel simultaneously to nucleon 1, to replace it. A nucleon can emit a real pion only receiving simultaneously an amount of energy equal to the pion mass or directly another pion.

The probability amplitude of simultaneous exchange of the two pions between the nucleons is the product of two probability amplitude of one pion exchange (relation 15):

$$A_{2\pi} \sim \mathrm{e}^{-R/\lambda\!\!\!^{-}_\pi} \cdot \mathrm{e}^{-R/\lambda\!\!\!^{-}_\pi} = \mathrm{e}^{-2R/\lambda\!\!\!^{-}_\pi} . \tag{16}$$

As it is well known the current masses of quarks (antiquarks) u (ū) and d (đ) are very small; the nucleon and pion masses are mainly of dynamical origin

("kinetic" energy). More than 95 % of the mass of the Universe is "kinetic energy".

The long range structure of the nucleon is given roughly by the pion [12] [13] [14] [15], which also gives the range of nuclear forces. In particular, the Compton wavelength of the pion $\lambda\!\!\!^{-}_{\pi} = \hbar/m_{\pi}c$, which has a value of 1.41 fm for the charged pion, gives the range r_N of the nucleon extension. The pion in nucleon can be represented by a degree of freedom of current mass ≅0, localized into a region of radius $r_N = \lambda\!\!\!^{-}_{\pi}$. This localization into the nucleon associates an energy (dynamical mass) $E \cong pc = \hbar c/\lambda\!\!\!^{-}_{\pi}$, given by the uncertainty relation, which is just the mass of the pion [9] [10].

When two nucleons approach each other to form a bound state, in particular the deuteron, they put in common some pion degrees of freedom, this means just the two pion exchange process. This is equivalent with a slight delocalization of each pion degree of freedom from a region of linear dimension r_N to a region of linear dimension $r_N + \Delta(R)$, where $\Delta(R)$ is direct proportional to the distance R between the two bound nucleons and is strongly dependent on the probability amplitude of the two pions to tunnel simultaneously the potential barrier between the two nucleons (Formula (16)). The dynamical mass of each pion degree of freedom gets:

$$E_{\Delta} = \frac{\hbar c}{\lambda\!\!\!^{-}_{\pi} + \Delta(R)}. \tag{17}$$

and is lower than the initial one (that in the free nucleon: $E = \hbar c/\lambda\!\!\!^{-}_{\pi}$). The total decrease of the dynamical masses of the two exchanged pions:

$$\Delta E = 2\left(\frac{\hbar c}{\lambda\!\!\!^{-}_{\pi} + \Delta(R)} - \frac{\hbar c}{\lambda\!\!\!^{-}_{\pi}}\right). \tag{18}$$

is just the main contribution to the central nucleon-nucleon potential $V_{2\pi}(R)$ due to two pion exchange.

As in the case of covalent bonding (see Formula (6)), $\Delta(R)$ is the product between the inter-nucleon distance R and the probability amplitude of the two pions to simultaneously tunnel the potential barrier between the two nucleons (Formula (16)):

$$\Delta(R) = R\mathrm{e}^{-2R/\lambda\!\!\!^{-}_{\pi}}. \tag{19}$$

From (18) and (19) it results the following expression of the central nucleon-nucleon potential due to two pion exchange:

$$V_{2\pi}(R) = 2\left(\frac{\hbar c}{\lambda\!\!\!^{-}_{\pi} + R\mathrm{e}^{-2R/\lambda\!\!\!^{-}_{\pi}}} - \frac{\hbar c}{\lambda\!\!\!^{-}_{\pi}}\right) = -2\frac{\hbar c}{\lambda\!\!\!^{-}_{\pi}} \cdot \frac{R\mathrm{e}^{-2R/\lambda\!\!\!^{-}_{\pi}}}{\lambda\!\!\!^{-}_{\pi} + R\mathrm{e}^{-2R/\lambda\!\!\!^{-}_{\pi}}}. \tag{20}$$

One notes that $\hbar c/\lambda\!\!\!^{-}_{\pi} = m_{\pi}c^2$.

This nucleon-nucleon potential $V_{2\pi}(R)$ due to two pion exchange is shown in **Figure 3**, for $\lambda\!\!\!^{-}_{\pi} = 1.41\,\mathrm{fm}$ (Compton wavelength of charged pion). The minimum value of the potential is −44 MeV, at R = 0.7 fm, which is comparable to the minimum value (−50 MeV) of the CD Bonn potential [2]. The fall of the

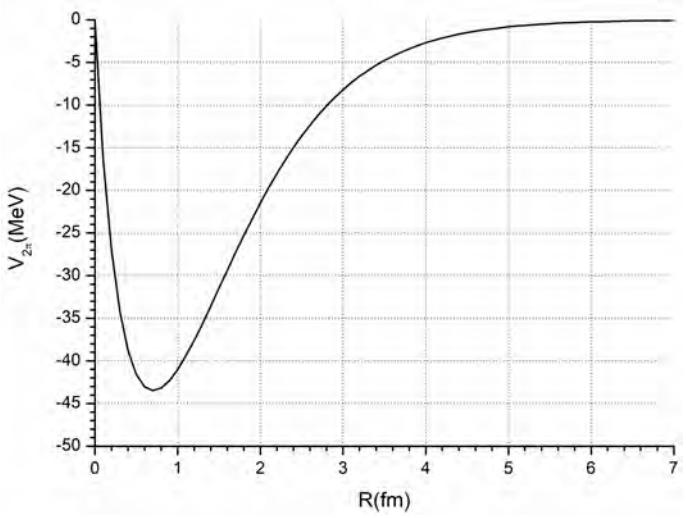


Figure 3. The nucleon-nucleon potential $V_{2\pi}(R)$ due to two pion exchange as a function of inter-nucleon distance R for $\hat{\lambda}_{\pi} = 1.41$ fm .

potential for small R ($R < 0.6$ fm) is compatible with the beginning of the hard core repulsion region [4] [5] [6] [7] [13] [16] [17] which gets dominant at short range. The fall of the potential for high values of R is slower than in the case of CD Bonn potential. The width at half maximum is less than 2 fm.

Formula (20) of the central nucleon-nucleon potential due to two pion exchange is identical with that obtained in [10]. A hard core repulsion given by an infinite potential for $R \le r_0$, where r_0 is the hard core radius, was added to the potential $V_{2\pi}(R)$. The Schrodinger equation for the two nucleons interacting by this two pion exchange potential with hard core repulsion was solved numerically for different values of the hard core radius r_0 [10]. For a value of the hard core repulsion radius equal to 0.5 fm, a typical value for nucleon-nucleon interaction [4] [5] [6] [7] [16] [17], the calculated binding energy is equal to the deuteron binding energy [10]. For pion mass values higher than the real pion mass, with pion mass increase the binding energy decreases, and gets zero at a value of about 190 MeV [10], a result comparable to that obtained in [6]. This underlines the central role of the pions as main player in the production of nuclear attraction.

4. Discussion and Conclusions

Feynman made a simple and unitary analysis of the mechanism of both covalent bonding and nucleon-nucleon interaction by particle exchange. In particular, he used the probability amplitude of a particle to get from one place to another a distance R away to describe the behavior of the exchanged particle, in fact its slight delocalization, and assumed that the interaction energy must be proportional to this amplitude.

Starting from this analysis, a unitary quantitative approach to the attraction mechanism in molecular covalent bonding and in central nucleon-nucleon inte-

raction was made. In our quantitative estimations, only the main ingredients of the interaction mechanism were considered to highlight the main attraction mechanism. Some degrees of freedom, electrons or pions, are put in common between the two protons of the molecule and between the two nucleons respectively. This delocalization, which is strongly limited by the probability amplitude of tunneling of the exchanged particles, determines a decrease in the dynamical mass (kinetic energy) of the system compared to the unbound state.

The decrease in the kinetic energy (dynamical mass from a relativistic point of view) of the exchanged electron in the H_2^+ molecular ion as function of the distance between the protons calculated in the present work is similar to that derived by detailed quantum mechanics calculus. The relation (9), which represents the attractive part of the potential for the molecular ion H_2^+, can be written as follows:

$$T(R) = -\frac{\hbar^2}{2m_e a_0^2}\left(x\mathrm{e}^{-x}\right)\frac{2+x\mathrm{e}^{-x}}{\left(1+x\mathrm{e}^{-x}\right)^2} \equiv -E_{kin,H}\left(x\mathrm{e}^{-x}\right)\frac{2+x\mathrm{e}^{-x}}{\left(1+x\mathrm{e}^{-x}\right)^2} \tag{21}$$

where $x = R/a_0$ and $E_{kin,H}$ is given by relation (8). Since the maximum value of $x\mathrm{e}^{-x}$ is 0.368, for $x = 1$ ($R = a_0$), the main contribution to the variation of T with distance is given by the term in parenthesis $x\mathrm{e}^{-x}$. This dependence is similar to the attractive part of the Rydberg potential used to describe the molecular covalent bond [18].

The decrease in the dynamical mass of the two pions exchanged simultaneously between the two nucleons is the principal mechanism responsible for the nucleon-nucleon attraction. A pion from nucleon 1 tunnels to nucleon 2 and simultaneously a pion from nucleon 2 tunnels to nucleon 1. The central nucleon-nucleon potential obtained in this approach (Formula (20)) has a minimum value comparable to that of the CD Bonn potential and has a fall towards zero value for small values of R compatible with the beginning of the hard core repulsion region. Relation (20) can be written as follows:

$$V_{2\pi}(R) = -2\frac{\hbar c}{\lambda\!\!\!^{-}_{\pi}} \cdot \frac{\frac{R}{\lambda\!\!\!^{-}_{\pi}}\mathrm{e}^{-2R/\lambda\!\!\!^{-}_{\pi}}}{1+\frac{R}{\lambda\!\!\!^{-}_{\pi}}\mathrm{e}^{-2R/\lambda\!\!\!^{-}_{\pi}}} \equiv -m_\pi c^2\left(x\mathrm{e}^{-x}\right)\frac{1}{1+\frac{x}{2}\mathrm{e}^{-x}} \tag{22}$$

where $x = 2R/\lambda\!\!\!^{-}_{\pi}$. Since the maximum value of $x\mathrm{e}^{-x}/2$ is 0.18, for $x = 1$ ($R = \lambda\!\!\!^{-}_{\pi}/2$), the same dependence $x\mathrm{e}^{-x}$ (the term in parenthesis) is dominant in the nucleon-nucleon attraction mechanism, like in the case of covalent bond. It looks like a universal attraction mechanism for the two interactions.

The pion is part of the nucleon structure. In particular, the Lattice QCD calculations have shown that the nucleon mass displays a linear dependence on the pion mass, known as the "ruler approximation" [19]:

$$m_N(\mathrm{MeV}) = 800 + m_\pi \tag{23}$$

This means that the decrease in the mass of the exchanged pions in the nucleon-nucleon interaction, actually means the decrease in the mass of the

nucleons. But this nucleon mass decrease is a manifestation of the partial chiral symmetry restoration in nuclear matter [20]. There is a direct relationship between the slight delocalization of the pion degrees of freedom and the mass decrease of the nucleons which exchange those pions.

Let's analyze at the quark level. Because the pions are made up of quarks and antiquarks, by pion exchange some quark degrees of freedom are implicitly exchanged between the two nucleons and, consequently, are slightly delocalized. This means that the confinement region of some quarks slightly increases and accordingly their dynamical mass decreases. It could be said that there is a slight deconfinement of some quark degrees of freedom and, therefore, the mass of nucleons decreases.

Acknowledgements

The author expresses his sincere thanks to E. Dudas, S. Nordholm and D. I. Palade for helpful discussions and suggestions.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Bacskay, G.B. and Nordholm, S. (2013) *The Journal of Physical Chemistry A*, **117**, 33. https://doi.org/10.1021/jp403284g

[2] Rioux, F. (1997) *Journal of Chemical Education*, **2**, 1.

[3] Feynman, R., Leighton, R. and Sands, M. (1964) Feynman Lectures on Physics. Vol. 3, Quantum Mechanics. Addison-Wesley, Boston, Chapters 3, 7 and 10.

[4] Machleidt, R. (2007) Lectures 1-4, "Nuclear Forces".

[5] Ericson, T. and Weise, W. (1988) Pions and Nuclei. Clarendon Press, Oxford.

[6] Epelbaum, E., Hammer, H.W. and Meissner, U.-G. (2008) Modern Theory of Nuclear Forces.

[7] Machleidt, R. (2013) Origin and Properties of Strong Inter-Nucleon Interactions.

[8] Mandache, N.B. (2009) Anomalous Delta-Type Electric and Magnetic Two-Nucleon Interactions. https://arxiv.org/abs/0904.1080

[9] Mandache, N.B. (2012) *Romanian Reports in Physics*, **64**, 1307.

[10] Mandache, N.B. and Palade, D.I. (2018) *Journal of Modern Physics*, **9**, 1459. https://doi.org/10.4236/jmp.2018.98090

[11] Jackson, J.D. (1998) Classical Electrodynamics. John Wiley & Sons, Inc., Hoboken.

[12] Thomas, A.W., Theberge, S. and Miller, G.A. (1981) *Physical Review D*, **24**, 216. https://doi.org/10.1103/PhysRevD.24.216

[13] Brown, G.E. and Rho, M. (1983) *Physics Today*, **36**, 24. https://doi.org/10.1063/1.2915491

[14] Thomas, A.W. and Weise, W. (2001) The Structure of the Nucleon. Wiley-VCH, Hoboken. https://doi.org/10.1002/352760314X

[15] Vanderhaeghen, M. and Walcher, T. (2010) Long Range Structure of the Nucleon.

[16] Ishii, N., Aoki, S. and Hatsuda, T. (2007) *Physical Review Letters*, **99**, Article ID: 022001. https://doi.org/10.1103/PhysRevLett.99.022001

[17] Aoki, S., Doi, T., Hatsuda, T., Ikeda, Y., Inoue, T., Ishii, N., Murano, K., Nemura, H. and Sasaki, K. (2012) *Progress of Theoretical and Experimental Physics*, **2012**, 01A105. https://doi.org/10.1093/ptep/pts010

[18] Varshni, Y.P. (1957) *Reviews of Modern Physics*, **29**, 664. https://doi.org/10.1103/RevModPhys.29.664

[19] Walker-Loud, A. (2013) Baryons in/and Lattice QCD.

[20] Birse, M. and McGovern, J. (1995) *Physics World*, **10**, 35. https://doi.org/10.1088/2058-7058/8/10/29

Journal of Modern Physics, 2021, 12, 1248-1266
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Active-Sterile Neutrino Oscillations and Leptogenesis

Bruce Hoeneisen

Universidad San Francisco de Quito, Quito, Ecuador
Email: bhoeneisen@usfq.edu.ec

How to cite this paper: Hoeneisen, B. (2021) Active-Sterile Neutrino Oscillations and Leptogenesis. *Journal of Modern Physics*, **12**, 1248-1266.
https://doi.org/10.4236/jmp.2021.129077

Received: May 18, 2021
Accepted: July 9, 2021
Published: July 12, 2021

Abstract

We study coherent active-sterile neutrino oscillations as a possible source of leptogenesis. To this end, we add 3 gauge invariant Weyl_R neutrinos to the Standard Model with both Dirac and Majorana type mass terms. We find that the measured active neutrino masses and mixings, and successful baryogenesis via leptogenesis, may be achieved with fine-tuning, if at least one of the sterile neutrinos has a mass in the approximate range 0.14 to 1.1 GeV.

Keywords

Leptogenesis, Majorana Neutrino, Matter-Antimatter Asymmetry

1. Introduction and Overview

We present a study of coherent active-sterile neutrino oscillations as a possible source of baryogenesis via leptogenesis. Consider a reaction of the form $e^{\mp}W^{\pm} \to \nu_i \to e^{\pm}W^{\mp}$ that produces a lepton asymmetry that is partially converted to a baryon asymmetry before the sphaleron freeze-out temperature $T_{\text{sph}} = 131.7 \pm 2.3\,\text{GeV}$ [1]. The present baryon-to-photon ratio of the universe is measured to be $\eta = (6.12 \pm 0.04) \times 10^{-10}$ [2]. Let us consider a bench-mark scenario with all numbers calculated at a reference temperature T_{sph}. We define the electron asymmetry $\delta_e \equiv \left(n_{e^-} - n_{e^+}\right) \Big/ \left(n_{e^-} + n_{e^+}\right)$, and similarly for δ_μ, δ_τ and δ_B. n_{e^-} is the number density of electrons. The asymmetry at T_{sph} from neutrino oscillations required to obtain η is

$$\delta_l \equiv \delta_e + \delta_\mu + \delta_\tau = -(37/12)\delta_B$$
$$= -(37/12)\eta(2/3)\cdot(385\times 22/(43\times 8)) \approx -3.1\times 10^{-8}$$ [3]. At T_{sph} the age of the universe is $t_u = 1.4\times 10^{-11}\,\text{s}$, and the time between collisions of active neutrinos in the reaction $\nu_e e^+ \to \nu_e e^+$ is $t_c = 1/(\sigma n_{e^-} c) \approx 7\times 10^{-22}\,\text{s} \approx 1/(0.001\,\text{GeV})$, where σ is the cross-section. We

note that at T_{sph} neutrinos are of short wavelength relative to t_c, *i.e.* $t_c \gg 2\pi/T_{\text{sph}}$.

Observed neutrino oscillations require that at least two neutrino eigenstates have mass. To this end, we add at least $n' = 2$ gauge singlet Weyl_R neutrinos ν_R to the Standard Model. To obtain lepton number violation, we assume the neutrinos are of the Majorana type, *i.e.* we add both Dirac and Majorana mass terms to the Lagrangian [4].

Let us consider the reaction $e^- W^+ \to \nu_i \to e^{\mp} W^{\pm}$, with neutrino mass eigenstates ν_i oscillating coherently during time t_c. The condition for coherent oscillations is that ν_i has mass $\lesssim 6\,\text{GeV}$. (The physics described in this overview will be developed in the following Sections.) The cross-section for the lepton number violating reaction is reduced relative to the lepton conserving reaction by a factor $m_i m_j / \left(2E^2\right)$ due to polarization miss-match, where m_i is the neutrino eigenstate mass, and E is the neutrino energy in the laboratory frame.

One mechanism to obtain CP violation is to have two interfering amplitudes with different "strong" phases and different "weak" phases [5]. A "strong" phase (the name is borrowed from B-physics) is a phase that does not change sign under CP-conjugation. A "weak" phase changes sign under CP-conjugation. Here, the "weak" phases are the CP-violating phases in the weak mixing matrix U. The "strong" phases are the propagation phases of the interfering ultra-relativistic neutrinos, $2X_{ij} = \left(m_i^2 - m_j^2\right) L / \left(2E\right)$, with energy $E \approx 2.8 T_{\text{sph}}$, and $L = t_c$. To obtain a sizable CP violation asymmetry, the relative propagation phase difference $2X_{ij}$ between two neutrinos in time t_c should be of order $\pi/2$ or less. This requires two neutrinos to satisfy $\sqrt{m_i^2 - m_j^2} \lesssim 1.1\,\text{GeV}$.

There are cosmological constraints, mainly from Big Bang Nucleosynthesis (BBN), that require the mass of sterile neutrinos to be $m_s \gtrsim 0.14\,\text{GeV}$. Thus, the interesting mass range for sterile neutrinos contributing to leptogenesis is approximately 0.14 GeV to 1.1 GeV.

From the following studies, we conclude that nature may have added, to the Standard Model, two or more gauge singlet Weyl_R Majorana neutrinos, with fine tuned parameters, as the source of neutrino masses and mixing, and successful baryogenesis via leptogenesis. This scenario is not new, yet is not mentioned in several leading leptogenesis reviews. Here, we emphasize analytic solutions, and an understanding of several delicate issues related to Majorana neutrinos, lepton number violation, CP-violation, polarization miss-match, and coherence. In the following Sections, we develop, step-by-step, the physics behind the preceding comments.

2. Dirac Neutrinos

In the following sections we consider a neutrino experiment with a source at the origin of coordinates, and a detector at a distance $z = L$. We assume $L \gg 2\pi/p_z$, so the neutrinos are almost on mass-shell. p_z is the neutrino momentum. At first let us consider a single neutrino flavor, and the reaction

$e^- W^+ \to \nu_e \to e^- W^+$.

Before electroweak symmetry breaking (EWSB) at $T_{\text{EWSB}} \approx 159 \pm 1\,\text{GeV}$ [1], the neutrino field ν_L is massless, carries the 2-dimensional "Weyl_L" representation of the proper Lorentz group, and satisfies the wave equation

$$i\bar{\sigma}_\mu \partial_\mu \nu_L = 0, \tag{1}$$

where $\sigma_0 \equiv 1_{2\times 2}$, $\bar{\sigma}_0 \equiv \sigma_0$, $\bar{\sigma}_k \equiv -\sigma_k$, and σ_k are the Pauli matrices [4]. Summation over repeated indices is understood. $k = 1,2,3$, and $\mu = 0,1,2,3$. Multiplying on the left by $-i\sigma_\nu \partial_\nu$, obtains the Klein-Gordon wave equation of a massless field:

$$\eta^{\mu\nu} \partial_\nu \partial_\mu \nu_L \equiv \partial^\mu \partial_\mu \nu_L = 0, \tag{2}$$

where $\eta^{\mu\nu} = \text{diag}(1,-1,-1,-1)$ is the metric.

After EWSB the Higgs boson acquires a vacuum expectation value v_h [4]. The field ν_L forward scatters on v_h with amplitude $Y^N v_h/\sqrt{2}$ becoming a ν_R (Y^N is a Yukawa coupling [4]), that forward scatters on v_h with amplitude $Y^{N*} v_h/\sqrt{2}$ becoming a ν_L, etc. The field ν_R transforms as "Weyl_R" [4]. These scatterings are forward because v_h does not depend on the space-time coordinates x^μ. These scatterings are described by the Dirac equation,

$$i\sigma_\mu \partial_\mu \nu_R = m\nu_L, \quad i\bar{\sigma}_\mu \partial_\mu \nu_L = m\nu_R, \tag{3}$$

with $m = |Y^N| v_h/\sqrt{2}$. In this way the field ν_R is created (arguably) after EWSB, on a time scale $1/m$, and the fields ν_L and ν_R couple together forming a 4-dimensional field ψ that carries the reducible Dirac = Weyl_L $\oplus$ Weyl_R representation of the proper Lorentz group. The solution of (3) proportional to $\exp(-iEt + ip_z z)$, in a Weyl basis, is [4]

$$\psi_u \equiv \begin{pmatrix} \nu_{Lu} \\ \nu_{Ru} \end{pmatrix} = \begin{pmatrix} \sqrt{E - p_z}\,\xi_1 \\ \sqrt{E + p_z}\,\xi_2 \\ \sqrt{E + p_z}\,\xi_1 \\ \sqrt{E - p_z}\,\xi_2 \end{pmatrix} \exp(-iEt + ip_z z), \tag{4}$$

corresponding to a particle of mass m, and momentum $\vec{p} = p_z \vec{e}_z$ with $p_z = +\sqrt{E^2 - m^2}$. This is the "stepping stone" mechanism of mass generation [6]. Alternatively, consider (3): the ν_L creates ν_R on a time scale $1/m$, which in turn creates ν_L, etc. Solutions for other $\vec{p}$ can be obtained with Lorentz transformations. ξ_1 and ξ_2 are complex numbers that define the polarization of the neutrino (to be discussed in Section 5).

The solution of (3) proportional to $\exp(iEt - ip_z z)$ is

$$\psi_v \equiv \begin{pmatrix} \nu_{Lv} \\ \nu_{Rv} \end{pmatrix} = \begin{pmatrix} \sqrt{E - p_z}\,\eta_1 \\ \sqrt{E + p_z}\,\eta_2 \\ -\sqrt{E + p_z}\,\eta_1 \\ -\sqrt{E - p_z}\,\eta_2 \end{pmatrix} \exp(iEt - ip_z z). \tag{5}$$

The charge conjugate of ψ_v is [4]

$$\left(\psi_v\right)^c = -i\gamma^2\psi_v^* = \begin{pmatrix}\left(\nu_{Rv}\right)^c\\ \left(\nu_{Lv}\right)^c\end{pmatrix} = \begin{pmatrix}-i\sigma_2\nu_{Rv}^*\\ i\sigma_2\nu_{Lv}^*\end{pmatrix} = \begin{pmatrix}\sqrt{E-p_z}\,\eta_2^*\\ -\sqrt{E+p_z}\,\eta_1^*\\ \sqrt{E+p_z}\,\eta_2^*\\ -\sqrt{E-p_z}\,\eta_1^*\end{pmatrix}\exp\left(-iEt+ip_z z\right). \quad (6)$$

Note that $-i\sigma_2\nu_{Rv}^*$ transforms as Weyl_L, while $i\sigma_2\nu_{Lv}^*$ transforms as Weyl_R [4]. $\nu_L^\dagger\nu_R$, $\nu_R^\dagger\nu_L$, $\nu_R^{\mathrm{T}}\sigma_2\nu_R$, and $\nu_R^\dagger\sigma_2\nu_R^*$ are scalars with respect to the proper Lorentz group. The Dirac Equations (3) can be summarized as $\left(i\gamma^\mu\partial_\mu - m\right)\psi = 0$. We work in the Weyl basis with γ matrices

$$\gamma^0 = \begin{pmatrix}0 & \sigma_0\\ \sigma_0 & 0\end{pmatrix},\quad \gamma^k = \begin{pmatrix}0 & \sigma_k\\ -\sigma_k & 0\end{pmatrix},\quad \gamma^5 = \begin{pmatrix}-\sigma_0 & 0\\ 0 & \sigma_0\end{pmatrix}. \quad (7)$$

The Weyl_L and Weyl_R projectors are $\gamma_L \equiv \left(1-\gamma^5\right)/2$, and $\gamma_R \equiv \left(1+\gamma^5\right)/2$. For example, the Weyl_L component of ψ is $\gamma_L\psi$. Note that $W^\pm$ and Z only "see" the Weyl_L fields $\gamma_L\psi_u$ or $\tilde{\psi}_v\gamma_R$. Neutrinos may, or may not, have a conserved $U(1)$ charge q such as lepton number. In quantum field theory, the fields are interpreted as follows:

- ψ_u creates a particle with charge $+q$, and spin angular momentum component $s_z = +\frac{1}{2}$ with amplitude $\sqrt{E-p_z}\,\xi_1$, and $s_z = -\frac{1}{2}$ with amplitude $\sqrt{E+p_z}\,\xi_2$;
- $\tilde{\psi}_u \equiv \psi_u^\dagger\gamma^0$ annihilates this particle;
- $\tilde{\psi}_v \equiv \psi_v^\dagger\gamma^0$ creates an antiparticle with charge $-q$, and spin $s_z = +\frac{1}{2}$ with amplitude $\sqrt{E+p_z}\,\eta_2^*$, and $s_z = -\frac{1}{2}$ with amplitude $\sqrt{E-p_z}\,\eta_1^*$;
- ψ_v annihilates this antiparticle.

This interpretation is needed to avoid unstable particles with negative energy. These particles and antiparticles have mass m, spin $\frac{1}{2}$, *positive* energy E, and momentum $p_z\vec{e}_z = +\sqrt{E^2-m^2}\,\vec{e}_z$. Note that antiparticles have the opposite charge of the corresponding particle.

Let us now consider two neutrino flavors, ν_e and ν_μ. The field ν_{Le} may forward scatter on ν_h with amplitude $Y_{ee}^N\nu_h/\sqrt{2}$ becoming a ν_{Re}, which may forward scatter on ν_h with amplitude $Y_{e\mu}^{E*}\nu_h/\sqrt{2}$ becoming a $\nu_{L\mu}$, etc. As a result, two mass eigenstates acquire masses:

$$\psi_1 = \cos\theta\psi_e + \sin\theta\psi_\mu, \quad \text{with mass } m_1, \quad (8)$$

$$\psi_2 = -\sin\theta\psi_e + \cos\theta\psi_\mu, \quad \text{with mass } m_2. \quad (9)$$

For simplicity, we have suppressed the sub-indices u for neutrinos, or v for anti-neutrinos. For example, the interaction $e^-W^+ \to \nu_{Le}$ producing a weak state has $\psi_\mu(0) = 0$, $\nu_{Re}(0) = 0$, and $\left[\left|\nu_{Le}^{(1)}(0)\right|^2 + \left|\nu_{Le}^{(2)}(0)\right|^2\right]^{1/2}$ is normalized to 1. An observation at distance L obtains e^-W^+ with probability $P_{ee} \propto \left|\psi_{Le}^{(1)}(L)\right|^2 + \left|\psi_{Le}^{(2)}(L)\right|^2$, or μ^-W^+ with probability $P_{e\mu} \propto \left|\psi_{L\mu}^{(1)}(L)\right|^2 + \left|\psi_{L\mu}^{(2)}(L)\right|^2$, where

$$P_{e\mu} = 1 - P_{ee} = 4\cos^2\theta\sin^2\theta\sin^2\left(X_{e\mu}\right), \quad (10)$$

with $X_{e\mu} \equiv \Delta m_{e\mu}^2 L/(4E)$, and $\Delta m_{e\mu}^2 \equiv m_e^2 - m_\mu^2$. This is the phenomenon of neutrino oscillations.

3. Majorana Neutrinos

If neutrinos have no additive conserved charge (such as lepton number), it is possible to add Majorana type mass terms to (3):

$$i\sigma_\mu \partial_\mu \nu_{Ru} = m\nu_{Lu} + M(\nu_{Rv})^c, \quad i\sigma_\mu \partial_\mu (\nu_{Lv})^c = m(\nu_{Rv})^c, \tag{11}$$

$$i\bar{\sigma}_\mu \partial_\mu (\nu_{Rv})^c = m^*(\nu_{Lv})^c + M^*\nu_{Ru}, \quad i\bar{\sigma}_\mu \partial_\mu \nu_{Lu} = m^*\nu_{Ru}. \tag{12}$$

$$i\bar{\sigma}_\mu \partial_\mu (\nu_{Ru})^c = m^*(\nu_{Lu})^c + M^*\nu_{Rv}, \quad i\bar{\sigma}_\mu \partial_\mu \nu_{Lv} = m^*\nu_{Rv}, \tag{13}$$

$$i\sigma_\mu \partial_\mu \nu_{Rv} = m\nu_{Lv} + M(\nu_{Ru})^c, \quad i\sigma_\mu \partial_\mu (\nu_{Lu})^c = m(\nu_{Ru})^c. \tag{14}$$

Here, with one generation, the masses can be made real by re-phasing the fields. The charge conjugate fields are $(\nu_{Lu})^c \equiv i\sigma_2 \nu_{Lu}^*$, $(\nu_{Lv})^c \equiv i\sigma_2 \nu_{Lv}^*$, $(\nu_{Ru})^c \equiv -i\sigma_2 \nu_{Ru}^*$, and $(\nu_{Rv})^c \equiv -i\sigma_2 \nu_{Rv}^*$. Majorana mass terms for fields ν_L are not added, at tree level, because such terms are not gauge invariant. Note that the Majorana mass terms link ν_{Ru} with $(\nu_{Rv})^c$, etc. Then, a created ν_{Lu} may forward scatter on v_h (with amplitude $Y^N v_h/\sqrt{2}$) becoming a ν_{Ru}, that may forward scatter on M^* (whatever it is, e.g. a dimension 5 operator containing v_h) becoming a $(\nu_{Rv})^c$, that may forward scatter on v_h (with amplitude $Y^N v_h/\sqrt{2}$) becoming a $(\nu_{Lv})^c$, etc, see Figure 1. Equations (13) and (14) are the charge conjugate of Equations (11) and (12), respectively.

Note that before EWSB, the fields ν_{Lu} and ν_{Lv} are in statistical equilibrium due to their interactions with the gauge bosons W^μ and B. From (11) to (14) we conclude that after EWSB, the fields ν_{Lu}, ν_{Ru}, $(\nu_{Rv})^c$, $(\nu_{Lv})^c$, and their conjugates, become linked together so that the Majorana character of neutrinos may emerge dynamically after EWSB on a time scale $1/m$ (for the case of interest $M \gg m$).

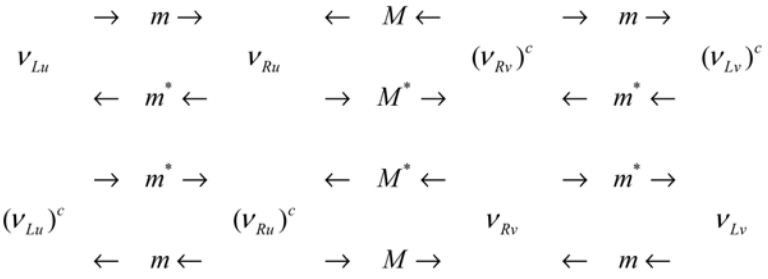


Figure 1. Graphical representation of (11), (12), (13), and (14), corresponding, respectively, to the four rows of arrows. Weyl_L and Weyl_R fields forward scatter on m and M. The lepton number conserving reactions are $e^- W^+ \to \nu_{Lu} \to e^- W^+$ and $e^+ W^- \to (\nu_{Lv})^c \to e^+ W^-$. The lepton number violating reactions are $e^- W^+ \to \nu_{Lu} \leftrightarrow \nu_{Ru} \leftrightarrow (\nu_{Rv})^c \leftrightarrow (\nu_{Lv})^c \to e^+ W^-$, and $e^+ W^- \to (\nu_{Lv})^c \leftrightarrow (\nu_{Rv})^c \leftrightarrow \nu_{Ru} \leftrightarrow \nu_{Lu} \to e^- W^+$. Ultra-relativistic ν_{Lu} and $(\nu_{Lv})^c$ have a polarization miss-match, see (4), (6) and Section 5.

Equations (11) to (14) are linear and homogeneous, and their general solution is a superposition of mass eigenstates. Each term of (11) transforms as Weyl_L, and is proportional to $\exp\left(-iEt+ip_z z\right)$, and so they may be mixed. Equations (11) can be re-written as

$$\begin{pmatrix} i\sigma_\mu\partial_\mu\left(\nu_{Lv}\right)^c \\ i\sigma_\mu\partial_\mu\nu_{Ru} \end{pmatrix} = \begin{pmatrix} 0 & m \\ m & M \end{pmatrix}\begin{pmatrix} \nu_{Lu} \\ \left(\nu_{Rv}\right)^c \end{pmatrix}. \tag{15}$$

Equations (12) can be re-written as

$$\begin{pmatrix} i\bar{\sigma}_\mu\partial_\mu\nu_{Lu} \\ i\bar{\sigma}_\mu\partial_\mu\left(\nu_{Rv}\right)^c \end{pmatrix} = \begin{pmatrix} 0 & m^* \\ m^* & M^* \end{pmatrix}\begin{pmatrix} \left(\nu_{Lv}\right)^c \\ \nu_{Ru} \end{pmatrix}. \tag{16}$$

Both (15) and (16) can be diagonalized simultaneously with a unitary matrix U and its complex-conjugate to obtain the equation in the mass eigenstate basis:

$$\begin{pmatrix} i\sigma_\mu\partial_\mu\nu_a \\ i\sigma_\mu\partial_\mu\nu_s \end{pmatrix} = \begin{pmatrix} m_a & 0 \\ 0 & m_s \end{pmatrix}\begin{pmatrix} \nu_a \\ \nu_s \end{pmatrix}, \tag{17}$$

$$\vec{\nu} \equiv \begin{pmatrix} \nu_{Lu} \\ \left(\nu_{Rv}\right)^c \end{pmatrix} \equiv U\vec{\nu}_{\text{mass}}, \quad \vec{\nu}_{\text{mass}} \equiv \begin{pmatrix} \nu_a \\ \nu_s \end{pmatrix} \equiv U^\dagger\vec{\nu}, \tag{18}$$

$$U^*\begin{pmatrix} m_a & 0 \\ 0 & m_s \end{pmatrix}U^\dagger = \begin{pmatrix} 0 & m \\ m & M \end{pmatrix}. \tag{19}$$

The unitary matrix U that satisfies (15), (17), and (19), with real and positive m_a, m_s, and M, is

$$U = \begin{pmatrix} \mathrm{e}^{i\alpha}\cos\theta & i\mathrm{e}^{i\alpha}\sin\theta \\ i\sin\theta & \cos\theta \end{pmatrix}, \tag{20}$$

where $m = |m|\exp\left\{-i\left(\alpha+\pi/2\right)\right\}$, $\tan^2\theta = m_a/m_s$, and $\tan\left(2\theta\right) = 2|m|/M$. The eigenvalues are

$$m_a = \frac{1}{2}\left[-M+\sqrt{M^2+4|m|^2}\right], \quad m_s = \frac{1}{2}\left[+M+\sqrt{M^2+4|m|^2}\right]. \tag{21}$$

From here on we take the Majorana masses $M_i \gg |m_{Dj}|$, so $\theta \ll 1$. Then ν_a is an "active" neutrino that is mostly ν_{Lu}, while ν_s is a "sterile" neutrino that is mostly $\left(\nu_{Rv}\right)^c$.

According to (18), the fields evolve as follows:

$$\vec{\nu}\left(z,t\right) = U\,\mathrm{diag}\left\{\exp\left(-iEt+i\sqrt{E^2-m_i^2}\,z\right)\right\}U^\dagger\vec{\nu}\left(0,0\right). \tag{22}$$

Consider a source that produces neutrinos in a weak state, e.g. $e^-W^+ \to \nu_e$. At the source, $\left(\nu_{Rv}\left(0\right)\right)^c = 0$ and $\left(\left|\nu_{Lu}^{(1)}\left(0\right)\right|^2 + \left|\nu_{Lu}^{(2)}\left(0\right)\right|^2\right)^{1/2}$ is normalized to 1. Define $\Delta m^2 \equiv m_s^2 - m_a^2$. We obtain

$$P_{aa} = \left[\left|\nu_{Lu}^{(1)}\left(L\right)\right|^2 + \left|\nu_{Lu}^{(2)}\left(L\right)\right|^2\right]^{1/2} = 1-P_{as} = 1-4\cos^2\theta\sin^2\theta\sin^2\left(\frac{\Delta m^2}{4E}L\right). \tag{23}$$

The lepton violating reaction has probability $P_{a\bar{a}}$ that differs from P_{aa} by polarization miss-match factors (to be discussed in Section 5):

$$P_{a\bar{a}} = \frac{1}{2E^2}\left[\left(m_a\cos^2\theta+m_s\sin^2\theta\right)^2 - 4m_a m_s\cos^2\theta\sin^2\theta\sin^2\left(\frac{\Delta m^2}{4E}L\right)\right]. \tag{24}$$

The interpretation of these equations is discussed in Section 4.

Let us generalize to $n_g = 3$ generations of weak $SU(2)_L$ doublets, and $n' = 3$ gauge singlet Weyl_R neutrinos. We introduce the notation

$$\vec{\nu} \equiv \begin{pmatrix} \vec{\nu}_{Lu} \\ (\vec{\nu}_{Rv})^c \end{pmatrix}, \quad \vec{\nu}_{Lu} \equiv \begin{pmatrix} \nu_{eLu} \\ \nu_{\mu Lu} \\ \nu_{\tau Lu} \end{pmatrix}, \quad (\vec{\nu}_{Rv})^c \equiv \begin{pmatrix} (\nu_{1Rv})^c \\ (\nu_{2Rv})^c \\ (\nu_{3Rv})^c \end{pmatrix}. \tag{25}$$

These fields are related to the mass eigenstates as follows:

$$\vec{\nu}_{\text{mass}} = U^\dagger \vec{\nu}, \quad \vec{\nu} = U \vec{\nu}_{\text{mass}}. \tag{26}$$

The $(n_g + n') \times (n_g + n')$ weak mixing matrix U is unitary: $UU^\dagger = 1$, $U^\dagger U = 1$. The generalization of (15) is

$$\begin{pmatrix} i\sigma_\mu \partial_\mu (\vec{\nu}_{Lv})^c \\ i\sigma_\mu \partial_\mu \vec{\nu}_{Ru} \end{pmatrix} = \begin{pmatrix} 0 & m_D^{\mathrm{T}} \\ m_D & M \end{pmatrix} \begin{pmatrix} \vec{\nu}_{Lu} \\ (\vec{\nu}_{Rv})^c \end{pmatrix}. \tag{27}$$

$m_D = Y^N v_h / \sqrt{2}$ is the complex $n' \times n_g$ Dirac mass matrix, and M is the symmetric $n' \times n'$ Majorana mass matrix chosen real. The symmetric mass matrix is diagonalized as follows:

$$U^{\mathrm{T}} \begin{pmatrix} 0 & m_D^{\mathrm{T}} \\ m_D & M \end{pmatrix} U = \mathrm{diag}\left(m_1, m_2, \cdots, m_{n_g + n'}\right). \tag{28}$$

The masses m_i of the mass eigenstates are chosen real and positive. The fields $\vec{\nu}$ evolve as (22). Then, for ultra-relativistic neutrinos, the probability of a lepton conserving event, e.g. $e^- W^+ \to \nu_i \to \mu^- W^+$, is [2]

$$\begin{aligned} P_{\alpha\beta} = \delta_{\alpha\beta} &- 4 \sum_{i<j}^{n_g + n'} \mathrm{Re}\left[U_{\alpha i} U_{\beta i}^* U_{\alpha j}^* U_{\beta j}\right] \sin^2 X_{ij} \\ &+ 2 \sum_{i<j}^{n_g + n'} \mathrm{Im}\left[U_{\alpha i} U_{\beta i}^* U_{\alpha j}^* U_{\beta j}\right] \sin\left(2X_{ij}\right), \end{aligned} \tag{29}$$

$$\equiv P_{\alpha\beta\mathrm{CPC}} + P_{\alpha\beta\mathrm{CPV}}, \tag{30}$$

where $X_{ij} = \left(m_i^2 - m_j^2\right) L / (4E)$. The sums in (29) are over mass eigenstates i, j with masses m_i and m_j that cannot be discriminated in the experiment, and are coherent, see Section 6. Under CP conjugation, $U \to U^*$. The first two terms on the right hand side of (29), denoted $P_{\alpha\beta\mathrm{CPC}}$, are CP conserving, while the last term, denoted $P_{\alpha\beta\mathrm{CPV}}$, may be CP violating. Note that to obtain CP violation, at least one physical phase in U is needed, in addition to the propagation phase $2X_{ij}$ (that requires $m_i \neq m_j$). Since U is unitary, $\sum_\alpha P_{\alpha\beta} = \sum_\beta P_{\alpha\beta} = 1$.

The probability to observe a lepton violating event, e.g. $e^- W^+ \to \nu_i \to \mu^+ W^-$, is

$$\begin{aligned} P_{\alpha\bar{\beta}} = \frac{1}{2E^2} \Bigg\{ &\left| \sum_i^{n_g + n'} m_i U_{\alpha i}^* U_{\beta i}^* \right|^2 - 4 \sum_{i<j}^{n_g + n'} m_i m_j \,\mathrm{Re}\left[U_{\alpha i}^* U_{\beta i}^* U_{\alpha j} U_{\beta j}\right] \sin^2 X_{ij} \\ &+ 2 \sum_{i<j}^{n_g + n'} m_i m_j \,\mathrm{Im}\left[U_{\alpha i}^* U_{\beta i}^* U_{\alpha j} U_{\beta j}\right] \sin\left(2X_{ij}\right) \Bigg\}, \end{aligned} \tag{31}$$

$$\equiv P_{\alpha\bar{\beta}\mathrm{CPC}} + P_{\alpha\bar{\beta}\mathrm{CPV}} \ll 1, \tag{32}$$

where we have included the polarization miss-match factors discussed in Section 5. The probability $P_{\bar{\alpha}\beta}$ for the CP-conjugate event, e.g. $e^+W^- \to \bar{\nu}_i \to \mu^-W^+$, is obtained by $U \to U^*$. Note that the first two terms, denoted $P_{\alpha\bar{\beta}\mathrm{CPC}}$, violate lepton number but conserve the CP symmetry. Note that the last term, denoted $P_{\alpha\bar{\beta}\mathrm{CPV}}$, is lepton number violating and may be CP violating, and is the source of the leptogenesis studied in this article. Considering the width of the neutrino energy distribution, strong CP violation requires $2X_{ij} \lesssim \pi/2$. The last two terms in (29) and (31) are due to coherent interference of the mass eigenstates.

Note that for $L \to 0$, $P_{\alpha\bar{\alpha}}$ is proportional to the square of the "effective mass" $\sum_i m_i \left|U_{\alpha i}\right|^2$. Neutrino-less double beta decay experiments constrain the effective mass of electrons to be less than 0.165 eV [2].

Equations (29) and (31) assume neutrinos are nearly on mass shell, *i.e.* $L \gg \lambda = 2\pi/p_z$, and that the neutrino mean energy $E \gg m_i$. The sums in these equations only include neutrinos that are coherent, see Section 6.

4. Interpretation

In a neutrino oscillation experiment, most neutrinos traverse the detector without interacting. In the limit $L \to 0$, $P_{\alpha\beta} \to \delta_{\alpha\beta}$. The probabilities $P_{\alpha\beta}$ are defined as the number of $\nu_\beta l_{\bar{\beta}} \to W^+$ counts at the far detector, divided by the number of $\nu_\alpha l_{\bar{\alpha}} \to W^+$ counts at the near detector, corrected for acceptance and detector efficiencies. If the efficiency of the detector for ν_β is negligible, $P_{\alpha\beta}$ may still be "measured" as a disappearance in the sum of other channels. For the probability $P_{\alpha\bar{\beta}}$, the detector efficiency factor due to polarization miss-match has already been included.

The interpretation of the preceding equations needs an understanding of the entire experiment. In particular we need to consider polarization miss-match (Section 5), and coherence (Section 6). If at the source the neutrino mass is sufficiently uncertain, then a weak state is produced, *i.e.* a coherent superposition of mass eigenstates. If at the source the neutrino mass is sufficiently well determined, then a mass eigenstate is produced. Even if the neutrino mass eigenstates are produced coherently, they may lose coherence before being detected, either in transit, and/or at the detector. If this is the case, then an "observation" has been made, and we need to pass from amplitudes to probabilities, *i.e.* interference terms are lost.

For simplicity, we consider a single generation, *i.e.* (15) to (24). If production is coherent, and the mass eigenstates have become incoherent, then the probability for ν_a is $\cos^2\theta$, and the probability for ν_s is $\sin^2\theta$. In the case of coherent production and incoherent detection, e.g. the neutrino mass is measured at the detector with a resolution sufficient to discriminate between m_a and m_s, the combined probability to have a ν_a and a lepton conserving event $e^-W^+ \to e^-W^+$ is $\cos^4\theta$, the combined probability to have a ν_s and a lepton

conserving event is $\sin^4\theta$, the combined probability to have a ν_a and a lepton violating event $e^-W^+ \to e^+W^-$ is $m_a m_s \cos^4\theta/(2E^2)$, and the combined probability to have a ν_s and a lepton violating event is $m_a m_s \sin^4\theta/(2E^2)$. The factor $m_a m_s/(2E^2)$ is due to polarization miss-match (Section 5) that reduces the cross-section by this factor, and applies to a real, *i.e.* finite extent, unpolarized detector. We assume $m_a/E \ll 1$ and $m_s/E \ll 1$. E is the energy of the neutrino in the laboratory frame, *i.e.* the detector.

The case of interest to the leptogenesis scenario studied in this article is coherent production and coherent detection, since interference is needed for CP violation. If production is coherent, and the mass eigenstates remain coherent at detection, then P_{aa} of (23) is the probability to observe the lepton conserving event $e^-W^+ \to e^-W^+$. In any practical neutrino oscillation experiment, with a finite detector, $P_{a\bar{a}}$ of (24) is the probability to observe the lepton violating event $e^-W^+ \to e^+W^-$.

5. Polarization Miss-Match

Consider the decay $W^+ \to e^+\nu_e$ in the rest frame of W^+. The neutrino ν_e carries the field ν_{Lu}, see (4). Let Θ be the angle between the W^+ spin and the ν_e momentum. Then $\xi_1 = \cos(\Theta/2)$ and $\xi_2 = i\sin(\Theta/2)$ [4]. Therefore, the amplitude for W^+ to create a right-handed (*i.e.* helicity +1/2) ν_e is $\propto \cos(\Theta/2)\sqrt{E-p_z}$, and the amplitude to create a left-handed ν_e is $\propto i\sin(\Theta/2)\sqrt{E+p_z}$. Helicity is the projection of the spin angular momentum in the direction of the momentum, *i.e.* $\vec{s}\cdot\hat{p} = (1/2)\vec{\sigma}\cdot\hat{p}$. Note that most ultra-relativistic ν_e are left-handed.

Consider the CP-conjugate decay $W^- \to e^-\bar{\nu}_e$ in the rest frame of W^-. The anti-neutrino $\bar{\nu}_e$ carries the field $(\nu_{Lv})^c$, see (6). Let Θ be the angle between the W^- spin and the $\bar{\nu}_e$ momentum. Then $\eta_2^* = \cos(\Theta/2)$ and $-\eta_1^* = i\sin(\Theta/2)$. Therefore, the amplitude for W^- to create a right-handed (*i.e.* helicity +1/2) $\bar{\nu}_e$ is $\propto \cos(\Theta/2)\sqrt{E+p_z}$, and the amplitude to create a left-handed $\bar{\nu}_e$ is $\propto i\sin(\Theta/2)\sqrt{E-p_z}$. Note that most ultra-relativistic $\bar{\nu}_e$ are right-handed.

Note that for ultra-relativistic Majorana neutrinos we can still distinguish neutrinos (lepton number ≈+1 and helicity ≈−1/2) from anti-neutrinos (lepton number ≈−1 and helicity ≈+1/2), since lepton number is conserved to a high degree of accuracy, see Section 10 for a numerical example.

Consider the lepton-conserving sequence of events $W^+ \to e^+\nu_e$ followed by $\nu_\mu\mu^+ \to W^+$. We assume $E \gg m_e$ and $E \gg m_\mu$. The probability, after averaging over $\cos\Theta$ and $\cos\Theta'$, is $\propto \left|\sum_i a_i\right|^2 E^2$, where a_i is the amplitude corresponding to mass eigenstate i, and Θ' is the angle with respect to the final W^+. Consider the lepton-violating sequence of events $W^+ \to e^+\nu_e$ followed by $\bar{\nu}_\mu\mu^- \to W^-$. The probability, after averaging over $\cos\Theta$ and $\cos\Theta'$, is $\propto \left|\sum_i a_i m_i\right|^2/2$. The probability for the lepton conserving events (29) is norma-

lized to $\left|\sum_i a_i\right|^2 = 1$, so the relative polarization miss-match factor for lepton violating events is $\left|\sum_i a_i m_i\right|^2 / \left(2E^2\right)$. This factor has been included in (24) and (31). We note that in the limit $m_i / E \to 0$, the experimental distinction between Dirac and Majorana neutrinos fades away, and lepton violation vanishes.

6. Coherence

The sums in (29) and (31) only include coherent neutrinos. To obtain coherent oscillations between two neutrinos of masses m_i and m_j it is necessary that the energy uncertainty σ_E of the produced and detected neutrinos be sufficiently large, *i.e.* $\sigma_E \gtrsim \sqrt{\Delta m^2 / 8}$, where $\Delta m^2 \equiv \left|m_j^2 - m_i^2\right|$ [7]. If this condition is met, what is created is a flavor eigenstate, *i.e.* a coherent superposition of mass eigenstates, and oscillations remain coherent while the wave packets of the two components overlap. The overlap ceases after the "coherence time" [7]

$$t_{\text{coh}} = 2\sqrt{2}\frac{2E^2}{\left|\Delta m^2\right|}\sigma_t, \tag{33}$$

where σ_t is the Gaussian wave packet duration. The combined coherence factor after time Δt is [7]

$$\varepsilon_{\text{coh}} = \exp\left[-\Delta m^2 / \left(8\sigma_E^2\right)\right] \cdot \exp\left[-\Delta t^2 / t_{\text{coh}}^2\right]. \tag{34}$$

In the present application we take, arguably, $\sigma_t \approx 1/\Gamma_W$, $\sigma_E \approx \Gamma_W$, and $\Delta t \approx t_c$, where t_c is the mean time for a neutrino to collide with a charged lepton. Consider the reference temperature $T_{\text{sph}} = 131.7\ \text{GeV}$, and $E \approx 2.8 T_{\text{sph}}$. The cross-section σ for $e^+\nu_e \to e^+\nu_e$ is given by (50.25) of [2]. We obtain $t_c = 1/\left(n_{e^-}\sigma c\right) \approx 7\times 10^{-22}\,\text{s} = 1/\left(0.001\,\text{GeV}\right)$, where n_{e^-} is the electron number density at T_{sph}. Consider an active-sterile neutrino oscillation with $\Delta m_{sa}^2 = m_s^2 - m_a^2 \approx m_s^2$. Requiring coherent oscillations, $\Delta m_{sa}^2 / \left(8\sigma_E^2\right) \lesssim 1$ obtains the bound $m_s \lesssim 6\ \text{GeV}$. For $E \approx 2.8 T_{\text{sph}}$, $\Delta t = t_c = t_{\text{coh}}$ corresponds to $m_s = 19\ \text{GeV}$. Therefore, for $m_s \lesssim 6\ \text{GeV}$, oscillations remain coherent for $\gtrsim 10 t_c$. We verify also that $t_c \gg 2\pi / p$, with $p = \left(E^2 - m^2\right)^{1/2}$, so the neutrinos of interest are nearly on mass shell.

7. Asymmetry Build-Up

So far we have been studying a neutrino oscillation experiment with baseline L. In this Section we apply the results to the universe when it has the reference temperature $T_{\text{sph}} = 131.7\ \text{GeV}$. Consider a single ν_e. The ν_e lifetime is t_c. The probability that the interaction $\nu_e e^\pm \to W^\pm$ occurs in the time interval from t to $t + \text{d}t$ is $\text{d}P = \text{e}^{-t/t_c}\text{d}t/t_c$. The lepton number violating and CP violating asymmetry, $P_{\alpha\bar{\beta}\text{CPV}}(t)$, is proportional to t for the case of interest $2X_{ij} \lesssim \pi/2$. The mean of $P_{\alpha\bar{\beta}\text{CPV}}(t)$ is then $P_{\alpha\bar{\beta}\text{CPV}}(t_c)$.

Consider the contribution of the channel $\alpha\bar{\beta} + \bar{\alpha}\beta$ to δ_l. Let $n_{\beta,i}$ be the comoving number density of l_β at time t_i, where $l_\beta = e^-, \mu^-, \tau^-$. Then, at time $t_{i+1} = t_i + t_c$,

$$\begin{aligned} n_{\beta,i+1} &= n_{\beta,i} + n_{\alpha,i}P_{\alpha\beta} + n_{\bar{\alpha},i}P_{\bar{\alpha}\beta,i}, \\ n_{\bar{\beta},i+1} &= n_{\bar{\beta},i} + n_{\alpha,i}P_{\alpha\bar{\beta}} + n_{\bar{\alpha},i}P_{\bar{\alpha}\bar{\beta},i}. \end{aligned} \tag{35}$$

Taking the difference of these two equations, and dividing by $n_{\beta,i+1} + n_{\bar{\beta},i+1} \approx n_{\beta,i} + n_{\bar{\beta},i} \approx n_{\alpha,i} + n_{\bar{\alpha},i}$, obtains

$$\delta_{\beta,i+1} \approx \delta_{\beta,i} + \left(P_{\alpha\beta\mathrm{CPV}} - P_{\alpha\bar{\beta}\mathrm{CPV}}\right) + \delta_{\alpha,i}\left(P_{\alpha\beta\mathrm{CPC}} - P_{\alpha\bar{\beta}\mathrm{CPC}}\right). \tag{36}$$

Summing over α and β obtains

$$\delta_{l,i+1} \approx \delta_{l,i} - \sum_{\alpha\bar{\beta}} P_{\alpha\bar{\beta}\mathrm{CPV}} - \sum_{\alpha\bar{\beta}} \delta_{\alpha,i} P_{\alpha\bar{\beta}\mathrm{CPC}}. \tag{37}$$

The last term is the "wash-out" term that tends to restore the equilibrium value $\delta_l = 0$. Note that δ_l decreases by $\sum_{\alpha\bar{\beta}} P_{\alpha\bar{\beta}\mathrm{CPV}}$ in time t_c until it reaches either

$$\delta_{l\mathrm{MAX}} = -\frac{t_u}{t_c}\sum_{\alpha\bar{\beta}} P_{\alpha\bar{\beta}\mathrm{CPV}}, \tag{38}$$

or until wash-out sets in at

$$\delta_{l\mathrm{WO}} \sum_{\alpha\bar{\beta}} P_{\alpha\bar{\beta}\mathrm{CPC}} \equiv \sum_{\alpha\bar{\beta}} \delta_{\alpha} P_{\alpha\bar{\beta}\mathrm{CPC}} = -\sum_{\alpha\bar{\beta}} P_{\alpha\bar{\beta}\mathrm{CPV}}. \tag{39}$$

We note that $P_{\alpha\bar{\beta}\mathrm{CPV}}$ and $P_{\alpha\bar{\beta}\mathrm{CPC}}$ are proportional to $L = t_c$ if $2X_{ij} \lesssim \pi/2$, so, in this case of interest, $\delta_{l\mathrm{MAX}}$ and $\delta_{l\mathrm{WO}}$ are independent of t_c.

8. Constraints from Cosmology

Constraints from Big Bang Nucleosynthesis (BBN), Baryon Acoustic Oscillations (BAU), and direct searches, limit the mass of sterile neutrinos to be greater than 0.14 GeV [8], so the interesting sterile neutrino mass range, for the leptogenesis scenario being considered, is approximately 0.14 GeV to 1.1 GeV. The lifetimes of these neutrinos range from approximately 10^{-5} s at $m_s = 1\,\mathrm{GeV}$, to 0.1 s for $m_s = 0.13\,\mathrm{GeV}$ [8].

Big Bang Nuleosynthesis and cosmic microwave background (CMB) measurements do not allow one additional ultra-relativistic degree of freedom at $T_{\mathrm{BBN}} \approx 1\,\mathrm{MeV}$ [2]. For the Standard Model, the equivalent number of neutrinos (for BBN) is $N_\nu = 3.045$ [2]. The Planck CMB result gives $N_\nu = 2.92^{+0.36}_{-0.37}$ at 95% confidence [2]. So an extra stable neutrino, that was once in statistical equilibrium with the Standard Model sector, needs to decouple at $T > T_C \approx 0.14\,\mathrm{GeV}$, where T_C is the confinement-deconfinement temperature. Such an extra neutrino contributes ≤0.12 to N_ν. Therefore, sterile neutrinos either 1) never reached statistical equilibrium with the Standard Molel sector, or 2) reached statistical equilibrium, but decoupled at $T > T_C$ and hence are sufficiently cooler than active neutrinos at T_{BBN} [2], or 3) the sterile neutrino mass is $m_s > T_{\mathrm{BBN}}$ and these neutrinos decayed before T reached T_{BBN}.

As an example, for $m_a = 0.005\,\mathrm{eV}$, and m_s in the range of interest 0.14 GeV to 1.1 GeV, we find that sterile neutrinos never reach statistical equilibrium with the Standard Model sector, or reach equilibrium but decouple at $T > 0.14\,\mathrm{GeV}$, and hence do not affect BBN.

9. Leptogenesis with $n_g = 1$ and $n' = 2$

Let us study the simplest case with lepton number violation and CP violation. We take $n_g = 1$ generations of active neutrinos and $n' = 2$ gauge singlet Weyl_R neutrinos. For $m_a \ll M_1$ and $m_a \ll M_2$, the unitary mixing matrix U from (28), to order $|m_{Di}|^2 / M_i^2$, is

$$U_{\alpha i} \approx \begin{pmatrix} -i\left(1 - \frac{|m_{D1}|^2}{2M_1^2} - \frac{|m_{D2}|^2}{2M_2^2}\right) & \frac{m_{D1}^*}{M_1} & \frac{m_{D2}^*}{M_2} \\ i\frac{m_{D1}}{M_1} & 1 - \frac{|m_{D1}|^2}{2M_1^2} & -\frac{m_{D1}m_{D2}^*}{2M_1M_2} \\ i\frac{m_{D2}}{M_2} & -\frac{m_{D1}^* m_{D2}}{2M_1M_2} & 1 - \frac{|m_{D2}|^2}{2M_2^2} \end{pmatrix}, \tag{40}$$

and the mass eigenstates are

$$m_a = \frac{m_{D1}^2}{M_1} + \frac{m_{D2}^2}{M_2}, \tag{41}$$

$$m_{s1} = \left(1 + \frac{|m_{D1}|^2}{M_1^2}\right) M_1, \tag{42}$$

$$m_{s2} = \left(1 + \frac{|m_{D2}|^2}{M_2^2}\right) M_2. \tag{43}$$

The active neutrino mass m_a, and sterile neutrino masses M_1 and M_2 are real and positive. The Dirac terms m_{D1} and m_{D2} are complex.

Let us write (38) for the present case $n_g = 1$, $n' = 2$. To order $m_a^2 M^2$ we obtain

$$\delta_e = \frac{n_{e^-} - n_{e^+}}{n_{e^-} + n_{e^+}} = \frac{-t_u}{2 \cdot \left(2.8 T_{\text{sph}}\right)^3} \left[m_a M_1^2 \,\text{Im}\left(\frac{m_{D1}^{*2}}{M_1}\right) + m_a M_2^2 \,\text{Im}\left(\frac{m_{D2}^{*2}}{M_2}\right) + \left(M_1^2 - M_2^2\right) \text{Im}\left(\frac{m_{D1}^2 m_{D2}^{*2}}{M_1 M_2}\right) \right]. \tag{44}$$

This equation assumes the approximation $\sin\left(2X_{ij}\right) \approx 2X_{ij}$ valid for $2X_{ij} \lesssim \pi/2$, corresponding to masses $\lesssim 1.1\,\text{GeV}$. The three terms correspond to interference of neutrinos $\nu_a \leftrightarrow \nu_{s1}$, $\nu_a \leftrightarrow \nu_{s2}$, and $\nu_{s1} \leftrightarrow \nu_{s2}$, respectively. We find that all terms of order $m_a^2 M_i^2$ cancel. In conclusion, if both sterile neutrinos have masses in the approximate range 0.14 to 1.1 leptogenesis is negligible.

Let us consider the case $0.14 \lesssim M_1 \lesssim 1.1\,\text{GeV}$, and $M_2 \gg 6\,\text{GeV}$ so that ν_{s2} is incoherent. In this case we keep only the first term in (44). Leptogenesis can be successful if we are able to find m_{D1} and m_{D2} that satisfy (41) and (44) (omitting the terms with M_2) with the required $\delta_l = \delta_e = -3.1 \times 10^{-8}$. Note that (41) and (44) are under-constrained: they have multiple solutions. We therefore use the Casas-Ibarra procedure [9], and write the solution to (41) in the form

$m_{Di} = \sqrt{M_i} R_i \sqrt{m_a}$, where R is any orthogonal, *i.e.* $R^{\mathrm{T}} R = 1$, $n' \times n_g$ matrix. In the present case

$$R = \begin{pmatrix} R_1 \\ R_2 \end{pmatrix}. \tag{45}$$

Substituting in (44) we obtain

$$\delta_e = \frac{-t_u m_a^2 M_1^2}{2 \cdot \left(2.8 T_{\mathrm{sph}}\right)^3} \mathrm{Im}\left(R_1^{*2}\right). \tag{46}$$

As an example, we take $M_1 = 1\,\mathrm{GeV}$, $M_2 = 50\,\mathrm{GeV}$, and $m_a = 0.015\,\mathrm{eV}$. Also $t_u = 1/\left(4.6 \times 10^{-14}\,\mathrm{GeV}\right)$, and $T_{\mathrm{sph}} = 131.7\,\mathrm{GeV}$. We obtain $\mathrm{Im}\left(R_1^{*2}\right) = 6.4 \times 10^8$. A solution then needs fine tuning, e.g.

$$R_1 = \mathrm{e}^{-i\pi/4} K, \quad R_2 = \mathrm{e}^{i\pi/4} K \sqrt{1 - i/K^2}, \tag{47}$$

where $K = \sqrt{6.4 \times 10^8}$. Since we have chosen $\mathrm{Re}\left(R_1^{*2}\right) = 0$, wash-out remains negligible.

Equation (44) can be generalized to $n' > 2$ by inspection.

10. Leptogenesis with $n_g = 3$ and $n' = 3$

Without loss of generality we work in a basis that diagonalizes the $n_g \times n_g$ charged lepton mass matrix, and the $n' \times n'$ Majorana mass matrix M. The $\left(n_g + n'\right) \times \left(n_g + n'\right)$ weak mixing matrix U, defined in (28), to lowest order in m_D / M, has the form [2]

$$U \approx \begin{pmatrix} \left(1 - \frac{1}{2} m_D^{\dagger} M^{*-1} M^{-1} m_D\right) V_l & m_D^{\dagger} M^{*-1} V_h \\ -M^{-1} m_D V_l & \left(1 - \frac{1}{2} M^{-1} m_D m_D^{\dagger} M^{*-1}\right) V_h \end{pmatrix}. \tag{48}$$

To the present order of approximation, we take $V_h = 1$. The 3 × 3 PMNS weak mixing matrix of active neutrinos V_l depends on three angles and one Dirac CP-violating phase δ_{CP}, that have been measured, and two Majorana CP-violating phases η_1 and η_2 (with the notation of [2]) that have not been measured. Unphysical phases of V_l, that can be canceled by re-phasing fields, have already been fixed. We take the central measured values of these parameters for normal (NO) or inverse (IO) neutrino mass ordering from the first column of Table 14.7 of [2]. The two active neutrino mass-squared differences are also obtained from this Table. η_1 and η_2 are free parameters until measured. Neutrino oscillation experiments show that at least 2 neutrino eigenstates have mass, so (arguably) at least $n' = 2$ gauge singlet Weyl_R neutrinos need to be added to the Standard Model. For the case $n' = 2$, the lightest active neutrino mass is zero. For the case $n' = 3$, the lightest active neutrino mass is a free parameter until measured.

The diagonal mass matrix of active neutrinos, obtained from (28) and (48), is [2]

$$m^l \equiv \mathrm{diag}\left(m_1, m_2, m_3\right) \approx -V_l^{\mathrm{T}} m_D^{\mathrm{T}} M^{-1} m_D V_l. \tag{49}$$

Successful leptogenesis is possible if we are able to solve (49), (38), and (39) with the needed lepton asymmetry $\delta_l = \delta_e + \delta_\mu + \delta_\tau = -3.1\times10^{-8}$. We focus on the case $n' = 3$. The problem is under-constrained, so again we follow the Casas-Ibarra procedure [9]. From (49), with the notation $\sqrt{M} \equiv \mathrm{diag}\left(\sqrt{M_1}, \sqrt{M_2}, \cdots\right)$, we obtain $R^{\mathrm{T}} R = 1$ with $R \equiv i\sqrt{M^{-1}} m_D V_l \sqrt{m^{l-1}}$. Finally,

$$m_D = Y^N \frac{v_h}{\sqrt{2}} = -i\sqrt{M} R \sqrt{m^l} V_l^\dagger, \tag{50}$$

where R is any orthogonal, *i.e.*

$$R^{\mathrm{T}} R = 1, \tag{51}$$

$n' \times n_g$ matrix. Equation (49) is satisfied by (50). To satisfy (38), the matrix R needs to be complex.

Successful leptogenesis requires fine tuning of the unknown parameters. As a proof of principle we present the following example: normal neutrino mass ordering, $m_1 = 0.001\,\mathrm{eV}$, $M_1 = 1.04\,\mathrm{GeV}$, and $M_2 = 1.06\,\mathrm{GeV}$. We also set $\eta_1 = 0$, $\eta_2 = 0$, and $M_3 = 60\,\mathrm{GeV}$, and note that the results depend negligibly on these last three parameters (for the texture of the matrix R chosen below). At $T = T_{\mathrm{sph}}$, the age of the universe is $t_u = 1.4\times10^{-11}\,\mathrm{s} = 1/\left(4.6\times10^{-14}\,\mathrm{GeV}\right)$, and $t_c = 7\times10^{-22}\,\mathrm{s} = 1/\left(0.001\,\mathrm{GeV}\right)$. Equation (51) has many solutions. Successful leptogenesis needs $\left|R_{\alpha i} R_{\beta i}\right| \gg 1$ and $R_{\alpha i} R_{\beta i}$ imaginary to high accuracy, as in (47). To obtain a solution that satisfies $\delta_l = -3.1\times10^{-8}$ we choose a particular texture of R (that makes the results insensitive to M_3), and obtain:

$$R = K\begin{pmatrix} -\mathrm{e}^{i\pi/4} & -\mathrm{e}^{-i\pi/4} & 0 \\ -\mathrm{e}^{-i\pi/4} & \mathrm{e}^{i\pi/4} & 0 \\ 0 & 0 & \dfrac{1}{K} \end{pmatrix} + \frac{1}{4K}\begin{pmatrix} -\mathrm{e}^{-i\pi/4} & -\mathrm{e}^{i\pi/4} & 0 \\ -\mathrm{e}^{i\pi/4} & \mathrm{e}^{-i\pi/4} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{52}$$

with $K = 52540$. The probabilities for a "neutrino oscillation experiment" with $L = t_c$, from terms in (29) and (31) are respectively:

$$P_{e^-e^-} = 1 - 3.4\times10^{-2} - 4.3\times10^{-19},$$

$$P_{e^-e^+} = 6.5\times10^{-32} + 1.1\times10^{-12} + 2.0\times10^{-19},$$

$$P_{e^-\mu^-} = 6.2\times10^{-4} - 8.2\times10^{-15},$$

$$P_{e^-\mu^+} = 6.3\times10^{-32} + 1.2\times10^{-12} - 1.7\times10^{-19},$$

$$P_{e^-\tau^-} = 5.1\times10^{-4} - 8.7\times10^{-15},$$

$$P_{e^-\tau^+} = 5.3\times10^{-32} + 9.5\times10^{-13} + 4.2\times10^{-19},$$

$$P_{\mu^-e^-} = 6.2\times10^{-4} + 8.2\times10^{-15},$$

$$P_{\mu^-e^+} = 6.3\times10^{-32} + 1.2\times10^{-12} - 1.7\times10^{-19},$$

$$P_{\mu^-\mu^-} = 1 - 3.6\times10^{-2} - 8.7\times10^{-19},$$
$$P_{\mu^-\mu^+} = 8.0\times10^{-32} + 1.2\times10^{-12} + 2.0\times10^{-18},$$
$$P_{\mu^-\tau^-} = 5.4\times10^{-4} - 2.0\times10^{-15},$$
$$P_{\mu^-\tau^+} = 7.1\times10^{-32} + 1.0\times10^{-12} - 1.2\times10^{-18},$$
$$P_{\tau^-e^-} = 5.1\times10^{-4} + 8.7\times10^{-15},$$
$$P_{\tau^-e^+} = 5.3\times10^{-32} + 9.5\times10^{-13} + 4.2\times10^{-19},$$
$$P_{\tau^-\mu^-} = 5.4\times10^{-4} + 2.0\times10^{-15},$$
$$P_{\tau^-\mu^+} = 7.1\times10^{-32} + 1.0\times10^{-12} - 1.2\times10^{-18},$$
$$P_{\tau^-\tau^-} = 1 - 2.9\times10^{-2} + 0.0,$$
$$P_{\tau^-\tau^+} = 6.3\times10^{-32} + 8.2\times10^{-13} + 1.3\times10^{-18}. \quad (53)$$

We note that the lepton number violating reactions are suppressed with respect to the lepton conserving ones, and the CP violating terms are suppressed with respect to the CP conserving ones. We note that the terms $P_{\alpha\bar{\beta}\mathrm{CPC}}$ are of order 10^{-12} and positive, while the terms $P_{\alpha\bar{\beta}\mathrm{CPV}}$ are of order 10^{-18} and can be positive or negative.

From the first term in $P_{e^-e^+}$ we obtain the effective mass of neutrino-less double beta decay experiments for this example:
$\sqrt{6.5\times10^{-32}\times2\cdot2.8\cdot T_{\mathrm{sph}}} = 0.00013\ \mathrm{eV}$. The current limit is 0.165 eV [2].

Asymmetries per channel are presented in **Table 1** (from (38) and (39) without the sums over α and $\bar{\beta}$). Note that, in this example, we do not reach saturation due to wash-out, *i.e.* $|\delta_{l\mathrm{MAX}}| < |\delta_{l\mathrm{WO}}|$ in each channel. Summing the asymmetries in the last row of **Table 1** obtains $\delta_l = \delta_{l\mathrm{MAX}} = -3.1\times10^{-8}$ as required. In conclusion, successful baryogenesis via leptogenesis may be achieved with the fine-tuning shown in (52), plus the fine tuning of the masses (so that the positive and negative terms in the last sum of (39) cancel to one part in 10^2).

Table 1. Lepton number asymmetries per channel $\delta_{l\mathrm{MAX}}$ (upper numbers), and wash-out asymmetries $\delta_{l\mathrm{WO}}$ (in parenthesis) for individual channels. $|\delta_l|$ is the least of $|\delta_{l\mathrm{MAX}}|$ and $|\delta_{l\mathrm{WO}}|$.

$\bar{\alpha}$	δ_e	δ_μ	δ_τ
e^+	-4.0×10^{-9} (-1.8×10^{-7})	3.4×10^{-9} (1.4×10^{-7})	-8.6×10^{-9} (-4.4×10^{-7})
μ^+	3.4×10^{-9} (1.4×10^{-7})	-4.0×10^{-8} (-1.6×10^{-6})	2.5×10^{-8} (1.2×10^{-6})
τ^+	-8.6×10^{-9} (-4.4×10^{-7})	2.5×10^{-8} (1.2×10^{-6})	-2.7×10^{-8} (-1.6×10^{-6})
Sum of δ_β	-9.2×10^{-9}	-1.2×10^{-8}	-1.0×10^{-8}

Several tests with modifications of this example follow:

- Setting $\delta_{CP}=0$, or $\eta_1=0.785$, or $\eta_2=0.785$, obtains $\delta_{l\mathrm{MAX}}=-3.1\times10^{-8}$ as before, so the CP-violating phases in the PMNS matrix V_l contribute negligibly to leptogenesis (in this scenario of active-sterile neutrino oscillations). Leptogenesis is mostly due to the CP-violating asymmetries in *R*, or equivalently m_D.
- Choosing a real *R* obtains $\delta_{l\mathrm{MAX}}=1.4\times10^{-34}$, so, again, the Dirac phase δ_{CP} of V_l contributes negligibly to leptogenesis.
- Setting $\delta_{CP}=\eta_1=\eta_2=0$ and a real *R* obtains $\delta_l=0$, as expected. This cross-check is satisfied for 0, 1, 2 or 3 coherent sterile neutrinos.
- Setting $M_2=50\,\mathrm{GeV}$, to test a case with one coherent sterile neutrino, obtains $\delta_{l\mathrm{MAX}}=-3.0\times10^{-8}$ and $\delta_{l\mathrm{WO}}=-2.8\times10^{-10}$. Note that wash-out dominates.
- Setting $M_1=1\,\mathrm{GeV}$, $M_2=0.5\,\mathrm{GeV}$ and $M_3=0.2\,\mathrm{GeV}$, to test a case with three coherent neutrinos, obtains $\delta_{l\mathrm{MAX}}=8.8\times10^{-9}$ and $\delta_{l\mathrm{WO}}=1.6\times10^{-10}$. Note that the signs are now wrong, and wash-out dominates.
- Results for inverse neutrino mass ordering are similar. However, we were unable to reach successful leptogenesis, *i.e.* $\delta_l=-3.1\times10^{-8}$.

11. Sterile Neutrino Dark Matter?

Detailed dark matter properties have recently been obtained by fitting spiral galaxy rotation curves, and, *independently*, by fitting galaxy stellar mass distributions [10]. These measurements imply that dark matter was in thermal and diffusive equilibrium with the Standard Model sector in the early universe, and decoupled (from the Standard Model sector and from self-annihilation) at a temperature $T\gtrsim0.2\,\mathrm{GeV}$. If dark matter particles are fermions, the measurements obtain their mass $m_h=107^{+37}_{-20}\,\mathrm{eV}$ [10]. This mass is disfavored by the Tremaine-Gunn limit (that applies to fermion dark matter) [11], which however needs revision [12] [13] [14] [15]. Fermion dark matter is also disfavored, relative to boson dark matter, by spiral galaxy rotation curves and by galaxy stellar mass distributions with a significance of 3.5σ [10].

Nevertheless, let us see if sterile neutrinos of mass m_h could have reached statistical equilibrium with the Standard Model sector by the time of the confinement-deconfinement temperature $T_C\approx0.14\,\mathrm{GeV}$. This is the minimum temperature at which dark matter in equilibrium with the Standard Model sector can decouple without spoiling the agreement with Big-Bang Nucleosynthesis. At this temperature the age of the universe is $t_u=1.7\times10^{-5}\,\mathrm{s}$, and the neutrino lifetime is $t_c=2.9\times10^{-10}\,\mathrm{s}$. The number of baryons per electron is $\eta\cdot(2/3)\cdot\left(205\times22/(43\times8)\right)=5.3\times10^{-9}$. The number P_{as} of sterile neutrinos of mass m_h that need to be produced per electron is $5.3\times10^{-9}\cdot\left(m_p/m_h\right)\cdot\left(\Omega_{\mathrm{CDM}}/\Omega_b\right)=0.25$. (We use the standard notation in cosmology [2].) For $n_g=n'=1$, *i.e.* from (23), we obtain $P_{as}=0.02$ (for $m_a=0.05\,\mathrm{eV}$), insufficient to produce the observed density of dark matter (as reported in [16]).

Let us consider $n_g = n' = 3$. The example of Section 10 has the matrix R with a texture that makes the results insensitive to M_3. We may set $M_3 = 107$ eV with no significant change in the results reported in Section 10. For that fine-tuned example we obtain from (29), before subtracting reverse conversions,

$$P_{e6} = 1.2, \quad P_{\mu 6} = 30.4, \quad P_{\tau 6} = 24.1. \tag{54}$$

The sub-index "6" stands for ν_6 with mass $M_3 = m_h$. Reverse conversion limits these numbers to the statistical equilibrium value 1. In conclusion, sufficient sterile neutrino dark matter production is possible. Such dark matter is disfavored by observations but not ruled out.

12. Sterile Neutrino Search?

Consider a neutrino experiment that reconstructs the detected neutrinos in all-charged final states with the capability to discriminate a sterile neutrino mass from the active neutrino masses. In this case there is no interference, and the probability to detect the sterile neutrino ν_i, relative to the probability to detect any neutrino in the $\alpha\alpha$ channel is

$$\frac{P_s}{P_a} = \left| U_{\alpha i} U_{\alpha i}^* \right|^2, \tag{55}$$

with no sum implied. For $n_g = n' = 1$, *i.e.* from (20), we obtain $P_s/P_a = \sin^4\theta = \left(m_a/m_s\right)^2$, which is experimentally hopeless. For $n_g = 1$ and $n' = 2$, *i.e.* from (40), we obtain $P_s/P_a = \left|R_1\right|^4 \left(m_a/M_1\right)^2$, which is very interesting! For the example in Section 9 we obtain $P_s/P_a \approx 4\times 10^{17}\left(m_a/M_1\right)^2 \approx 9\times 10^{-5}$, which is less hopeless. For $n_g = n' = 3$, and the example in Section 10, we obtain $P_4/P_{\mu\mu} = 9\times 10^{-5}$, where "4" stands for ν_4 of mass M_1. This is the maximum for all channels, and is experimentally challenging. A search for sterile neutrinos in the approximate mass range 0.14 GeV to 2.0 GeV may be considered. A study has been presented in [17]. In conclusion, the same factor $K \approx 5\times 10^4$ that makes the model fine-tuned, enters to the fourth power, and may allow the model to be tested experimentally!

13. Conclusions

We have studied coherent active-sterile neutrino oscillations as a possible source of leptogenesis. To this end, we add n' gauge invariant Weyl_R neutrinos to the Standard Model with both Dirac m_D and Majorana M mass terms. We find that for $n' = 3$ we can obtain the measured active neutrino masses and mixings, and successful baryogenesis via leptogenesis, with, however, the fine tuning described in Sections 9 and 10, see (47) and (52). The Dirac CP-violating phase δ_{CP}, and Majorana CP-violating phases η_1 and η_2 of the 3 × 3 PMNS weak mixing matrix V_l contribute negligibly to this scenario of leptogenesis. The major contribution comes from the phases of the Dirac mass matrix m_D that links the Weyl_L and Weyl_R neutrinos. The Dirac nature of charged particles and the possible Majorana nature of neutrinos, emerge dynamically after elec-

troweak symmetry breaking, *i.e.* just before sphaleron freeze-out. The possible Majorana nature of neutrinos allows lepton number violation, with, however, a cross-section reduced by a factor $m_i m_j / \left(2E^2\right)$ due to polarization miss-match. This penalty renders lepton number violation beyond the reach of current laboratory experiments. CP-violation is the result of coherent interference of neutrinos with two clashing phases: a phase from the weak mixing matrix U (mainly from the Dirac mass matrix m_D), and a phase $2X_{ij}$ from neutrino propagation. The interference is coherent if the sterile neutrino mass m_s is less than approximately 6 GeV. The condition $2X_{ij} \lesssim \pi/2$, for strong CP violation, implies $m_s \lesssim 1.1\,\text{GeV}$. Constraints from Big Bang Nucleosynthesis require $m_s \gtrsim 0.14$. We find that at least one of the sterile neutrinos needs to have a mass in the approximate range 0.14 to 1.1 GeV to obtain successful leptogenesis.

With $n' = 3$, we may include in the model sterile neutrino dark matter with the measured mass $m_h = 107^{+37}_{-20}$ eV [10]. However, such dark matter is disfavored by observations (but not ruled out [10]).

The present scenario of leptogenesis requires a fine tuning parameter $K \approx 5 \times 10^4$, and a pattern of R such as (52). Why should nature select such a pattern (reminiscent of patterns in chemistry and biology)? It is interesting to note that with this fine tuning parameter K, the neutrino Yukawa coupling magnitudes become comparable to the ones of charged leptons and quarks. It is also interesting to note that K may bring sterile neutrino search within experimental reach.

The scenario studied in this article is similar to the model νMSM [8], where calculations have been carried out numerically in full detail. The search for sterile neutrinos with $m_s \lesssim 2\,\text{GeV}$ with sufficient sensitivity may be possible in a dedicated experiment [17].

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] DOnofrio, M., Rummukainen, K. and Tranberg, A. (2014) *Physical Review Letters*, **113**, Article ID: 141602. https://doi.org/10.1103/PhysRevLett.113.141602

[2] Zyla, P.A., *et al.* (2020) *Progress of Theoretical and Experimental Physics*, **2020**, 083C01. https://doi.org/10.1093/ptep/ptaa104

[3] Davidson, S., Enrico Nardi, E. and Nir, Y. (2008) *Physics Reports*, **466**, 105. https://doi.org/10.1016/j.physrep.2008.06.002

[4] Schwartz, M.D. (2014) Quantum Field Theory and the Standard Model. Cambridge University Press, Cambridge. https://doi.org/10.1017/9781139540940

[5] Branco, G.C., Lavoura, L. and Silva, J.P. (1999) CP Violation. Claredon Press, Oxford.

[6] Hoeneisen, B. (2006) Trying to Understand Mass. arxiv:0609080.

[7] Akhmedov, E. (2019) Quantum Mechanics Aspects and Subtleties of Neutrino Os-

cillations. *International Conference on History of the Neutrino*: 1930-2018, Paris, 5-7 September 2018. arxiv:1901.05232.

[8] Canetti, L., Drewes, M., Frossard, T. and Shaposhnikov, M. (2012) *Physical Review D*, **87**, Article ID: 093006. https://doi.org/10.1103/PhysRevD.87.093006

[9] Casas, J.A. and Ibarra, A. (2001) *Nuclear Physics B*, **618**, 171-204. https://doi.org/10.1016/S0550-3213(01)00475-8

[10] Hoeneisen, B. (2020) *International Journal of Astronomy and Astrophysics*, **10**, 203-223. https://doi.org/10.4236/ijaa.2020.103011

[11] Tremaine, S. and Gunn, J.E. (1979) *Physical Review Letters*, **42**, 407-410. https://doi.org/10.1103/PhysRevLett.42.407

[12] Hammer, F., Yang, Y.B., Arenou, F., Puech, M., Flores, H. and Babusiaux, C. (2019) *The Astrophysical Journal*, **883**, 171. https://doi.org/10.3847/1538-4357/ab36b6

[13] Hammer, F., Yang, Y., Arenou, F., Wang, J., Li, H., Bonifacio, P. and Babusiaux, C. (2020) *The Astrophysical Journal*, **892**, 3. https://doi.org/10.3847/1538-4357/ab77be

[14] Yang, Y., Hammer, F., Fouquet, S., Flores, H., Puech, M., Pawlowski, M.S. and Kroupa, P. (2014) *MNRAS*, **442**, Article ID: 24192433. https://doi.org/10.1093/mnras/stu931

[15] Hammer, F., Yang, Y.B., Arenou, F., Babusiaux, C., Puech, M. and Flores, H. (2018) *ApJ*, **860**, 76. https://doi.org/10.3847/1538-4357/aac3da

[16] Hoeneisen, B. (2021) *International Journal of Astronomy and Astrophysics*, **11**, 59-72. https://doi.org/10.4236/ijaa.2021.111004

[17] Gninenko, S.N., Gorbunov, D.S. and Shaposhnikov, M.E. (2012) *Advances in High Energy Physics*, **2012**, Article ID: 718259. https://doi.org/10.1155/2012/718259

Journal of Modern Physics, 2021, 12, 1267-1280
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

New Procedure to Obtain Specific and High Absorbent Silicon Nanotextures: Inverted Pyramids, Cubic Nano-Microholes, Spiroconical Nano-Microholes and Rhombohedral-Stared Nanosheet Bouquets (Nanobuckets)

Ndeye Coumba Y. Fall[1], Moussa Touré[1], Remi Ndioukane[1], Abdoul Kadri Diallo[1], Diouma Kobor[1], Marcel Pasquinelli[2]

[1]Laboratoire de Chimie et de Physique des Matériaux (LCPM), Université Assane Seck de Ziguinchor, Ziguinchor, Senegal
[2]Aix Marseille Université, Domaine Universitaire de Saint Jérôme, Marseille, France
Email: n.fall1530@zig.univ.sn

How to cite this paper: Fall, N.C.Y., Touré, M., Ndioukane, R., Diallo, A.K., Kobor, D. and Pasquinelli, M. (2021) New Procedure to Obtain Specific and High Absorbent Silicon Nanotextures: Inverted Pyramids, Cubic Nano-Microholes, Spiroconical Nano-Microholes and Rhombohedral-Stared Nanosheet Bouquets (Nanobuckets). *Journal of Modern Physics*, **12**, 1267-1280.
https://doi.org/10.4236/jmp.2021.129078

Received: January 22, 2021
Accepted: July 16, 2021
Published: July 19, 2021

Abstract

The present work relates to a process for silicon surface texturing for preparing large-area, silicon nanotextures on silicon substrates at ambient temperature by assisted chemical etching. A novel strategy comprises of two fundamental steps (metal-assisted chemical etching (MACE) and solution post-treatment) of using the silver catalyst to obtain specific nano- or micro-textures. The strategy is based on metal-induced (Ag) local oxidation and dissolution of a silicon substrate in three different concentrations of aqueous fluoride solution with the post-treatment solution. The etching technique is dependent on the etching time and concentration of aqueous fluoride solution. Therefore, detailed scanning electron microscopy observations reveal specifics shapes as inverted pyramids, cubic nano-microholes, spiroconical nano-microholes and rhombohedral-stared nanosheet bouquets (called Nanobukets), obtained for the first time on a (100) silicon surface by this new variant of the MACE method named Double Etching Method (DEM). Silicon nanostructures are used in many nanotechnology applications such as nano-microelectronics, optoelectronics or biomedical applications. UV-Visible spectrometry measurements carried out made it possible to obtain the lowest reflectance and highest absorbance values who are 3% and 97%, respectively for the rhomboedral-stared nanosheet bouquets on (100) crystalline silicon substrates in the UV-visible-NIR wavelength range from 300 to 1200 nm.

Keywords

Silicon Nanotextures, New Procedure, Etching, Reflectance, Absorption

1. Introduction

Black silicon light trapping structures, including nanowires [1] [2] and nanoholes [3] [4], have demonstrated their excellent anti-reflectance properties, providing the opportunities for enhancing light-harvesting. However, despite of the superior light absorption enhancement, the efficiency of textured silicon solar cells remains relatively low compared to those of conventional c-Si solar cells. Therefore, carefully designed new nanostructures such as inverted pyramids (IPs) [5] [6] and nanocones [7] [8] with low surface area enhancement by lithography method, have be done to balance optical gain. Of all these structures, IPs are mostly adopted for its excellent light absorption and the successful fabrication of IPs based solar cells with an efficiency 24.7% further confirms its superiority for high efficiency solar cells [9]. However, the lithography method used for fabricating such structures is costly and time-consuming, which hinders its further implement for mass production. Recently, a cost-effective and efficient approach called metal-assisted chemical etching (MACE) method has been developed to avoid the above disadvantages for industrial mass production. Today, the MACE technique has advanced to a level where desired architectures and surface quality can be readily fabricated; hence, it is becoming one of the main technologies in preparing silicon nanostructures. However, the number of nanostructures or nanotextures shapes is limited with this method and with silver as catalyst. Only nanowires, nanoholes and nanocubes are fabricated using this technique and this catalyst under laboratory applications. To obtain other shapes such as pyramids, inverted pyramids and other morphologies, the addition of post nanostructure rebuilding (post treatment) solution containing additives is needed. In 2014, Ye *et al.* reported an IPs (size of ~260 nm) based 156 * 156 mm^2 multi-crystalline Si solar cell with an efficiency of 18.45% (Jsc = 36.68 mA/cm^2) by MACE process and a post treatment with low concentration NaOH solution [10]. Until now, two step methods including MACE Black Silicon fabrication and a post structure reshaping process are believed to be a feasible way to connect nanostructure texture with high efficiency Si solar cells [11].

The present work proposes a new method for nanotexturing the surface of a (100) oriented p type silicon substrate in a specific manner with the post-processing solution. More precisely, the new method makes it possible to obtain novel silicon nanotextures shapes for the first time and using only silver as catalyst with any other additive or chemical solution and free external energy at ambient temperature. The new procedure is a variant of the MACE method, called Double Etching Method (DEM) as described in [12].

2. Experimental Method

To develop specific micro- or nanostructures an original procedure based on an over-etching of nanotextures allowing to obtain inverted pyramids, cubic nano-microholes, spiroconical nano-microholes and rhombohedral-stared nanosheets bouquets (nanobukets), was created. The process is named Double Etching Method (DEM).

A boron doped p-type <100> silicon wafer with a thickness between 600 - 650 μm, with a resistivity of 1 - 5 Ωcm and with a polished surface was purchased from BT Electronics. Before etching, the substrates are degreased and cleaned for 15 min in ultrasonic machine by using acetone, ethanol and de-ionized water solutions, respectively. This step is followed by cleaning with a piranha solution (H_2SO_4/H_2O_2) for 10 min to remove any impurities of organics and rinsed with de-ionized water. Substrates, once rinsed with de-ionized water, are immersed in the etching solution containing HF (4.8, 9.6 and 22.8 M)/$AgNO_3$ (0.02 M) mixture in a volume ratio of v: v = 6:19, depending on the desired nanotexture, for a preselected time (30, 60 and 90 min). These concentration ratios, as well as the etching time are varied. The electrochemical reactions involved in this etching step are as follows:

$$Ag^+ + e^- = Ag \quad E^\theta = 0.79\ V$$

$$Si + 2H_2O = SiO_2 + 4H^+ + 4e^- \quad E^\theta = 0.91\ V$$

$$SiO_2 + 6HF = SiF_6^{2-} + 2H_2O + 2H^+$$

The global reaction is summarized in the following equation:

$$Si + 6HF + 4Ag^+ = 4Ag + SiF_6^{2-} + 6H^+$$

The samples were placed directly, after rinsing with deionized water, in the HF (40%)/H_2O_2 solution with a volume ratio of v: v = 20:4 for 30 min. During this step, the nanostructures are created and a nanotextured surface composed mainly of nanowires and nanoholes remains. Their length and depth depend to the first etching time (MACE method). The electrochemical reactions involved are as follows:

$$H_2O_2 + 2H^+ \rightarrow 2H_2O + 2h^+ \quad E^\theta = 1.76\ V$$

$$Si + 6HF + nh^+ \rightarrow H_2SiF_6 + nH^+ + \left[n/2\right]H_2$$

The global reaction is summarized in the following equation:

$$Si + 6HF + n/2\,H_2O_2 \rightarrow H_2SiF_6 + nH_2O + \left[2 - n/2\right]H_2$$

The 3rd step is the innovative and fundamental step responsible for the shapes of the different fabricated nanotextures depending on the pre-nanostructures formed. In this step, the silicon sample thus nanostructured or textured with nanowires, nanoholes, nanopits and/or other nanotextures randomly produced are immersed in a post-treatment solution containing Ag ions for 15 min see Figure 1. Then, the Ag nanoparticles deposited are removed by using a concentrated solution of HNO_3. Surface morphology and chemical composition of the

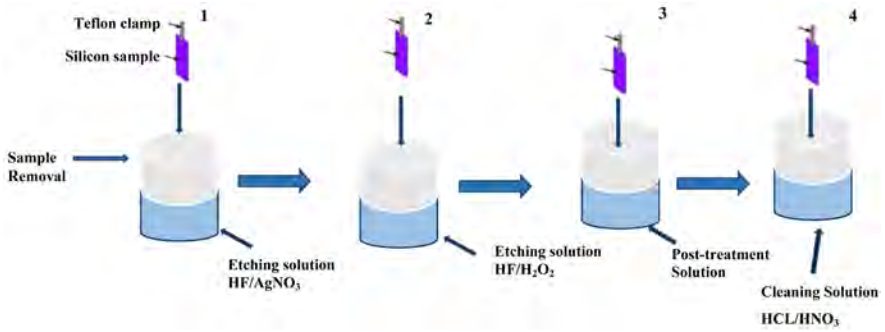


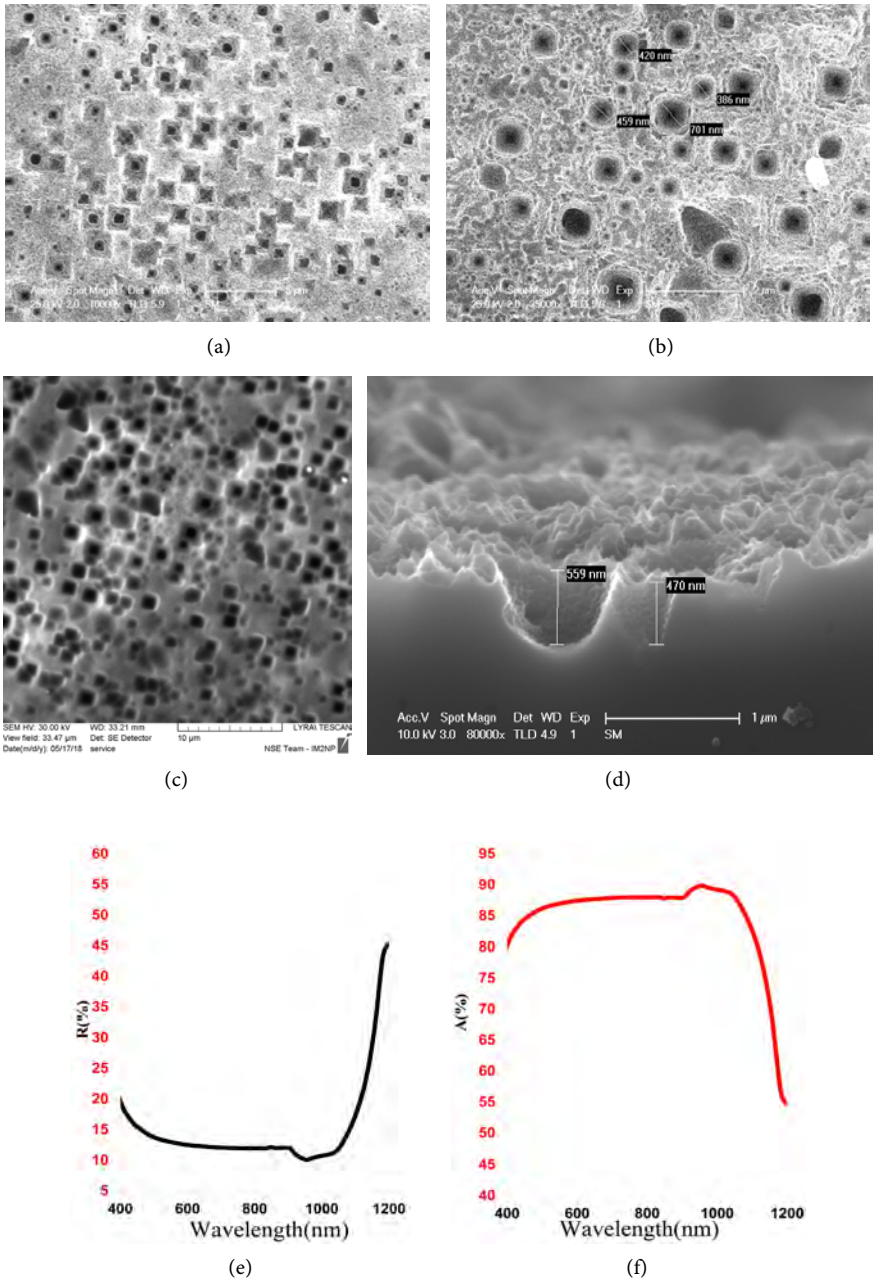
Figure 1. Different basic steps of the double etching method procedure.

Si samples are then characterized by using a MERLIN FEG of Zeiss scanning electron microscopy and energy dispersive X-ray spectrometry (EDX) model in ICMPE (Aix-Marseille-Université), respectively. The total reflectance at normal incidence was measured using a spectrophotometer with an integrating sphere model in LAMBDA 950S integrating sphere in *IM2NP* (Aix-Marseille-Université).

3. Results and Discussions

Figures 2(a)-(d) show SEM images of dispersed inverted pyramids with a 30 min etching time on a (100) oriented p type silicon substrate. The post-treatment with a solution, after the first etching step, made it possible to obtain sparse inverted pyramids (with the catalyst Ag) with sizes varying from 100 to 700 nm approximately (for the HF/$AgNO_3$ ratio of 4.8/0.02 for 30 min) making it possible to reduce the percentage of reflectance to 10% and thus increasing its absorbance to approximately 90% in the wavelength range between 300 and 1200 nm (Figure 2(e) and Figure 2(f)). The different obtained inverted pyramids seem to show that the immersion in the post treatment solution is a key step depending on the shape of the initial texture created during the two first steps. Thus for non-symmetrical nanotextures (straits, shallow holes, dots, pits etc.) obtained in those steps with an etching time up to 30 min and a low HF concentration of around 4.8 M, sparse inverted nanopyramids are formed on the silicon surface with a depth varying between 100 and 560 nm.

By increasing the HF concentration from 4.8 to 9.6 M (in HF/$AgNO_3$ ratio) for the same etching time (30 min), the whole silicon surface is completely covered with dense inverted pyramids (Figures 3(a)-(d)). Their sizes varied between 100 nm and more than 2000 nm confirming etchant concentration effects on the holes depth. The change in inverted pyramids density compared to the first ones could be explained by the increase of etching sites on the surface by increasing the HF/$AgNO_3$ ratio. Optical measurements showed reflection values of less than 10% corresponding to more than 90% absorption in the wavelength range between 300 and 1200 nm (Figure 3(e) and Figure 3(f)). These values are in agreement with the results from literature for inverted pyramids [13] [14]. The cross-section SEM images of the inverted pyramids structures shown in Figure 2(d) and Figure 3(d) reveal that the angle between the structured



(a) (b) (c) (d) (e) (f)

Figure 2. SEM images of sparse inverted pyramids for 30 min etching time in 4.8 M HF concentration (a) front view 10 kX magnetisation, (b) inverted pyramids sizes view, (c) 33.47 μm field view and (d) transverse view and UV-Visible-NIR (e) reflection and (f) absorption for <100> oriented p type Silicon.

surfaces and the (100) Si surface is around 55˚, indicating the Si (111) plane termination of the facets as indicated by [14]. The upper sides of the inverted pyramids show in some cases angle values of more than 120˚ (polygons, **Figure 3(c)**) indicating the presence of the over etching in this new procedure.

Figure 4 shows SEM images and UV-visible spectra with an HF molar concentration of 4.8 M (in HF/$AgNO_3$ ratio) for 90 min etching time. Sparse spiro-conical nanoholes with bottom and top diameters between 100 and 400 nm were

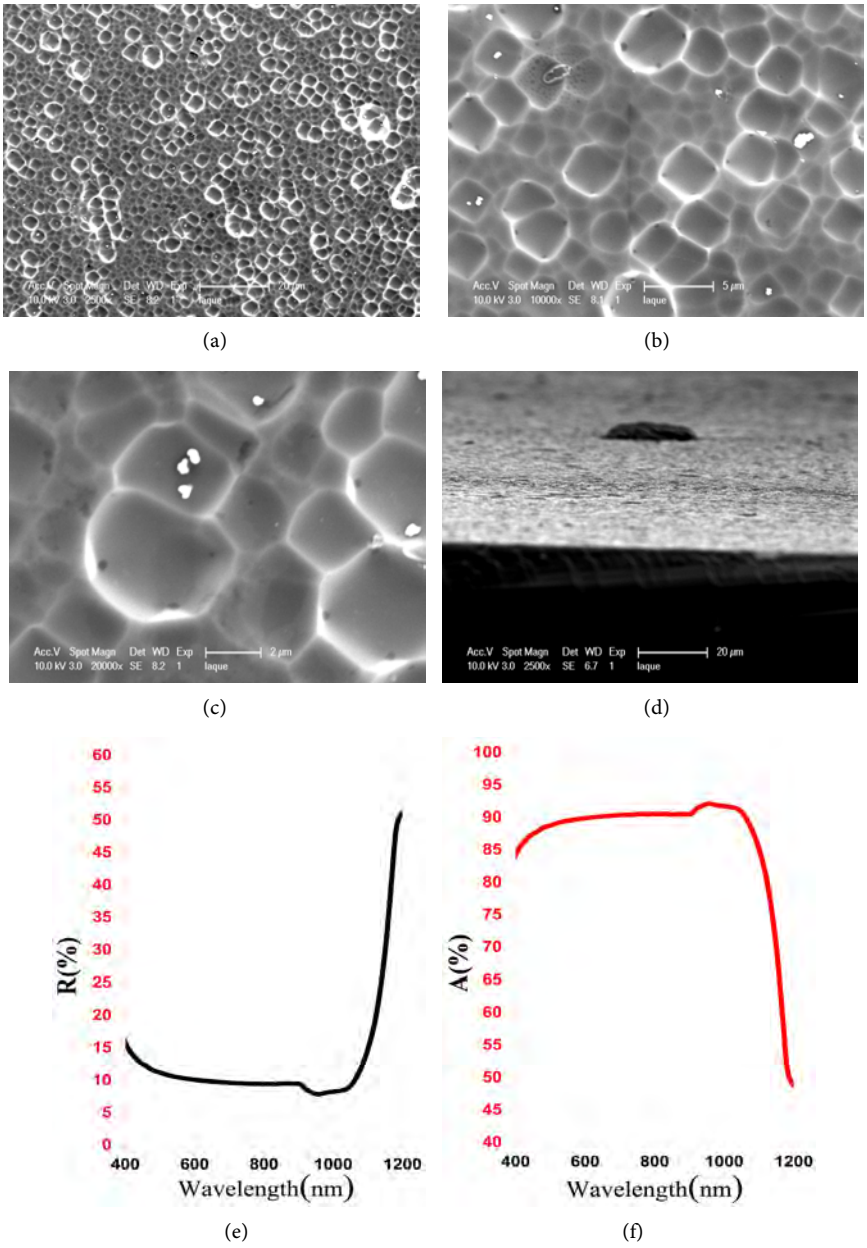


Figure 3. SEM images of p type Silicon with very dense and homogenously distributed inverted pyramids for 30 min etching time and 9.6 M HF (a) front view 2.5 kX magnetisation, (b) 10 kX magnetisation, (c) 20 kX magnetisation, (d) transverse view, UV-Visible (e) % reflexion and (f) % absorption.

obtained. The ratio between top and bottom diameter is around 4. For the second time such specific nanotexture shapes were fabricated in the laboratory of the corresponding authors (the first fabrication has been reported in [12]) (**Figures 4(a)-(d)**). To obtain such shapes using this new procedure, it needs to be considered that two combined phenomena will take place. One is a physical phenomenon related to fluid mechanics. It is a kinetic mechanism of the movement of a liquid with a high flow rate penetrating a nanohole (vortex movement). The second one is an electrochemical reaction, which takes place in the

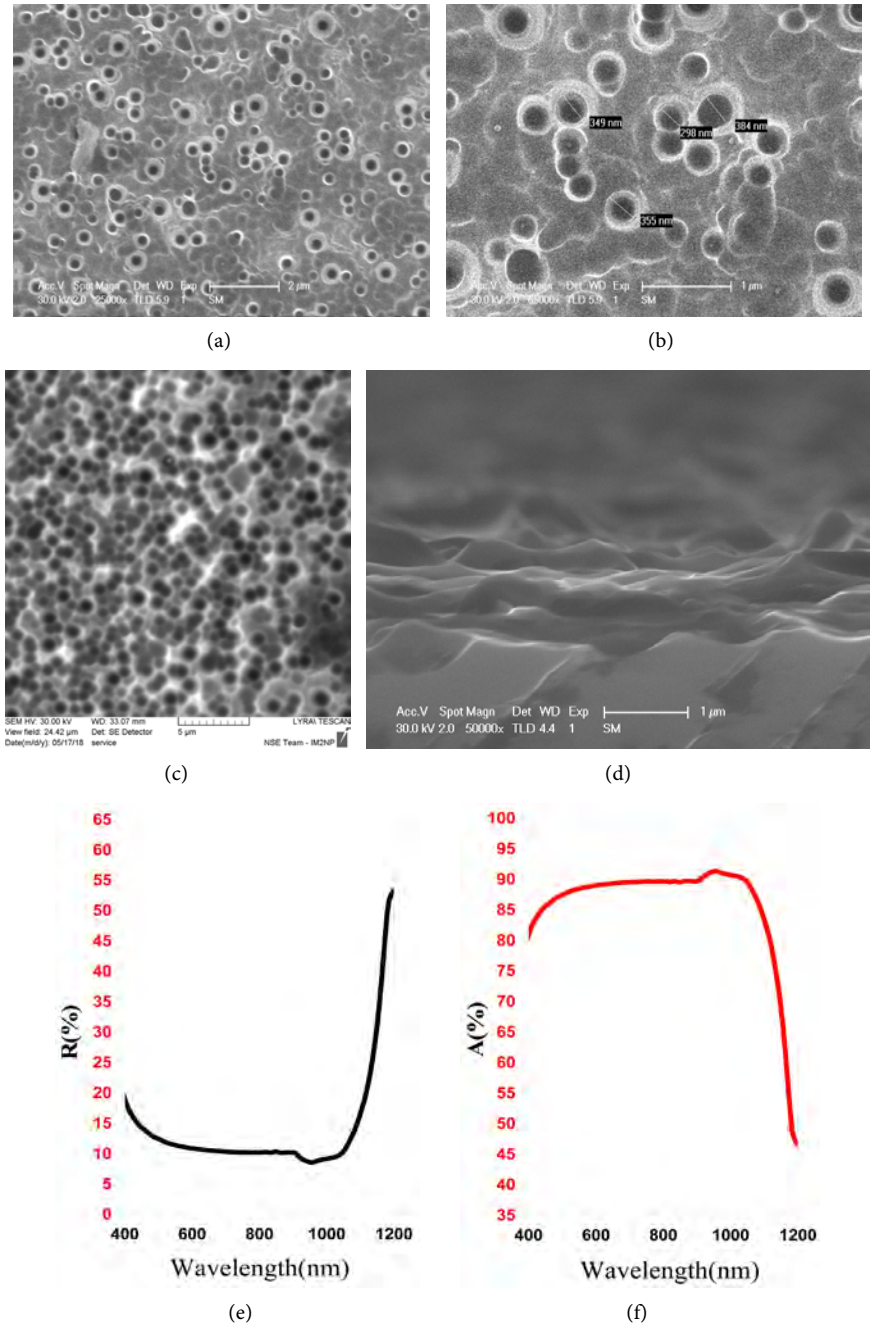


Figure 4. SEM images of p type Silicon spiroconical nanoholes for 90 min etching time and 4.8 M HF (a) front view 25 kX magnetisation, (b) 65 kX magnetisation for nanoholes sizes view, (c) 24.42 µm field view of, (d) transverse view, UV-Visible (e) % reflexion and (f) % absorption.

etching process. The Ag^+ ions will attack the silicon to oxidize it and dissolve it in the solution as detailed in the etching technique due to the presence of the HF/H_2O_2 solution residues. The use of this solution in the procedure seems fundamental because the presence of H_2O_2 as oxidant at a percentage higher than 30 would increase the horizontal etching speed to the detriment of the vertical etching by annihilating the gravity of the Ag^+ nanoparticles thus preventing it

from descending vertically. The etching phenomenon, linked to the post-treatment solution in a movement mechanism (vortex), result in spiroconical nanoholes (**Figures 4(a)-(d)**). The result can be explained by the relatively long etching time (90 min), permitting to etch more deeply the nanoholes formed in the two first steps of the procedure, compared to the 30 min etching time. This hypothesis is in agreement with the formation of the sparse inverted pyramids, which is explained above, consisting to the formation of straits, pits and pitches (or nanoholes) in the first steps (MACE) and their changing, in the last steps (post treatment solution), to inverted pyramids (**Figure 2**) (or spiroconical nanoholes thanks to nanofluid mechanics (**Figure 4**)). The specific shape of these nanotextures could be used in many applications such as nano- or microfluidic mechanisms in biomedical applications, nanofiltering and nano- or micro-mixing solutions before chemical reactions. These nanostructures gave a reflectance less than 10% and an absorbance more than 90% in the UV-Visible-NIR-range (**Figure 4(e)** and **Figure 4(f)**). FTIR measurement for spiroconical nanostructures gave a transmittance around 1% while the absorption is around 99% between 400 and 4000 cm^{-1} (results not shown here), indicating their interest for IR detectors [15]. On the other hand, the angle between the (100) face of the substrate and the nanocone walls is approximately equal to 54.7˚ like the one observed in inverted pyramids on (100) Silicon. This indicates that the etching is done along the same axes in addition to the vortex phenomenon.

Increasing an HF molar concentration between 4.8 and 9.6 (in $HF/AgNO_3$ ratio) makes it possible to obtain a mixture of spiroconical microholes and inverted nanopyramids in interlayers and this for a shorter time of at least 30 min (**Figures 5(a)-(d)**) with reflectance and absorbance values in the UV-visible-NIR of approximately 9% and more than 91%, respectively (**Figure 5(e)** and **Figure 5(f)**). The presence of residual inverted nanopyramids is due to the short etching time (as in **Figure 2**), although the increase in the HF molar concentration has made it possible to obtain a good part of spiroconical nano- and microholes, which are created by the presence of enough deep nanoholes for the vortex movement mechanism. The internal morphology of a selected microhole (**Figure 6(d)**), its uniformity and spiroconical shape, as well as the presence of silver nanoparticles, after etching, confirm the mechanism detailed above. The fabrication of such mixed shapes is in agreement with the hypothesis stated by the authors concerning the role of the pre-etching with help of the MACE method. The surface absorption is comparable to inverted pyramids and spiroconical nanoholes taken individually.

Figure 6 represent SEM images and UV-Visible spectra of cubic micro-holes distributed over the entire surface of (100) oriented p type silicon substrate. Increasing etchant concentration in ($HF/AgNO_3$ molar ratio from 9.6/0.02 to 22.8/0.02) and the etching time (60 min) makes it possible to obtain cubic microholes which sides varies between 500 and 1000 nm approximately as well as the presence of some residual inverted pyramids (**Figures 6(a)-(d)**). A minimum reflectance and maximum absorbance of around 9% and 91% respectively

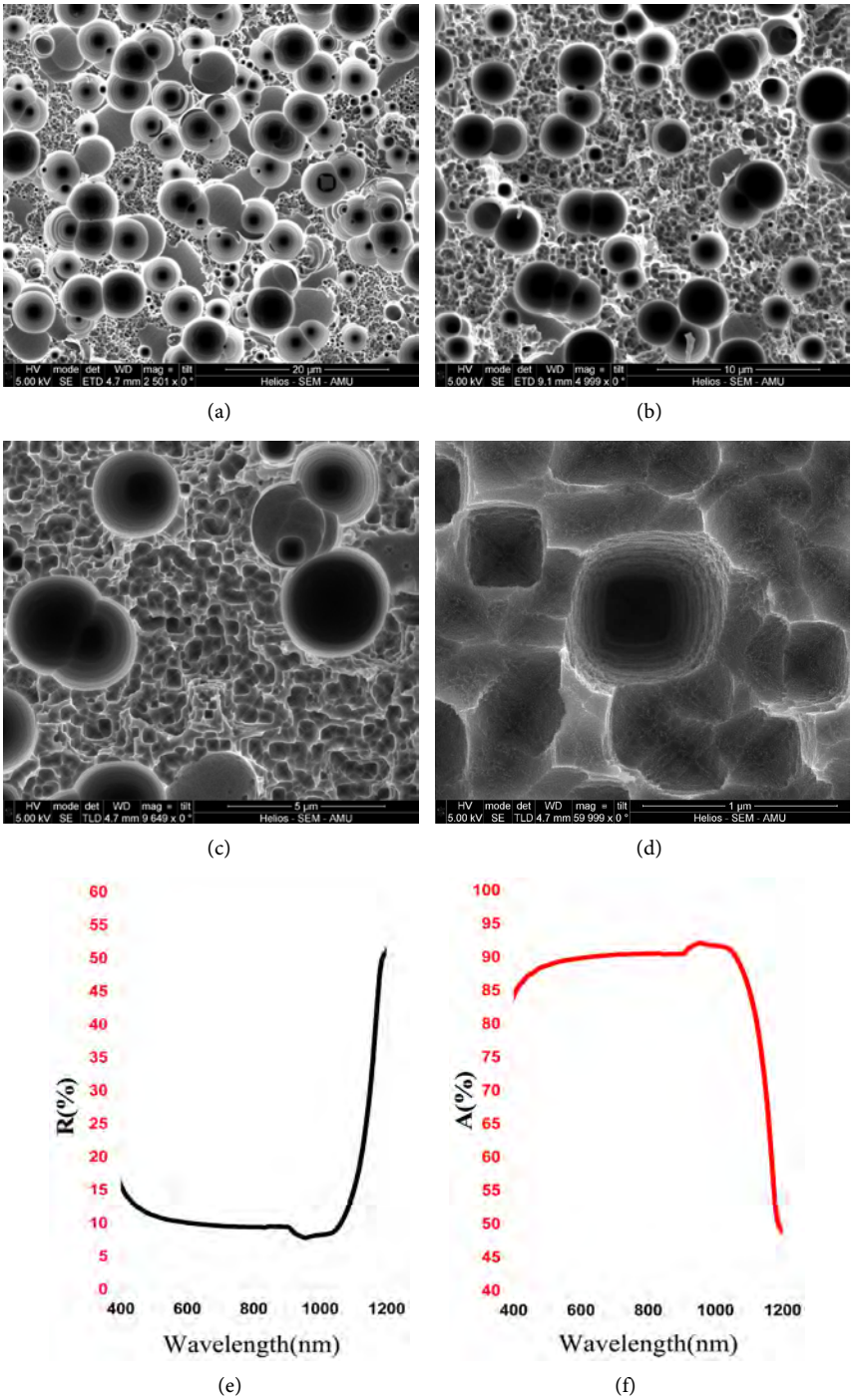


Figure 5. SEM images of p type Silicon mixte spiroconical and inverted pyramids nano/microholes for 30 min etching time and 9.6 M HF (a) front view 2.5 kX magnetisation, (b) 5 kX magnetisation, (c) 9.6 kX magnetisation, (d) 60 kX magnetization and zooming on spiroconical microholes, UV-Visible (e) % reflection and (f) % absorption.

(**Figure 6(e)** and **Figure 6(f)**) were obtained. The increase of HF concentration and the etching time permitted to transform the inverted pyramids (**Figures 3(a)-(d)**) into cubic microholes. The presence of few inverted pyramids confirms this hypothesis.

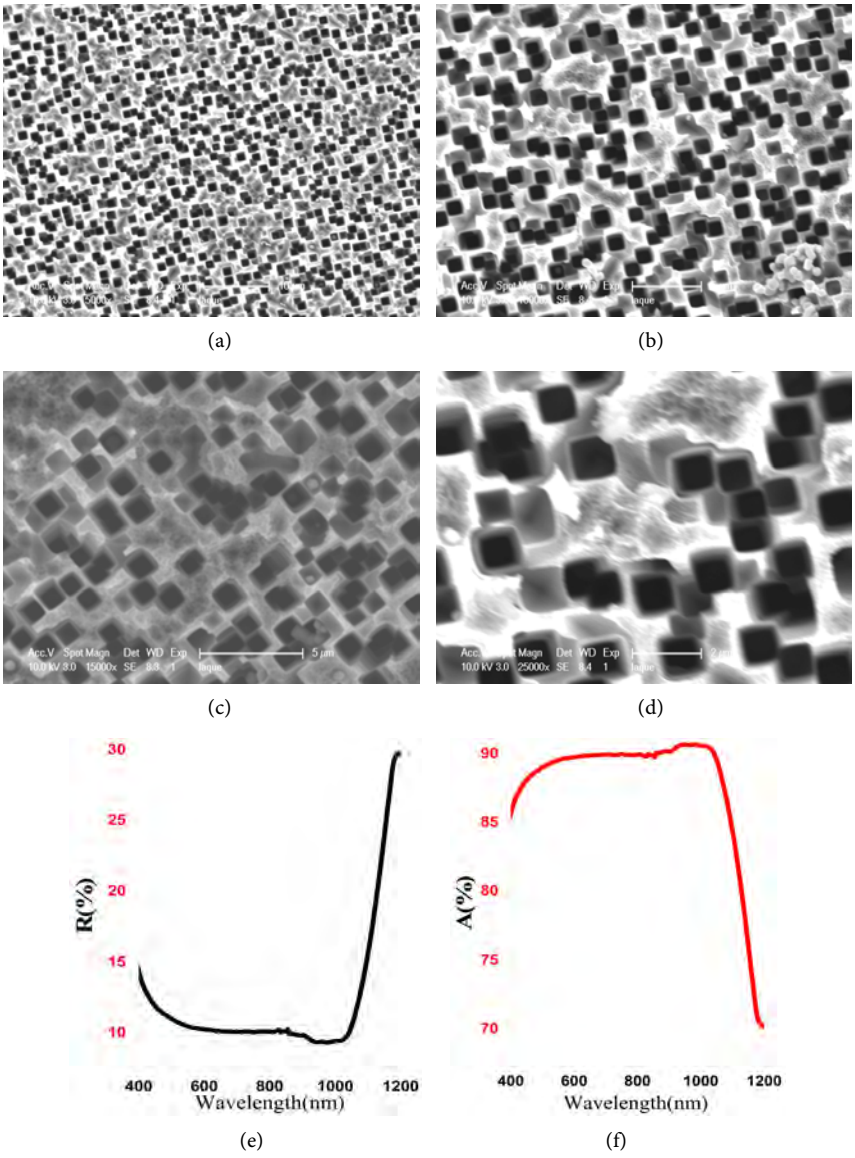
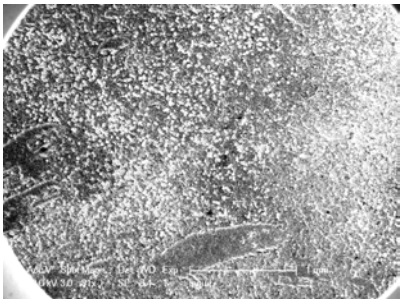
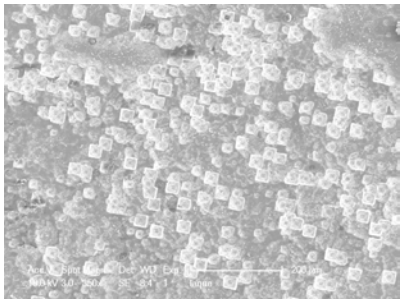


Figure 6. SEM images of <100> p type silicon cubic nano/microholes for 60 min etching time and 22.8 M HF (a) front view 5 kX magnetisation, (b) 10 kX magnetisation, (c) 15 kX magnetisation, (d) 25 kX magnetisation, UV-Visible (e) % reflection and (f) % absorption.

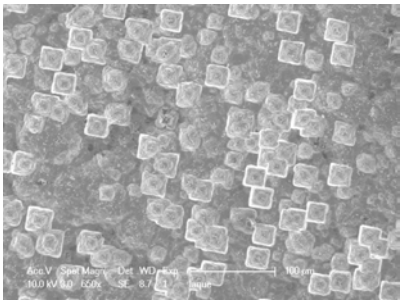
Finally, exceptionally and for the first time, increasing the etching time from 60 to 90 min and with a constant HF concentration (22.8 M), it is possible to obtain specific nanostructures in the form of Rhombohedral Stared Nanosheets Bouquets (Nanoflowers Bouquets named Nanobukets) (**Figures 7(a)-(h)**). To our knowledge, these rhombohedral superposed silicon nanosheets (**Figure 7(g)**) are a discovery. They have been fabricated on silicon substrates with a very low reflectance around 3% and a large absorbance of more than 97% in the UV - Visible - NIR wavelength (**Figure 7(j)** and **Figure 7(k)**). These results make them one of the best anti-reflective layers for photovoltaic applications. The shape of these nanobukets as well as the morphology of the silicon surface makes



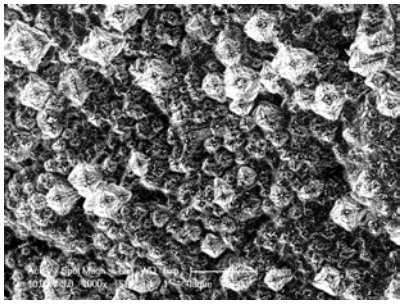
(a)



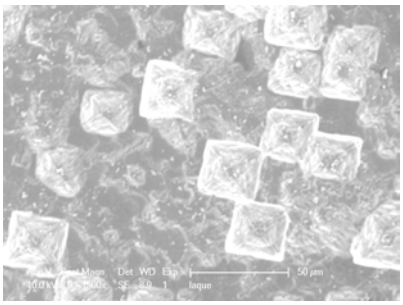
(b)



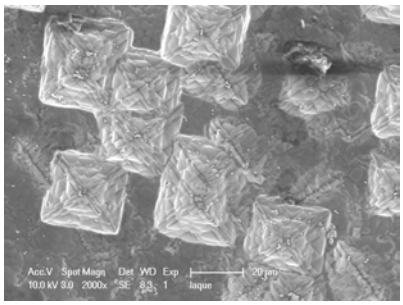
(c)



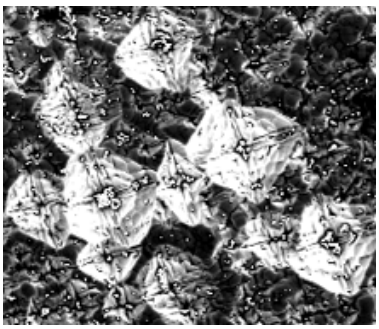
(d)



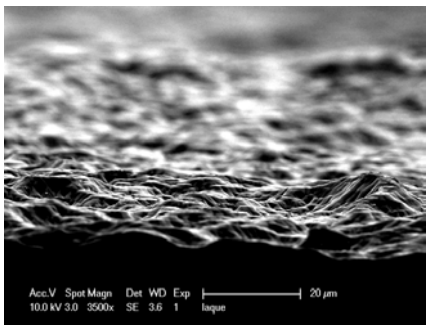
(e)



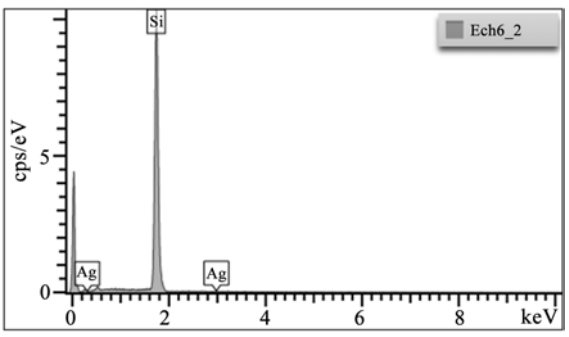
(f)



(g)

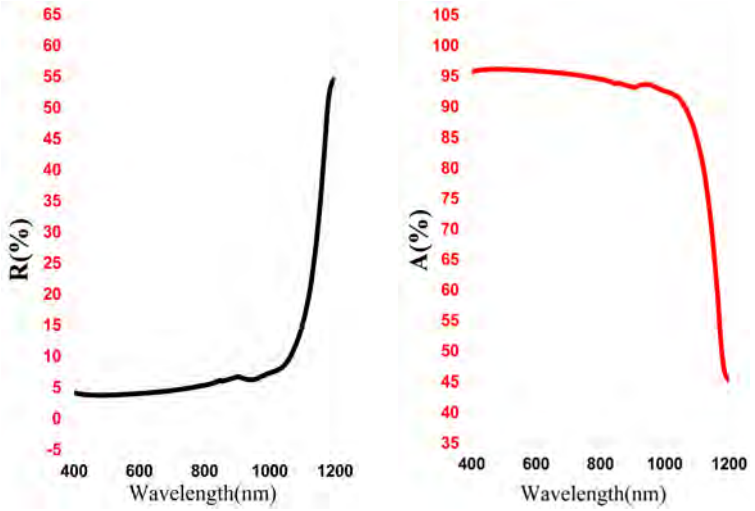

Acc.V Spot Magn Det WD Exp
10.0 kV 3.0 3500x SE 3.6 1 laque
20 µm

(h)


Ech6_2
Si
Ag
Ag
cps/eV
5
0
0
2
4
6
8
keV

Element	Raie Type	Apparent concentration	k ratio	% Mass	% Mass Sigma	Standard description
Si	Series K	9.00	0.07128	100.00	0.00	SiO_2
Ag	Series L	0.00	0.00000	0.00	0.00	Ag
Total:				100.00		

(i)



(j) (k)

Figure 7. SEM images of <100> p type silicon rhombohedral-stared nanosheets bouquets for 90 min etching time and 22.8 M HF (a) front view for 0.091 kX magnetisation, (b) 0.350 kX magnetisation, (c) 0.650 kX magnetisation, (d) 1 kX magnetisation, (e) 1.5 kX magnetisation, (f) 2 kX magnetisation, (g) zooming on bouquets for nanosheets view for 1 kX magnetisation, (h) transverse view of the bouquets, (i) EDX spectrum and UV-Visible (j) % reflection, (k) % absorption.

them potential materials for biomedical application. These specific nanosheets can be explained by an over-etching (by increasing the time from 60 to 90 min) of the microholes in preferred <111> directions thanks to their cubic form thus creating rhombohedral symmetry with axes on the main diagonals of the cube. The etching, done on the (100) faces of the microcubes in an oblique manner, makes it possible to obtain very thin sheets superposed along the C4 axis. The transversal mode shows this thin nanostructure with a quasi-fibrous surface (**Figure 7(h)**) explaining the excellent optical properties. The existence of such nanofibers (**Figure 7(d)** and **Figure 7(g)**) could indicate the growth of another component such as silver nanowires but the EDX analysis (**Figure 7(i)**) permitted to reject this hypothesis showing 100% silicon composition. These results indicate the possibility to obtain silicon nanosheets using such new innovative procedure applicable in industrial level.

4. Conclusion

New and specific silicon nanostructures shapes have been successfully fabricated using a new procedure of metal assisted chemical etching (MACE), using only

Ag as catalyst and post-treatment solution. This variant of the MACE method named Double Etching Method (DEM) permitted to fabricate with the same catalyst:

- Sparse and dense inverted pyramids, which is the second time to have such shapes with Ag catalyst.
- Cubic microholes.
- Spiroconical nano-microholes obtained thanks to the presence of a combination of nanofluidic and electrochemical mechanisms.
- Mixed inverted pyramids and spiroconical nano-/microholes, for the first time.
- And for the first time, exceptional rhombohedral-stared nanosheets bouquets (named here nanobukets), presenting one of the best optical properties (3% reflection) for silicon nanostructures.

The innovation from this procedure is the possibility to fabricate silicon nanosheets in form of bouquets that opens large possibilities of their use in biomedical and photovoltaic applications. The method is comparable to the nanostructure rebuilding procedure, but no additives are needed in the procedure presented in this work making it very simple and easy.

Acknowledgements

This work was financially supported by the French Cooperation Exchange. Thanks to Prof. Dr. Ulf Blieske of Cologne Institute for Renewable Energy for his help in re-reading the article.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Chen, J.M., Chen, C.Y., Wong, C.P., *et al.* (2017) Inherent Formation of Porous P-Type Si Nanowires Using Palladium-Assisted Chemical Etching. *Applied Surface Science*, **392**, 498-502. https://doi.org/10.1016/j.apsusc.2016.09.048

[2] Hung, Y.J. and Lee, S.L. (2014) Manipulating the Antireflective Properties of Vertically Aligned Silicon Nanowires. *Solar Energy Materials & Solar Cells*, **130**, 573-581. https://doi.org/10.1016/j.solmat.2014.08.004

[3] Cheng, S.L., Lin, Y.H., Lee, S.W., *et al.* (2012) Fabrication of Size-Tunable, Periodic Si Nanohole Arrays by Plasma Modified Nanosphere Lithography and Anisotropic Wet Etching. *Applied Surface Science*, **263**, 430-435. https://doi.org/10.1016/j.apsusc.2012.09.073

[4] Subramani, T., Hsueh, C.C., Syu, H.J., *et al.* (2016) Interface Modification for Efficiency Enhancement in Silicon Nanohole Hybrid Solar Cells. *RSC Advances*, **6**, 12374-12381. https://doi.org/10.1039/C5RA23109D

[5] Mavrokefalos, A., Han, S.E., Yerci, S., *et al.* (2012) Efficient Light Trapping in Inverted Nanopyramid Thin Crystalline Silicon Membranes for Solar Cell Applications. *Nano Letters*, **2**, 2792-2796. https://doi.org/10.1021/nl2045777

[6] Fan, Y., Han, P., Liang, P., *et al.* (2013) Differences in Etching Characteristics of TMAH and KOH on Preparing Inverted Pyramids for Silicon Solar Cells. *Applied Surface Science*, **264**, 761-766. https://doi.org/10.1016/j.apsusc.2012.10.117

[7] Jeong, S., Garnett, E.C., Wang, S., *et al.* (2012) Hybrid Silicon Nanocone-Polymer Solar Cells. *Nano Letters*, **12**, 2971. https://doi.org/10.1021/nl300713x

[8] Wang, K.X., Yu, Z., Liu, V., *et al.* (2012) Absorption Enhancement in Ultrathin Solar Cells with Antireflection and Light-Trapping Nanocone Gratings. *Nano Letters*, **12**, 1616. https://doi.org/10.1021/nl204550q

[9] Zhao, J., Wang, A. and Green, M.A. (1999) 24.5% Efficiency Silicon PERT Cells on MCZ Substrates and 24.7% Efficiency PERL Cells on FZ Substrates. *Progress in Photovoltaics*, **7**, 471-474. https://doi.org/10.1002/(SICI)1099-159X(199911/12)7:6%3C471::AID-PIP298%3E3.0.CO;2-7

[10] Ye, X., Zou, S., Chen, K., *et al.* (2015) 18.45% Efficient Multi Crystalline Silicon Solar Cells with Novel Nanoscale Pseudo Pyramid Texture. *Advanced Functional Materials*, **24**, 6708-6716. https://doi.org/10.1002/adfm.201401589

[11] Oh, J., Yuan, H.C. and Branz, H.M. (2012) An 18.2%-Efficient Black-Silicon Solar Cell Achieved through Control of Carrier Recombination in Nanostructures. *Nature Nanotechnology*, **7**, 743-748. https://doi.org/10.1038/nnano.2012.166

[12] Kobor, D. (2016) Procédé de Réalisation de Nano ou microtrous Spiro-coniques sur Substrat Silicium. OAPI Patent No. E03F5/00(06.01), Deposit No. 1201500089.

[13] Touré, M., Kobor, D., Ndiaye, L.G., Ndiaye, A. and Tine, M. (2016) Influence of Pyramids and Inverted Pyramids on Silicon Optical Properties. 2016 *International Energy and Sustainability Conference* (*IESC*), 1-6. https://doi.org/10.1109/IESC.2016.7569486

[14] Tang, Q., *et al.* (2018) Formation Mechanism of Inverted Pyramid from Sub-Micro to Micro Scale on c-Si Surface by Metal Assisted Chemical Etching Temperature. *Applied Surface Science*, **455**, 283-294. https://doi.org/10.1016/j.apsusc.2018.05.023

[15] Touré, M. (2017) Elaboration et caractérisation des nanostructures de silicium et hetero-structure 3C-SiC/Si pour cellules photovoltaïques. PhD Thesis, Université Assane Seck de Ziguinchor, Ziguinchor.

Journal of Modern Physics, 2021, 12, 1281-1294
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Gravitational Energy Levels: Part Two

Edward Tannous

Independent Researcher, Jaffa, Israel
Email: Etannouss@gmail.com

How to cite this paper: Tannous, E. (2021) Gravitational Energy Levels: Part Two. *Journal of Modern Physics*, **12**, 1281-1294.
https://doi.org/10.4236/jmp.2021.129079

Received: May 15, 2021
Accepted: July 17, 2021
Published: July 20, 2021

Abstract

We present here a model that explains in a simple, easy and summarized manner, the values, meaning and reasons for the force of gravity, using simple physical tools. According to this model, a gravitational field actually creates different energy levels, similar to the atom, around the center of mass of the gravitational source, and a transition between the energy levels results in the creation of the force of weight acting on each small body which is in the gravitational field. As the body approaches a gravitational field, its energy value decreases to a value of $m_0 u_{(R)}^2$, proportional to the distance R between the centers of the masses, when $u_{(R)}$ is the magnitude of the self-speed of light vector (the progression in the time axis) of the small body, and its value decreases as it approaches the center of the origin of the field. This change in the energy levels is the cause of the force of gravity. A formula is obtained for the concept of potential gravitational energy and the variables on which it depends, and for the time differences between two frames that are in the gravitational field, taking into account the motion and location of each frame. It is obtained from this model that the speed of light is also a variable value as a result of the effect of the gravitational field.

Keywords

Force of Gravity, Potential Energy, Kinetic Energy, Time Differences between Frames, Gravitational Curvature of Light Beams

1. Introduction

This article is a continuation of previous articles called "Negative Mass" ref. [1] and "Energetic Angle" ref. [2], which present a model of a body, in accordance with the theory of special relativity, which is in constant motion in space (at high velocities close to the speed of light), so that the body has a velocity equal to the speed of light in the space-time, has an energetic angle and a negative mass. This model shows that each body moves at the speed of light in the space-time, and in

a different direction, which is called the "self-speed of light vector", depending on the velocity of each body relative to another body.

The present research paper also refers to Einstein's general theory of relativity ref. [3] and gives diverse answers using the basic laws of physics, which are the cornerstones of this science, especially the Energy Conservation Law ref. [4] and Newton's Laws ref. [5], to various issues in physics, such as black hole ref. [6].

Albert Einstein's formula describing the rest energy of a body with mass (m) is given by the formula $E = mc^2$ ref. [7]. As mentioned in the previous articles ref. [1] [2], the energy is divided into two parts:

A. Self-time energy E_{st}, which determines the amount of energy left in the mass as a result of the velocity.

B. State kinetic energy E_{α}, which is the energy that the body carries within it in the reference frame, as a result of the velocity.

State kinetic energy also includes other forms of energy and not just kinetic energy, such as potential energy.

Therefore, the total energy of the body is: $E = E_{st} + E_{\alpha}$. In case the body is free, *i.e.* outside the gravitational field: $E = E_0 = E_{st} + E_{\alpha} = m_0 c^2$. In this article, we show the total energy of the body in a gravitational field.

2. A Body in a Gravitational Field

Initially, we refer to a small body which approaches a large body under the influence of a gravitational field. The body loses some of its self-time energy E_{st}, because it is exerted by a gravitational force $F_{(R)}$, which is a conservation force. The magnitude of the energy it loses is equal to the amount of work invested: $\int_R^{\infty} F(R)\mathrm{d}R$, where R is the distance between the center of the small body m and the center of the large body M, as shown in **Figure 1(a)**.

Therefore, the self-time energy (mass energy) E_{st} of the small body will decrease as it gets closer to the large body. This decrease in the value of the energy is not at the expense of the magnitude of the mass (which, naturally, does not change), but at the expense of the self-speed of light vector which we marked in

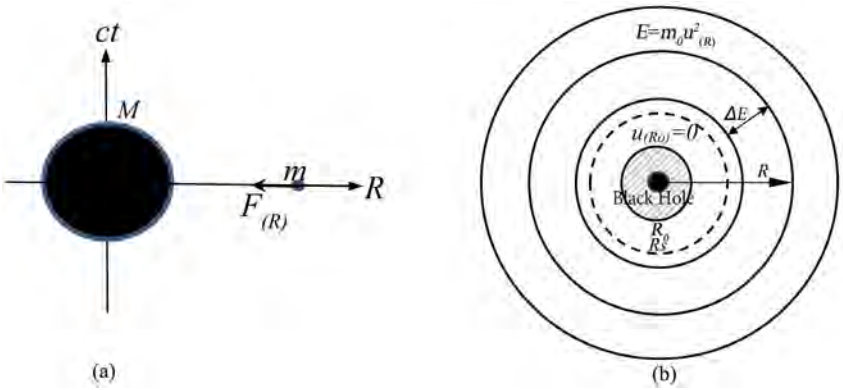


Figure 1. (a): Small body m under the influence of a gravitational field of a large body M; (b): Description of the energy levels of a black hole, while indicating the zero horizon R_0, and the event horizon R_s.

the previous article in $\boldsymbol{C}$ ref. [1]. Because of the effect of the gravitational field, we mark the self-speed of light vector by $\boldsymbol{U}_{(R)}$, since its magnitude is not equal to the speed of light, but is smaller as it gets closer to the center of mass of the gravitational field source, therefore it depends on the distance R, *i.e.*, $\left|\boldsymbol{U}_{(R)}\right| = u_{(R)} < c$.

We calculate the energies of the body under the influence of a gravitational field at a distance R, where the body m_0 is at rest relative to M, *i.e.*, $\alpha = 0$ as shown in **Figure 2(a)**.

Self-time energy (mass energy):

$$E_{st} = m_0 \boldsymbol{C}\boldsymbol{U}_{(R)} = m_0 c u_{(R)} \tag{1}$$

wherein:

$\boldsymbol{C}$—The self-speed of light vector of the large body.

$u_{(R)}$—The absolute value of the self-speed of light vector of the small body $\boldsymbol{U}_{(R)}$, *i.e.* $u_{(R)} = \left|\boldsymbol{U}_{(R)}\right|$.

The state kinetic energy, which is the potential energy only in this case, since there is no motion in the R direction:

$$E_\alpha = m_0 \boldsymbol{U}_{(R)} \underbrace{\left(\boldsymbol{U}_{(R)} - \boldsymbol{C}\right)}_{v_{ax}} = m_0 u_{(R)}^2 - m_0 c u_{(R)} \tag{2}$$

Therefore, the total energy of the small body under the influence of a gravitational field is:

$$E = E_{st} + E_\alpha = m_0 u_{(R)}^2 \tag{3}$$

Initially, if a small body m_0 is at rest at a distance of R from a large body M, the amount of work required, by a conservation and variable force accordingly, to move the body to a completely free state, *i.e.* out of gravitational influence, is calculated using Newton's law:

$$W_{(R\to\infty)} = \int_R^\infty F_{(R)} \mathrm{d}R = \int_R^\infty m_0 \frac{GM}{R^2} \mathrm{d}R = m_0 \frac{GM}{R} \tag{4}$$

wherein G is Newton's gravitational constant.

According to the law of conservation of energy, the amount of work we calculated in Formula (4) is the same amount of energy that the body loses,

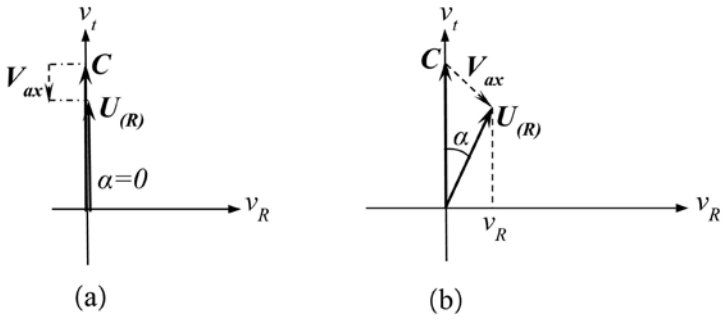


Figure 2. (a) An array of self-speed of light vector of the large body C and the small body $U_{(R)}$, at rest. (b) An array of self-speed of light vector of the large body C and the small body $U_{(R)}$, moving at velocity v_R.

approaching a gravitational field to a distance of R, *i.e.*, $W_{(R\to\infty)} = W_{(\infty\to R)}$. This amount of energy is the amount that the body loses from the self-time energy (mass energy) E_{st}.

Using Formula (1), we obtain that the difference between the self-time energy (mass energy) E_{st} in a completely free state (outside the gravitational field) $E_{st} = m_0c^2$ and the energy under gravitational influence $E_{st} = m_0cu_{(R)}$ is the amount of work we calculated in Formula (4):

$$\Delta E_{st} = m_0c^2 - m_0cu_{(R)} = W_{(R\to\infty)} = m_0\frac{GM}{R} \tag{5}$$

Thus we obtain the size of the self-speed of light vector of each small body, which is at a distance of R from a large body (star) with a gravitational field:

$$u_{(R)} = c - \frac{GM}{cR} \tag{6}$$

$u_{(R)}$ is the magnitude of the self-speed of light vector of the small body in space-time under the influence of a gravitational field, and is always smaller than the speed of light c in its absolute value. Therefore, and from Formula (6), a number of conclusions can be drawn:

1) **Energy**: The self-speed of light vector of each body at a distance R will be $\boldsymbol{U}_{(R)}$, therefore in any energetic state it will have total energy $E = m_0u_{(R)}^2$. That is, the sum of the self-time energy (mass energy) E_{st} and the state kinetic energy E_a will always be $E = E_{st} + E_\alpha = m_0u_{(R)}^2$. As the body gets closer to the large body, its energy will decrease accordingly.

2) **Horizon Zero** is the state where:

$$R_0 = \frac{GM}{c^2} \tag{7}$$

According to Formula (6), it seems that the self-speed of light vector of the body in this state will be equal to zero, *i.e.*, it cannot move in any direction at any energetic angle. Furthermore, its energies become zero $E = E_{st} = E_\alpha = 0$, on this horizon it will be in a state of freezing in time, *i.e.*, its time is not advancing.

If a zero horizon Ro exists outside the mass of the large body (the star), it is a black hole. This can be seen in **Figure 1(b)**.

3) **Time**: Each body at rest, at a distance R, will have a slower time than a free body at a difference of $\Delta T = \frac{cT - u_{(R)}T}{c} = \frac{GM}{c^2R}T = \frac{R_0}{R}T$

4) **Mass**: Although the mass m_o does not change in shape and composition, its measured value m (effective mass) is small under the influence of the gravitational field. Since the self-time energy (mass energy) is smaller $E_{st} = m_0\boldsymbol{CU}(R) = m_0cu_{(R)}\cos\alpha = mc^2$ (α is the energetic angle, assuming that the body moves in some direction in the Euclidean space (x,y,z), so that its velocity is $v = u_{(R)}\sin\alpha$ ref. [1] [2]), we obtain an effective mass in this case:

$$m = m_0\frac{u_{(R)}}{c}\cos\alpha = m_0\left(1 - \frac{GM}{c^2R}\right)\cos\alpha = m_0\left(1 - \frac{R_0}{R}\right)\cos\alpha \tag{8}$$

5) **Gravitational force**: The gravitational force [8], in fact, is created as a result of a transition between the different energy levels, see **Figure 1(b)**.

6) **Speed of light:** Below it appears that the speed of light under the influence of a gravitational field is equal to $u_{(R)}$.

3. Potential Energy

Many physicists have failed to explain the fact that potential energy $E_p = mg\Delta R$ depends proportionally on the magnitude R. That is, the farther away from the star, the greater the potential energy, but if we set the value of $g = \frac{GM}{R^2}$, we obtain that $E_p = m\frac{GM}{R^2}\Delta R$, where the potential energy is inversely proportional to R. In addition, another question is if the potential energy is positive or negative. Here we give an exact expression of the potential gravitational energy.

The state kinetic energy is an energy that contains within it the two energies, both the kinetic and the potential. If we take a body at rest at a distance R as shown in **Figure 2(a)**, in this case all the state kinetic energy is a potential energy, because the body is at rest, as we obtained in Formula (2). If we set the value of $u_{(R)}$ from Formula (6), we obtain the expression of the potential energy:

$$E_p = E_{\alpha=0} = m_0\left(\frac{G^2M^2}{c^2R^2} - \frac{GM}{R}\right) \tag{9}$$

The potential energy can also be written, using Formula (7), as follows:

$$E_p = m_0\frac{GM}{R}\left(\frac{R_0}{R} - 1\right) \tag{10}$$

If we analyze the function of the potential energy E_P, it seems that it always has a negative value (assuming that $R \geq R_0$). The potential energy is equal to zero in two places, $E_p = 0$, when $\begin{cases} R = R_0 \\ R = \infty \end{cases}$. The minimum value is obtained when R is equal to:

$$R_s = 2\frac{GM}{c^2} = 2R_0 \tag{11}$$

This is the Schwarzschild radius that represents the Event horizon ref. [9] [10].

It is easy to show that if we calculate the difference in the potential energy between two radii close to each other R_1 and R_2 (when $R_2 > R_1$), under the influence of a weak gravitational field (such as Earth), we obtain:

$$\Delta E_p \approx m_0\underbrace{\frac{GM}{R^2}}_{g}\underbrace{\left(R_2 - R_1\right)}_{h} = m_0gh$$

4. Kinetic Energy

As mentioned earlier, state kinetic energy E_α contains within it two types of

energy, kinetic energy and potential energy, *i.e.*:

$$E_\alpha = E_p + E_k \tag{12}$$

Figure 2(b) shows a body at a distance R moving with the velocity $v = u_{(R)} \sin\alpha$, having a self-speed of light vector $\boldsymbol{U}_{(R)}$. The state kinetic energy will be:

$$E_\alpha = m_0 \boldsymbol{U}_{(R)} \left(\boldsymbol{U}_{(R)} - \boldsymbol{C} \right) = m_0 u_{(R)}^2 - m_0 c u_{(R)} \cos\alpha$$

When we subtract from the state kinetic energy the value of the potential energy (Formula (2)), we obtain the exact value of the kinetic energy:

$$E_k = m_0 c u_{(R)} \left(1 - \cos\alpha\right) = 2 m_0 c u_{(R)} \sin^2\left(\frac{\alpha}{2}\right) \tag{13}$$

In a weak field and at low velocities, using approximations that $\sin^2\left(\frac{\alpha}{2}\right) \approx \left(\frac{\alpha}{2}\right)^2 \approx \frac{\sin^2\alpha}{4}$, $u_{(R)} = c$ and the formula $v = u_{(R)} \sin\alpha$, we obtain the known formula for kinetic energy: $E_k = \frac{m_0 v^2}{2}$.

5. Escape Energy and Escape Velocity

In order to be free from the effect of the gravitational field, we must give the small body an escape energy, which is a kinetic energy equal in its absolute value to the potential energy that we calculated earlier, but with the opposite sign, so that their sum (state kinetic energy) is equal to zero $E_\alpha = E_p + E_k = 0$, *i.e.*:

$$E_{k(\text{Escape})} = -E_p = m_0 \left(\frac{GM}{R} - \frac{G^2 M^2}{c^2 R^2} \right) = m_0 \frac{GM}{R} \left(1 - \frac{R_0}{R} \right) \tag{14}$$

R_0 is a zero horizon, and has a relatively small value existing outside the mass, in cases of black holes only. Therefore when R is much larger than R_0, it can be written approximately:

$$E_{k(\text{Escape})} = m_0 \frac{GM}{R} \tag{15}$$

If $E_{k(\text{Escape})}$ is equal to E_P, then we obtain that the state kinetic energy E_α, which is actually their sum, will be equal to zero, according to Formula (12). According to Formula (2), we see that the state kinetic energy is a Dot product of two vectors: $E_\alpha = m_0 \boldsymbol{U}_{(R)} \boldsymbol{V}_{ax} = m_0 u_{(R)} v_{ax} \cos\beta$, when $\boldsymbol{V}_{ax}$ is the separation velocity in space-time of two bodies ref. [1] [2]. The value of the state kinetic energy becomes zero when the angle between the two vectors $\boldsymbol{U}_{(R)}$ and $\boldsymbol{V}_{ax}$ is 90˚, as shown in Figure 3.

From the two formulas:

$$\cos\alpha = \frac{u_{(R)}}{c} = 1 - \frac{GM}{c^2 R}$$

$$v_{(R)(\text{Escape})} = u_{(R)} \sin\alpha$$

The escape velocity is obtained:

$$v_{(R)(\text{Escape})} \geq \left(1 - \frac{GM}{c^2 R}\right) \sqrt{2\frac{GM}{R} - \frac{G^2 M^2}{c^2 R^2}} \tag{16}$$

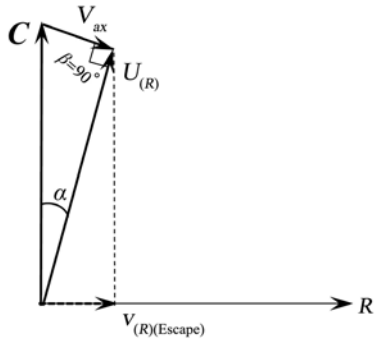


Figure 3. A description of the escape velocity in space-time, when the state kinetic energy is equal to zero.

It is clear that under the influence of a weak gravitational field, *i.e.*, far from zero horizon, we obtain that the escape velocity is $v_{(R)(\text{Escape})} \geq \sqrt{2\frac{GM}{R}}$.

As mentioned earlier, a body with an escape velocity causes the state kinetic energy to become zero, to any value of R. That is, on the way to the escape, the angle between $\boldsymbol{U}_{(R)}$ and $\boldsymbol{V}_{ax}$ in space-time will always be 90˚, until it reaches a distance large enough to detach from gravity, as shown in Figure 4. The separation velocity between the two bodies, the large body and the small body $\boldsymbol{V}_{ax}$ in the space-time, will decrease to a value of zero in the infinity, so the value of vector $\boldsymbol{U}_{(R)}$ will be equal to the speed of light $\boldsymbol{C}$ in an absolute manner. Another thing that is obtained from the fact that the state kinetic energy becoming zero is that the total energy E of the small body is equal to the energy of the self-time E_{st}, that is $E = E_{st}$, because $E_{\alpha} = 0$.

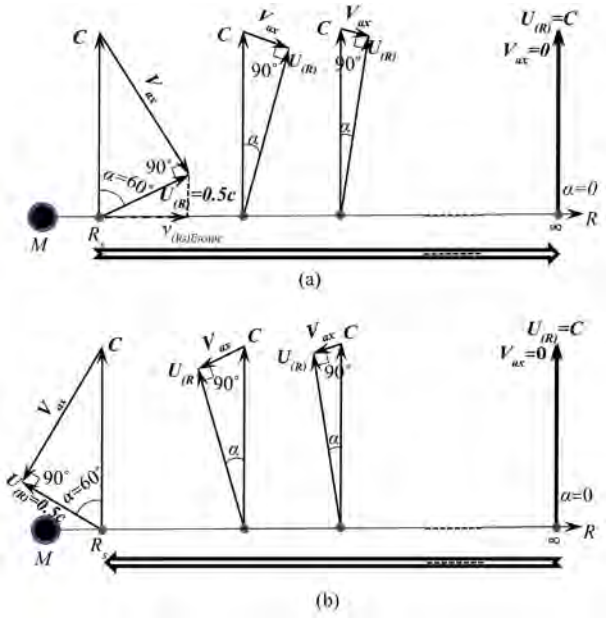


Figure 4. (a) A body with escape velocity maintains the state kinetic energy as zero along the way until it detaches from the gravitational field effect. (b) The process is reversible, *i.e.*, a body approaching a gravitational field receives the same values of energies.

In the inverted state, *i.e.*, when a small body is attracted to a large body from a great distance from a rest state, it undergoes the same process with the same parameters and magnitudes of the energies, in the opposite direction, as shown in Figure 4. When it reaches the event horizon R_s, its self-speed of light vector obtains a value of $u_{(R)} = 0.5c$, according to Formula (6), therefore the energetic angle will be $\alpha = 60^\circ$, and its velocity on the R axis will be at its maximum value, equal to: $v_{(Rs)\text{Escape}} = u_{(Rs)} \sin\alpha = 0.433c$. This is the escape velocity from the event horizon, as shown in Figure 4(a). In other words, the total energy and self-time energy of the small body on the event horizon is $E = E_{st} = 0.25m_0c^2$. The state kinetic energy is equal to zero $E_\alpha = E_k + E_p = 0$, but its components are of value $E_k = E_p = 0.25m_0c^2$. Therefore, in order to free the small body from the event horizon, we must invest in it kinetic energy equal to $E_k = 0.25m_0c^2$.

6. Satellite Time

As is well known, GPS satellites must be synchronized in time on Earth in order to get maximum accuracy. Furthermore, NASA's International Space Station (ISS) also needs to be synchronized to a clock on Earth in order to perform certain experiments. Therefore, we take these two examples as an example of our model test ref. [11].

Looking at the Earth, which is a relatively small planet, we try to test Formula (6). We obtain that the difference between the self-speed of light vector $\boldsymbol{C}$ of a free body (outside a gravitational field), which has an absolute value of the speed of light $|\boldsymbol{C}| = c$ and a body under the influence of the gravitational field of earth is so small, that it can reach a maximum value, on the surface of the earth, at a rate of $\Delta u = c - u_{(R)} = \dfrac{GM}{cR} \approx 0.21[\text{m/sec}]$. This is a velocity that seems negligible in relation to the speed of light, but this velocity creates a difference of $\Delta T = \Delta uT/c = 60$ μsec in one day, *i.e.* during a day of $T = 86{,}400$ sec, our time on Earth is slower by 60 μsec per day, relative to the free body, as shown in Figure 5.

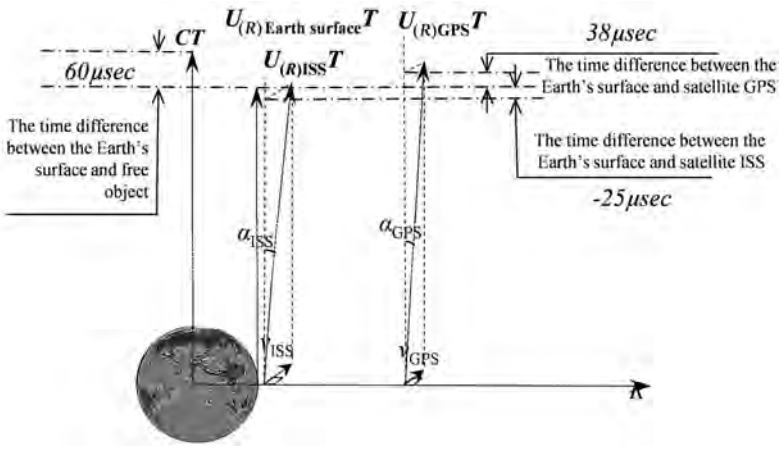


Figure 5. Description of time differences in a day, between ISS and GPS satellites and the time on Earth.

The time difference between a free body (in this case we take the Earth's core as reference, as a point mass in space) and a body under the influence of a gravitational field (satellite), moving at v in space x, y, z when the energetic angle α is obtained from the expression $v = u_{(R)} \sin\alpha$, we calculate the projection of the timeline, as shown in Figure 5: $c\Delta T = cT - u_{(R)} T\cos\alpha$

We insert Formula (6) and we obtain:

$$\Delta T = T\left(1 - \cos\alpha + \underbrace{\frac{GM}{c^2 R}}_{\frac{R_0}{R}}\cos\alpha\right) \tag{17}$$

Formula (17) expresses the time differences between two reference frames of two bodies, one is a free stationary body (without gravitational effect), and the other is a mobile body under gravitational field influence. Therefore, it can be seen that this formula contains within it the two theories of private and general relativity together.

A table was constructed to calculate the time differences in a day, according to Formula (17) and the above data.

The results of Table 1 can be seen schematically in Figure 5, which depicts the self-speed of light vectors of each one of the frames, and its projection on the frame axis.

Table 1. The time differences of the ISS and GPS satellites in a day, taking into account the gravitational field and satellite motion.

	Distance R (m)	Velocity V (m/sec)	The energetic angle α (rad)	Time difference in a day ΔT (sec)	Time difference in a day relative to the surface of the earth
Free body (Earth Core)	0	0	0	0	+60 μsec
The surface area of the Earth	6,357,000	0	0	60 μsec	0
Satellite ISS	6,767,000	7700	2.56838×10^{-5}	85 μsec	−25 μsec
Satellite GPS	26,541,000	3874	1.29219×10^{-5}	22 μsec	+38 μsec

7. The Speed of Light under the Influence of a Gravitational Field

Figure 6(a) depicts a self-speed of light vector of a small body moving to the center of the large body (star), at point B, at a velocity $v_{(R)}$ with an energetic angle α such that $v = u_{(R)} \sin\alpha$ Figure 6(b) depicts an ICF (Integral Couple Frame) ref. [1]. Figure 6(c) depicts the line of light from point B to point A. In space-time, two very close points are chosen, as is well known $\beta = (\pi/2 - \alpha)/2$ ref. [1]. Therefore, the speed of light in this small interval is:

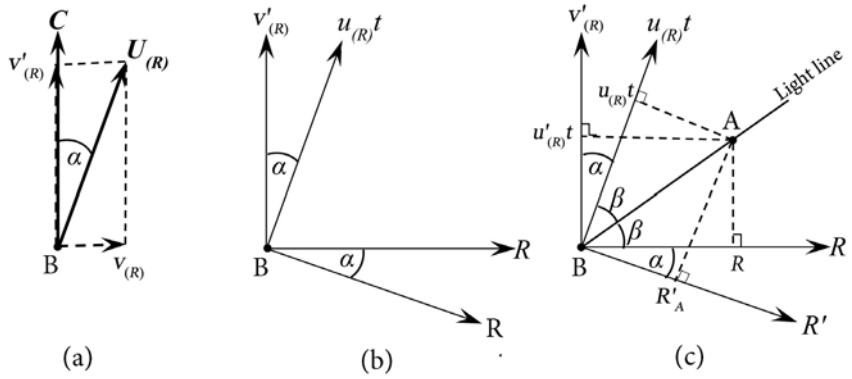


Figure 6. (a) Description of the self-speed of light vectors of the small body $U_{(R)}$ moving at a velocity $v_{(R)}$, and of the large body C; (b) ICF (Integral Couple Frame); (c) Description of the light path in the ICF frame.

$$\begin{aligned}\text{Gravitational light speed} &= \frac{R_A}{t_A} = \frac{u_{(R)}t_A}{t_A} \\ &= u_{(R)} = c - \frac{GM}{cR}\end{aligned} \tag{18}$$

Therefore, the speed of light, under the influence of a gravitational field, at a distance R, is equal to $u_{(R)}$.

As another example, two spaced apart points A, B are chosen. A beam of light traveling from point A to point B passes through a gravitational field, as shown in **Figure 7**. The trajectory that the beam will travel will be the shortest optical path length, according to Fermat's principle ref. [12], *i.e.*, the shortest time to travel between the two points:

$$T_{[\min]} = \int_A^B \frac{|\mathrm{ds}|}{u_{(R)}} \tag{19}$$

As can be seen in **Figure 7**, the trajectory of the light beam is not straight, and it depends on the magnitude of the gravitational field through which it passes. That is, it is a gravitational curvature of the beams of light. Optically, in order to calculate the trajectory of the light motion, it can be assumed that the refractive index at level R is: $n_{(R)} = \dfrac{c}{u_{(R)}}$.

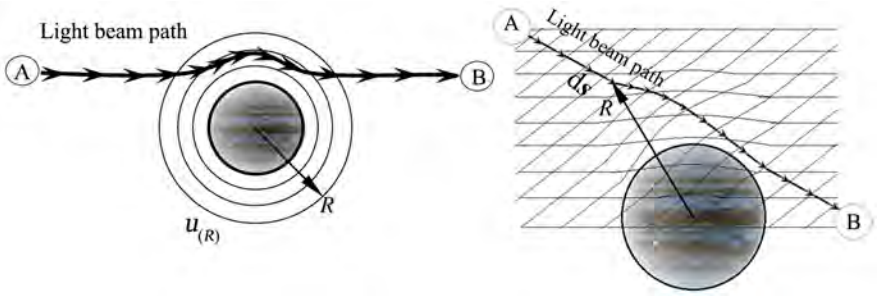


Figure 7. Description of the trajectory of a beam of light moving from point A to point B, through a gravitational field.

8. The Gravitational Force

It can be said that a gravitational field creates different energy levels, which depend directly on the distance R and the mass M of the field source. In the transition between the energy levels, force of gravity is obtained.

Force of gravity acts on a small body at **rest** under the influence of a gravitational field of a large body M, at a distance R, called the body weight, and is equal to the gradient of the potential energy: $\boldsymbol{F} = -\text{grad}E_p = -\nabla \cdot E_p$. The potential energy is actually the state kinetic energy when the body is at rest, which we calculated in Formulas (9) and (10). Therefore force of gravity will be:

$$\boldsymbol{F}_{(R)} = -\frac{\mathrm{d}E_p}{\mathrm{d}R} = -m_0\left(\frac{GM}{R^2} - 2\frac{G^2M^2}{c^2R^3}\right) \tag{20}$$

Formula (20) can also be written as follows:

$$\boldsymbol{F}_{(R)} = -m_0\frac{GM}{R^2}\left(1 - \frac{2R_0}{R}\right) \tag{21}$$

The value $R_s = 2R_0$ is the event horizon, in this case.

According to Formula (21), in the case of a black hole, we see that force of gravity becomes zero on the event horizon, and it even changes direction in the area of $R_s > R > R_0$, *i.e.*, in this area it becomes a repulsive force, as shown in Figure 8. It is important to note that very strong forces are included in these areas. It can be said that the event horizon behaves like a (bouncing) trampoline on which the body will swing.

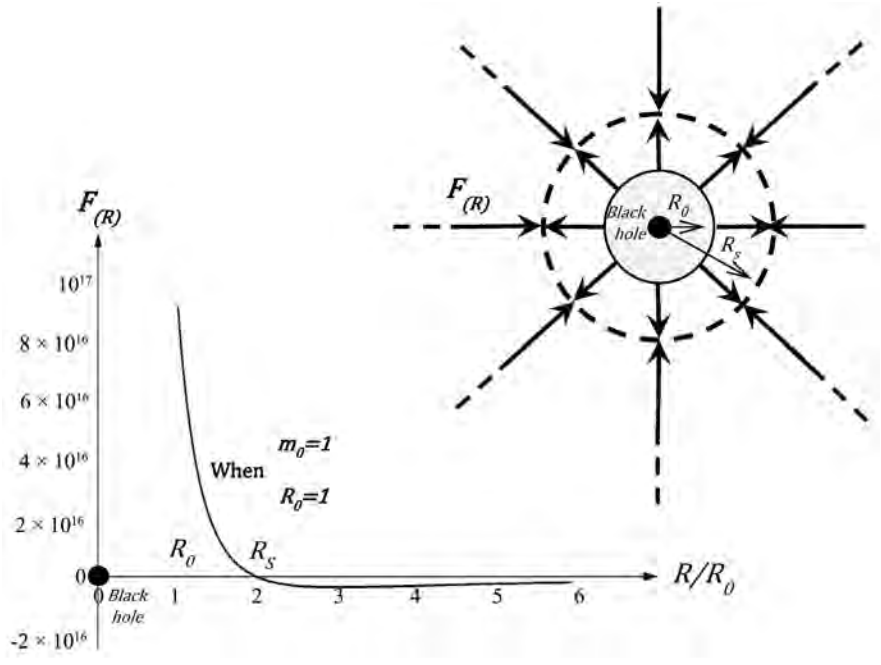


Figure 8. Description of the direction of the gravitational force in a black hole. Up to the event horizon, the force is a force of gravity, between the event horizon and the zero horizon there is a repulsive force.

For example, a body which stabilizes at the end of a process in the event horizon of a black hole. In this case, his energy is $E = m_0 u_{(R)}^2 = 0.25 m_0 c^2$. It is clear that forces acting on the body are conservative forces, so the energy difference, which is $\Delta E = 0.75 m_0 c^2$, transfers this energy to the core of the black hole. Therefore, the self-speed of light vector of the black hole core will be greater than the speed of light, since in the process of energy transfer there is no mass transfer, when the mass of the black hole core remains constant in this case. Here we got for the first time a higher velocity than the speed of light, in a limited range between $R_s > R > 0$.

9. Discussion and Summary

9.1. Discussion

We show the total energy of a small body in a gravitational field, its kinetic energy and potential energy. We show that as the body gets closer to the large body, its energy will decrease accordingly. We calculate and explain the horizon zero, the time difference between two bodies, the effective mass of the body, the gravitational force created by this change in the energy levels, and the effect of a gravitational field on the speed of light. Furthermore, we calculate its escape energy and escape velocity, showing that a body with escape velocity maintains the state kinetic energy as zero along the way until it detaches from the gravitational field effect and that the process is reversible, *i.e.*, a body approaching a gravitational field receives the same values of energies.

Later on, we show that there is a time difference between the ISS and GPS satellites, caused by a time difference two reference frames of two bodies: one is a free stationary body (without gravitational effect), and the other is a mobile body under gravitational field influence.

We continue with the calculation of the speed of light, and we show the influence of a gravitational field on it. Finally, we show that gravitational field creates different energy levels, which depend directly on the distance R and the mass M of the field source. In the transition between the energy levels, force of gravity is obtained. We reach a higher velocity than the speed of light, in a limited range between $R_s > R > 0$.

9.2. Summary

This model presents the force of gravity in a different and innovative way, thus giving answers to issues on many topics in physics that bother many scientists. From looking at a body with a large mass that forms envelopes of energy levels, this model shows that these energy levels are the reason for the formation of the gravitational force. The gravitational force, in fact, is created as a result of the transfer between the different energy levels. It is possible to refer to a single mass or to several masses (for example, a galaxy, or the entire universe) which at their center there is an imaginary mass equivalent to them. In addition, this model shows that these energy levels are also the reason for the decrease in the speed of

light, resulting in time differences, which create potential gravitational energy which is accurately calculated in this model. These energy levels are also the reason for bending the movement of light that passes through the gravitational field they create.

9.3. Future Studies

This research paper is in fact a gateway to other research papers, in which we will examine additional topics, including:

1) What does the force of gravity depend on, when there are different initial conditions, for example initial velocity, or negative mass, *i.e.* a small body with a large energetic angle.

2) Is the event horizon a fixed value or does it change according to the initial conditions?

3) Does a gravitational field affect the state kinetic energy?

4) This research paper is a true theory for large bodies, such as stars and black holes, and also for tiny particles such as the various components of the atom. Therefore, a general idea can be found from it for all four known forces in physics ref. [13].

5) What are the astrophysical consequences of the self-speed of light vector and the core of gravitational mass?

Acknowledgements

R. B. G thanks always to my teacher and Rabbi Prof. Aharon Peled, as well as my thanks to Noya Epelstain.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Tannous, E. (2019) Negative Mass. *Journal of Modern Physics*, **10**, 861-880. https://doi.org/10.4236/jmp.2019.107057

[2] Tannous, E. (2021) Energetic Angle. *Newest Updates in Physical Science Research*, **1**, 45-64. https://doi.org/10.9734/bpi/nupsr/v1/7088D

[3] O'Connor, J.J. and Robertson, E.F. (1996) General Relativity. History Topics: Mathematical Physics Index, School of Mathematics and Statistics, Scotland.

[4] Feynman, R. (1970) The Feynman Lectures on Physics. Pearson P T R, 1st Edition, Volume 1, Chapter 4. https://www.feynmanlectures.caltech.edu/I_04.html

[5] Newton, I. (1966) Philosophiae Naturalis Principia Mathematica (Mathematical Principles of Natural Philosophy, 1687). Translated in English by A. Motte, Revised and Annotated by F. Cajori (University of California Press). https://doi.org/10.5479/sil.52126.39088015628399

[6] Wald, R.M. (1997) Gravitational Collapse and Cosmic Censorship. Black Holes, Gravitational Radiation and the Universe. Springer, Dordrecht, 69-86.

[7] Günther, H. and Müller, V. (2019) Einstein's Energy-Mass Equivalence. The Special Theory of Relativity: Einstein's World in New Axiomatics. Springer, Singapore, 97-105. https://doi.org/10.1007/978-981-13-7783-9_7

[8] Overbye, D. (2015) Black Hole Hunters. NASA. Archived from the Original.

[9] Hamilton, A. (2020) Journey into a Schwarzschild Black Hole, JILA: A Joint Institute of NIST and the University of Colorado Boulder. https://jila.colorado.edu/~ajsh/insidebh/schw.html

[10] Thorne, K.S. and Hawking, S. (1994) Black Holes and Time Warps: Einstein's Outrageous Legacy. W. W. Norton & Company, New York, 134-135.

[11] Narayankar, S. (2019) *International Journal of Advanced Research* (*IJAR*), **7**, 782-789. https://doi.org/10.21474/IJAR01/8697

[12] Born, M. and Wolf, E. (1975) Principles of Optics. Cambridge University Press, Cambridge, p. 740.

[13] Davies, P. (1986) The Forces of Nature. 2nd Edition, Cambridge University Press, Cambridge.

Journal of Modern Physics, 2021, 12, 1295-1345
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Non-Uniqueness of Einstein's Special Relativity, and the Inconclusiveness of High Energy (Relativistic) Physics

Georg von Brzeski, Vadim von Brzeski

Helios Labs, 945 Hoxett St., Gilroy, CA, USA
Email: numberjoy@gmail.com

How to cite this paper: von Brzeski, G. and von Brzeski, V. (2021) Non-Uniqueness of Einstein's Special Relativity, and the Inconclusiveness of High Energy (Relativistic) Physics. *Journal of Modern Physics*, **12**, 1295-1345.
https://doi.org/10.4236/jmp.2021.129080

Received: May 9, 2021
Accepted: July 18, 2021
Published: July 21, 2021

Abstract

In this paper, we present a new form of "special relativity" (BSR), which is isomorphic to Einstein's "special relativity" (ESR). This in turn proves the non-uniqueness of Einstein's "special relativity" and implies the inconclusiveness of so-called "relativistic physics". This work presents new results of principal significance for the foundations of physics and practical results for high energy physics, deep space astrophysics, and cosmology as well. The entire exposition is done within the formalism of the Lorentz $SL(2C)$ group acting via isometries on **real 3-dimensional Lobachevskian (hyperbolic) spaces** L^3 regarded as quotients $SL(2C)/SU(2)$. We show via direct calculations that both ESR and BSR are parametric maps from Lobachevskian into Euclidean space, namely a **gnomonic** (central) map in the case of ESR, and a **stereographic** map in the case of BSR. Such an identification allows us to link these maps to relevant models of Lobachevskian geometry. Thus, we identify ESR as the physical realization of the Beltrami-Klein (non-conformal) model, and BSR as the physical realization of the Poincare (conformal) model of Lobachevskian geometry. Although we focus our discussion on ball models of Lobachevskian geometry, our method is quite general, and for instance, may be applied to the half-space model of Lobachevskian geometry with appropriate "Lorentz group" acting via isometries on (positive) half space, resulting yet in another "special relativity" isomorphic with ESR and BSR. By using the notion of a **homotopy** of maps, the identification of "special relativities" as maps from Lobachevskian into Euclidean space allows us to justify the existence of an uncountable infinity of hybrid "special relativities" and consequently an uncountable infinity of "relativistic physics" built upon them. This is another new result in physics and it states that so called "relativistic physics" is unique only up to a homotopy. Finally, we show that "paradoxes" of "special relativities" in either ESR or BSR are simply common distortions of

maps between non-isometric spaces. The entire exposition is kept at elementary level accessible to majority of students in physics and/or engineering.

Keywords

Lobachevskian (Hyperbolic) Geometry, Lorentz Group $SL(2C)$ Action, "Special Relativity", High Energy (Relativistic) Physics, "Paradoxes", Deep Space Astrophysics, Cosmology

1. Introduction

The work we present here deals with maps. The meaning of a map used in mathematics and physics is slightly different; however, its essence is the same. There is an opinion that mathematics studies sets (with additional structures) and maps between those sets. This is true also in physics. However, while the maps studied in mathematics are usually between some abstract mathematical structures, in physics, due to its experimental nature, the range of a map is always some subset of the real numbers.

The subsets of the **real numbers** the physicists are concerned with **are called experimental data**. A map from a physical space into real numbers is called a **direct map** which, in scientific practice, is realized as **data acquisition and reduction**. The process of data acquisition is itself quite complex. For instance, in High Energy Physics, there are situations where some event of interest has to be singled out from billions of similar events.

Data acquired due to a direct map is an "easy" part of a scientist's job. The more difficult part is to **interpret** this experimental data, and this involves an **inverse map**. An inverse map deals with **our understanding of what actually happens in physical space(s) from which the direct map data come from.** The relation between maps in Mathematics and Physics can be presented as follows:

1) Direct map in Mathematics corresponds to data acquisition in Physics.

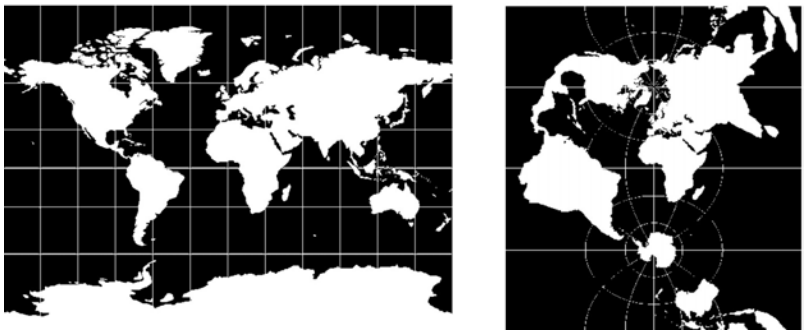
2) Inverse map in Mathematics corresponds to data interpretation in Physics.

In this paper, we are concerned with two issues:

1) With the **non-uniqueness of maps**, *i.e.* when several, alternative, **homotopy equivalent** maps exist, resulting in alternative **non-unique interpretations** of physical phenomena.

2) With **distortions** introduced in the mapping process, which result in **apparent deformations** of physical entities, which in turn are seen as **paradoxes**. See for example, Figure 1 below.

We will address the situation when experimental data come with distortions introduced by certain maps. These distortions are independent of informational noise, but they do depend on the particular map with which the experimental data are interpreted. This results in non-unique **quantitative** interpretations of physical phenomena.



(a) (b)

Figure 1. The figures above show two well known ([Source: Wikipedia Commons]) Mercator maps $S^2 \to E^2$, standard in (a) and transverse in (b), which are related by homotopy. (The notion of homotopy is discussed in detail in Sections 2.3 and 6). Distortions present in the standard projection are continuously changed into distortions in the transverse projection, as the angle between an axis of the projection cylinder and North-South axis of the sphere S^2 changes from 0 to 90 degrees. In this paper we discuss analogous mappings and their distortions, however between Lobachevskian and Euclidean spaces, and show how **relative apparent sizes** of objects due to mapping distortions, particularly in the case of Lobachevskian space mapping, are incorrectly viewed in physics literature as **"paradoxes"**. (a) Standard mercator projection; (b) Transverse mercator projection.

More to the point, we are interested in maps between spaces of the **same dimension but with different** (constant) **curvatures**. The non-uniqueness of such interpretations is **inherently present** in High Energy (high relative velocities) Physics (HEP), and they arise naturally in maps of Lobachevskian negatively curved spaces into a Euclidean flat space.

Lobachevskian (hyperbolic) geometry was developed by several mathematicians, Gauss, Schweikart, Bolay father and son, Beltrami and Lobachevski, just to mention a few. After a **two thousand year** struggle to prove Euclid's fifth postulate, it appeared around 1835 in closed form due to N.I. Lobachevski, a Russian mathematician of Polish ancestry. A good introduction to Lobachevskian geometry can be found in Anderson [1].

It is needless to say that High Energy Physics (HEP), otherwise known in the literature as "relativistic physics" and "relativistic astrophysics", where relative velocities range from fractions of c to nearly c, is based on Einstein's "special relativity" (ESR). Since 1905, it has been generally accepted that phenomena occurring at high relative velocities (with respect to c) are modeled in **unique** way (*i.e.* **in the only way possible**) by Einstein's "special relativity". Such beliefs resulted in the confidence that the numerical information represented by data gained from ESR indeed reflects the truth about the Nature. This belief, as it will be shown, is misguided.

Since we present a "special relativity" that is mathematically **isomorphic but not isometric** (or numerically different) from Einstein's "special relativity", in order to avoid confusion, we will label Einstein's "special relativity" as **ESR** and the authors' "special relativity" as **BSR**. Both ESR and BSR presented here are

based on **the same group of symmetries**—the Lorentz group $SL(2C)$, acting on its own homogeneous space.

In this paper, we present the following new scientific results, which beyond practical significance, are of utmost importance to the foundation of physics:

1) An alternative to Einstein's "special relativity" (ESR) namely the **authors' "special relativity" (BSR)** (based on the same Lorentz group).

a) It is shown that BSR corresponds to the **Poincare** model of Lobachevskian geometry, while Einstein's ESR corresponds to the **Beltrami-Klein** model of Lobachevskian geometry.

b) It is shown that BSR is equivalent to a **stereographic** projection from Lobachevskian into Euclidean space, while Einstein's ESR is equivalent to **gnomonic** (central) projection from Lobachevskian space into Euclidean space.

c) It is shown that both projections, the Poincare (stereographic) and the Beltrami-Klein (gnomonic), result from different **actions of the Lorentz group** $SL(2C)$ **on its own homogeneous space** $SL(2C)/SU(2)$ isomorphic to a real 3-dimensional Lobachevskian space.

2) Due to above results, we prove the **non-uniqueness** of Einstein's "special relativity" and the non-uniqueness of any Lorentz group based "relativity" as well.

3) On the basis the **homotopy** theory, an existence of an uncountable **infinity** (*i.e.* continuum) of alternative "special relativities" is proved.

4) On the basis of (3) above, we prove the **inconclusiveness of High Energy Physics**, or more precisely, the **conclusiveness of HEP up to a homotopy only**.

5) The so called *Twin Paradox* is mathematical solved in a symmetric setting showing the **apparent** distortions introduced by various maps called "special relativities".

6) The problem of images of **fast moving circular objects** is shown to be **a-priori undecidable**, being non-unique and dependent on a particular map.

How This Paper Is Organized

Our work is functionally divided into the following sections:

- Section 2: We discuss the properties of maps, specifically of maps between spaces of constant curvature and Euclidean space. This is important because expertise gained from spherical cartography, $K>0$, will be applied to cartography from Lobachevskian space (hyperboloids) $K<0$.
- Section 3: We discuss Lorentz group actions on homogeneous spaces. From the single concept of Lorentz group action, we arrive at our equation of our "special relativity" BSR as well as equations of Einstein's "special relativity" ESR.
- Sections 4 and 5: We show that various "Special Relativities" are simply maps from Lobachevskian Space into Euclidean Space. In particular we study the gnomonic (central) and stereographic maps from spheres and hyperboloids, and establish their relation to Beltrami-Klein model and to Poincare model of Lobachevskian geometry respectively. The isomorphism of the Beltrami-Klein and Poincare models is shown, thus showing the isomorphism between

Einstein's "Special Relativity" ESR and our "Special Relativity" BSR.

- Section 6: We discuss the homotopy of maps, resulting in a continuously infinite set of possible "special relativities". The existence of infinitely many uncountable "special relativities" is proved via the concept of homotopy.
- Section 7: We discuss several "paradoxes" from relativistic and high energy physics. We show the apparent nature of so-called "relativistic effects" and/or "paradoxes" as distortions resulting from maps between non-isometric spaces of the same real dimension.

A word on notation. We use quite standard notation, however, following Gelfand, Grayev, and Vilenkin, everywhere in this work, we call $SL(2C)$ the Lorentz group. In older literature, the Lorentz group is $SO(1,3)$, for which $SL(2C)$ is its double cover.

2. Maps from Spaces of Constant Curvature into Euclidean Space and Their Properties

In this section we will give the reader an easy introduction to effects of distortions which are present in maps between non isometric spaces. As we will later see, these distortions in the context of "special relativity" are misunderstood as real phenomena and are represented by variety of "paradoxes".

2.1. Distortions of Data Due to Maps between Non-Isometric Spaces

The physics in this work takes place in Lobachevskian negatively curved spaces. Unfortunately we do not perceive negative curvature in the way we perceive the spectral content of a light (colors), or music, or motion around us. **Thus there is a need to "translate", to map, to project the results internal to Lobachevskian spaces into the 3-dimensional piece of Euclidean (flat) space of our laboratory**. This is a step of utmost importance, since during the mapping we introduce (out of necessity) all kinds of **distortions which even today are not properly understood**.

Exploration of the world around us, and exploration of the Earth in particular, would be hardly possible without maps. It is essential to understand the different ways or methods of making maps, and the **distortions of the images** produced by these maps or models.

In ancient times, as long as people did not wander "too far" from their homes, maps drawn on a flat piece of paper were quite faithful. The Earth was believed to be flat just like the piece of a paper. Thus image of the Earth's flat surface on a flat piece of paper was perfect with no distortions of any kind, neither angular **nor in relative sizes**. Mapping of a flat space onto flat space (of the same dimension) with **no distortions** is possible because:

Remark 1 *All Euclidean spaces of the same dimension are* ***isometric****. It follows that maps between Euclidean spaces* (*of the same dimension*) *are globally distortion-less on the entire space*.

Due to the above property of Euclidean spaces of the same dimension, we

have in fact **only one Euclidean geometry**. Therefore, **maps between formally different representations of Euclidean geometry via different spaces (of the same dimension) are in fact done "in the same space"**.

This situation changed when people started to navigate open seas and realized that what they considered as flat was in fact **positively curved**. Thus for navigational purposes, new map making methods were developed. These methods applied mathematics **to get images of the Earth's curved surface on a piece of flat paper, *i.e.* onto a piece of the Euclidean plane** E^2.

It is important to understand how the distorted image of the Earth's surface we see on a flat map is related to the positively curved surface S^2, $K>0$, of the Earth which is a **model of the non-Euclidean - spherical geometry** in two dimensions. It is important because using examples of positively curved spaces, which intuitively are better perceived than negatively curved ones, we will demonstrate some general mathematical concepts like **isometry, isomorphism and conformality**. These concepts apply equally well to maps from positively curved spaces and maps from **negatively** curved spaces. The experience acquired from mapping spaces of **constant positive curvature** into Euclidean space will make it easy to comprehend mappings of spaces of **constant negative curvature** into Euclidean space. As we will show, at the base of Einstein's "special relativity" (ESR), our "relativity" (BSR), and all possible other future "relativities" based on the Lorentz group $SL(2C)$, lies the **problem of mapping of non-Euclidean** (negatively curved) **spaces into Euclidean flat spaces**.

In general, when the curvature of the space in question is not constant, the problem of mapping is quite complicated. However, in the case of constant positive curvature, for instance, several ways of mapping have been developed. Many models (images) of the spherical geometry of the Earth's surface in a Euclidean (flat space) exist. These maps, known also as **projections**, give images of a curved geometry via a flat (Euclidean) geometry. Widely known examples include **orthographic**, **stereographic, gnomonic, and Mercator** projections. Each particular map has its own advantages and disadvantages depending on its applications.

Different maps (viewed as sets) from a **curved** $K=const.\neq 0$ space into a **flat** $K=0$ space all are **isomorphic**. However they are **not isometric**, and some are **non-conformal**, which means that images resulting from those maps will show **distortions in relative sizes and angular relations** of mapped objects. As a rule, distortions are larger the larger the chunk of curved space that is mapped into a flat space, so the mapping is faithful only locally. The general rule is:

Remark 2 *Maps between spaces of constant nonzero curvature and Euclidean spaces cause distortions, and only locally are they approximately distortion-free. This is because spaces having different constant nonzero curvatures are not isometric, assuming that dimensions of all spaces are equal.*

In the case of an $S^2 \to E^2$ mapping, distortions are determined by the ratio

$\frac{l}{R}$ of the curved domain's linear size l to the radius R of the ball B^3 bounded by S^2. The smaller the ratio $\frac{l}{R} \ll 1$ is (recall the $\frac{\upsilon}{c} \ll 1$ regime in ESR), the more faithful the image of that curved surface will be on flat a surface. It is obvious that such **maps** are neither regarded as **"laws of nature"** nor as **"theories"** (as in the "theory of special relativity"), but merely as convenient ways of mapping curved spaces (surfaces) into flat spaces (surfaces).

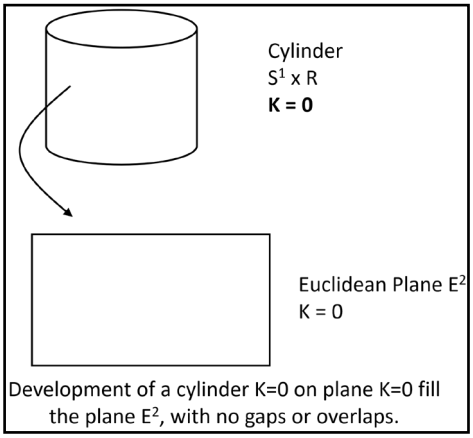
If we were to regard map making as the business of mathematics, and exploratory journeys due to those maps as physical experiments, then it is obvious that not every map (from all available maps) will be useful for a particular experiment, since some of those maps will produce highly distorted information that is far from reality. For instance, in the case of spherical geometry applied to mapping the Earth's surface, it is quite relevant which map (flat model of the Earth surface) will be used as the tool to navigate the globe. The standard Mercator map is **entirely useless** to navigate around the poles (e.g. around Greenland or Antarctica) due to **excessive distortions** (non-linearities); see Figure 1. The Mercator map shows Greenland being as large as Africa, despite Africa actually being 15 times larger than Greenland. Nevertheless, the Mercator map is quite a good model of the Earth's geometry if navigating around low latitudes (*i.e.* close to the equator).

Methods of mapping the Earth's surface have developed quite well over time. This is because **geometric forms of positive curvature** are abundant around us, and were known to man since ancient times. Maps of objects of constant positive curvature into Euclidean space were known to artists, map makers, mathematicians for centuries. On the other hand, **objects or forms of negative curvature** went unknown or unnoticed up to the 19th century, and there were no relevant maps (at least in physics) from spaces of negative curvature into Euclidean spaces. In physics, the first such map was Einstein's ESR, and in art, Escher's paintings. It is no surprise that when ESR appeared in 1905 and was interpreted (unfortunately) via Minkowski's flat geometry—with all its paradoxes—it was quite a shock then and is still a major misunderstanding today.

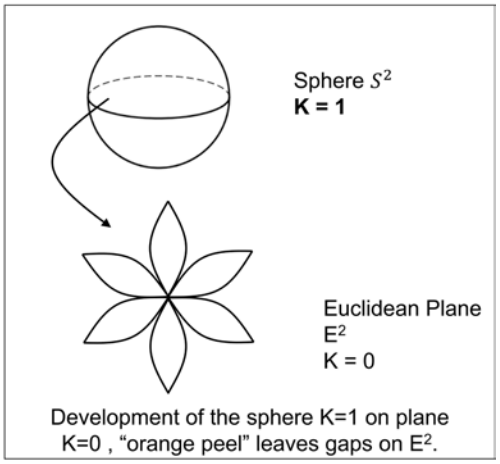
As in maps from positively curved spaces, we expect that when we map from negatively curved spaces into Euclidean space, **we will also experience distortions of images**. But since the curvature K is **negative** this time, the character of **distortions will be opposite to those seen in the positive curvature case**. **Distortions of maps from negatively curved spaces into flat spaces will appear as contractions**, instead of expansions.

It is well known in mathematics (see Figure 2) but overlooked by physicists and astronomers, that **negatively curved spaces are more volumetric**, and positively curved spaces are less volumetric than Euclidean spaces. Specifically:

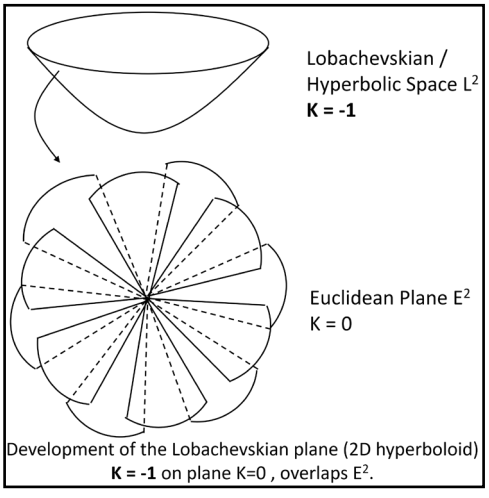
1) An inhabitant of a Lobachevskian space (a hyperboloid) with $K < 0$, who interprets other geometries in terms of his own, will see Euclidean space as **contracting** and a spherical space as **contracting even more**.



(a)



(b)



(c)

Figure 2. Figures (a)-(c) show the dependence of the volumetric content of spaces of constant Gaussian curvature K versus the sign of their curvatures. We see that spherical space, $K>0$, is less volumetric than Euclidean space $K=0$, while Lobachevskian space $K<0$ is more volumetric than Euclidean space $K=0$. It is interesting to note that this property of Lobachevskian spaces already found an application in data packing techniques and digital signal design (more space - more data) in the Internet domain.

2) On the other hand, an inhabitant of a spherical universe with $K>0$, who interprets other spaces via the geometry of his own space, will conclude that Euclidean spaces **expand** and Lobachevskian spaces **expand even more.**

3) An inhabitant of an Euclidean world with $K=0$, who believes that any other geometry (and physics) has to be interpreted in Euclidean terms, will see a spherical space as **Euclidean-contracting**, and see the **Lobachevskian (hyperbolic) space as Euclidean-expanding**.

We live in a world which locally, due to limited resolving power of our instruments, can be regarded as Euclidean. However, experiments clearly show that the physics and geometry of **high relative velocities** and **large distances** do not agree with the rules of Euclidean geometry [2] [3] [4]. Since we are dealing with velocities close to c and with distances on cosmological scales, in order to avoid confusion and common misinterpretations, we always need to remember the relation between the **sign of the curvature of the spaces and the relative volumes**. This is summarized in Table 1 with assumption that the dimension of all entries is the same.

Table 1. The "contraction"/"expansion" effects of images of physical objects or even entire spaces result from **comparing incomparable metric relations in non-isometric spaces**. The nine possible cases above (read as maps from rows $\rightarrow$ columns) show the apparent "effects" from the lack of consideration of non-Euclidean geometries. For instance, the third-row, second-column entry, "contraction": this is what an observer from a negatively curved space $K<0$ will conclude about a flat space $K=0$. Also, the second-row, third-column entry, "expansion": this is what observer from flat space $K=0$ will conclude about a negatively curved space $K<0$, precisely illustrating the misconception of the so-called "expanding universe" which has been plaguing cosmology for over a hundred years.

	$K=+1$	$K=0$	$K=-1$
$K=+1$	**isometry**	"expansion"	"expansion"
$K=0$	"contraction"	**isometry**	"expansion"
$K=-1$	"contraction"	"contraction"	**isometry**

2.2. Non-Uniqueness of Data Interpretation Due to Maps from Spaces of Constant Curvature into Euclidean Space

The aim of this section is to show how a continuous and uncountable **infinite family of parametric maps**, dependent on a real parameter, arise from maps of spaces of constant curvature into Euclidean space. The real parameter a on which the map depends produces a **gnomonic** (central) map and a **stereographic** map at its two particular extreme values. All other values of the parameter in between the extremes correspond to an uncountable **infinity of a mixed map**. The uncountable infinity of possible maps is expressed in a natural way by the notion of homotopy [3] which, in an informal way, is introduced in Subsection 2.3.

2.2.1. Spherical Cartography

To get more insight into cartography from a sphere via different **projections**, it will be natural to use the language of **projective geometry**. (Projective geometry was developed in the 19-th century for the needs of pure geometry; today it is the basic tool for computer graphics.)

Regarding our discussion, projective geometry encompasses both cases of spherical and Lobachevskian (hyperbolic) geometry. It gives an elegant and uniform way to treat spherical and Lobachevskian cartography by appropriate (to each case) normalization of projective coordinates. We present the spherical cartography case below, whereas the Lobachevskian cartography case is discussed in 2.2.2.

For the sake of simplicity only, we will consider the one dimensional case, namely the real projective line RP^1, and its mapping into an affine line E^1.

On the projective line RP^1 we have to deal with **projective coordinates**. Since in a real projective space of dimension n there are $n+1$ projective (homogeneous) coordinates, it follows that on a projective line which has dimension **one**, there will be **two** projective coordinates which we denote as ξ_0 and ξ_3. Subscripts are irrelevant but later on we will see why they are chosen in this way. Of course in any space of dimension n, there can only be n **linearly independent** coordinates. That redundancy, in the case of projective coordinates, is handled by imposing some kind of normalization condition which we choose as:

$$\xi_0^2 + \xi_3^2 = 1 \tag{1}$$

Normalization condition (1) results in **only one independent coordinate** and (in projective coordinates) is the equation of the unit circle centered at $o(0,0)$.

The single (local) coordinate on an affine line into which the projection is done is denoted by x. Affine x and projective coordinates ξ are related as:

$$x = \frac{\xi_3}{\xi_0 - a} \tag{2}$$

The parameter a in (2) has a simple meaning. From the **point of view of geometry**, it is equal to the **Euclidean distance** $\mathbf{d}(\mathbf{o},\mathbf{p})$ between the center of the unit sphere o and the center of projection p, so $a = d(o,p)$. The parameter a determines the **kind of projection** or a **type of a map**. Since in the present work we are interested in gnomonic and stereographic projections (and hybrid projections as well) the parameter a will be limited to values listed below. From the **point of view of topology**, parameter a is a **variable homotopy parameter** which determines the **continuous deformation** of one map into another [5]. As a changes **continuously** through its range, the maps change (deform) accordingly, something known in mathematics as the **homotopy of maps**, which we will discuss in more detail later on.

In this paper we concern ourselves with the following cases of the parameter a:

- $a = 0 \Rightarrow$ gnomonic (central) projection.
- $a \in (0,1) \Rightarrow$ continuum of hybrid (mixed) projections, mixed maps.
- $a = 1 \Rightarrow$ stereographic projection, stereographic map.

a = 0: **The case of gnomonic (central) projection**, from a projective line into an affine line.

We can express projective coordinates via coordinate(s) x by solving Equations (1) and (2) with $a=0$. This gives:

$$\xi_0 = \frac{1}{\sqrt{1+x^2}}, \quad \xi_3 = \frac{x}{\sqrt{1+x^2}} \tag{3}$$

In general, in an n-dimensional case of a gnomonic projection from an n-dimensional projective space into an n-dimensional affine space:

$$\xi_0 = \frac{1}{\sqrt{1+|x|^2}}, \quad \xi_i = \frac{x_i}{\sqrt{1+|x|^2}}, \quad i=1,\cdots,n \tag{4}$$

We call **projective** (homogeneous) coordinates is the form (4) **Weierstrass coordinates** since they are analogous to those used by Weierstrass in his work on Lobachevskian (hyperbolic) geometry [6] which we will explicitly derive in the next section.

a = 1. **The case of stereographic projection**, from a projective line into an affine line.

Solving Equations (1) and (2) with $a=1$, we find the projective coordinates expressed via affine coordinate(s) due to stereographic map from a projective to an affine line. Those are:

$$\xi_0 = \frac{1-x^2}{1+x^2}, \quad \xi_3 = \frac{2x}{1+x^2} \tag{5}$$

In general, in an n-dimensional case of stereographic projection from an n-dimensional projective space into an n-dimensional affine space:

$$\xi_0 = \frac{1-|x|^2}{1+|x|^2}, \quad \xi_i = \frac{2x_i}{1+|x|^2}, \quad i=1,\cdots,n \tag{6}$$

In both cases, in angular coordinates:

$$\xi_0 = \cos\alpha, \quad \xi_3 = \sin\alpha \tag{7}$$

Projective coordinates in a form of (6) are called **rational coordinates** and are related to a **stereographic projection**. They are also referred to as the **rational parametrization of a circle**. It is obvious that in all three cases (3), (5), and (7), $\xi_1^2 + \xi_3^2 = 1$.

In Figure 3 and Figure 4, the gnomonic and stereographic projections from the unit radius sphere into Euclidean space are shown. The aim of Figure 3 and Figure 4 is to show that the data exchange between a spherical space $K>0$ and a Euclidean space $K=0$ **cannot be interpreted in unique way**. Moreover, we see that the data due to a gnomonic projection are $D_g = \tan\alpha$, while data due to stereographic projection are $D_s = 2\tan\frac{\alpha}{2}$. Since:

$$\tan\alpha = \frac{2\tan\frac{\alpha}{2}}{1-\tan^2\frac{\alpha}{2}} \tag{8}$$

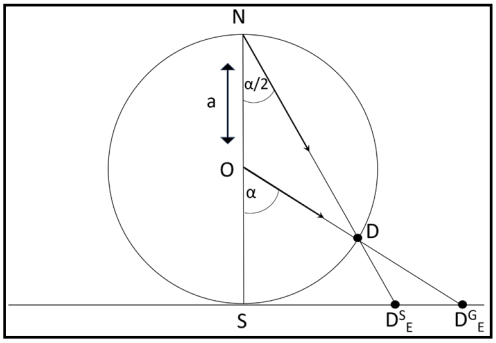


Figure 3. Spherical Space Mapping, $S^2(K>0)\to E^2(K=0)$. The single data D from spherical space are mapped onto **two different data** D_E^S and D_E^G in Euclidean space of a physicist's laboratory. Data D_E^S are due to a stereographic map while data D_E^G are due to a gnomonic map. Between D_E^S and D_E^G (one the segment (D_E^S, D_E^G)) there is a continuum of Euclidean data corresponding to hybrid maps when the projection point takes an arbitrary position on the segment *ON*.

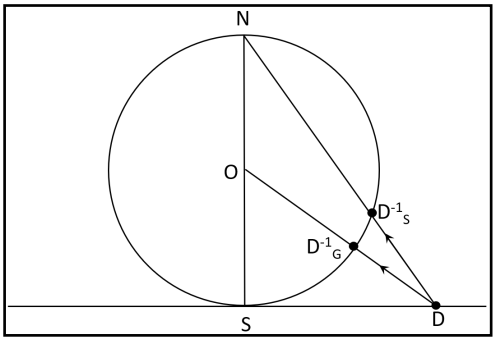


Figure 4. Spherical Space Mapping, $E^2(K=0)\to S^2(K>0)$, the inverse map. The single data D in the Euclidean space of a physicist's laboratory is mapped onto two different data D_G^{-1} and D_S^{-1} in spherical space. Data D_S^{-1} are due to a (inverse) stereographic map, while data D_G^{-1} result from a (inverse) gnomonic map. Between D_S^{-1} and D_G^{-1} on the segment [D_S^{-1}, D_G^{-1}] there is a continuum of spherical space data corresponding to a single Euclidean datum due to (inverse) hybrid maps. This illustrates the state of the **undecidable** situation when physicists are unable to decide which data in spherical space correspond (in a unique way) to the data they acquired in Euclidean space. In the language of cause and effect, it is impossible to single out a particular cause which results in a detected effect. This indeterminate situation is typical of quantum mechanics. We return to this below when we discuss "relativities" as maps.

or:

$$D_g=\frac{D_s}{1-\left(\frac{D_s}{2}\right)^2} \tag{9}$$

The inverse map is given as:

$$\tan\frac{\theta}{2}=\frac{\tan\theta}{1+\sqrt{1-\tan^2\theta}}=\frac{\left(1-\sqrt{1-\tan^2\theta}\right)}{\tan\theta} \tag{10}$$

Equations (8) and (10) give an **isomorphism** between a **gnomonic map** and a **stereographic map** of a unit sphere onto the Euclidean plane.

$$Gnomonic \Leftrightarrow \tan\theta \ \overleftrightarrow{ISO} \ \tan\frac{\theta}{2} \Leftrightarrow Stereographic \tag{11}$$

These facts will appear again when we discuss "special relativities" as gnomonic and stereographic maps.

Conclusion 3 *From the formulas of two alternative parametrizations of the unit circle* (*3*) *and* (*5*) *and from* ***Figure 3*** *and* ***Figure 4****, we note that the parametrization of a circle* (*a 1-dimensional* ***spherical space***) *by a* ***full parameter*** α *corresponds to a* ***gnomonic*** (*central*) *map, while a parametrization of a unit circle by a* ***half parameter*** $\alpha/2$ *corresponds to a* ***stereographic map****.*

At the end of this section we would like the reader to remember the following:

1) A central or gnomonic projection (map) from a sphere, $K>0$, is related to the Weierstrass parametrization of a unit sphere.

2) A stereographic projection from the a sphere, $K>0$, is related to the rational parametrization of the unit sphere.

2.2.2. Lobachevskian (Hyperbolic) Cartography

We now repeat the derivation from Section 2.2.1 but this time for Lobachevskian space; see **Figure 5**. Again, consider the projective line RP^1 equipped with projective coordinates u_0, u_3 subject to normalization condition (12):

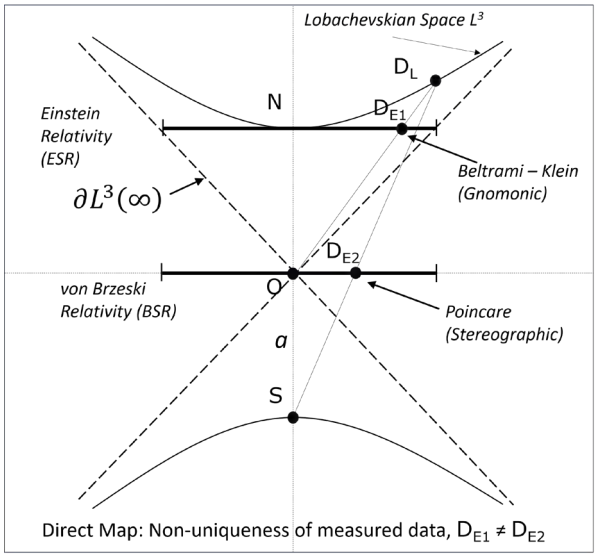


Figure 5. Lobachevskian (hyperbolic) Space Mapping, $L^2(K<0)\to E^2(K=0)$. The same single data D_L in Lobachevskian space are mapped onto two different data D_{E1} and D_{E2} into Euclidean space of physicist laboratory. Data D_{E1} are due to gnomonic map while data D_{E2} are due to stereographic map. Between D_{E1} and D_{E2} there is a continuum of Euclidean data corresponding to hybrid maps when projection point takes arbitrary position on $[-1,0]$ segment. Data D_{E1} and D_{E2} are related by isomorphism as $D_{E1}=\dfrac{2D_{E2}}{1+D_{E2}^2}$.

$$u_0^2 - u_3^2 = 1 \tag{12}$$

which in projective coordinates u_0 and u_3 is a two-branched unit hyperbola centered $O(0,0)$.

In order to be closer to what will follow, we take the Lobachevskian line as a 1-dimensional Lobachevskian velocity space, in which points represent velocities and distances between points represent **relative velocities**. This time the affine coordinate (depending on the projection from a Lobachevskian line to an affine line) will be denoted as v for a gnomonic (central) projection and as υ for a stereographic projection. The affine (local) coordinates **v** and υ are the **measurable data** which relativistic physics acquires from experiments. The parameter *a* represents the distance between the center of the hyperbola and the center of projection, and takes values $a = 0$, $0 < a < 1$, $a = -1$, yielding:

- $a = 0 \Rightarrow$ a gnomonic projection.
- $a \in (-1,0) \Rightarrow$ a continuum of hybrid projections (mixed maps).
- $a = -1 \Rightarrow$ a stereographic projection.

Depending on the parameter *a*, the affine coordinates v and υ are expressed, via projective coordinates, as:

$$\mathrm{v} \text{ or } \upsilon = \frac{u_3}{u_0 - a} \tag{13}$$

v in the case of $a = 0$, and υ in the case of $a = -1$, respectively; see Figure 5.

The reader is encouraged to repeat the calculations from previous section to see that:

a = 0. **The case of gnomonic projection**, from a Lobachevskian line into an affine line. Solutions of Equations (12) and (13) give:

$$u_0 = \frac{1}{\sqrt{1-\mathrm{v}^2}}, \quad u_3 = \frac{\mathrm{v}_3}{\sqrt{1-\mathrm{v}^2}} \quad |\mathrm{v}| \in (0,1) \tag{14}$$

and in the general *n*-dimensional case of a gnomonic (central) projection from an *n*-dimensional Lobachevskian space into an *n*-dimensional affine space:

$$u_0 = \frac{1}{\sqrt{1-|\mathrm{v}|^2}}, \quad u_i = \frac{\mathrm{v}_i}{\sqrt{1-|\mathrm{v}|^2}}, \quad i = 1,\cdots,n, \quad |\mathrm{v}| \in (0,1) \tag{15}$$

Projective coordinates in the form (15) are known as Weierstrass coordinates as mentioned earlier, which Weierstrass developed and used at least 50 years before Einstein's ESR. An alternative parametrization via hyperbolic functions yields:

$$u_0 = \cosh\theta, \quad u_3 = \sinh\theta \tag{16}$$

It is easy to check that both (14) and (16) obey the normalization condition (12). From (13) we see that the affine coordinate v at $a = 0$ is equal to:

$$\frac{u_3}{u_0} = \mathrm{v} = \tanh\theta \tag{17}$$

where θ is the (signed) distance in the Lobachevskian line, meaning Loba-

chevskian relative velocity. Note that while $0<|\theta|<\infty$, $0<v<1$. Values $\theta=\infty$ and $v=1$ belong to the boundary at infinity of a Lobachevskian line and are interpreted as velocities of photons.

a = −1. **The case of stereographic projection**, from a Lobachevskian line into an affine line. Similarly to the previous paragraph, we obtain:

$$u_0=\frac{1+\upsilon^2}{1-\upsilon^2}=\cosh\theta,\quad u_3=\frac{2\upsilon}{1-\upsilon^2}=\sinh\theta \tag{18}$$

Coordinates (18) are **rational projective coordinates**, which give the rational parametrization of a unit hyperbola and are analogous to the rational parametrization of a unit circle (5). Rational projective coordinates in the Lobachevskian case are related to **stereographic projection from a hyperboloid** as rational coordinates in the spherical case are related to a stereographic projection from a sphere.

The generalization of Formula (18) in an n-dimensional Lobachevskian space is:

$$u_0=\frac{1+|\upsilon|^2}{1-|\upsilon|^2},\quad u_i=\frac{2\upsilon_i}{1-|\upsilon|^2},\quad i=1,\cdots,n,\quad |\mathrm{v}|\in(0,1) \tag{19}$$

Next, from (17) and (18) we see that:

$$\frac{u_3}{u_0}=\tanh\theta=\mathrm{v}=\frac{2\upsilon}{1+\upsilon^2} \tag{20}$$

The formula

$$\mathrm{v}=\frac{2\upsilon}{1+\upsilon^2} \tag{21}$$

shows a well known relation of **isomorphism** between the Beltrami-Klein and Poincare models of Lobachevskian geometry, which is represented here by Lobachevskian velocity space.

Next, since:

$$\tanh\frac{\theta}{2}=\left(\frac{\cosh\theta-1}{\cosh\theta+1}\right)^{1/2} \tag{22}$$

from (22), we find the Euclidean velocity υ versus Lobachevskian velocity θ, in the case of a stereographic map, is:

$$\upsilon=\tanh\frac{\theta}{2} \tag{23}$$

Thus the relations of isomorphism between a gnomonic and a stereographic projection are:

1) In terms of Lobachevskian velocity space:

$$\tanh\theta=\frac{2\tanh\dfrac{\theta}{2}}{1+\tanh^2\dfrac{\theta}{2}} \tag{24}$$

$$\tanh\frac{\theta}{2}=\frac{\tanh\theta}{1+\sqrt{1-\tanh^2\theta}}=\frac{1-\sqrt{1-\tanh^2\theta}}{\tanh\theta} \tag{25}$$

See the analogy with spherical cartography.

2) In terms of Euclidean velocity space (that the physicist deals with):

$$\mathrm{v}=\frac{2\upsilon}{1+\upsilon^2} \tag{26}$$

$$\upsilon=\frac{\mathrm{v}}{1+\sqrt{1-\mathrm{v}^2}}=\frac{1-\sqrt{1-\mathrm{v}^2}}{\mathrm{v}} \tag{27}$$

$$\text{Beltrami-Klein} \ni \tanh\theta \ \overrightarrow{ISO} \ \tanh\frac{\theta}{2} \in \text{Poincare} \tag{28}$$

The arrow over *ISO* indicates both ways of mapping due to the isomorphism between gnomonic and stereographic images from the (unit) hyperboloid into the plane E^2.

The above simple calculations have led to a very important and well known conclusion:

1) The Beltrami-Klein model corresponds to a gnomonic (central) projection from Lobachevskian into Euclidean space.

2) The Poincare model corresponds to a stereographic projection from Lobachevskian to Euclidean space.

So far, performing the calculations in spherical and Lobachevskian cartographies, we have not mentioned anything at all about the Lorentz group or Lorentz transformations. In the following exposition (Section 3), we will obtain **the same results via Lorentz group action (Lorentz transformations) on Lobachevskian space**, resulting in "Special Relativities" which will be identified as maps (gnomonic, stereographic, or mixed).

At the end of this section we would like the reader to remember that:

1) The central or gnomonic projection (map) from Lobachevskian spaces (hyperboloids), $K<0$, into Euclidean space is related to the Weierstrass parametrization of the unit hyperboloid.

2) The stereographic projection (map) from Lobachevskian spaces (hyperboloids), $K<0$, into Euclidean space is related to the rational parametrization of a unit hyperboloid.

A nice discussion of hyperbolic cartography can be found in Reynolds [7].

2.3. Homotopy and an Infinity of Maps

A homotopy is a topological notion. It is quite important and useful for our exposition. It is analogous to the notion of **homeomorphism**, but it is coarser than homeomorphism. A formal definition will be given in Section 6. While a homeomorphism relates topological spaces, a homotopy relates continuous maps, which can be regarded as "points" in a **space of continuous mappings**. In terms of equivalence, a homotopy is an equivalence relation on a space of continuous mappings, similarly like homeomorphism is an equivalence relation on topolog-

ical spaces.

Mappings, or projections in general, depend on the position of the projection point, also known as the center of projection. Since we can **continuously** change the projection point, it follows that all such maps, viewed as a points in a so-called **mapping space** are **uncountably infinite** and can be changed one into another one **continuously**. In other words, they are all **homotopic**. The reader should note that these topological properties of maps apply to maps from spheres as well as from hyperboloids (Lobachevskian spaces).

We believe it is worth looking at the general picture via a **homotopy of maps**, which we will later apply to maps from Lobachevskian space $K<0$ into Euclidean space $K=0$. A homotopy of maps, simply put, is the continuous deformation of one continuous map into another continuous map. Spaces due to homotopic maps are of the same homotopy type [3], which roughly means that one space can be continuously deformed onto another.

Consider the gnomonic and stereographic projections in case of the sphere S^2 into the Euclidean plane E^2. The point from which the projection is done in a gnomonic map is the center of a sphere S^2, while in a stereographic map the projection is done from "the North Pole" point N, Figure 3. If we start with a gnomonic map and move the point of projection continuously up from the center of the sphere along the radius toward the North Pole N we will have a **continuous family of hybrid maps**. When the point of projection reaches to the North Pole N we will have a stereographic projection. You may imagine the whole process as a **continuous deformation of the gnomonic map into stereographic map**. We see that points p along the unit segment **continuously parametrize** maps of S^2 into E^2. Since on the unit segment there is an **continuum of points** (uncountable infinity), there is therefore a **continuum of maps** of a spherical space, $K>0$, into the Euclidean space E^2. In regards to our exposition, we can say that images of the Earth's surface S^2 due to gnomonic and stereographic maps are of the same homotopy type, *i.e.* **images of S^2 in E^2 are homotopy equivalent**.

Regarding the distortions of maps from curved into flat spaces, we recall that:

1) Maps from curved spaces into Euclidean space are **not unique and not isometric**. They will show **distortions** depending on the curvature of the initial space, on the method of mapping, and on the linear size of the mapped domain.

2) Maps (images) from positively curved spaces, $K>0$, e.g. from the Earth's surface, into the Euclidean plane $K=0$ will show **images of distortions of enlargement**, or **images of expansion** of objects, since a space of zero curvature (flat) is more volumetric than a positively curved space.

3) Conversely, maps from a Euclidean space into spherical space will show **distortions as compression**, and an contraction of images.

4) Maps (images) from Lobachevskian negatively curved spaces (hyperboloids), $K<0$, into an Euclidean space, $K=0$, will show **distortions of compression**, or an **contraction** of images of objects ("Fitzgerald contraction", "time dilation"). This is because Euclidean space is less volumetric than a hyperbolic

negatively curved (Lobachevskian) space.

5) Conversely, maps from an Euclidean space into a Lobachevskian space will show **distortions of enlargement**, or **distortions of expansion**.

3. Lobachevskian Homogeneous Spaces Related to the Lorentz Group and Lorentz Group Action

This section contains some already known material [8] [9] [10]. It is included as a convenience for the reader only.

Definition 4 *A space X with a given group acting on X is called a* ***homogeneous*** *space if any two points $x, y \in X$ can be joined by some $g \in G$, meaning $y = gx$ for any $x, y \in X$ and some $g \in G$. The group G is called the group of motions of the space X.*

Homogeneous spaces are very useful in mathematics and physics. Due to ideas going back to the German geometer Feliks Klein [11], homogeneous space X can be described solely in terms of the symmetry group G acting on it. Due to this construction, a homogeneous space inherits many useful properties which belong to a group associated with it.

The procedure itself is as follows. Take some arbitrary point $o \in X$ (call it the origin) and find a subgroup $H \subset G$ which leaves point o unchanged, $Ho = o$. That subgroup H is called a **stabilizer** of the point o. The choice of a point o is irrelevant. If one thinks, for instance, about a homogeneous space as of a sphere S^2 then any point on the sphere may be regarded as the "origin".

After the stabilizer has been found, the quotient space G/H is constructed in which "points" are identified with copies of H shifted by elements of G, namely $x_0 = eH$, $x_1 = g_1 H$, $x_2 = g_2 H, \cdots$ This one to one correspondence between points x_i and cosets $g_i H$ establishes an isomorphism between the spaces X and G/H.

We are interested in a real Lobachevskian 3-dimensional space(s) and a group of motions associated with it which is the Lorentz group $SL(2C)$. First, a few definitions:

Definition 5 *The Lorentz group $SL(2C)$ is the group of 2×2 complex matrices $g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$, $\alpha, \beta, \gamma, \delta \in C$, with unit determinant $\det g = 1$. Since $\det g = 1$, as a topological space, $SL(2C)$ may be viewed as a hyperboloid which is a generic representation of Lobachevskian space. $SL(2C)$ is a double cover of $SO(1,3)$ which is used in orthodox "special relativity" in Minkowski pseudo-Euclidean real 4-dimensional flat geometry.*

Definition 6 *Lobachevskian (real) space is a simply connected, non-compact, locally compact, metric space of constant Gaussian curvature $K < 0$. The generic value of K is usually set to -1, $K = -1$.*

Note that **compact** negatively curved spaces are not Lobachevskian spaces in the sense of Definition 6 and are not discussed here.

We will use two parallel representations of Lobachevskian space:

1) As an upper sheet of a $3d$-hyperboloid equipped with projective (homogeneous) coordinates

$$\xi_i,\ i=0,1,2,3,\ \xi_0^2-\xi_1^2-\xi_2^2-\xi_3^2=1,\ \xi_0\geq 1.$$

2) As a set H^+ of **positive definite Hermitian matrices** in $SL(2C)$, $H^+\subset SL(2C)$. Positive definite matrices are those which have all eigenvalues positive.

In the second representation above, points in the real 3-dimensional Lobachevskian space are identified with complex 2×2 positive definite Hermitian matrices (with unit determinant), $\tilde{\xi}=\begin{pmatrix}\xi_0-\xi_3 & \xi_2-i\xi_1\\ \xi_2+i\xi_1 & \xi_0+\xi_3\end{pmatrix}$, where ξ $(\xi_0,\xi_1,\xi_2,\xi_3)$ are projective (homogeneous) coordinates in Lobachevskian real $3d$-space normalized as $\det\tilde{\xi}=1=\xi_0^2-\xi_1^2-\xi_2^2-\xi_3^2$.

A one-to-one (bijective) correspondence between projective coordinates ξ and matrices in H^+ is given as:

$$\tilde{\xi}=\begin{pmatrix}\xi_0-\xi_3 & \xi_1-i\xi_2\\ \xi_1+i\xi_2 & \xi_0+\xi_3\end{pmatrix}=\xi_\alpha\sigma^\alpha \tag{29}$$

and

$$\xi_\alpha=\frac{1}{2}Tr\left(\tilde{\xi}\sigma^\alpha\right) \tag{30}$$

where $\sigma^0=e$ is the identity 2×2 matrix and σ^k, $k=1,2,3$ are Pauli matrices, and summation is over the same indexes.

Remark 7 *The advantage of working with 2×2 positive definite Hermitian matrices $\tilde{\xi}$, $\xi=u$ or $\xi=x$ of unit determinant is that: first, matrices $\tilde{u}$ and $\tilde{x}$ represent points u and x in real 3-dim Lobachevskian space L_U^3 and L_X^3 respectively. Projective* (*homogeneous*) *coordinates of points u or x are given explicitly via* Formula (*30*). *Matrices $\tilde{u}$ and $\tilde{x}$ are also viewed as Hermitian operators acting via isometries* (*Lorentz group, $SL(2C)$ actions*) *on a real 3-dim Lobachevskian space. We do not introduce any special notation to distinguish between the above two cases since the meaning of $\tilde{u}$ and $\tilde{x}$ is clear from the context.*

Before we show how to find a stabilizer for the Lobachevskian space $H^+\subset SL(2C)$ we need to define the group actions, called transformations in the physics literature. We describe two types of actions we use in this work.

1) The **left action**, also called a **left translation** of the Lorentz group on Lobachevskian space L_X^3 [10] [12].

$$h=gh',\quad h,h'\in L_X^3,\quad g\in SL(2C) \tag{31}$$

2) The **double sided** action, or **two sided translation**, (sometimes called a similarity transformation) on Lobachevskian space L_X^3 [10].

$$h=g^*h'g,\quad h,h'\in L_X^3, g\in SL(2C) \tag{32}$$

The star superscript denotes Hermitian conjugation $g\to g*=\overline{g}^T$.

Remark 8 *In the case of double sided action* (*32*), *we see that matrices* g *and* $-g$ *result in the same motion and have to be identified. This is done by taking the quotient* $SL(2C)/Z(e,-e)=PSL(2C)$, *resulting in the projective Lorentz group* $PSL(2C)$. *The center* (*of the* $SL(2C)$) Z *is also called the kernel of non-effectiveness. Since* Z *is discrete,* $SL(2C)\to PSL(2C)$ *is a covering map. Note that* $PSL(2C)$ *is isomorphic with the* (*proper*) *Lorentz group* $SO(1,3)$ *acting on a flat pseudo-Euclidean space. Keeping in mind that two matrices in* $SL(2C)$ *which differ only by a sign induce the same Lorentz transformation, we still will use* $SL(2C)$ *notation instead of* $PSL(2C)$ *for this case.*

Definition 9 *Motion due to action* $g\in SL(2C)$ *on a real 3-dim Lobachevskian space* L_ξ *given either by* Formula (*31*) *or* Formula (*32*), *sends a point* $\xi'\in L^3_\xi$ *having homogeneous coordinates* $\xi'_0,\xi'_1,\xi'_2,\xi'_3$ *represented by matrix* $\tilde{\xi}'$ *onto a point* $\xi\in L^3_\xi$ *having homogeneous* (*projective*) *coordinates* ξ_0,ξ_1,ξ_2,ξ_3 *represented by matrix* $\tilde{\xi}$.

Now it is easy to find a stabilizer. Take a unit matrix $e\in H^+$ as the "center" of Lobachevskian space, $e=\begin{pmatrix}1&0\\0&1\end{pmatrix}$ and apply the Lorentz group motion (Lorentz transformation) to it. From the definition of a stabilizer, we have:

$$e=g^*eg=g^*g \tag{33}$$

which says that matrix g is **unitary**. Thus we have the following conclusion:

Conclusion 10 *The stabilizer of the "origin" e in Lobachevskian* 3*-dim real space is a group of unitary* 2×2 *matrices* $SU(2)$, *or any conjugates to it in* $SL(2C)$ *if another "origin" than e has been selected.*

Thus we arrive at another definition of Lobachevskian space:

Definition 11 *Real* 3*-dim Lobachevskian space* H^+ *is isomorphic with the* ***coset space*** *of the Lorentz group* $SL(2C)$ *with respect to* $SU(2)$. *The group* $SL(2C)$ *is the group of rigid motions* (*isometries*) *of* H^+.

$$SL(2C)\supset H^+\simeq SL(2C)/SU(2) \tag{34}$$

where $\simeq$ denotes isomorphism which we will later, with some abuse of notation, denote with an equal sign.

In this representation, Lobachevskian space can be viewed as covered by (collapsed to a point) copies of $SU(2)$ and translated over the entire Lobachevskian space H^+ by the action of a group $SL(2C)$. Note that $SU(2)$ is locally homeomorphic to the rotation group $SO(3)$.

Further in this paper we use two physical representations of Lobachevskian real 3-dim space, either expressed in projective coordinates or matrix coordinates. We call one the **coordinate or position Lobachevskian space**, which we denote as L^3_X, and the other one the **Lobachevskian velocity/momentum space**, which we denotes as L^3_U. The geometry of both is identical, and **with respect to EM fields, both representations will cause the same effects**. For instance, the redshift in EM spectra may result either from L^3_X (in which case it is called cosmological redshift), and/or from L^3_U (in which case it is called the

Doppler shift), or from both at the same time; see also Section 4.2.4. Therefore, we discuss these two representations of Lobachevskian space in more detail 3.1 and 3.2 below.

At the end of this section, the reader is encouraged to remember that:

1) The Lorentz group is the group $SL(2C)$.

2) Lobachevskian (hyperbolic) real 3-dim space is a **coset space** of Lorentz $SL(2C)$ group with respect to $SU(2)$. $L^3 = SL(2C)/SU(2)$.

3) The Lorentz group acts on its own homogeneous space $SL(2C)/SU(2)$ via isometries (31) and (32), also called rigid motions.

4) In physics, the action of a Lorentz group is called a Lorentz transformation.

3.1. Representation of Lobachevskian Geometry as a 3-Dim Position Space $L_X^3 \sim SL(2C)/SU(2)$

In the case of Lobachevskian position space L_X^3, $\xi = x$, we will represent points in 3-dim Lobachevskian position space either via projective (homogeneous) coordinates x, $x_0^2 - x_1^2 - x_2^2 - x_3^2 = 1$, $x_0 \geq 1$, or via 2×2 **positive definite Hermitian matrices** $\tilde{x}$, with a unit determinant, assembled from homogeneous coordinates.

$$\tilde{x} = \begin{pmatrix} x_0 - x_3 & x_1 - ix_2 \\ x_1 + ix_2 & x_0 + x_3 \end{pmatrix}, \quad \det \tilde{x} = 1, \quad x_0 \geq 1 \tag{35}$$

There are several equivalent definitions of positive definite Hermitian matrices. We adopt the one in which a positive definite (Hermitian) matrix is a one with all positive eigenvalues. Thus, we have the following isomorphism between real Lobachevskian 3-dim space L^3, the set of positive definite Hermitian matrices $H \subset Sl(2C)$, and the quotient space $SL(2C)/SU(2)$ [10]:

$$L_X^3 = SL(2C)/SU(2) \tag{36}$$

Equation $\det \tilde{x} = x^2 = 1$, $x_0 \geq 1$ is the equation of the upper sheet of a two sheet hyperboloid, which in turn is the generic model of a real 3-dim Lobachevskian space. It follows that **transformations executed by the Lorentz group**, due to conditions $x^2 = invariant$, **are restricted to the Lobachevskian space** $x^2 = constant > 0$ (constant is usually set to unity—a common choice in curvature normalization).

In a projective representation (Beltrami - Klein ball model), Lobachevskian space is realized on the hyperplane $x_0 = 1$; see **Figure 5**. The hyperplane $x_0 = 1$ intersects the cone $x^2 = 0$, along the Euclidean sphere $S^2(0,r)$ which represents the boundary at infinity $\partial L_X^3(\infty)$, for L_X^3 having Euclidean radius $r = \tanh(\infty) = 1_X$. In this realization, Lobachevskian space is viewed as the interior of a 3-dim ball $B^3(0,1_X)$ with its boundary at infinity represented by the 2-dimensional sphere $S^2(0,1_X)$.

The reader should note that the real (un-normalized) value of $\mathbf{1}_X$ (per astrophysical evidence) when compared with our geometric experience is **incredibly high**. It follows that on **local distances**, in all means, we experience Loba-

chevskian 3-dim space as Euclidean 3-dim space. "All means" above should be understood as: our experimental techniques are unable to detect any departures from theorems of Euclidean geometry (e.g. the Pythagorean theorem) at distances up to about 10^6 light years - ten times the diameter of Milky Way (~10^5 light years). Moving from a local to a global point of view, the entire Lobachevskian space L_X^3 is assembled from 3-dimensional Euclidean "patches" glued together with a mathematical object called a **connection**; however, we will not need to use this construction in the text.

In homogeneous spaces, loosely speaking, the neighborhood of any point $x \in L_X^3$ "looks the same" as the neighborhood of any other point. In the other words **all points of** L^3 **are equivalent**, and are related by the equivalence relation $\xi R \xi'$, executed by Lorentz $SL(2C)$ group of motion $\xi = g\xi'$ or $\xi = g^* \xi' g$.

We say that the homogeneous space L^3 consists **only of one orbit of any arbitrary point**. It follows that physical processes in homogeneous spaces **do not depend on a particular location** since all points and their neighborhoods are "the same", or more precisely, equivalent. This is interpreted in physics as a postulate of no preferred frame. Homogeneous space(s) are very convenient since any process (properties) can be studied in the neighborhood of its neutral element (the "center" of homogeneous space) and then translated to any location by the group action.

In our discussion, in order to connect to experimental data, distances in Lobachevskian space L_X^3 need to be mapped onto distances in the relevant model of L_X^3 in Euclidean space. There are infinitely many such maps. We will use **two maps** which map **internal Lobachevskian distances** l (and functions of those distances, e.g. kinetic energy) from Lobachevskian position space L_X^3 onto **distances** d **or** δ **in the Euclidean space** E. Note that Lobachevskian distance l is the **only two-point invariant** in Lobachevskian space with respect to Lorentz group $SL(2C)$ action.

In the **Beltrami - Klein** model, distances d and l are related as:

$$d = \tanh l \tag{37}$$

In the **Poincare** model, distances δ and l are related as:

$$\delta = \tanh \frac{l}{2} \tag{38}$$

A prompt calculation of the relation between Euclidean distances (Euclidean images) d and δ (of Lobachevskian (hyperbolic) distance l) from Formulas (37) and (38) gives the following equation:

$$d = \frac{2\delta}{1+\delta^2} \tag{39}$$

Formula (39) above represents a well known, explicit **isomorphism** between the **Poincare and Beltrami-Klein models** of Lobachevskian geometry [12]. All of the above also applies to the representation of Lobachevskian geometry via position space L_X^3, or to the representation of Lobachevskian space via velocity

space L_U^3, which we discuss next.

3.2. Representation of Lobachevskian Geometry as a 3-Dim Velocity/Momentum Space $L_U^3 = SL(2C)/SU(2)$

The geometric properties of Lobachevskian real, 3-dim velocity/momentum space are precisely the same as the geometric properties of Lobachevskian 3-dim position space discussed above. However we elaborate on it a bit more.

Velocity space is regarded as a 3-dim real Lobachevskian (hyperbolic) space L_U^3. The Gaussian (negative) curvature has a value of $K = -c^{-2}$, where the constant $c > 0$, $c \in R$, is regarded as the velocity of light in a vacuum. The generic value of K is set to -1, which translates to a choice of physical units in which $c = 1$. The **signed Lobachevskian distance** θ between two points $p_1, p_2 \in L_U^3$, $\pm\theta(p_1, p_2)$ is interpreted as **relative** Lobachevskian velocity θ. Note that θ is in fact dimensionless. This is because the actual distance is $\theta\sqrt{-K}$, but since we set $K = -1$, the curvature K is not present numerically. Curvature is of dimension inverse-squared-length, which makes the product $\theta\sqrt{-K}$ a pure real number with no physical label. Note also that the distance $|\theta|$, the relative velocity, is the **only invariant between two points** (two frames in relative uniform motion), with the respect to $SL(2C)$ isometries in Lobachevskian space.

In a typical experimental High Energy Physics (HEP) experimental situation Lobachevskian velocities have to be mapped onto Euclidean velocities in the Euclidean space we live in. Depending on the choice of the map, the relation between Euclidean (relative) velocities v, the velocities we measure experimentally υ, and the internal Lobachevskian (hyperbolic) relative velocity θ, are given by Formulas (40) and (41) below:

$$\mathrm{v} = \tanh\theta \tag{40}$$

due to the Beltrami-Klein model (shown in Section 4 to correspond to ESR), and

$$\upsilon = \tanh\frac{\theta}{2} \tag{41}$$

due to the Poincare model (shown in Section 5 to correspond to BSR). The isomorphism between the two maps (40) and (41) (the Beltrami-Klein and Poincare models) is given by:

$$\mathrm{v} = \frac{2\upsilon}{1+\upsilon^2} \tag{42}$$

It is straightforward to see that in the case of Lobachevskian velocity/momentum space, L_U^3, the isomorphism between the Beltrami-Klein and the Poincare models of Lobachevskian geometry is related to the homotopy between a gnomonic and a stereographic map discussed in Section 2.

The internal Lobachevskian relative velocity θ can take any real values $\theta \in [0, \infty)$. In physics, maps (40) and (41), $\theta \to \mathrm{v}$, or $\theta \to \upsilon$ **map the non-compact half line** $[0, \infty)$ **onto segment** $[0,1)$. In applications to physics,

in order to incorporate velocities of photons at the boundary at infinity, $\partial L_U^3(\infty)$ is added to Lobachevskian L_U^3 space, resulting in the so called extended Lobachevskian space $L_U^3 \bigcup \partial L_U^3$. This way, $0 \to 0$, and $\infty \to c = 1$. Relations $\theta + \infty = \infty$ and $\infty + \infty = \infty$, under maps (40) and/or (41) become $v + c = c$, $\upsilon + c = c$, and $c + c = c$, or symbolically "1+1=1", which says that **the velocity of light does not depend on the state of motion of a source** (Michelson-Morley experiment), and that speed of light is the maximum speed possible.

Points u in Lobachevskian 3-dim velocity space are represented by 2×2 complex Hermitian positive definite matrices, $\tilde{u} = \begin{pmatrix} u_0 - u_3 & u_1 - iu_2 \\ u_1 + iu_2 & u_0 + u_3 \end{pmatrix}$. The set of complex 2×2 positive definite Hermitian matrices is **not a group** since the product of two such a matrices not need to be a positive definite matrix.

Note that the determinant of matrix $\tilde{u}$ is equal to 1, $\det \tilde{u} = 1$, which means that as a topological space, the Lobachevskian velocity space, in homogeneous coordinates $u(u_1, u_2, u_3, u_4)$, is modeled in a unit ($c = 1$) hyperboloid $[u,u] = u_0^2 - u_1^2 - u_2^2 - u_3^2 = 1$, $u_0 \geq 1$.

Remark 12 *Since* $[u,u] = 1$ *is equivalent to* $\det \tilde{u} = 1$, ***only three components of velocity*** u ***are independent and consequently there is no such thing in nature as*** 4***-velocity. Representation of*** u ***via four homogeneous (projective) coordinates*** u_α, $\alpha = 0,1,2,3$ ***is merely a mathematical tool of convenience***.

The representation of velocity space as a Lobachevskian space has the following properties:

1) Points at a **finite** Lobachevskian distance, $\theta < \infty$, from any internal point $u \in L_U^3$ represent velocities of massive $m_0 > 0$ particles.

2) Points at an **infinite** Lobachevskian distance, $\theta = \infty$, from any internal point $u \in L_U^3$ belong to the boundary at infinity $\partial L_U^3(\infty)$. They are interpreted as velocities of massless $m_0 = 0$ particles - photons (and perhaps neutrinos).

3) Space beyond the boundary at infinity is called imaginary Lobachevskian space. Its model is the one-sheet hyperboloid. Points beyond the boundary at infinity $\partial L_U^3(\infty)$ are thought as velocities of (so far) hypothetical particles - tachions. Distances in imaginary Lobachevskian space need not be real numbers.

At the end of this section, the reader is encouraged to remember that:

1) Lorentz group is the group $SL(2C)$

2) Lobachevskian (hyperbolic) real 3 dim space is a **coset space** of Lorentz $SL(2C)$ group with respect to $SU(2)$. $L^3 = SL(2C)/SU(2)$.

3) Lobachevskian space $SL(2C)/SU(2)$, in our particular discussion, has two isomorphic representations: one, as a coordinate position space $L_X^3 = SL(2C)/SU(2)$, and two, as a velocity/momentum space $L_{V(P)}^3 = SL(2C)/SU(2)$.

4) Lorentz group $SL(2C)$ acts via isometries given by Formula (31) or (32) on either representation of L^3.

4. "Special Relativities" as Maps from Lobachevskian Space into Euclidean Space

We have arrived at the point when we are ready to present "special relativities" and their true meaning as parametric maps. However, before we present our version of "special relativity", BSR, it will be beneficial for the reader to first see how the above mathematical exposition produces Einstein's "special relativity", ESR. Since the methodology applied to obtain either ESR or BSR is identical in both cases, the reader should gain confidence that the $SL(2C)$ group action approach works and produces viable physics. So this section on ESR may be regarded as mini "warm up" before the next section on BSR.

4.1. Einstein's "Special Relativity" (ESR) as a Gnomonic Map

In this section, we present how we understand Einstein's "special relativity" (ESR) as a kind of a map, namely a gnonomic map. The reader will have the opportunity to compare our exposition of Einstein's ESR with an orthodox treatment via Minkowski flat 4-geometry as is commonly seen in the literature and to make his/her own conclusions.

Einstein's "special relativity" results from the left-action of the Lorentz Group $SL(2C)$ on the Lobachevskian coordinate space $L_X^3 = SL(2C)/SU(2)$:

$$h = gh',\ h,h' \in L_X^3,\ g \in SL(2C) \tag{43}$$

Remark 13 *The action* (*43*) *has simple meaning for experimental high energy physics and astrophysics. It describes a* ***single sided motion*** *typical for experiments on accelerators with fixed target, e.g. Stanford's Linear Accelerator.*

To extract some physics of interest from (43), we proceed as follows. The positive definite 2×2 Hermitian matrices $\tilde{x},\tilde{x}'$ represent points in Lobachevskian space $L_X^3 = SL(2C)/SU(2)$, and the $SL(2C)$ matrix $g = \tilde{u} = \begin{pmatrix} u_0 - u_3 & u_2 - iu_1 \\ u_2 + iu_1 & u_0 + u_3 \end{pmatrix}$, viewed as a Hermitian operator, represents the **motion operator**, which sends the point $\tilde{x}'$ onto point $\tilde{x}$. Therefore we come to the following matrix equation:

$$\tilde{x} = \tilde{u}\tilde{x}',\ \tilde{x},\tilde{x}' \in L_X^3 = SL(2C)/SU(2),\ \tilde{u} \in SL(2C), \tag{44}$$

The simple matrix equation $\tilde{x} = \tilde{u}\tilde{x}'$ (44) **contains the entirety of Einstein's "special relativity"** in the sense that all of ESR can be reproduced from it. Mathematically it is a left translation (**isometry**) on the coordinate Lobachevskian space executed by the **motion operator** $\tilde{u}$. From the point of view of physics, it relates two points $\tilde{x}'$ and $\tilde{x}$ in Lobachevskian position space (hyperboloid) which are in relative uniform motion represented by the operator $\tilde{u}$ acting on Lobachevskian space - position hyperboloid $\det\tilde{x} = 1 = x^2$. The coordinates of both points x and x' are projective (homogeneous) coordinates in Lobachevskian position space. That is the essence of ESR.

For reasons of simplicity alone, we represent 2×2 matrices $\tilde{u}$ in diagonal form, which from the point of view of physics, means that the velocity u is con-

fined to the x_0, x_3 plane and consequently has only two u_0, u_3 homogeneous components. In non-homogeneous affine coordinates $\frac{u_3}{u_0} = \mathrm{v}_3$, relative velocity has only one component v_3.

Remark 14 *Note that in the one-dimensional model we are analyzing, the complex Lorentz group* $SL(2C)$ *reduces to a group of real diagonal matrices, meaning the subgroup of* $SL(2R)$ *—the Lorentz group of real* 2×2 *matrices with unit determinant. Note also that in the diagonal representation the matrix operator is in fact expressed by its eigenvalues, which we will calculate explicitly in Section* 4.1.3.

Equation (44) now yields:

$$\begin{pmatrix} x_0 - x_3 & 0 \\ 0 & x_0 + x_3 \end{pmatrix} = \begin{pmatrix} u_0 - u_3 & 0 \\ 0 & u_0 + u_3 \end{pmatrix} \begin{pmatrix} x_0' - x_3' & 0 \\ 0 & x_0' + x_3' \end{pmatrix} \tag{45}$$

We can obtain the transformed x_0 coordinate from the **trace property**, observing that for diagonal matrices a, b, c, if $a = bc \Rightarrow Tra = Tr(bc)$. We obtain the transformed coordinate x_3 by comparing the appropriate matrix elements in both sides of (45) and then subtracting them. This gives:

$$x_0 = u_0 x_0' + u_3 x_3' \tag{46}$$

$$x_3 = u_3 x_0' + u_0 x_3' \tag{47}$$

Or in the matrix form:

$$\begin{pmatrix} x_0 \\ x_3 \end{pmatrix} = \begin{pmatrix} u_0 & u_3 \\ u_3 & u_0 \end{pmatrix} \begin{pmatrix} x_0' \\ x_3' \end{pmatrix} \tag{48}$$

Remark 15 *For* $U = \begin{pmatrix} u_0 & u_3 \\ u_3 & u_0 \end{pmatrix}$, $U^{-1} = \begin{pmatrix} u_0 & -u_3 \\ -u_3 & u_0 \end{pmatrix}$. *Thus if* $x = Ux'$, *then* $U^{-1}x = x'$. *From the point of view of physics, the velocity sign is* ***reversed*** *in the inverse matrix* U^{-1}. *This obviously means that if the system A' is in motion with velocity v, with respect to the system A, then A is moving with velocity -v, with respect to A'.*

We recall that in Lobachevskian velocity space $u_0^2 - u_3^2 = 1 = u_0^2\left(1 - \frac{u_3^2}{u_0^2}\right) \Rightarrow$ $u_0 = \frac{1}{\sqrt{1-\mathrm{v}_3^2}}$. The non-homogeneous (local) coordinates are related to projective (homogeneous) coordinates as: $\mathrm{v}_3 = \frac{u_3}{u_0}$. Velocity is aligned along u_3, and in non-homogeneous coordinates relative velocity v has **only one component**.

If we regard the **projective coordinate** x_0 as "time" (measured in meters of a light-path), then Equations (46) and (47) can be rewritten in terms of a temporal x_0 and a spatial coordinate *x*.

Now we abandon subscripts, and keeping in mind that the normalization $c = 1$ affects only physical units and that $x_0 = ct$, we end up with the well know formulas from Einstein's "special relativity" for temporal and spatial coordinate transformation in systems being in uniform relative motion:

$$t = \frac{t' + \mathrm{v}x'}{\sqrt{1-\mathrm{v}^2}} \tag{49}$$

$$x = \frac{\mathrm{v}t' + x'}{\sqrt{1-\mathrm{v}^2}} \tag{50}$$

Equations (49) and (50) were introduced by Albert Einstein in 1905 in his work [13], however **due an entirely different form of reason from our own**.

In the Euclidean limit, Equations (49) and (50) are:

$$t = t' \tag{51}$$

$$x = \mathrm{v}t + x' \tag{52}$$

Equations (51) and (52) are known as Galilean transformations due to **single sided motion**. In physics they correspond to **fixed target scattering**, like in the Stanford Linear Accelerator.

Transformations (49) and (50) in compact form are:

$$\begin{pmatrix} t \\ x \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{1-\mathrm{v}^2}} & \frac{\mathrm{v}}{\sqrt{1-\mathrm{v}^2}} \\ \frac{\mathrm{v}}{\sqrt{1-\mathrm{v}^2}} & x\frac{1}{\sqrt{1-\mathrm{v}^2}} \end{pmatrix} \begin{pmatrix} t' \\ x' \end{pmatrix} \tag{53}$$

Here relative velocity v is in fractions of the velocity of light c. If c has to be present explicitly, then we need to re-scale: $\mathrm{v} \to \frac{\mathrm{v}}{c}$ and substitute ct for x_0.

Since the matrix in (53) is uni-modular, there is a unique parametrization of the matrix elements as:

$$\cosh\theta = \frac{1}{\sqrt{1-\mathrm{v}^2}}, \quad \sinh\theta = \frac{\mathrm{v}}{\sqrt{1-\mathrm{v}^2}} \tag{54}$$

$$u_0 = \frac{1}{\sqrt{1-\mathrm{v}^2}}, \quad u_i = \frac{\mathrm{v}_i}{\sqrt{1-\mathrm{v}^2}}, \quad i = 1,2,3, |\mathrm{v}| < 1 \tag{55}$$

The coordinates (55) are called **Weierstrass coordinates** due to the German mathematician Weierstrass [8]. Weierstrass coordinates were already discussed in Section 2 and the reader is encouraged to review. It is easy to see that Weierstrass homogeneous coordinates (55) are normalized to unity $u^2 = u_0^2 - u_1^2 - u_2^2 - u_0^2 = 1$. Weierstrass used coordinates (55) in his work on Lobachevskian (hyperbolic) geometry more than 50 years before ESR, however, Einstein never mentioned or referred to them.

From Formula (54) we conclude that the **internal Lobachevskian velocity** θ and **local Euclidean velocity v** measured in experiments are related by:

$$\tanh\theta = \mathrm{v} \tag{56}$$

which is the equation by which distances in Lobachevskian space are mapped onto Euclidean distances due to the Beltrami-Klein model, here representing Lobachevskian space by the space of velocities L_U^3. It should be noted that θ, the internal Lobachevskian velocity is incorrectly (in our opinion) called the "hyperbolic angle" or "rapidity" in the literature on high energy particle physics.

Conclusion 16 *Comparing* Formulas (*55*) *and* (*56*) *resulting from Lorentz transformations* (*in Einstein's* "*special relativity*") *with* Formulas (*14*) *and* (*17*) *which we derived from the* ***gnomonic (central) projection from hyperboloids****, we see that they are identical. This means* ***Einstein's "special relativity" is nothing more than a gnomonic map from Lobachevskian space into Euclidean space****. This is the true meaning of ESR.*

4.1.1. Addition of Velocities along the Same Direction

From the map (56) between Lobachevskian and Euclidean velocities we easily see that:

$$\tanh\left(\theta_1+\theta_2\right)=\frac{\tanh\theta_1+\tanh\theta_2}{1+\tan\theta_1\tanh\theta_2}=\frac{\mathrm{v}_1+\mathrm{v}_2}{1+\mathrm{v}_1\mathrm{v}_2} \tag{57}$$

Recall that θ in (57) is **not an angle**. It is a real non-negative number $\theta\in(0,\infty)$ representing the unit-less length of a Lobachevskian linear segment in 1-dim Lobachevskian space, *i.e.* a Lobachevskian line. A Lobachevskian line is a one dimensional, real, metric, non-compact space of constant negative curvature. An image of θ in Euclidean space is what is measured by physicists due to various maps; in this case due to the Beltrami-Klein map, $\mathrm{v}=\tanh\theta$. It seems that at the time he introduced ESR, Einstein was not familiar with the work of Lobachevski on non-Euclidean geometry; nevertheless he deduced the correct formula for addition of velocities.

4.1.2. Distortions

From Equations (49) and (50) we can obtain the distortions resulting from Einstein's ESR map. These distortions, due to historical reasons, are called **"time dilation"** and **"length contraction"**, or **"Fitzerald contraction"**.

$$\Delta t=\Delta t'\frac{1}{\sqrt{1-\mathrm{v}^2}}=\Delta t'\cosh\theta \tag{58}$$

$$\Delta x=\Delta x'\frac{1}{\sqrt{1-\mathrm{v}^2}}=\Delta x'\cosh\theta \tag{59}$$

From (58) and (59), we see that when the curvature approaches zero, $-K=c^{-2}\to 0$ equivalent to $\mathrm{v}<<c$ (Euclidean maps), the apparent distortions vanish and $\Delta t=\Delta t',\Delta x=\Delta x'$.

4.1.3. Physical Meaning of Diagonal Entries (Eigenvalues) of Motion Operator $\tilde{u}$

To see the physical meaning of the diagonal entries $\Lambda_1=u_0-u_3$ and $\Lambda_2=\Lambda_1^{-1}=u_0+u_3$ of the motion operator $\tilde{u}$ we need to express them in local coordinates $\frac{u}{u_o}$. Using Weierstrass coordinates for u yields:

$$\Lambda_1=\sqrt{\frac{1+\mathrm{v}}{1-\mathrm{v}}},\quad \Lambda_2=\Lambda_1^{-1}=\sqrt{\frac{1-\mathrm{v}}{1+\mathrm{v}}} \tag{60}$$

Thus the diagonal entries of the motion operator are easily recognized as **Lobachevski - Doppler** blue Λ_1 and red Λ_2 **frequency shifts** for a single sided

motion expressed in local coordinates v. Elements of $SL(2R)$ with reciprocal values like in (60) in the geometry of the Mobius group are called hyperbolic transformations.

4.2. Lobachevski - Poincare - Von Brzeski "Special Relativity" (BSR) as a Stereographic Map

In this section, we show that our version of special relativity, *i.e.* BSR, is a **stereographic projection from Lobachevskian space (a hyperboloid)** discussed in Section 2.2.2. A stereographic projection from a hyperboloid results in the Poincare representation of Lobachevskian geometry in the unit ball, or in our case, in the unit disc.

As we discussed in Section 3, the Lorentz group may act on Lobachevskian space L_X^3 in several ways. Our choice is the action (33) via the automorphism $h = g^* h' g$ of L_X^3 executed by a **double sided motion** (double sided translation), which we now compute in detail.

Remark 17 *The action* (*32*) *is of direct and fundamental significance for experimental high energy physics. It represents a typical case of* ***center of momentum scattering*** *of identical particles in accelerators with counter rotating beams*; *see* Figure 6.

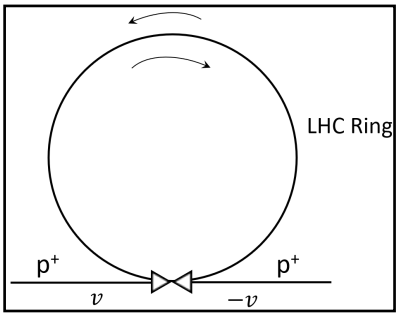


Figure 6. Two protons with opposite velocities v and $-v$ collide in the LHC ring.

The point $\tilde{x}$ in Lobachevskian space L_X^3 is represented by a uni-modular positive definite Hermitian matrix

$$\tilde{x} = \begin{pmatrix} x_0 - x_3 & x_2 - ix_1 \\ x_2 + ix_1 & x_0 + x_3 \end{pmatrix} \tag{61}$$

$\tilde{x} \in L^3 = H^+$ and for matrix $g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ in (32) we take, as in Section 3, the Hermitian motion operator matrix:

$$\tilde{u} = \begin{pmatrix} u_0 - u_3 & u_2 - iu_1 \\ u_2 + iu_1 & u_0 + u_3 \end{pmatrix} \tag{62}$$

We see that $\det \tilde{u} = u^2 = u_0^2 - u_1^2 - u_2^2 - u_3^2 = 1$, which tells us that if $u_0 \geq 1$, the point u belongs to the upper sheet the of hyperboloid $u^2 = 1$.

Substituting in $h = \tilde{x}$, $h' = \tilde{x}'$, and $g = \tilde{u}'$ into Equation (32) yields:

$$\tilde{x} = \tilde{u}^* \tilde{x}' \tilde{u} \tag{63}$$

As we have already said, matrices $\tilde{u}$ and $-\tilde{u}$ correspond to the same

transformation and are liable for identification.

The action (63) leaves the **quadratic form** $Q(x,x)=Q(x',x')=\det\tilde{x}'=\det\tilde{x}=x^2=x'^2=1$ **invariant**, which is due to the **invariance of the matrix determinant** $\det\tilde{x}$ under transformations (63). Geometrically, this means that the hyperboloid $x'^2=1$ is invariant (is mapped onto itself) under all motions (63). This simple matrix Equation (63), $\tilde{x}=\tilde{u}^*\tilde{x}'\tilde{u}$, contains the entirety of the authors' "special relativity" BSR, in the sense that BSR can be derived from it.

For the purposes of the present work, just for simplicity, we will limit ourselves to one dimension. One dimensional Lobachevskian space or the **Lobachevskian line** has **two homogeneous** (projective) coordinates x_0 and x_3. The Lobachevskian line is aligned along x_3. Equation (63) in the one dimensional case is as follows. We write the coordinates $x=x(x')$, due to motion (63) in L_X^1 executed by $g\in SL(2C)$ ***i.e.*** **the Lorentz transformations**. In one dimensional space, the matrices (61) and (62) become diagonal which greatly simplifies the calculations.

Since in the one dimensional case, velocity u and displacement x are aligned along the same line, the matrices (61) and (62) now are **real diagonal** and hence **commutative**.

The one dimensional motion operator is now:

$$\tilde{u}=\begin{pmatrix}u_0-u_3 & 0\\ 0 & u_0+u_3\end{pmatrix},\quad \tilde{u}\in SL(2C) \tag{64}$$

acting on one dimensional Lobachevskian space L_X^1:

$$\tilde{x}=\begin{pmatrix}x_0-x_3 & 0\\ 0 & x_0+x_3\end{pmatrix},\quad \tilde{x}\in SL(2C)/SU(2) \tag{65}$$

according to the double sided action as:

$$\tilde{x}=\begin{pmatrix}u_0-u_3 & 0\\ 0 & u_0+u_3\end{pmatrix}\begin{pmatrix}x_0'-x_3' & 0\\ 0 & x_0'+x_3'\end{pmatrix}\begin{pmatrix}u_0-u_3 & 0\\ 0 & u_0+u_3\end{pmatrix} \tag{66}$$

We can compute the double sided action (66) instantly by observing that for 2×2 diagonal matrices a,b,c,d, if $a=bcd$ then $Tr(a)=Tr(bcd)$, and from (66) we find that:

$$x_0=\left(u_0^2+u_3^2\right)x_0'+2u_0u_3x_3' \tag{67}$$

$$x_3=2u_0u_3x_0'+\left(u_0^2+u_3^2\right)x_3' \tag{68}$$

We write transformations (67) and (68) in a slightly different form by converting homogeneous u velocity into affine υ velocity which we measure in experiments. Recall that in homogeneous coordinates in Lobachevskian space $u^2=1$, and $1=u_0^2\left(1-\upsilon\frac{u_3^2}{u_0^2}\right)$. In homogeneous coordinates υ_3 is: $\upsilon_3=\frac{u_3}{u_0}$. It follows that,

$$u_0^2+u_3^2=\frac{1+\upsilon_3^2}{1-\upsilon_3^2} \tag{69}$$

$$u_0 u_3 = \frac{\upsilon_3}{1-\upsilon_3^2} \tag{70}$$

Therefore, **the transformations of homogeneous coordinates on a Lobachevskian line due to two sided translations are**:

$$x_0 = \frac{\left(1+\upsilon^2\right)x_0' + 2\upsilon x_3'}{1-\upsilon^2} \tag{71}$$

$$x_3 = \frac{2\upsilon x_0' + \left(1+\upsilon^2\right)x_3'}{1-\upsilon^2} \tag{72}$$

Therefore, if we assume that the **projective zero coordinate** x_0 in (71) and (72) is **understood as "time"** ct (measured in meters of a light-path) then taking $c=1$ we can relabel $x_0 = t$ and $x_0' = t'$, we arrive at the following equations:

$$t = \frac{\left(1+\upsilon^2\right)t' + 2\upsilon x'}{1-\upsilon^2} \tag{73}$$

$$x = \frac{2\upsilon t' + \left(1+\upsilon^2\right)x'}{1-\upsilon^2} \tag{74}$$

Equations (73) and (74) are authors' **transformations for temporal** t **and spatial coordinates** x **in systems being in relative uniform motion**. The Equations (73) and (74) of the authors' "special relativity" (BSR) are an alternative to Einstein's transformations (49) and (50) in his ESR.

The reader should note that from the point of view of physics, our transformations correspond to the **center of momentum frame**, which is the bread and butter of particle scattering experiments for identical particles in circular accelerators with counter-rotating beams. The reader should also be aware that in (73) and (74), the velocity $|\upsilon| \in [0,1)$ is dimensionless, and that the speed of light factor $c=1$ is not present explicitly and it only affects physical units. To get (73) and (74) in common physical units, substitute $\upsilon \to \frac{\upsilon}{c}$ and $x_0 \to ct$.

In the **Euclidean limit**, transformations (73) and (74) yield:

$$t = t' \tag{75}$$

$$x = 2\upsilon t + x' \tag{76}$$

Remark 18 *We'd like to explain the term "Euclidean limit" used above to linearize (74) and (75). In common MKS units, for example, the term* $2\upsilon x$ *in the numerator of (74) is* $2\frac{\upsilon}{c}\frac{x}{c}$ *and it is on the order of* c^{-2}. *We note that the Gaussian curvature of Lobachevskian velocity space is* $K = -c^{-2}$. *Therefore, we are* ***neglecting the curvature*** *of Lobachevskian space. Neglecting the curvature of Lobachevskian geometry implies that locally (in local coordinates* υ*), the geometry can be approximated by Euclidean geometry to an arbitrary degree of precision* [13]. *In other words, we use the fact that around any of its points, velocity space is approximately flat (Euclidean). The procedure itself is called* ***linearization*** *or* ***Euclideanization*** *of physics modeled on curved, non-Euclidean,*

geometries.

Transformations (73) and (74) can be written in a compact form:

$$\begin{pmatrix} t \\ x \end{pmatrix} = \begin{pmatrix} \frac{1+\upsilon^2}{1-\upsilon^2} & \frac{2\upsilon}{1-\upsilon^2} \\ \frac{2\upsilon}{1-\upsilon^2} & \frac{1+\upsilon^2}{1-\upsilon^2} \end{pmatrix} \begin{pmatrix} t' \\ x' \end{pmatrix} \tag{77}$$

A quick look at entries of a matrix (77) tells us that in Lobachevskian velocity space L_U^3, we use projective (homogeneous) coordinates $u(u_0,u_1,u_2,u_3)$ of the **form already introduced in Section 2 for stereographic projection from Lobachevskian space**.

$$u_0 = \frac{1+\upsilon^2}{1-\upsilon^2},\ u_i = \frac{2\upsilon}{1-\upsilon^2},\ i=1,2,3,\ |\upsilon|<1 \tag{78}$$

normalized such that $u^2 = u_0^2 - u_1^2 - u_2^2 - u_3^2 = 1$. Note that our projective (homogeneous) coordinates are different from Weierstrass projective (homogeneous) coordinates we encountered in Einstein's ESR, which should be no surprise since ours and Einstein's "relativity" refer to different models of Lobachevskian velocity space, namely Poincare and Beltrami-Klein, respectively.

Since the matrix in (77) is **unimodular**, $\det(.)=1$, there is a unique real parameter θ such that:

$$\frac{1+\upsilon^2}{1-\upsilon^2} = \cosh\theta = u_0 \tag{79}$$

and

$$\frac{2\upsilon}{1-\upsilon^2} = \sinh\theta = u \tag{80}$$

From (79) and (80) we find that:

$$\tanh\theta = \frac{2\upsilon}{1+\upsilon^2} \tag{81}$$

which is the **isomorphism between the Poincare and the Beltrami-Klein models of Lobachevskian geometry**, in (27) and (28), we mentioned in Section 2.2.2.

Since $\tanh\frac{\theta}{2} = \left(\frac{\cosh\theta-1}{\cosh\theta+1}\right)^{\frac{1}{2}}$, using (79) and (80) we can find υ versus θ:

$$\upsilon = \tanh\frac{\theta}{2} \tag{82}$$

Formula (82) is the map of **Lobachevskian velocity θ onto local Euclidean velocity υ due to the Poincare model** of Lobachevskian geometry. The parameter θ in (82) is the **intrinsic Lobachevskian, dimensionless, relative velocity** $|\theta| \in [0,+\infty)$. In particle physics it is often (in our view) incorrectly called the "velocity parameter" or "rapidity". Velocity $|\upsilon| \in [0,1)$ is the image of Lobachevskian velocity θ, as it appears to us, in Euclidean space. Note that θ is dimensionless because the Gaussian curvature of Lobachevskian velocity space is set to $K=-1$. The function $\tanh(.)$ relates Lobachevskian and Euclidean dis-

tances. The value $\theta=\infty$ and $\upsilon=1$ is restricted to photons (and perhaps to neutrinos) which are points at the boundary at infinity for Lobachevskian velocity space.

Comparing BSR Formulas (79) and (80) resulting from the $SL(2C)$ **Lorentz group action** (Lorentz transformations) to Formulas (19) and (20) resulting from Lobachevskian cartography in the case of a **stereographic projection from hyperboloids**, we see that seemingly unrelated areas of science, cartography and "special relativity" are in fact two sides of the same coin. Both are just **certain maps** of infinitely many possible maps (as we will see below).

Conclusion 19 *The image of the double sided action* (*63*) *of the Lorentz group* $SL(2C)$, *or the image of the double sided motion, on a real 3-dim Lobachevskian space* $SL(2C)/SU(2)$ (*viewed as a hyperboloid* $x^2=1$), *is isomorphic with the Poincare ball model representation of Lobachevskian space resulting from a* ***stereographic*** *projection of Lobachevskian space* (*hyperboloid*) *into Euclidean space*; *see* ***Figure 5***.

4.2.1. Addition of Velocities along the Same Direction

In accordance with (82) we have the following equation:

$$\tanh\left(\frac{\theta_1}{2}+\frac{\theta_2}{2}\right)=\frac{\tanh\frac{\theta}{2}+\tanh\frac{\theta}{2}}{1+\tanh\frac{\theta}{2}\tanh\frac{\theta}{2}}=\frac{\upsilon_1+\upsilon_2}{1+\upsilon_1\upsilon_2} \tag{83}$$

which is the same as in Einstein's ESR. This is a **model independent** result. Recall that θ in (83) is **not an angle**. It is a real non-negative number $\theta\in(0,\infty)$ representing a unit-less length of a Lobachevskian linear segment in one dimensional Lobachevskian space, *i.e.* a Lobachevskian line. As we mentioned above, a Lobachevskian line is one dimensional, real, metric, non-compact space of constant negative curvature. An image of θ in Euclidean space is measured by physicists due to various maps. In this case due to the Poincare map: $\upsilon=\tanh\frac{\theta}{2}$.

4.2.2. Distortions of Maps in BSR and Their Geometric Meaning

From the transformations laws or maps (73) and (74) it is easy to see that the apparent **distortions** for spatial and temporal intervals (for a two sided uniform motion, *i.e.* at fixed distance in Lobachevskian velocity space) in BSR, given by the diagonal terms of matrix (77), will be:

$$\Delta t=\Delta t'\frac{1+\upsilon^2}{1-\upsilon^2}, or\ \frac{\Delta t}{\Delta t'}=\cosh\theta \tag{84}$$

$$\Delta x=\Delta x'\frac{1+\upsilon^2}{1-\upsilon^2}, or\ \frac{\Delta x}{\Delta x'}=\cosh\theta \tag{85}$$

Equations (84) and (85) are obviously distortions of a **stereographic projection** from a hyperboloid into in the Euclidean plane seen as the Poincare disc model.

The reader should note that the dimensionless velocity υ in (84) and (85) in

the MKS system of units will be $\frac{\upsilon^2}{c^2} = \upsilon^2 c^{-2} = \upsilon^2 |K|$ where $-K = c^{-2}$ is the Gaussian curvature of Lobachevskian velocity space. Evidently, as $K \to 0$, and velocity space becomes Euclideanized, the distortions (84) and (85) of images under map (73) and (74) vanish and $\Delta t = \Delta t', \Delta x = \Delta x''$, meaning this is a case of a mapping between two flat spaces. Maps in such cases will be distortion-less. No apparent "time dilation" and no apparent "length contraction" will exist. In the literature, such maps are referred to as Galilean.

4.2.3. Physical Meaning of Diagonal Entries (Eigenvalues) of Motion Operator $\tilde{u}$

It is easy to see that the diagonal entries of the double sided motion operator are:

$$\Lambda_1 = \frac{1+\upsilon}{1-\upsilon} \tag{86}$$

$$\Lambda_2 = \frac{1-\upsilon}{1+\upsilon} \tag{87}$$

These are Lobachevski-Doppler blue (86) and red (87) shifts for two **sided motion**. Examples are:

1) A frequency shift from the reflection from a moving mirror recorded by a transceiver when both, transceiver and mirror (target) are in motion (e.g. when radar and target are both in flight).

2) Frequency shifts recorded in reflective telescopes.

3) Experiments in circular accelerators with counter rotating beams, see **Figure 6**.

4.2.4. Practical Applications to Astrophysics of Formulas for Frequency Shifts in Beltrami-Klein (ESR) and in Poincare (BSR) Models

Frequency shifts resulting from Lobachevskian geometry are given either by Formulas (60) in Beltrami-Klein representation of Lobachevskian geometry (ESR), or by Formulas (86) and (87) in the Poincare representation of Lobachevskian geometry (BSR). These formulas apply either to a Lobachevskian large scale vacuum resulting in cosmological redshift, or to a Lobachevskian velocity space resulting in known Doppler shift. Recall that both position and velocity Lobachevskian spaces are metric spaces. The signed distance in Lobachevskian velocity space is called relative velocity.

Since the formulas for frequency shifts are widely used in astrophysics, radar techniques, warfare, medicine, and in nuclear physics (e.g. Mossbauer effect), and since the numbers representing spectral shifts derived from those models differ, it is **very relevant** which "special relativity" Formulas (ESR or BSR) will be used to draw conclusions about the physics represented by the measured frequency shifts z. **Figure 7** shows the uncertainties in conclusions due to ESR and BSR in a typical astrophysics scenario. We illustrate it via simplified three examples below.

Suppose, for instance, that the measured redshift is $z = 2$ and let us analyze what this real number 2 tells us about physics.

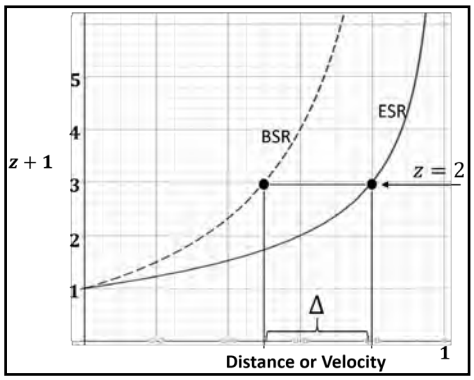


Figure 7. In the figure above, the dashed line shows the spectral shift z due to BSR, and the solid line shows the spectral shift due to ESR. The horizontal axis is calibrated in units of distance in position space, $R=1=15$ billion light-years, in which case $\Delta=\Delta x$ refers to the difference in position due to BSR and ESR, or in units of distance in velocity space, $c=1=3\times10^8$ m/s, in which case $\Delta=\Delta u$ refers to the difference in velocity due to BSR and ESR.

- Example 1. Let us assume that the measured redshift $z=2$ from some source is a result of the **distance in position space**, meaning a negatively curved Lobachevskian Universe and there is no relative motion of the source with respect to observer. In this case, assuming arbitrarily that the radius of the Lobachevskian Universe regarded as an interior of Euclidean ball is say 15 billion light years (bly), normalized to $R=1$, we see in **Figure 7**, that applying ESR we obtain the distance to the source of 7.5 bly = 0.5R, while applying BSR we obtain a distance to the source 12 bly = 0.8R. We see that the uncertainty Δ_X in location of a luminous object is $\Delta_X=4.5$ bly.
- Example 2. Let's assume now that a luminous object is "close", so we neglect the spectral shift due to the distance in position space, and all of the redshift $z=2$ is due to **distance in velocity space**, *i.e.* relative (receding) velocity. We view Lobachevskian velocities space as an interior of unit radius $R=c=1$ equipped with hyperbolic metric. The signed distance here is simply relative velocity. In this case, the uncertainty in velocity determination Δ_U due to the different maps ESR and BSR is $\Delta_U=0.8c-0.5c=90000$ km/sec.
- Example 3. Taking into consideration both velocity and position, we come to the notion of a **phase space**. In the phase space picture the uncertainty of position and velocity is simply $\Delta_X\times\Delta_U$ and the only thing we can say is that an object is **somewhere** in an **uncertainty cell of phase space** having size $\Delta_X\times\Delta_U$ = 4.5 bly × 90,000 km/sec. The uncertainty relation $\Delta_X\times\Delta_U$ is quite close analogous to Heisenberg's uncertainty relation $\Delta x\times\Delta p$ in quantum mechanics.

If we recall that Lobachevskian geometry affects not only color of light (spectrum) but also its intensity, which makes Euclidean photometry non-applicable and misleading [4], we find ourselves in situation where the entire deep space astrophysics and cosmology cannot be trusted as providing viable picture of the

Universe.

4.3. Advantages and Generality of Our Approach

The formalism presented here is intuitively simple and **does not require more than three real dimensions**. Recall that since $g \in SL(2C)$, $\det \tilde{x} = 1$, hyperboloid $x^2 = const. > 0$, $x_0 \geq 1$, is mapped onto itself under motions (31) or (32). It follows that the physical phenomena as described in Einstein's original ESR are due to isometries (rigid motions) within a Lobachevskian 3-dim real space. ESR **does not extend to the entire** R^4 **but is restricted to only a Lobachevskian real 3-dim space (hyperboloid)** $x^2 = x_0^2 - x_1^2 - x_2^2 - x_3^2 = 1$, $x_0 \geq 0$, invariant under the action of the Lorentz group $SL(2C)$. Hence, **there is no point in paying attention to what is going on "outside" of Lobachevskian space** $x^2 = 1$.

The main point which is totally missed in orthodox treatments of "special relativity" based on $SO(1,3)$ is the **importance of the Lorentz group** $SL(2C)$ **in physics, which is the group of isometries of real 3-dim Lobachevskian space, and it is the foundation of Lobachevskian geometry and Lobachevskian physics**. Note that non-integer spinor fields (e.g. electron fields) are associated with the representations of $SL(2C)$ instead of the (proper) Lorentz group $SO(1,3)$.

It is worth noting that our method is quite general. Take for example, a Lobachevskian plane regarded as the upper-half plane $Im(z) > 0$ in complex variables. Then the "Lorentz group" acting via isometries on the upper-half plane is the group of fractional linear transformations $z \to \frac{az+b}{cz+d}$, which is the Mobius group, and we can obtain the transformation of "time" and position coordinates in yet another "special relativity" in this case.

Regarding ESR, the Minkowski treatment focuses on a **pseudo-Euclidean (flat) embedding** space instead on **negatively curved Lobachevskian embedded space** where physics happens. The Minkowski 4-dim pseudo-Euclidean interpretation of "special relativity" allows us to compute various entities via ESR in Minkowski's picture, but the **reasons for physical mechanisms are entirely absent, paradoxes are present, and generalizations are impossible.**

Obviously physics in a manifold does not depend on the ambient **embedding** space in which a manifold is embedded. An observer looking at a globe from the ambient three dimensional Euclidean space in his room will see a model airplane flying from San Francisco to Frankfurt along a path having three coordinates x, y, z in **flat three dimensional space**. However, the autopilot which actually guides the plane from San Francisco to Frankfurt uses only **two coordinates** on the surface of Earth, longitude and latitude. The autopilot "thinks" exclusively in terms of a two dimensional (albeit curved) space, and knows nothing about the fact that the Earth's surface S^2, is embedded into some higher dimensional space E^3 because this information is entirely irrelevant. Studies of geometry and physics from the point of view of higher dimensional embedding spaces

were typical for the 18th and 19th centuries. The modern approach is via internal geometry and internal physics in the embedded spaces.

The more relevant view on ESR, as a physical 3-dim Lobachevskian geometry, as pointed out by Barrett [14], was promoted by Vladimir Varicak, a Croatian mathematician and physicist in as early as 1908 [15]. Unfortunately, it appears that his papers were not understood by the scientific community and went mostly unnoticed. Physicists fascinated with 4-dim Minkowski flat ambient geometry were seemingly unable or unwilling to pay attention to Varicak's work. It is interesting to note that while developing and publishing his "special relativity" (ESR) [9] in 1905, Albert Einstein, influenced perhaps by his teacher Minkowski, focused his attention on the linear (pseudo-Euclidean) 4-dim space instead of directly mapping 3-dim Lobachevskian geometry (known since 1835) into a 3-dim Euclidean space. It is remarkable that a whole generation of physicists during the past 100+ years has focused on Minkowski's embedding space, and has missed the very essence that "special relativity" (ESR) is just a **gnomonic map** and as such **cannot be unique**.

At this point it is worth reviewing what we have covered so far:

1) Lorentz transformations, induced by elements of the Lorentz group $SL(2C)$, are isometries of a real 3-dim Lobachevskian (hyperbolic) space and as such are distortion-less.

2) Various parametric maps from real 3-dim Lobachevskian space into Euclidean 3-dim space are are "special relativities", and there are as many "special relativities" as there are parametric maps; in fact, an uncountable infinity of them.

3) Distortions introduced by "special relativities" are common distortions of maps between not isometric spaces, in our case between Lobachevskian negatively curved space $K<0$ and Euclidean space $K=0$.

4) Since "special relativities" are not unique, it implies that **any relativistic physics (High Energy Physics) based on a particular "special relativity" is not unique as well**.

Now, we will discuss consequences for science and for physics in particular.

5. Isomorphism between ESR and BSR "Special Relativities"

To review, the coordinates in Lobachevskian velocity space that are used in this work are:

1) **Projective (homogeneous) Weierstrass coordinates** $u\left(u_0^2,u_1^2,u_2^2,u_3^2\right)$, in the form $u_0=\frac{1}{\sqrt{1-\mathrm{v}^2}}$, $u_i=\frac{\mathrm{v}_i}{\sqrt{1-\mathrm{v}^2}}$, $|\mathrm{v}|<1$, $i=1,2,3$, normalized $u^2=u_0^2-u_1^2-u_2^2-u_3^2=1$, as per Formula (55).

2) **Projective (homogeneous) rational coordinates** used by authors to parametrize unit hyperboloid $u^2=u_0^2-u_1^2-u_2^2-u_3^2=1$, $u_0=\frac{1+\upsilon^2}{1-\upsilon^2}$, $u_i=\frac{2\upsilon}{1-\upsilon^2}$, $|\upsilon|<1$, $i=1,2,3$, normalized $u^2=1$.

For instance, in an abstract Lobachevskian plane L^2 creates images in the Euclidean plane E^2 by specifying coordinates in E^2 within the range of ad-

missible values, see e.g. [16]. Thus the coordinates in an open disc representation (without a boundary at infinity) of a Lobachevskian plane are:

1) Coordinates in the disc D^2 of the form $\mathrm{v}_x = \tanh\theta\cos\alpha$, $\mathrm{v}_y = \tanh\theta\sin\alpha$, $\mathrm{v}_x^2 + \mathrm{v}_y^2 < 1$, $0 \le \alpha \le 2\pi$ result in the Beltrami-Klein model of the Lobachevskian velocity plane.

2) Coordinates in the disc D^2 of the form $\upsilon_x = \tanh\frac{\theta}{2}\cos\alpha$, $\upsilon_y = \tanh\frac{\theta}{2}\sin\alpha$, $\upsilon_x^2 + \upsilon_y^2 < 1$, $0 \le \alpha \le 2\pi$ result in the Poincare model of the Lobachevskian velocity plane.

The angular coordinate α is irrelevant to our discussion and will be omitted.

As we have seen, different models (maps) of Lobachevskian geometry in Euclidean space show different distortions. All effects of "special relativities", either ESR or BSR, are just distortions caused by the mapping of a Lobachevskian, negatively curved (velocity) space into a flat space. We already noted that in ESR velocity $\mathbf{v}$ and intrinsic Lobachevskian velocity θ are related as $\mathbf{v} = \tanh\theta$, while in our BSR this relation is $\upsilon = \tanh\frac{\theta}{2}$.

Since:

$$\textit{Beltrami-Klein} \ni \mathbf{v} = \tanh\theta = \frac{2\tanh\frac{\theta}{2}}{1+\tanh^2\frac{\theta}{2}} = \frac{2\upsilon}{1+\upsilon^2}, \upsilon \in \textit{Poincare} \tag{88}$$

and since (inverse map)

$$\upsilon = \tanh\frac{\theta}{2} = \frac{\tanh\theta}{1+\sqrt{1-\tanh^2\theta}} = \frac{\mathbf{v}}{1+\sqrt{1-\mathbf{v}^2}} \tag{89}$$

we see that the velocity $\mathbf{v}$ measured in experiments due to ESR and the velocity υ measured in experiments due to BSR are mutually related by (88) and (89), which are known as **isomorphism maps** between the **Beltrami-Klein ball model (v)** of Lobachevskian geometry and the **Poincare ball model** of Lobachevskian geometry. It follows that **all formulas** of Einstein's ESR regarding velocities and their functions (or metric relations and their functions) due to the Beltrami-Klein model can be converted to formulas in our BSR due to the Poincare model (and vice-versa) via respective substitutions (88) and (89).

Proposition 20 *The Poincare model of Lobachevskian geometry corresponds to the double sided action of the Lorentz group on Lobachevskian space while Beltrami-Klein model of Lobachevskian geometry corresponds to the left action of the Lorentz group on Lobachevskian space. The isomorphism between these two actions is realized by* $\tanh\theta \Leftrightarrow \tanh\frac{\theta}{2}, \theta \in [0,\infty)$ *given explicitly by the correspondence* Formula (*90*) *below:*

$$\begin{pmatrix} \frac{1+\upsilon^2}{1-\upsilon^2} & \frac{2\upsilon}{1-\upsilon^2} \\ \frac{2\upsilon}{1-\upsilon^2} & \frac{1+\upsilon^2}{1-\upsilon^2} \end{pmatrix} \Leftrightarrow \begin{pmatrix} \frac{1}{\sqrt{1-\mathrm{v}^2}} & \frac{\mathrm{v}}{\sqrt{1-\mathrm{v}^2}} \\ \frac{\mathrm{v}}{\sqrt{1-\mathrm{v}^2}} & \frac{1}{\sqrt{1-\mathrm{v}^2}} \end{pmatrix} \tag{90}$$

The **LHS** of (90) is the matrix from the "special relativity" of Lobachevski - Poincare - von Brzeski (77), $(t', x') \to (t, x)$, based on the present work. The **RHS** of (90) is the matrix from the "special relativity" of Einstein (Lobachevski - Beltrami-Klein). The double sided arrow in the middle of (90) shows the isomorphism of the mappings, and can be shown as follows:

1) Mapping from Einstein to von Brzeski, or ESR $\Rightarrow$ BSR.

Take Equation (42) or (88) of the isomorphism $\mathrm{v} = \frac{2\upsilon}{1+\upsilon^2}$ relating the Beltrami-Klein and Poincare models of Lobachevskian geometry and substitute it into all entries in the RHS of (92), *i.e.* into ESR. Simple calculations will result in the LHS (von Brzeski) matrix, *i.e.* BSR.

For example:

$$\frac{1}{\sqrt{1-\mathrm{v}^2}} = \frac{1}{\sqrt{1-\dfrac{4\upsilon^2}{\left(1+\upsilon^2\right)^2}}} = \frac{1}{\sqrt{\left(\dfrac{1-\upsilon^2}{1+\upsilon^2}\right)^2}} = \frac{1+\upsilon^2}{1-\upsilon^2} \tag{91}$$

and:

$$\frac{\mathrm{v}}{\sqrt{1-\mathrm{v}^2}} = \frac{1+\upsilon^2}{1-\upsilon^2}\frac{2\upsilon}{1+\upsilon^2} = \frac{2\upsilon}{1-\upsilon^2} \tag{92}$$

2) Mapping from von Brzeski to Einstein, or BSR $\Rightarrow$ ESR:

Take Equation (89) of the isomorphism $\upsilon = \frac{\mathrm{v}}{1+\sqrt{1-\mathrm{v}^2}}$ between the Poincare and Beltrami - Klein models of Lobachevskian geometry and substitute it into all entries in the LHS of (90), *i.e.* into BSR. This will convert it into the RHS of (90), *i.e.* ESR. We leave the calculation to the reader.

It is quite remarkable that in the course of the present work on physics, the isomorphism Formulas (88) and (89), which are in the domain of pure geometry, were "rediscovered" by means of physics. It shows the deep and amazing interconnection between the abstract world of (Lobachevskian) geometry and the material world of physics.

6. Homotopy of Maps. Uncountable Infinity of "Special Relativities". Undecidability of High Energy Physics

First, if we are interested in High Energy Physics (HEP), we have to say what kind of physics this is exactly. To do so, we will first define "low energy physics".

Definition 21 *Low Energy* ("*Non Relativistic*") *Physics is physics modeled on Euclidean geometry.*

Low energy physics deals in **small distances** in Lobachevskian velocity/momentum space $SL(2C)/SU(2)$. Next, we need to explain what we mean by "large distances" and "small distances". In Euclidean space adjectives such as "large", "small" are **meaningless**. They carry zero information associated with them. However, in Lobachevskian spaces they have a very definite meaning. The key is the value of the negative curvature. Note that the Gaussian curvature of

Lobachevskian velocity space is $K=-c^{-2}$. It follows that **low distances** will be those distances represented by relative velocities much less in comparison with c, or equivalently when the curvature K is close to zero. In other words, when viewed locally, the curvature of Lobachevskian space may be disregarded. Physics in such domains may be regarded as Euclidean, *i.e.* "low energy", physics.

Definition 22 *High Energy* ("*Relativistic*") *Physics is physics modeled on Lobachevskian geometry*.

This is because energy is determined by distance in Lobachevskian velocity/momentum space $SL(2C)/SU(2)$. The larger the distance, the larger the energy associated with it. Thus HEP is the physics of **large distances** in $SL(2C)/SU(2)$. Accordingly, high energy physics is physics when the negative curvature of Lobachevskian space $SL(2C)/SU(2)$, in either coordinate representation or in velocity/momentum representation, must be taken into account.

We already introduced the notion of a homotopy of maps in Section 2.3 in an informal way. In this section, homotopy will be used to prove the existence of an **uncountable infinity** (continuum) of "special relativities" and consequently an uncountable infinity of high energy relativistic) physics. We begin with the definition of homotopy.

Definition 23 *Two continuous mappings f and g are homotopic* (*form a homotopy*) *if there exists a continuous mapping* $h_t\times[0,1]$ *such that* $h(t=0)=f$ *and* $h(t=1)=g$.

The mapping h as per definition 23 is represented by the Equation (93) below.

$$h(t)=(1-t)f+tg, t\in[0,1] \tag{93}$$

Since the unit segment $[0,1]$ of the real line contains an uncountable infinity (continuum) of points, it follows that there is an uncountable infinity of maps which are in one-to-one correspondence with points $t\in[0,1]$. Maps related by a homotopy are called homotopy equivalent, in a similar way as topological spaces are equivalent when related by a homeomorphism. Equivalence in this context is understood in its standard mathematical sense. Equation (93) has simple intuitive meaning. Starting with the map f at some real parameter $t=0$ we **continuously deform the map** f until it becomes the map g at some other value of a real parameter $t=1$.

With respect to "special relativities" BSR and ESR given by stereographic and gnomonic maps respectively, and represented by Poincare and Beltrami-Klein models of Lobachevskian geometry, respectively, we have the following theorem. Recall that the Poincare model is conformal while the Beltrami-Klein is not.

Theorem 24 *There exists an* ***uncountable infinity of non-isometric and non-conformal models of Lobachevskian geometry*** *and the* ***uncountable infinity*** **(*continuum*) *of "special relativities"*** *built upon them. Any such hybrid* (*mixed*) *model is in one-to-one correspondence with the some point* $t\in[0,1]$ *in* Equation (*93*).

Proof. Directly from definition of homotopy.

$$h(t)=(1-t)\tilde{u}^{*}\tilde{x}'\tilde{u}+t\tilde{u}\tilde{x}',\quad t\in[0,1] \tag{94}$$

We see that the above mapping (94) obeys the definition of homotopy. At $t=0$, $h(0)$ we have a **stereographic map** $\tilde{u}^{*}\tilde{x}'\tilde{u}$, while at $t=1$, $h(1)$ we have a **gnomonic map** $\tilde{u}\tilde{x}'$, and there is an **uncountable infinity** of mixed (hybrid) maps, neither stereographic nor gnomonic, in between. In this way, the Poincare model is continuously deformed into the Beltrami-Klein model, or in terms of special relativities, our "relativity" BSR is continuously deformed into Einstein's "relativity" ESR.

The message from theorem 24 is very disturbing. If there were only two options, we could run an experiment to see which option might better fit a particular condition. If the number of options would be (arbitrarily) finite, we could "in principle" verify which map best fit the experiment. But if the number of options is **uncountably infinite**, we unfortunately cannot run an **infinity** of experiments **even in principle** since it obviously would require infinite time. Thus experiments in HEP show only **one of infinitely many** faces of reality, and one point of infinitely many possibilities cannot be regarded as the ultimate truth.

Conclusion 25 (*Incompleteness of HEP*) *Based on assumptions*:
$x^2=1=invariant$, $c^2=1=invariant$, *under the isometries of 3-dim real Lobachevskian spaces* L_X^3 *and* L_V^3 *executed by the Lorentz group* $SL(2C)$, *the knowledge acquired from "relativistic" or high energy physics is inconclusive and/or incomplete*.

To see the validity of the above conclusion, consider a typical experiment in "relativistic" particle physics: the decay of unstable particles. Let take the well known decay of π^0 into two photons, $\pi^0\to 2\gamma$. The experiment, in an **Euclidean** laboratory frame, results in some lifetime data, *i.e.* some **real number** t_{π^0}. Now we need to calculate back (**interpret**): what is the lifetime of π^0 in its frame, *i.e.* in the momentum hyperboloid which is **Lobachevskian** space. If the calculations are done in Beltrami-Klein model (or ESR), we obtain some value, say t_{BK}. If the calculations are done using the Poincare model (BSR), we will obtain a **different** value $t_P\neq t_{BK}$. So what we end up with is a π^0 **meson with two different lifetimes** which is obviously **impossible**. Furthermore, the same argument applies to momenta, energy, polarizations, reaction cross-sections, angular distribution of reaction products (since ESR is non-conformal while BSR is conformal), and so on. As we noted above, the situation is even more vague since there is in fact an infinity of mixed (hybrid) "relativities".

Remark 26 (*Analogy with Quantum Mechanics*) *It is well known that quantum mechanics deals with states which can be either* ***pure states*** *or* ***mixed states***. *If we label, for instance the Beltrami-Klein model as a pure state* Φ_{BK} *and Poincare model as a pure state* Ψ_P, *then the mixed state, or* ***hybrid model*** *mentioned here will be,* $F=p_1\Phi_{BK}+(1-p_1)\Psi_P$, *where* p_1 *and* $1-p_1=p_2$ *are real, non-negative numbers* $0\le p_1,p_2\le 1$ *representing* ***probability amplitudes*** *of finding the mixed state in one of the pure states* Φ_{BK} *or* Ψ_P. *It is*

clear that there exists a continuum of such hybrid models or mixed states. For example, if $p_1 = \frac{1}{2}$ *and* $p_2 = 1 - \frac{1}{2} = \frac{1}{2}$ *then we have a* ***mixed state representing the hybrid model, which is neither a Poincare model nor a Beltrami-Klein model*** *of Lobachevskian geometry. It follows that maps viewed as a physical states, which are homotopy related or which are of* ***the same homotopy type****, will produce* ***isomorphic*** *physics. We know that in quantum mechanics mixed states are represented by* ***density matrices****. Hybrid "relativities" can be represented in the same way which makes* ***"classical physics" as probabilistic as quantum physics****. This is an extremely interesting insight into* ***"quantum classical worlds"****, which dismisses the common belief that quantum physics is probabilistic while classical physics is deterministic.*

7. Paradoxes as Distortions of Maps between Non-Isometric Spaces

When humans explore the world around them, they encounter new phenomena, and use mathematical tools, e.g. maps, to make sense of the new phenomena. But if incorrect maps are used for such exploratory experiments, bizarre conclusions will result. Unfortunately new phenomena are often judged and interpreted in terms of already existing knowledge, which is frequently incompatible with experimental results. Sometimes, as was the case with quantum mechanics, a new fruitful approach emerges, but in many cases, the incompatibility between existing science and new theoretical or experimental facts leads to bizarre and false conclusions.

In this section, we prove the apparent nature of the so called **Twin Paradox**, and we resolve (in general) the problem of the shape of **a circular fast moving object**.

7.1. The "Twin Paradox", Its Origin, and Its Trivial Solution

The Twin Paradox is as old as ESR itself and so far there is **no satisfactory and reasonable solution** of the problem. It is interesting that Einstein himself never gave any solution to Twin paradox in the "special relativity" he authored. The solution presented here is the first and only one in the literature on the subject which is mathematically sound and does not involve any subjective factors.

We start with the standard scenario, but since we have already introduced the isomorphism between ESR and BSR, we proceed with a **symmetrical treatment** of both twins A and B, referenced to the same stationary clock C. Doing so removes all kinds of arguments based on asymmetry, e.g. traveling versus non-traveling twin, still present in literature.

At a fixed time, say "time zero" t_0 on an Earth clock C, both twin A and twin B start their journey in the same fast rocket and are subjected precisely to the same conditions. Upon returning to Earth, and **before seeing clock C**, the twins use isomorphic relativities and calculate what each will see on clock C. Note that

all three objects, twin A, twin B, and clock C are in the same place and are mutually motionless.

Twin A makes his calculation due to the Lobachevski-Beltrami-Klein model (ESR), and expects that the clock will show the time as $t_0 + \Delta t_{ESR}$. Twin B, who knows of Poincare model, makes his calculations of temporal distortions due to Lobachevski-Poincare-von Brzeski model (BSR), and expects that the clock will show the time as $t_0 + \Delta t_{BSR}$. Since $\Delta t_{ESR} \neq \Delta_{BSR}$ it implies that $t_0 + t_{ESR} \neq t_0 + \Delta_{BSR}$, which in turn implies that the clock should display **two different readings - an impossible outcome**, *i.e.* a "paradox". Obviously, the clock will display a **single reading** at the same instance when twins A and B look at it. The **state of the clock, *i.e.* the real number on the clock's display, does not depend on an identical copy of an observer (twin) who is looking at it**. Needless to say, this will cause much confusion for both twins.

Next, when twin A and twin B attempt to calculate their ages due to different isomorphic "relativities" (maps they used in their journeys), they will discover that **their ages are no longer the same**, despite the fact that they underwent **precisely the same journey and were subjected to precisely to the same conditions**.

Using Formulas (58) and (84), we can summarize the situation as follows:

1) If $\Delta t'_A = \Delta t'_B$ it implies that $\Delta t_A \neq \Delta t_B$,

2) If $\Delta t_A = \Delta t_B$ it implies that $\Delta t'_A \neq \Delta t'_B$.

In other words, after the trip(s), we will obtain quite bizarre results: either the clock will show double readings or twins A and B have two different ages. This is obviously an impossible outcome showing the apparent nature of distortions resulting from non-isometric maps called "special relativities".

The so called **Twin Paradox is not real and it is due to apparent distortions only**. The paradox is apparent in the same way as the shape and size of Greenland is different on two (homotopic) maps; see Figure 1. Distortions introduced by various maps of curved space into a flat space are of **apparent character only** and should not be viewed as something real.

The following example is even more evident in its outcome. It is in essence what high energy particle physicists confront routinely in their work (we will return to this when we discuss problem with lifetimes of unstable particles.

Since we have alternative "relativities", we do not need twins, and we limit our consideration to one traveler only. A traveler takes a trip in a rocket. After the trip (returning to Earth), the traveler is asked about his or her age. The traveler has many options for the answer. One option is to calculate the age due to ESR (gnomonic map); another option is to calculate the age due to BSR (stereographic map); furthermore, the traveler has an **infinity** of options to calculate the age due to the hybrid maps in between. As a result, the traveler is **unable to give a definite answer**. Then, the traveler is asked again: who might know your age? The answer is, that after the trip **nobody really knows** the actual age of the traveler. Problems of this sort are called **undecidable problems**. A theory which contains undecidable problems is incomplete.

Remark 27 *A reader familiar with quantum mechanics is familiar with the dilemma known as "**Schrodinger's cat**" regarding the incompleteness of information due to a quantum mixed state. If we replace the Earth clock C in the above analysis with the cat* (*biological clock*), *we will come to the analogous situation of **inconclusiveness** of the state of the cat prior to actually seeing it. One of the twins will claim that the cat is already dead while the other will claim that the cat is still alive. Since both claims are due to isomorphic "relativities" of equal logical value, we arrive at a **state of the cat which is half dead and half alive**. The only possibility is to observe the cat after returning from the trip* (*collapse of the wave function by breaking the cat box*) *and checking the state of the cat.*

Conclusion 28 *Relativistic physics contains undecidable problems, and therefore is an incomplete theory.*

The close resemblance here with quantum mechanics (QM) is striking. The **information about the traveler's age is lost** due to the geometry of mixed maps (mixed states in QM). This clearly shows that what we have here is the quantum mechanical - **probabilistic** case rather than "classical" deterministic case. It is ironic that Einstein who was hostile of quantum mechanics developed his ESR not realizing that ESR is just a "pure state" (or "pure map"), one of infinitely many, in a more general scheme which is inherently probabilistic and in which the **outcome of a physical process is specified not by a single real number but by probability amplitudes for alternative outcomes**.

7.2. "Wheels of a Fast Moving Bicycle" and Undecidable Questions in High Energy Physics

Regarding the shape of fast moving circular objects ("bicycle wheels") it is easy to see that all definite conclusions up to now, in one way or the other, are equally wrong. This means that all authors who conclude that a fast moving circular object will appear as circular, or those who conclude they will appear as not circular (e.g. elliptical), are equally wrong. This is because the question about the shape of fast moving spherical/circular object is **undecidable**, which means that **it is impossible a-priori to give a definitive answer to this question**. Undecidability here is of the same sort as the undecidability about the traveler's age we discussed in Section 7.1, however, it follows from the conformality or non-conformality properties of different maps.

Conformal maps are maps which preserve angular relations, so shapes of objects in images remain unchanged however the size of the objects change. Non-conformal maps do not preserve angular relations and consequently shapes of objects in images will be deformed. It is well known that the stereographic map is conformal while the gnomonic is not. The Poincare model of Lobachevskian geometry, and BSR which is a result of a stereographic projection are conformal, while the Beltrami-Klein model of Lobachevskian geometry and ESR which results from a gnomonic projection are non-conformal. Therefore, **there is no way to conclude a-priori anything definite about the shape of a fast moving circular object** because the definite "truth" **depends on which map of**

Lobachevskian velocity space or which model of "relativity" is used.

Suppose you have two cameras, one producing Poincare images or maps, and the other one producing Beltrami-Klein images or maps onto a piece of flat paper. ("Camera" here does not refer to a piece of hardware that takes pictures, but as a mapping or algorithm which maps objects between spaces of different curvatures). If you use the Beltrami-Klein camera, the images of a fast moving circle **will be ellipses and not circles**. This is because **Beltrami-Klein model of Lobachevskian geometry (used in ESR) is a non-conformal model**. On the other hand, if you use the Poincare camera, the **images of fast moving circles will be circles**, not ellipses. This is because **Poincare model is conformal**. Lobachevskian circles (spheres) in the Poincare model are also the Euclidean circles (spheres); there is distortion in size but not in a shape, a fact well known in non-Euclidean geometry. We have the following theorem:

Theorem 29 ***It is impossible a-priori to determine the shape of fast moving circular (spherical) objects. The shape of fast moving circular objects is model dependent****. In a Lobachevski-Beltrami-Klein-Einstein map, the image of a fast moving circle* ***will be an ellipse****. This is because the Beltrami-Klein model used in ESR is a* ***non-conformal model****. In the Lobachevski-Poincare-von Brzeski map, the image of a fast moving circle* ***will be a circle****. This is because the Poincare model is a* ***conformal model****.*

Proof. Directly from properties of non-conformality/conformality of Beltrami-Klein and Poincare models (Figure 8 and Figure 9).

Conclusion 30 *The inability to answer the question regarding the shape of a fast moving circular object is a result of incompleteness of information. In fact we proved that questions about the world of high relative velocities, or questions about the* ***world of high energy physics****,* ***are a-priori undecidable****. Answers to those questions are decidable only* ***up to a homotopy of maps****.*

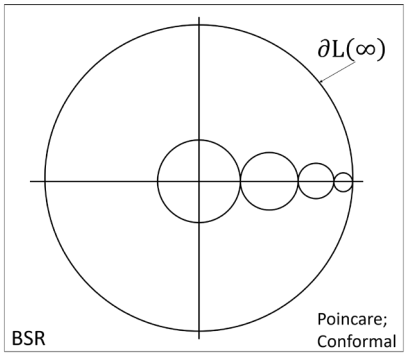


Figure 8. The apparent shape of "fast moving" spherical (circular) objects due to Poincare model of Lobachevskian geometry and the associated with it version of our "special relativity" - BSR. Since the model is conformal, spheres (circles) will appear in preserved shapes but in decreased sizes as distance (relative velocity) in Lobachevskian velocity/momentum space increases. The **apparent** size of objects close to the boundary at infinity, which is the sphere S^2 (circle S^1 in the above figure), will be arbitrarily small - point-like. The same reasoning applies to Lobachevskian position space - Lobachevskian Universe, and can be viewed as a real example of Lobachevskian cartography.

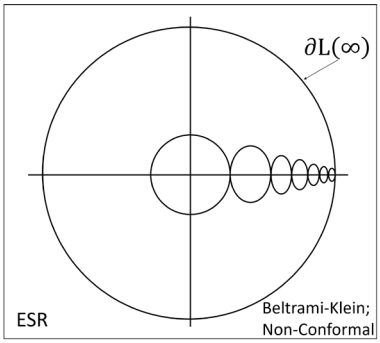


Figure 9. The apparent shape of "fast moving" spherical (circular) objects due to Beltrami-Klein model of Lobachevskian geometry and the associated Einstein's "special relativity" - ESR. Since this model is non-conformal, as relative velocity (with respect to the center) increases, spheres (circles) will appear: a) flattened (ellipsoidal) and b) smaller in size. The **apparent** shape and size of objects close to the boundary at infinity (circle S^1 in the above figure) will be very "thin" and very "small", respectively. The same argument holds for a Lobachevskian Universe if perceived via the Beltrami-Klein model, and can be viewed as a real example of Lobachevskian cartography.

Application to Astrophysics

Deep space astrophysics shows that geometry of large scale vacuum (background geometry) of the Universe is Lobachevskian [4] [5] [6] [7]. Therefore we can ask a question. How do we perceive the Lobachevskian Universe around us, via the Beltrami-Klein model or via the Poincare model, or via a mixed model? This is an interesting question which can be resolved experimentally.

As we already mentioned, Lobachevskian geometry in physics is represented by velocity space L_U and by position (coordinate) space L_X which we identify with the large scale vacuum - Lobachevskian Universe. We know that galaxies in the Universe are basically of two shapes, spherical and elliptical. If the Universe appears to us as in the Poincare model, then **statistical distribution of galactic shapes will not depend on distance**. Independently of how far we can see, there will on average be the same count of spherical and elliptical galaxies. That is because Poincare model is conformal, spheres are mapped onto spheres, ellipsoids are mapped onto ellipsoids.

On the other hand if space appears to us as the Beltrami-Klein model, the distant spherical galaxies will look more and more elliptical. So the **number of elliptical galaxies will rise with distance**. However things are more complicated since depending on galaxy orientation with respect to an observer, some elliptical galaxies will look even more elliptical while others will appear less elliptical, or maybe even as spheres.

Moreover, there is still an infinity of mixed models which makes things even more complicated. The problem is not easy to resolve but we hope that astrophysicists with help of advanced statistical analysis will be able to solve it.

7.3. The Meaning of Distortions for Physics. What Is Real and What Is Apparent?

Everybody has their own sense of reality and we will not comment on that.

However, the adjective "apparent" has more of an ambiguous meaning for each person. Therefore we take a definition of "apparent" from Webster's New Collegiate Dictionary.

Apparent (illusory, seeming, ostensible): *manifest to the senses or mind as real or true on the basis of evidence that may or may not be factually valid.*

The main source of confusion about what is real and what is apparent is that we are comparing **metric relations (sizes) in non-isometric spaces**, e.g. comparing "apples and oranges". **We cannot reach conclusions about sizes between spaces of different curvature which are not isometric**. While we can go directly to Greenland to verify our data regarding cartography from spheres, we cannot go to a Lobachevskian negatively curved space and to make direct measurements there. What we can account for at most is that we have to work with distorted images in our 3-dim Euclidean world. **Maps (projections) between curved and flat spaces result in distortions and there is no way to circumvent this misfortune**. The reader can now appreciate why we devoted a considerable portion of this paper to maps (projections) from curved to flat spaces.

The fundamental misunderstanding about Lorentz transformations which plagues all of physics since 1905 is that they introduce distortions. This is false. Lorentz transformations **are isometries of Lobachevskian space and hence are distortion-less**. Distortions are introduced when we map (project) information from a hyperboloid into our Euclidean laboratory. "Paradoxes" result from incorrect interpretation of "special relativities" via flat Minkowski pseudo-Euclidean geometry. In the interpretation of "special relativities" as maps from Lobachevskian into Euclidean space no "paradoxes" are present.

We've come to the point where we have to explain which effects of "special relativities" can be regarded as **real**, and which ones are only **apparent**. Recall that since it's publication in 1905, some phenomena related to "special relativity" have been viewed as paradoxical. Moreover, due to formulation of "special relativity" via a Minkowski flat space, the origin and nature of such paradoxes were entire obscure. Our exposition makes it easy to understand the origin of such paradoxical phenomena, and below we give our view on what is real and what is not when dealing with "special relativities".

If we admit that the velocity of light in a vacuum is a limiting value, *i.e.* $c=1$, and we admit the independence of the velocity of light (in a vacuum) with the state of motion, $c \pm u = c$, then space of velocities has a natural mathematical model as Lobachevskian (hyperbolic) 3 dimensional space, and its group of isometries is the Lorentz group $SL(2C)$. From here, due to various Lorentz group actions on Lobachevskian, space we obtain various "special relativities", each identified as some map from Lobachevskian $K<0$ into Euclidean space $K=0$. Thus it is natural to discuss the distortions or deformations such maps introduce. Since each map is between a pair of spaces of different curvature which are not isometric, the situation with distortions is more subtle than it might initially appear.

In fact **two kinds of distortions of different origin** have to be considered:

1) Since Lobachevskian space is more volumetric than Euclidean space (see Table 1), images from Lobachevskian space recorded in Euclidean space, by necessity, will be **compressed**. It should be noted that the effect of compression is regarded as the compression of images, and not as a compression of material bodies. This effect is **independent of the map used, and is a consequence of pure geometry**. For this reason we think that **effects of image compression might be regarded as real** since they are given by Nature itself and they cannot be eliminated by man in any way. On the other hand, it has to be clearly understood that a **comparison of metric relations (sizes) in two non-isometric spaces having different volumetric content and having two different units of length, is highly misleading**.

2) In addition to distortions (deformations) of images resulting from different curvatures, there are "technical" (man-made) distortions resulting from the **particular type of map** employed in the projection of images from curved into Euclidean space. It has been shown that for instance gnomonic and stereographic maps introduce different distortions when mapping the same curved space into a flat space; see Figure 1(a) and Figure 1(b). The distortions resulting from the type of a map are **purely apparent**, *i.e.* not real, and may be altered by choosing an appropriate (to an experiment) map.

As a result, in any experiment at **high energies** (large distances in Lobachevskian space), we record a **mixture** of such **real** and **apparent** distortions, which cannot be untangled—the only option we have in order to obtain information from a negatively curved space is to employ some kind of map! Thus, images (data) which we record in "relativistic physics" should be regarded as "real/apparent" images (data). To which degree the images are real or cannot be separated in a unique way.

8. Summary

In this paper, we presented the **discovery of the non-uniqueness of Einstein's "special relativity" and inconclusiveness of the High Energy Physics ("relativistic physics") resulting from different maps from Lobachevskian into Euclidean space**. Different, yet homotopy equivalent, maps result in mathematically equivalent (isomorphic) but physically not-equivalent "special relativities" in that they will produce **different numerical predictions**. Another new result is the identification of "special relativities" as common maps from Lobachevskian spaces into Euclidean space, and their association with well known models of Lobachevsian geometry.

Einstein's "special relativity" (ESR) is **not a theory**. It is a **gnomonic** map from Lobachevskian space into Euclidean space, is one of infinitely many possible maps, and as such is **not unique.** From a mathematical point of view, it is based on the Beltrami-Klein model of Lobachevskian geometry. This Lobachevski-Beltrami-Klein-Einstein "relativity" **is isomorphic, but it is not isome-**

tric and not conformal with the authors' BSR, what we call the Lobachevski-Poincare-von Brzeski "relativity". From the point of view of physics, Einstein's "special relativity" map corresponds to **single sided motion**. In physical applications, it is a natural choice for fixed target scattering experiments like those on accelerators with a fixed target.

The "special relativity" presented in this paper, namely Lobachevski-Poincare-von Brzeski (BSR) is **not a theory** either. It is a **stereographic** map from Lobachevskian space into Euclidean space and as such **is also not unique**. From mathematical point of view, it is based on a Poincare model of Lobachevskian geometry, which is isomorphic but **not isometric** with the ESR "relativity". From the point of view of physics, it describes **two sided motion**. It corresponds to **center of momentum frame** scattering, in accelerators with counter rotating beams. In astrophysics, it applies to so called "relativistic beams" ejected from galactic nuclei.

Between ESR and BSR, there is a **continuum of hybrid "relativities"** due to homotopy between maps, as was explained in the Section 6. Therefore, **results acquired in High Energy "Relativistic" Physics are not conclusive**. More precisely, they are conclusive up to homotopy only.

The results of experiments in High Energy Physics and deep space astrophysics depend on maps called "special relativities", which translate non-Euclidean reality into Euclidean data in our laboratory, data which are inevitably deformed due to different distortions introduced by those different maps.

Our "special relativity" BSR is equally valid from the logical, the mathematical, and the physical point of view. There are no criteria of any kind to discriminate BSR versus ESR since there are no criteria (beyond the matter of convenience in a particular situation) to discriminate the Poincare model versus Beltrami-Klein model of Lobachevskian geometry.

To our knowledge, the BSR alternative to ESR is presented here for the first time in scientific literature. "Alternative" has to be understood in the same sense as alternative maps are used in the practice of mapping the Earth's surface. As we said in the introduction, map making is mathematics. Navigation and exploration of the physical world, using those maps, is experimental physics.

Unfortunately, we cannot experience (perceive) global Lobachevskian geometry directly, but only via our local Euclidean reality. Our Euclidean experience is only projections—images or maps due to the Poincare model or due to the Beltrami-Klein model, or due to an infinity of homotopy equivalent hybrid models. Different maps or projections or images will result in different data from the world of high energy physics, and we do not have any "solid reference" to know their true nature (pre-images). We have no direct access to the Lobachevskian geometry "source code" of Nature and we must work with the Lobachevskian world in terms of its Euclidean images. The situation is reminiscent of Plato's *Allegory of the Cave*. The bird-eye's view of the presented paper is summarized in Table 2.

Table 2. Bird's eye view of Einstein's and the authors' "relativities".

	Einstein "Special Relativity" (ESR)	**von Brzeski "Special Relativity" (BSR)**
Lobachevskian Real 3-dim Space	$SL(2C)/SU(2)$ Quotient of Lorentz Group	$SL(2C)/SU(2)$ Quotient of Lorentz Group
Type of Lorentz Group Action on $SL(2C)/SU(2)$	Single sided (left) $x = gx'$	Double sided $x = g^{*}x'g$
Type of Projection (Type of Map)	Gnomonic (central), non-conformal	Stereographic, conformal
Type of Coordinates	Projective Weierstrass	Projective rational
Resulting Model of Lobachevskian Geometry	Beltrami-Klein, non-conformal	Poincare, conformal
Relation between BSR and ESR	Isomorphic as models, Homotopic as maps	Isomorphic as models, Homotopic as maps

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Anderson, W.A. (1999) Hyperbolic Geometry. Springer, London, Berlin, Heidelberg. https://doi.org/10.1007/978-1-4471-3987-4

[2] von Brzeski, G. and von Brzeski, V. (2018) *Journal of Modern Physics*, **9**, 1326-1359. https://doi.org/10.4236/jmp.2018.96081

[3] von Brzeski, G. (2008) *Acta Physica Polonica*, *B*, **39**, 1501-1520.

[4] von Brzeski, G. (2007) *Russian Journal of Mathematical Physics*, **14**, 366-369. https://doi.org/10.1134/S1061920807030107

[5] Borisovich, Yu., Bliznyakov, N., Izrailevich, Y.A. and Fomenko, T. (1985) Introduction to Topology (English edition). Mir Publishers, Moscow.

[6] Buseman, H. and Kelly, P.J. (1953) Projective Geometry and Projective Metrics. Academic Press, New York.

[7] Reynolds, W. (1993) *The American Mathematical Monthly*, **100**, 442-455. https://doi.org/10.1080/00029890.1993.11990430

[8] Gelfand, I.M., Grayev, M.I. and Vilenkin, I.Y. (1966) Integral Geometry and Representations Theory. Translated from Russian, Academic Press, New York.

[9] Gelfand, I.M., Minlos, R.A. and Shapiro, Z.Y. (1963) Representations of the Rotation and Lorentz Groups and their Applications. Pergamon, New York.

[10] Gorbatsevich, V.V., Onischik, A.L. and Vinberg, E.B. (1997) Foundations of Lie Theory and Lie Transformations Groups. Springer Verlag, Providence, Rhode Island.

[11] Klein, F. (2000) Forlesungen Uber Nicht Euklidische Geometries. American Mathematical Society, New York, Berlin, Heidelberg.

[12] Greenberg, M. (1993) Euclidean and Non-Euclidean Geometries. 3rd. Edition, Development and History, W.H. Freeman and Company, New York.

[13] Einstein, A. (1905) *Annalen der Physik*, **322**, 891-921.

https://doi.org/10.1002/andp.19053221004

[14] Barrett, J.F. (1994) On Varicack's Interpretation of Special Relativity in Hyperbolic Space with Application to the Redshift. *PIRT Conference*, Imperial College, London, September 1994, 17-20.

[15] Varicak, V. (1912) On the Non Euclidean Interpretation of the Theory of Relativity. Translation from German. Jahresbericht der Deutschen Mathematiker Vereinigung 21.

[16] Ramsay, A. and Richtmayer, R.D. (1995) Introduction to Hyperbolic Geometry. Springer, New York, Berlin, Heidelberg. https://doi.org/10.1007/978-1-4757-5585-5

Journal of Modern Physics, 2021, 12, 1346-1365
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Mutual Influence of the Atmosphere and the Ocean under Wave Processes

Vladimir G. Kirtskhalia[1], Konstantin R. Ninidze[2]

[1]Vekua Sokhumi Institute of Physics and Technology, Tbilisi, Georgia
[2]Sokhumi State University, Tbilisi, Georgia
Email: v.kirtskhalia@gmail.com

How to cite this paper: Kirtskhalia, V.G. and Ninidze, K.R. (2021) Mutual Influence of the Atmosphere and the Ocean under Wave Processes. *Journal of Modern Physics*, **12**, 1346-1365.
https://doi.org/10.4236/jmp.2021.129081

Received: May 30, 2021
Accepted: July 23, 2021
Published: July 26, 2021

Abstract

The article solves the problem of surface gravitational waves using the theory of tangential discontinuity between media: air-water. Using the improved equation of mass continuity and taking into account the atmosphere inhomogeneity in the gravitational field of the Earth, it is shown that during wave processes, these two media mutually influence each other, which explains the reason for the formation of a stormy condition over the ocean and the drop in atmospheric pressure before the storm. The mechanism of the formation of the "killer wave" has been established and thus the "greatest mystery of nature" has been solved. The scale of wind and tsunami wavelengths has been established.

Keywords

Atmosphere, Ocean, Gravitational Waves, Waves of Wind, Tsunami Waves, Killer Wave

1. Introduction

In early works (see e.g. [1]), we argued that linear theory describes only capillary waves. As it turned out, this statement was erroneous, but we came to this conclusion thanks to the existing theory of capillary waves, which is based on the assumptions of the potentiality motion of a liquid in the gravitational field of the Earth and its incompressibility. In these assumptions, the linear theory of tangential discontinuity is used in solving the problem, and in the dynamic boundary condition on the liquid surface, the surface tension force is used as the only stabilizing factor of the pressure perturbation ([2], §62). Nevertheless, in the dispersion equation of capillary waves, a gravitational acceleration is present along with the surface tension coefficient and therefore, these waves are called

capillary-gravitational. After we showed that the liquid cannot be incompressible, *i.e.* $\nabla\vec{V} \neq 0$ [3], where $\vec{V}$ -is the velocity of the liquid particle, the gravitational acceleration from the dispersion equation disappeared [1]. This fact led us to believe that gravitational waves should be described by a nonlinear theory. However, a simple analysis shows that in most cases gravitational waves are linear. Indeed, the maximum perturbation of pressure on the water surface in a gravitational wave is equal to $P'_{\max} = \rho_0 ga$, where a is the amplitude of the wave and ρ_0 is the undisturbed density of water. At $a = 3$ m, which is approximately equal to the maximum value of the amplitude of the wind wave away from the coast and is greater than the amplitude of the tsunami wave, this disturbance is equal to 2.7×10^4 Pa, while the equilibrium pressure is $P_0 = 1$ atm. $= 1.013 \times 10^5$ Pa. Thus, the pressure on the surface of water can be represented in the following form $P = P_0 + P'$, where $P'/P_0 < 1$, and therefore, linear theory can be used.

A surface gravitational wave is generated and propagated at the interface between two media, *i.e.* it is a typical problem of tangential discontinuity. Therefore, the dispersion equation must contain the thermodynamic parameters of both media. Despite this, when solving this problem, many authors do not take into account the influence of atmospheric parameters on the phenomenon under study (see for example. [3] [4] [5]), but nevertheless, they get results that match the results of observations. For example, in work [3] the two-dimensional problem of the surface gravitational wave was solved, in which the following are used:

1) Equation of motion of fluid— $\rho\dfrac{\mathrm{d}\vec{V}}{\mathrm{d}t} = -\nabla P + \rho\vec{g}$ (1)

2) Equation of incompressibility of liquid— $\nabla\vec{V} = 0$ (2)

Here: $\vec{V}$ is velocity of a liquid particle, P—pressure, ρ—density, $\vec{g}$—gravitational acceleration. It is assumed that the velocity is small and after linearization, the system of Equations (1), (2) takes the form:

$$\frac{\partial u}{\partial t} = -\frac{1}{\rho}\frac{\partial P}{\partial x}, \tag{3}$$

$$\frac{\partial w}{\partial t} = -\frac{1}{\rho}\frac{\partial P}{\partial z} - g, \tag{4}$$

$$\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} = 0. \tag{5}$$

By denoting $\xi(x,t) = \xi_0 \cos(\omega t - kx)$, the displacement of the liquid surface along the Z axis and writing the boundary conditions on the free surface and on the bottom as

$$w\big|_{z=0} = \frac{\partial \xi}{\partial t} \quad \text{and} \quad w\big|_{z=-H} = 0, \tag{6}$$

solutions of the system of Equations (3), (4), (5) will be:

$$u = \xi_0 \omega \frac{\cosh\left[k\left(z+H\right)\right]}{\sinh\left(kH\right)} \cos\left(\omega t - kx\right), \tag{7}$$

$$w = -\xi_0 \omega \frac{\sinh\left[k\left(z+H\right)\right]}{\sinh\left(kH\right)} \sin\left(\omega t - kx\right), \tag{8}$$

$$P = -\rho g z + \rho g \xi_0 \frac{\cosh\left[k\left(z+H\right)\right]}{\cosh\left(kH\right)} \cos\left(\omega t - kx\right). \tag{9}$$

Substituting solutions (7) and (9) in (3), or (8) and (9) in (4) at, the author obtains an expression for the phase velocity in the form:

$$\left|U_p\right| = \sqrt{\frac{g}{k}\tanh\left(kH\right)}. \tag{10}$$

For deep water, when $kH = 2\pi H/\lambda > 1 \Rightarrow \tanh\left(kH\right) \cong 1$, where λ is the wavelength, from (10) follows

$$\left|U_p\right| = \sqrt{\frac{g}{k}} \tag{11}$$

and for shallow water, when $kH = 2\pi H/\lambda < 1 \Rightarrow \tanh\left(kH\right) \cong kH$,

$$\left|U_p\right| = \sqrt{gH}. \tag{12}$$

Similar results were obtained in [4] [5] using the hydrostatic approximation method, in which Equation (4) is integrated under the following assumption $\partial w/\partial t = 0$. In the work [3] it is noted that the validity of this assumption is not proven and this doubt is well-founded, because firstly- $\partial w/\partial t = 0$ means that $w = const = 0$, and therefore, no vertical movement of the liquid particle occurs, and second-at $\partial w/\partial t = 0$, from Equation (4) follows $\nabla P = \rho \vec{g}$ which is a condition for the equilibrium of the liquid and therefore, vibrations are impossible.

As noted above, while analyzing the problem of surface gravity waves, both media should be taken into account—air and water, *i.e.* should be obtained in a linear approximation of the equations of gravitational waves in these media, and then, tied to their solutions on the boundary surface using boundary conditions.

When performing these procedures, the perturbed values of density and pressure in each medium are linked by the equation of the state of the medium $\rho = P'/C^2$, where C is the speed of sound in the medium. We called attention to the absurdity of the fact that the speed of sound in the entire atmosphere is calculated by the formula $C = \sqrt{\gamma k_B T/m_0}$ [6] [7], where $\gamma = c_p/c_\upsilon = 1.4$ is an adiabatic index of air and is equal to the ratio of heat capacities at constant pressure and volume, $k_B = 1.38\times10^{-23}$ J/K —the Boltzmann constant, $m_0 = 4.81\times10^{-26}$ kg —mass of one air molecule, T—absolute temperature. This means that on the altitude of 60 km and on the North Pole, where the temperatures are the same and equal to −40˚C, the speeds of sounds should have the same values.

We have found the answer to this paradox and it is that the abovementioned formula is just only for a homogeneous medium where the density depends only

on the pressure $\rho = \rho(P)$. In the inhomogeneous medium, which is the earth's atmosphere, due to the influence of the gravitational field on it, the density also depends on entropy $\rho = \rho(P,S)$. In that case the density perturbation equals to the following:

$$\rho' = \left(\frac{\partial \rho_0}{\partial P_0}\right)_s P' + \left(\frac{\partial \rho_0}{\partial S_0}\right)_p S', \tag{13}$$

where P_0, ρ_0, S_0 -represent unperturbed values of pressure, density and entropy accordingly. Considering that $S' = \left(\partial S_0 / \partial P_0\right)_T P'$, from (13) we easily receive [8] [9]:

$$\rho' = \left(\frac{1}{C_s^2} + \frac{1}{C_p^2}\right) P' = \frac{1}{C^2} P'. \tag{14}$$

Here:

$$C_s^2 = \left(\frac{\partial P_0}{\partial \rho_0}\right) = \gamma \frac{k_B T}{m_0} \text{ is the speed of adiabatic sound,} \tag{15}$$

$$C_P^2 = \left[\left(\frac{\partial \rho_0}{\partial S_0}\right)_P \left(\frac{\partial S_0}{\partial P_0}\right)_T\right]^{-1} = \frac{c_p \rho_0^2}{T\left(\partial \rho_0 / \partial T\right)_s^2} \text{ is the speed of isobaric sound,} \tag{16}$$

$$C^2 = \frac{C_s^2 C_p^2}{C_s^2 + C_p^2} \text{ is the true value of the speed of sound.} \tag{17}$$

Thus, the square of the true value of the speed of sound is reduced from the squares of the velocities of adiabatic and isobaric sounds. Substituting into (16) the value ρ_0 from the Laplace equation

$$\rho_0 = \rho_0^0 \exp\left(-m_0 g z / k_B T\right), \tag{18}$$

where ρ_0^0 -is unperturbed density value at sea level, for the speed of isobaric sound we receive the following:

$$C_p = \sqrt{\frac{c_p k_B^2 T^3}{m_0^2 g^2 z^2}}. \tag{19}$$

Using (15) and (19), from (17) we find:

$$C(z,T) = \sqrt{\frac{\gamma k_B T}{m_0\left(1 + \frac{\gamma m g^2 z^2}{c_p k_B T^2}\right)}}. \tag{20}$$

We can see that the true value of the speed of sound in the Earth's atmosphere truly depends on the altitude (density) and condition $z = 0$ is equivalent to the condition $g = 0$. Thus, at the sea level, air is a homogeneous medium and the speed of sound should be calculated by the Equation (15).

Figure 1 shows the temperature distribution over height in the Earth's atmosphere [10]. We can see that up to the altitude of approximately 11 km., the temperature drops according to strictly linear law $T = \alpha + \beta z$, where $\alpha = 288.15$ K

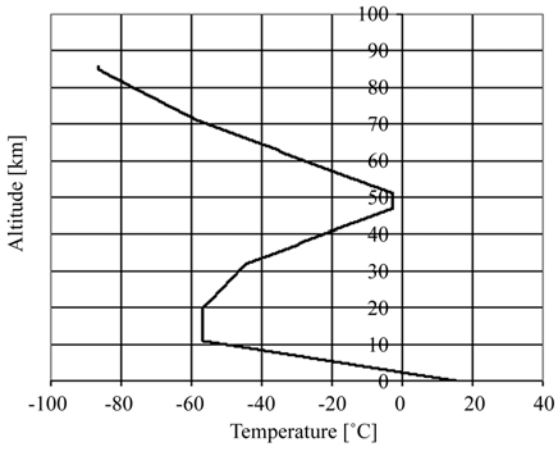


Figure 1. Temperature as a function of geometrical altitude.

and $\beta = -6.52\times10^{-3}\,\mathrm{K/m}$. This part of the atmosphere is called the troposphere, which has no heat source, hence the entropy S is constant, *i.e.* can be used the adiabatic equation

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{\partial S}{\partial t} + \left(\vec{V}\nabla\right)S = 0, \tag{21}$$

On the top boundary of the troposphere, the drop in temperature abruptly stops and remains constant till the altitude of approximately 20 km (tropopause), and then increases in the stratosphere. In this regard, the law of entropy constancy, which was used in our calculations, ceases to be true. Therefore, our theory is valid up to the height $z \cong 11\,\mathrm{km}$.

Substituting in (15), (19) and (20) $T = 288.15 - 6.53\times10^{-3}z$ and taking into account that $c_p = 10^3\,\mathrm{J/kg\cdot K}$, we get:

$$C_s(z) = \sqrt{401.66\left(288.15 - 6.53\times10^{-3}z\right)}, \tag{22}$$

$$C_p(z) = 925.70\frac{\sqrt{\left(288.15 - 6.53\times10^{-3}z\right)^3}}{z}, \tag{23}$$

$$C(z) = 20.05\sqrt{\frac{\left(288.15 - 6.53\times10^{-3}z\right)^3}{\left(288.15 - 6.53\times10^{-3}z\right)^2 + 4.69\times10^{-4}z^2}}. \tag{24}$$

In **Figure 2**, the graphs of expressions (22) and (23) are shown. Calculations show that the relative inaccuracy, between values $C_s(z)$ and $C(z)$ at heights $z = 1\,\mathrm{km}$ and $z = 10\,\mathrm{km}$ is equal to 0.3% and 33%, respectively, *i.e.* increases 110 times. Obviously, such an error cannot be ignored when calculating the Mach number [11].

In the work [12], the author suggests that at the upper boundary of the troposphere ($z \cong (10\text{-}11)\,\mathrm{km}$ which is often called the ozone layer), there is an exothermic reaction of ozone synthesis ($\mathrm{O_2 + O \rightarrow O_3 + 24\ k.cal/mol}$), which is the reason for such dynamics of the temperature distribution. Let us show that our theory fully confirms this hypothesis.

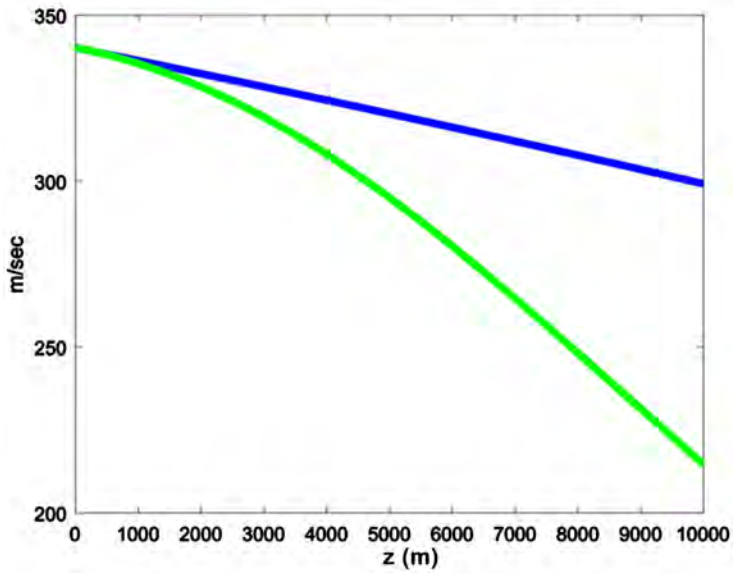


Figure 2. Dependence of isobaric ($C_s(z)$-the blue curve) and true ($C(z)$-the green curve) speeds of sounds on the altitude in the troposphere.

In **Figure 3**, the graphs of the height distribution of the adiabatic $C_s(z)$ and isobaric $C_p(z)$ velocities of sounds are shown. It is seen that these graphs intersect at a height of $z \cong 10200$ m. Equating $C_s^2(T)$ and $C_p^2(z,T)$ from Formulas (15) and (19), we obtain

$$\frac{\gamma kT}{m_0} = \frac{c_p k^2 T^3}{m_0^2 g^2 z^2} \Rightarrow z = \sqrt{\frac{c_p k}{\gamma m_0}} \frac{T}{g}. \tag{25}$$

Substituting in (25) $T = \alpha + \beta z$ we find:

$$z = \frac{\sqrt{c_p k \alpha^2 / \gamma m_0 g^2}}{1 + \sqrt{c_p k \beta^2 / \gamma m_0 g^2}} = 10230 \text{ m}. \tag{26}$$

This altitude almost exactly coincides with the upper boundary of the troposphere, presented in **Figure 1**, which proves the high reliability of our theory. Thus, it can be said with high confidence, that expressions (25) and/or (26) are equations of the upper border of the troposphere, where adiabatic and isobaric speeds of sound are equated, *i.e.* a resonance $\omega_s = \omega_p$ occurs. As it is known, the abrupt change in the dynamics of the process, which is observed in **Figure 1**, is as a general rule in connection with resonance. Thus, it can be assumed that the resonance of frequencies of the adiabatic and isobaric sounds is a trigger mechanism for the exothermic reaction of ozone synthesis and, as a consequence, the release of a large amount of heat.

The discovery of an isobaric sound leads to another important result. Let us transform the following expression:

$$\left[\frac{\rho_0}{(\partial\rho_0/\partial T)_p}\right]^2 = \left[\frac{m}{\upsilon}\left(\frac{\partial(m/\upsilon)}{\partial T}\right)_p^{-1}\right]^2_{m=const} = \left[\frac{1}{\upsilon}\left(\frac{\partial\upsilon}{\partial T}\right)_p\right]^{-2} = \frac{1}{\beta_p^2}, \tag{27}$$

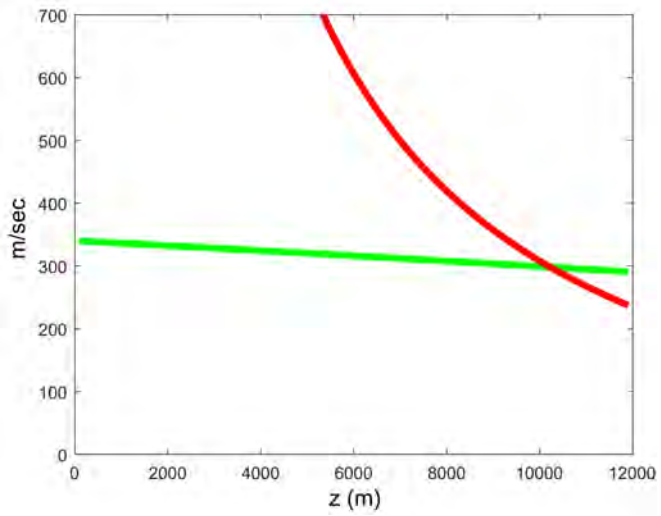


Figure 3. The dependences of adiabatic ($C_s(T)$ -green curve) and isobaric ($C_p(z,T)$ -red curve) sound velocities from the altitude in the troposphere.

where υ -is gas volume and β_p -is the coefficient of thermal expansion. Then from (16) we shall find that:

$$C_p(z,T)=\frac{1}{\beta_p}\left(\frac{c_p}{T}\right)^{1/2}. \tag{28}$$

From (28) it is derived, that $\beta_p=\beta_p(z,T)$, *i.e.* the coefficient of the thermal expansion depends not only on the temperature, as is usual to modern thermodynamics, but also on the altitude of the atmosphere. This fact invalidates the universality of the laws of the ideal gas. More detailed account of this please see the work [13].

Since the sound wave carries the density perturbation, an idea of generalization of the equation of mass continuity for an inhomogeneous medium arose. This equation was also received for the homogeneous medium and it determines the change of density as a result of substance mass change in the constant volume. However, the change in density is possibly also due to the change of volume of the constant mass of the substance, *i.e.*:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t}=\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{m}{\upsilon}\right)=\frac{\upsilon\dfrac{\mathrm{d}m}{\mathrm{d}t}-m\dfrac{\mathrm{d}\upsilon}{\mathrm{d}t}}{\upsilon^2}=\left(\frac{\mathrm{d}\rho}{\mathrm{d}t}\right)_{\upsilon}-\left(\rho\frac{\mathrm{d}}{\mathrm{d}t}\ln\upsilon\right)_m, \tag{29}$$

Equation (29) is reduced to the following [14]:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t}=-\rho\nabla\vec{V}-\frac{\vec{V}\nabla P}{C_p^2}. \tag{30}$$

The first summand in the right side of the Equation (30) determines the change of mass in the constant volume while the second summand is the isobaric change of the volume of the substance constant mass as a result of the change in temperature, which is caused by the change in entropy in the inhomogeneous medium. This work has radically altered the existing concepts of compressibility

and incompressibility of medium. It is considered, that these concepts have a mechanical meaning. In reality, they have thermodynamic meaning and characterize the homogeneity or non homogeneity of the medium. The homogeneous medium is always compressible. The incompressibility is a consequence of its non homogeneity and it manifests as strongly as inhomogeneous the medium is, as it occurs in the Earth's atmosphere with increasing altitude. The speed of sound in a homogeneous medium is adiabatic, and in an inhomogeneous medium it is a combination of the speeds of adiabatic and isobaric sounds.

The work [1] considers the problem of internal waves from the monograph [2]. The authors consider the water to be incompressible and contemplate, that the density perturbation is isobaric, *i.e.* only the second summand is considered in the Equation (13). As a result, they receive a so-called internal wave, the frequency of which depends only on the direction of the wave vector, the magnitude of which can be any. It is clear that such wave does not exist in nature and the reason for this paradox is the use of incompressibility condition towards the water. Hence, water is a compressible (homogeneous) medium. Though it may seem paradoxical, water is a much more compressible medium (in the thermodynamic sense), than the atmosphere in the higher layers. Work [14] has been dedicated to this problem.

Another paradox that we noticed is that the Euler equation in its current form contradicts the basic principle of physics. Indeed, let us consider the Euler equation in its generally accepted form: $\rho \mathrm{d}\vec{V}/\mathrm{d}t = \rho\left[\partial\vec{V}/\partial t + \left(\vec{V}\nabla\right)\vec{V}\right] = -\nabla P + \rho\vec{g}$. If $\vec{V}_0$ there is a stationary velocity of motion of the liquid and $\vec{V}'$ is its small perturbation, then after linearization its left side, which determines the acceleration of the liquid particle, will be equal to $\left[\partial\vec{V}'/\partial t + \left(\vec{V}_0\nabla\right)\vec{V}'\right]$. Thus, the acceleration of a liquid particle depends on the stationary velocity of the medium, which contradicts the principle of relativity. This contradiction is caused by the assumption $\mathrm{d}\rho/\mathrm{d}t = 0 \Rightarrow \rho = const$, *i.e.* the assumption of incompressibility of the liquid $\nabla\vec{V} = 0$ and neglect of the second term in Equation (30). In fact, if the first term is equal to zero (incompressible medium), remains the second term, and if the second term is equal to zero (compressible medium), the first remains. Thus, $\mathrm{d}\rho/\mathrm{d}t \neq 0$ and density ρ must be entered under the derivative

$$\frac{\mathrm{d}\left(\rho\vec{V}\right)}{\mathrm{d}t} = \vec{V}\frac{\mathrm{d}\rho}{\mathrm{d}t} + \rho\frac{\mathrm{d}\vec{V}}{\mathrm{d}t} = -\nabla P + \rho\vec{g}. \tag{31}$$

Let's substitute in the right side of Equation (31) the value $\mathrm{d}\rho/\mathrm{d}t$ from Equation (30)

$$\begin{aligned} &-\vec{V}\rho\nabla\vec{V} - \frac{V^2\nabla P}{C_p^2} + \rho\frac{\mathrm{d}\vec{V}}{\mathrm{d}t} = -\nabla P + \rho\vec{g} \\ &\Rightarrow \rho\left[\frac{\mathrm{d}\vec{V}}{\mathrm{d}t} - \left(V\nabla\right)\vec{V}\right] - \frac{V^2\nabla P}{C_p^2} = -\nabla P + \rho\vec{g} \end{aligned} \tag{32}$$

Neglect the term $V^2\nabla P/C_p^2$, the Euler equation takes the form:

$$\rho\frac{\partial\vec{V}}{\partial t} = -\nabla P + \rho\vec{g}. \tag{33}$$

As can be seen the nonlinear term dropped out of the Euler equation and, therefore, there is no contradiction. The Equation (30) along with the Equation (27) was used by us in the problem of the capillary waves [15], after which all existing contradictions were removed.

In our opinion, plasma is an incompressible medium, because it is a substantially inhomogeneous, due to the fact that each electron and ion are in a force field created by neighboring particles. It can be assumed that this is precisely the reason for the difficulties in implementing a controlled thermonuclear reaction, since it is impossible to compress the plasma to the required state and keep it in this state for a long enough period of time.

2. A New Approach to the Theory of Surface Gravitational Waves

Let us now apply these equations to the problem of surface gravitational waves:

$$\begin{cases} \rho \dfrac{\mathrm{d}\vec{V}}{\mathrm{d}t} = \rho\left[\dfrac{\partial \vec{V}}{\partial t} + \left(\vec{V}\nabla\right)\vec{V}\right] = -\nabla P + \rho\vec{g} \\ \dfrac{\mathrm{d}\rho}{\mathrm{d}t} = \dfrac{\partial \rho}{\partial t} + \left(\vec{V}\nabla\right)\rho = -\rho\nabla\vec{V} - \dfrac{\vec{V}\nabla P}{C_p^2} \end{cases} \tag{34}$$

Suppose that $V_0 = 0$ and let's represent all the variables as the sum of their stationary values and small perturbations:

$$P = P_0 + P';\rho = \rho_0 + \rho';\vec{V} = \vec{V}'.$$

Using under the linearization of system (34) the condition of equilibrium of the medium in the gravitational field of the Earth $\nabla P_0 = \rho_0\vec{g}$, which is true for both air and water, we get:

$$\begin{cases} \rho_0 \dfrac{\partial \vec{V}'}{\partial t} = -\nabla P' + \rho'\vec{g} \\ \dfrac{\partial \rho'}{\partial t} + \left(\vec{V}''\nabla\right)\rho_0 = -\rho_0\nabla\vec{V}' - \dfrac{\rho_0\vec{g}\vec{V}'}{C_p^2} \end{cases} \tag{35}$$

Let's apply the operator ∇ to the first equation of the system (32) and the operator $\partial/\partial t$ to the second, after which we have

$$\begin{cases} \nabla\rho_0 \dfrac{\partial \vec{V}'}{\partial t} + \rho_0\nabla\dfrac{\partial \vec{V}'}{\partial t} = -\Delta P' + \nabla\rho'\vec{g} \\ \dfrac{\partial^2 \rho'}{\partial t^2} + \left(\dfrac{\partial \vec{V}'}{\partial t}\nabla\right)\rho_0 = -\rho_0\nabla\dfrac{\partial \vec{V}'}{\partial t} - \dfrac{\partial \vec{V}'}{\partial t}\dfrac{\nabla P_0}{C_p^2} \end{cases} \tag{36}$$

Given that P_0 and ρ_0 in air and in water depend only on the vertical z coordinate, the following relations are valid:

$$z\nabla\rho_0 \frac{\partial \vec{V}'}{\partial t} = \left(\frac{\partial \vec{V}'}{\partial t}\nabla\right)\rho_0 = \frac{\mathrm{d}\rho_0}{\mathrm{d}z}\frac{\partial V_z'}{\partial t};\ \frac{\partial \vec{V}'}{\partial t}\frac{\nabla P_0}{C_p^2} = -\frac{\rho_0 g}{C_p^2}\frac{\partial V_z'}{\partial t}\nabla\rho'\vec{g} = -g\frac{\partial \rho'}{\partial z},$$

after that, the system (36) can be rewritten as

$$\begin{cases} \frac{\mathrm{d}\rho_0}{\mathrm{d}z}\frac{\partial V_z'}{\partial t}+\rho_0\nabla\frac{\partial \vec{V}'}{\partial t}=-\Delta P'-g\frac{\partial \rho'}{\partial z} \\ \frac{\mathrm{d}\rho_0}{\mathrm{d}z}\frac{\partial V_z'}{\partial t}+\rho_0\nabla\frac{\partial \vec{V}'}{\partial t}=-\frac{\partial^2\rho'}{\partial t^2}+\frac{\rho_0 g}{C_p^2}\frac{\partial V_z'}{\partial t} \end{cases} \tag{37}$$

Equating the right parts of the system (37) we get

$$\Delta P'+g\frac{\partial \rho'}{\partial z}+\frac{\rho_0 g}{C_p^2}\frac{\partial V_z'}{\partial t}-\frac{\partial^2\rho'}{\partial t^2}=0. \tag{38}$$

Using the equation of state of the medium $\rho'=\left(1/C^2\right)P'$ and expressing $\partial V_z'/\partial t$ from the first equation of the system (35) $\partial V_z'/\partial t=-\left(1/\rho_0\right)\left(\partial P'/\partial z+g\,P'/C^2\right)$, Equation (38) takes the form:

$$\Delta P'+g\frac{\partial}{\partial z}\left(\frac{P'}{C^2}\right)-\frac{g}{C_p^2}\left(\frac{\partial P'}{\partial z}+g\frac{P'}{C^2}\right)-\frac{1}{C^2}\frac{\partial^2 P'}{\partial t^2}=0. \tag{39}$$

C-is the speed of sound in the gravitational field of the Earth and is an important thermodynamic parameter of the medium, which, like other parameters, must depend on the vertical z coordinate. It is expressed in terms of the adiabatic C_s and isobaric C_p speeds of sounds by the following ratio [8]:

$$\frac{1}{C^2(z)}=\frac{1}{C_s^2(z)}+\frac{1}{C_p^2(z)}\Rightarrow C^2(z)=\frac{C_s^2(z)C_p^2(z)}{C_s^2(z)+C_p^2(z)}. \tag{40}$$

Taking into account (40), Equation (39) can be written as:

$$\Delta P'+\frac{g}{C_s^2(z)}\frac{\partial P'}{\partial z}-\frac{g}{C^2(z)}\left(\frac{1}{C^2(z)}\frac{\mathrm{d}C^2(z)}{\mathrm{d}z}+\frac{g}{C_p^2(z)}\right)P'-\frac{1}{C^2(z)}\frac{\partial^2 P'}{\partial t^2}=0. \tag{41}$$

We introduce the notation:

$$\begin{aligned} &\Gamma(z)=\frac{1}{C^4(z)}\frac{\mathrm{d}C^2(z)}{\mathrm{d}z}=\frac{1}{C_s^4(z)}\frac{\mathrm{d}C_s^2(z)}{\mathrm{d}z}+\frac{1}{C_p^4(z)}\frac{\mathrm{d}C_p^2(z)}{\mathrm{d}z}; \\ &\Sigma(z)=\frac{1}{C^2(z)C_p^2(z)} \end{aligned} \tag{42}$$

and then, Equation (39) finally takes the form:

$$\Delta P'+\frac{g}{C_s^2(z)}\frac{\partial P'}{\partial z}-g\left[\Gamma(z)+g\Sigma(z)\right]P'-\frac{1}{C^2(z)}\frac{\partial^2 P'}{\partial t^2}=0. \tag{43}$$

Equation (43) is the equation of mechanical waves in any medium in the earth's gravitational field, or the generalized equation of gravitational waves. We will search P' in the form:

$$P'(x,z,t)=P_a(z)\exp\left[i(kx-\omega t)\right] \tag{44}$$

after then from (43) we get:

$$\frac{\mathrm{d}^2P_a(z)}{\mathrm{d}z^2}+\frac{g}{C_s^2(z)}\frac{\mathrm{d}P_a(z)}{\mathrm{d}z}-\left\{k^2+g\left[\Gamma(z)+g\Sigma(z)\right]-\frac{\omega^2}{C^2(z)}\right\}P_a'(z)=0. \tag{45}$$

Let's denote the values related to air $(z>0)$ by index 1, and to water

$(z<0)$-by index 2 and consider this equation for air, where:

$$C_{s1}^2=\gamma\frac{k_BT}{m_0}\quad\text{and}\quad C_{p1}^2=\frac{c_pk_B^2T^3}{m_0^2g^2z^2}. \tag{46}$$

Substituting in (46) $T=\alpha+\beta z$, where $\alpha=288.15\ \mathrm{K}$ and $\beta=-6.52\times10^{-3}\ \mathrm{K/m}$, we get:

$$C_{s1}^2(z)=\gamma\frac{k_B(\alpha+\beta z)}{m_0}\quad\text{and}\quad C_{p1}^2(z)=\frac{c_pk_B^2(\alpha+\beta z)^3}{m_0^2g^2z^2}. \tag{47}$$

Thus, Equation (45) is a second-order differential equation with variable coefficients. Considering that $c_p=10^3\ \mathrm{J/kg\cdot K}$, to simplify the problem, we will average these coefficients in the range of heights from $z=z_0=10000\ \mathrm{m}$ after which the following relations are valid

$$\begin{aligned}
\frac{1}{\bar{C}_{s1}^2}&=\frac{1}{z_0}\int_0^{z_0}\frac{1}{C_{s1}^2(z)}\mathrm{d}z=9.80\times10^{-6}\frac{\sec^2}{\mathrm{m}^2},\\
\frac{1}{\bar{C}_{p1}^2}&=\frac{1}{z_0}\int_0^{z_0}\frac{1}{C_{p1}^2(z)}\mathrm{d}z=2.90\times10^{-6}\frac{\sec^2}{\mathrm{m}^2},\\
\frac{1}{\bar{C}_1^2}&=\frac{1}{z_0}\int_0^{z_0}\frac{1}{C_1^2(z)}\mathrm{d}z=12.70\times10^{-6}\frac{\sec^2}{\mathrm{m}^2},\\
\bar{\Gamma}_1&=\frac{1}{z_0}\int_0^{z_0}\Gamma(z)\mathrm{d}z=-1.30\times10^{-9}\frac{\sec^2}{\mathrm{m}^3},\\
\bar{\Sigma}_1&=\frac{1}{z_0}\int_0^{z_0}\Sigma_1(z)\mathrm{d}z=4.75\times10^{-11}\frac{\sec^4}{\mathrm{m}^4}.
\end{aligned} \tag{48}$$

Denoting $g(\bar{\Gamma}_1+g\bar{\Sigma}_1)=\bar{\Omega}_1=-8.19\times10^{-9}\ \mathrm{m}^{-2}$, Equation (45) for air takes the form:

$$\frac{\mathrm{d}^2P_{a1}(z)}{\mathrm{d}z^2}+\frac{g}{\bar{C}_{s1}^2}\frac{\mathrm{d}P_{a1}(z)}{\mathrm{d}z}-\left(k^2+\bar{\Omega}_1-\frac{\omega^2}{\bar{C}_1^2}\right)P_{a1}(z)=0. \tag{49}$$

We will seek a solution to Equation (49) in the form

$$P_{a1}(z)=A\exp(\gamma z), \tag{50}$$

which gives

$$P_{a1}(z)=A_1\exp(\gamma_1z)+A_2\exp(\gamma_2z), \tag{51}$$

where:

$$\gamma_1=-\frac{k}{\theta_{s1}}\left[1+\sqrt{1+\theta_{s1}^2\left(1+\frac{\bar{\Omega}_1}{k^2}-x^2\right)}\right], \tag{52}$$

$$\gamma_2=-\frac{k}{\theta_{s1}}\left[1-\sqrt{1+\theta_{s1}^2\left(1+\frac{\bar{\Omega}_1}{k^2}-x^2\right)}\right]. \tag{53}$$

Here $x=U_p/\bar{C}_1$ and $U_p=\omega/k$-is the phase velocity of the wave. θ_{s1}-dimensionless quantity that is equal to

$$\theta_{s1} = \frac{2k\bar{C}_{s1}^2}{g}. \tag{54}$$

It is easy to see that if the condition is met

$$1+\theta_{s1}^2\left(1+\frac{\bar{\Omega}_1}{k^2}-x^2\right)>1, \tag{55}$$

we have: $\gamma_1<0$, $\gamma_2>0$ and then, based on the wave surface condition ($P_{a1}(z)\to 0$ when $z\to\infty$), in (51) must be put $A_2=0$. On condition

$$0\le 1+\theta_{s1}^2\left(1+\frac{\bar{\Omega}_1}{k^2}-x^2\right)<1, \tag{56}$$

we have: $\gamma_1<0$, $\gamma_2<0$ a and then in the right part (51), both terms should be taken into account. Substituting the value $\bar{\Omega}_1=-8.19\times10^{-9}\ \mathrm{m}^{-2}$, the solutions of inequalities (56) and (56) are:

$$k>9.05\times10^{-5}\Big/\sqrt{1-x^2}\ \mathrm{m}^{-1}\Rightarrow\lambda<6.90\times10^4\sqrt{1-x^2}\ \mathrm{m}, \tag{57}$$

$$k<9.05\times10^{-5}\Big/\sqrt{1-x^2}\ \mathrm{m}^{-1}\Rightarrow\lambda>6.90\times10^4\sqrt{1-x^2}\ \mathrm{m}. \tag{58}$$

let's call the waves satisfying condition (57) wind waves, and condition (58) tsunami waves.

For water $(z<0)$ dependencies C_{s2} and C_{p2} on z are unknown, but we can assume with a high probability that they are very weak. As for their numerical values, they can be determined from the experimental data. Substituting in (28) the values of the coefficient of thermal expansion and specific heat capacity at constant pressure for water $\beta_p=1.50\times10^{-4}\ \mathrm{K}^{-1}$, $c_p=4.19\times10^3\ \mathrm{J/kg\cdot K}$ at a temperature of $T=288\ \mathrm{K}$ we obtain $C_{p2}=25210\ \mathrm{m/sec}$. On the other hand, the speed of sound in water, measured experimentally with great accuracy $C_2=1480\ \mathrm{m/sec}$, and then, from formula (17), we have $C_{s2}=C_2C_{p2}\Big/\sqrt{C_{p2}^2-C_2^2}=1482.60\ \mathrm{m/sec}$. As we can see, the speed of sound in water is practically equal to the adiabatic speed of sound, *i.e.* $C_2=C_{s2}$. This result irrefutably proves that the mechanism of sound propagation in water is adiabatic $(C_{p2}=\infty)$ and on the right-hand side of Equation (30) only the first term remains. Thus, Equation (45) for water has the form:

$$\frac{\mathrm{d}^2P_{a2}(z)}{\mathrm{d}z^2}+\frac{g}{C_2^2}\frac{\mathrm{d}P_{a2}(z)}{\mathrm{d}z}-\left(k^2-\frac{\omega^2}{C_2^2}\right)P_{a2}(z)=0 \tag{59}$$

Representing the amplitude of the pressure perturbation in the form $P_{a2}(z)=B\exp(\delta z)$ from (59), we obtain

$$P_{a2}(z)=B_1\exp(\delta_1 z)+B_2\exp(\delta_2 z), \tag{60}$$

where:

$$\delta_1=-\frac{k}{\theta_{s2}}\left[1+\sqrt{1+\theta_{s2}^2\left(1+\frac{\bar{\Omega}_2}{k^2}-\frac{U_{p2}^2}{C_2^2}\right)}\right]<0, \tag{61}$$

$$\delta_2 = -\frac{k}{\theta_{s2}}\left[1-\sqrt{1+\theta_{s2}^2\left(1+\frac{\bar{\Omega}_2}{k^2}-\frac{U_p^2}{C_2^2}\right)}\right] > 0, \tag{62}$$

$$\theta_{s2} = \frac{2kC_{s2}^2}{g} = \frac{2kC_2^2}{g}. \tag{63}$$

Obviously, since the region of water is bounded along the vertical coordinate, both terms in (60) must be preserved for it.

3. Waves of Wind

Let us first consider the wind waves, when the amplitudes of pressure perturbations in air and water are determined by a system of expressions:

$$\begin{cases} P_{a1}(z) = A\exp(\gamma_1 z) \\ P_{a2}(z) = B_1\exp(\delta_1 z) + B_2\exp(\delta_2 z) \end{cases} \tag{64}$$

Now it is necessary to determine the boundary condition connecting the perturbed pressures of two media on the water surface. Obviously, this condition must have the form

$$P_2\big|_{z=0} = P_1\big|_{z=0} + \rho_{02} g\varsigma(x,t), \tag{65}$$

where

$$\xi(x,t) = a\exp\left[i(kx-\omega t)\right] \tag{66}$$

is a displacement of the free surface of the liquid along the Z axis, a is the amplitude of the surface wave. Conditions for the continuity of the z components of perturbed air and water velocities at the air-water interface will give:

$$V_{z1}\big|_{z=0} = V_{z2}\big|_{z=0} = \frac{\partial\xi}{\partial t} \tag{67}$$

and at the bottom of the water $(z=-H)$ we will have:

$$V_{z2}\big|_{z=-H} = 0. \tag{68}$$

By representing $V_z(x,z,t)$ from the first equation of the system (35) in the form $V_z(x,z,t) = V_a(z)\exp\left[i(kx-\omega t)\right]$, we obtain:

$$V_z(x,z,t) = -\frac{i}{\rho_0\omega}\left[\frac{\mathrm{d}P_a(z)}{\mathrm{d}z} + \frac{g}{C^2}P_a(z)\right]\exp\left[i(kx-\omega t)\right]. \tag{69}$$

Taking into account expressions (64) and (69), the boundary conditions (47), (49) and (50) will have the form

$$\begin{cases} A_1 - B_1 - B_2 + \rho_{02} g a = 0 \\ \dfrac{1}{\rho_{01}\omega}\left(\gamma_1 + \dfrac{g}{\bar{C}_1^2}\right)A - \omega a = 0 \\ \dfrac{1}{\rho_{02}\omega}\left(\delta_1 + \dfrac{g}{C_2^2}\right)B_1 + \dfrac{1}{\rho_{02}\omega}\left(\delta_2 + \dfrac{g}{C_2^2}\right)B_2 - \omega a = 0 \\ \left(\delta_1 + \dfrac{g}{C_2^2}\right)\exp(-\delta_1 H)B_1 + \left(\delta_2 + \dfrac{g}{C_2^2}\right)\exp(-\delta_2 H)B_2 = 0 \end{cases} \tag{70}$$

By equating the determinant of the system (70) to zero, we obtain the dispersion equation for surface gravitational waves in the form

$$\tilde{\delta}_1 \exp\left(-\delta_1 H\right)\left[\frac{\tilde{\gamma}_1\tilde{\delta}_2 g}{\rho_{01}\omega^2}+\frac{\tilde{\delta}_2}{\rho_{02}}-\frac{\tilde{\gamma}_1}{\rho_{01}}\right]-\tilde{\delta}_2 \exp\left(-\delta_2 H\right)\left[\frac{\tilde{\gamma}_1\tilde{\delta}_1 g}{\rho_{01}\omega^2}+\frac{\tilde{\delta}_1}{\rho_{02}}-\frac{\tilde{\gamma}_1}{\rho_{01}}\right]=0, \quad (71)$$

where:

$$\tilde{\gamma}_1=\gamma_1+\frac{g}{\bar{C}_1^2}=\frac{2g}{\bar{C}_{p1}^2}+\frac{k}{\theta_{s1}}\left[1-\sqrt{1+\theta_{s1}^2\left(1+\frac{\bar{\Omega}_1}{k^2}-x^2\right)}\right], \quad (72)$$

$$\tilde{\delta}_1=\delta_1+\frac{g}{\bar{C}_2^2}=\frac{k}{\theta_{s2}}\left[1-\sqrt{1+\theta_{s2}^2\left(1+\frac{\bar{\Omega}_2}{k^2}-\frac{\bar{C}_1^2}{\bar{C}_2^2}x^2\right)}\right], \quad (73)$$

$$\tilde{\delta}_2=\delta_2+\frac{g}{\bar{C}_2^2}=\frac{k}{\theta_{s2}}\left[1+\sqrt{1+\theta_{s2}^2\left(1+\frac{\bar{\Omega}_2}{k}-\frac{\bar{C}_1^2}{\bar{C}_2^2}x^2\right)}\right]. \quad (74)$$

Considering that $x\le 1$, in formula (72) x^2 must be saved. As for the quantity $U_p^2/C_2^2=\left(\bar{C}_1^2/C_2^2\right)x^2\cong 0.03x^2$ it is of the second or greater order of smallness and in the linear approximation should be discarded in formulas (73) and (74). For the minimum value $k_{\min}=9.05\times10^{-5}\ \mathrm{m}^{-1}$ from (63) we have $\left(\theta_{s2}\right)_{\min}\cong 40$. Thus, we can neglect the unit in comparison with θ_{s2} and from (61), (62), (73) and (74) we have: $\delta_1=\tilde{\delta}_1=-k$, $\delta_2=\tilde{\delta}_2=k$. Then from (71) we easily obtain:

$$\left\{\frac{2}{\theta_{p1}}+\frac{1}{\vartheta_{s1}}\left[1-\sqrt{1+\theta_{s1}^2\left(1+\frac{\bar{\Omega}_1}{k^2}-x\right)}\right]\right\}\times\left[\frac{kg}{\omega^2}\tanh\left(kH\right)-1\right]$$
$$-k\frac{\rho_{01}}{\rho_{02}}\tanh\left(kH\right)=0, \quad (75)$$

where $\theta_{p1}=2k\bar{C}_{p1}^2/g$. Equation (75) is a dispersion equation for wind waves and, as we see, it contains the thermodynamic parameters of both water and air. The last summand in (75) can be neglected due to its smallness, and then, it splits into two equations:

$$\frac{2}{\theta_{p1}}+\frac{1}{\theta_{s1}}\left[1-\sqrt{1+\theta_{s1}^2\left(1+\frac{\bar{\Omega}_1}{k^2}-x^2\right)}\right]=0, \quad (76)$$

$$\frac{kg}{\omega^2}\tanh\left(kH\right)-1=0. \quad (77)$$

Equation (76) describes longitudinal waves in the air excited by perturbations on the surface of water and propagating along this surface. Let's consider this equation, ignoring the dependence of the sounds velocities on the coordinate z, *i.e.* take their values at sea level: $\bar{C}_1=C_1(0)=C_{s1}(0)=340\ \mathrm{m/sec}$, $C_{p1}=\infty$. Then $\bar{\Omega}_1(z)=0$ and it will take the for

$$1-\sqrt{1+\theta_{s1}^2\left(1-x^2\right)}=0\Rightarrow|x|=1\Rightarrow|U_p|=C_1(0). \quad (78)$$

Obviously, in this case, the condition (57) is optional and the wave vector k

can take any value. We see that in this assumption, the speed of the wave in the air does not depend on the wavelength and is equal to the speed of sound at sea level. There is no doubt that perturbations whose wave lengths are comparable to the lengths of surface gravitational waves cannot propagate in the air at the speed of sound, and thus it is impossible to ignore the dependence of the speed of sound on the z coordinate.

The solution to Equation (76) is:

$$|x| = \sqrt{1 + \frac{\vec{\Omega}_1}{k^2} - \frac{4}{\theta_{p1}\theta_{s1}}\left(1 + \frac{\theta_{s1}}{\theta_{p1}}\right)} = \sqrt{1 + \frac{1}{k^2}\left[\bar{\Omega}_1 - \frac{g^2}{\bar{C}_{p1}^2\bar{C}_{s1}^2}\left(1 + \frac{\bar{C}_{s1}^2}{\bar{C}_{p1}^2}\right)\right]}. \tag{79}$$

Substituting in (79) the numerical values of the parameters from (48), we get

$$|x| = \sqrt{1 - \frac{1.17\times10^{-8}}{k^2}} \Rightarrow |U_p| = \sqrt{1 - \frac{1.17\times10^{-8}}{k^2}}\bar{C}_1. \tag{80}$$

From (80) we find

$$k = \frac{1.08\times10^{-4}}{\sqrt{1-x^2}} \tag{81}$$

which is consistent with the condition (57) and therefore the roots (79) are not extraneous for any valid values k. Table 1 shows the roots of Equation (62) for those values $k\ (\lambda)$, that satisfy condition (57)

We see that for $k > 10^{-3}\ \mathrm{m}^{-1} \Rightarrow \lambda < 6.28\times10^{3}\ \mathrm{m}$, the phase velocity of the wave in the air is constant and equal to $U_p = \bar{C}_1 \cong 281.72\ \mathrm{m/sec}$, and then, with a decrease of k, it falls and at $k = 1.08\times10^{-4}\ \mathrm{m}^{-1} \Rightarrow \lambda = 5.80\times10^{4}\ \mathrm{m}$, we have $x = 0$, *i.e.*, the wave stops. This means that at this wavelength, two regions of about 30 km long, with high and low pressures are formed in the atmosphere above water. In the area of low atmospheric pressure, the amplitude of the surface wave will increase and the pressure difference between the two areas will lead to the appearance of wind. When is further reduced k to its minimum value, which is defined from (57) $k_{\min} \cong 9\times10^{-5}\ \mathrm{m}^{-1} \Rightarrow \lambda_{\max} \cong 7\times10^{4}\ \mathrm{m}$, the roots of Equation (76) and the corresponding frequencies of standing waves in the atmosphere become imaginary, which leads to a sharp increase in the pressure difference and, consequently, the amplitude of the surface wave and the wind force. This result clearly explains the reason for the drop in atmospheric pressure over the sea and ocean before the storm, as well as the reason for its strengthening.

Table 1. The dependence of the roots of Equation (76) on $k \geq 9.05\times10^{-5}\ \mathrm{m}^{-1} \Rightarrow \lambda \leq 6.94\times10^{4}\ \mathrm{m}$.

$k\left(\mathrm{m}^{-1}\right)$	10^{-1}	10^{-2}	10^{-3}	1.10×10^{-4}	1.08×10^{-4}	1.06×10^{-4}	9.05×10^{-5}
$\lambda(\mathrm{m})$	62.80	628.00	6280.00	5.71×10^{4}	5.81×10^{4}	5.92×10^{4}	6.94×10^{4}
x	1.00	1.00	0.99	0.17	0.00	0.09i	0.66i
$U_p\left(\mathrm{m/sec}\right)$	281.72	281.72	278.90	47.89	0.00	25.35i	186.59i

The solution to Equation (77), which describes surface gravitational waves on water, is:

$$\left|U_p\right| = \sqrt{\frac{g}{k}\tanh\left(kH\right)} = \sqrt{\frac{g\lambda}{2\pi}\tanh\left(\frac{2\pi H}{\lambda}\right)} \tag{82}$$

which coincides with (10) and thus, the existing theory for wind waves gives the correct result.

The phase velocity of a wave in the atmosphere does not depend on the depth of the ocean and decreases from the speed of sound to zero with increasing wavelength. In the ocean, on the contrary, the phase velocity increases from zero with increasing wavelength and depth. It is evident, that at certain wavelengths, which will depend on the depth of the ocean, these speeds will coincide and resonance of the frequencies will occur of waves in the atmosphere and the ocean. We can assume that this resonance is the cause of the "killer wave", especially since there is no other explanation yet. It should be noted, that the resonance alone is not enough for the appearance of a "killer wave"—it is essential that the oscillations in the air and in the water are occurring in antiphase.

The resonant wavelengths are obtained by equating the right sides of the Equations (80) and (82), *i.e.*

$$\sqrt{1-\frac{1.17\times10^{-8}\lambda^2}{4\pi^2}}\bar{C}_1 = \sqrt{\frac{g\lambda}{2\pi}\tanh\left(\frac{2\pi H}{\lambda}\right)} \tag{83}$$

For deep water $\left(2\pi H/\lambda > 1 \Rightarrow \tanh\left(2\pi H/\lambda\right) = 1\right)$, the Equation (83) obtains

$$1-\frac{1.17\times10^{-8}}{4\pi^2}\lambda^2 = \frac{g}{2\pi\bar{C}_1^2}\lambda \tag{84}$$

The Equation (84) does not depend on H and its solution is $\lambda \cong 6.4\times10^4$ m so therefore $H > 10^4$ m. Since there are practically no such depths with a flat relief, it can be said with great confidence that "killer waves" do not arise in deep water.

For the shallow water $\left(2\pi H/\lambda < 1 \Rightarrow \tanh\left(2\pi H/\lambda\right) = 2\pi H/\lambda\right)$ we have

$$\lambda = 2\pi\times10^4\sqrt{\frac{1}{1.17}\left(1-\frac{gH}{\bar{C}_1^2}\right)} \tag{85}$$

In order for the values λ from (85) to satisfy the shallow water condition, the following condition must be fulfilled

$$\left(1-\frac{gH}{\bar{C}_1^2}\right) > 1.17\times10^{-8}H^2 \tag{86}$$

which obtains the following $H < 5.3\times10^3$ m. Therefore, the "killer waves" arise in shallow water, the depth of which does not exceed 5.3 km.

The values of the killer wavelength for different depths and the corresponding values of the phase velocities, calculated by the formulas (85) and $U_p = \sqrt{gH}$ as well as the values of the periods and the ratio $2\pi H/\lambda$ are given in **Table 2** below.

Table 2. Values of lengths, phase velocities and periods of the "killer wave" for various depths.

$H(\mathrm{m})$	$\lambda(\mathrm{m})$	$U_p(\mathrm{m/sec})$	$T(\mathrm{sec})$	$2\pi H/\lambda$
200	57,331	44.3	1300	0.02
400	56,595	62.6	904	0.04
600	55,849	76.7	728	0.07
800	55,092	88.5	622	0.09
1000	54,325	99.0	549	0.11

Figure 4 presents the graphs of the dependences of the phase velocities of longitudinal waves in the air (80) and on the water surface (82) for the same depths. We can see that the values of the resonant wavelengths coincide with the tabulated values.

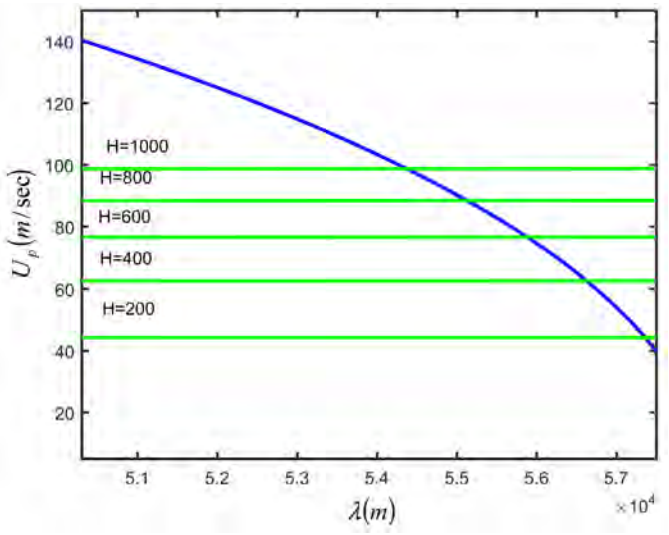


Figure 4. Plots of dependences of the phase velocities of longitudinal waves on the air (80) (blue curve) and on the water surface (82) (green curves) for different depths.

4. Tsunami Waves

Let us now consider tsunami waves, *i.e.* surface gravitational waves for those values k that satisfy condition (58): $k < 9.05\times10^{-5}\ \mathrm{m}^{-1} \Rightarrow \lambda > 6.90\times10^{4}\ \mathrm{m}$. In this case, the amplitude of longitudinal waves in the atmosphere is determined by the expression (51), where $A_1 \cong A_2 < 1$ and according to (52), (53) and (56), $\gamma_1 < 0$ and $\gamma_2 < 0$, *i.e.* both terms give the wave attenuation at $z \to \infty$. It can be seen that the wave corresponding to the first term decays faster than the wave corresponding to the second term, and therefore, to fulfill the condition of the wave's superficiality, it is sufficient to leave only the second term and thus we have:

$$\begin{cases} P_{a1}(z) = A\exp(\gamma_2 z) \\ P_{a2}(z) = B_1\exp(\delta_1 z) + B_2\exp(\delta_2 z) \end{cases} \tag{87}$$

Note that boundary conditions (65), (67), and (68) are also valid for tsunami waves, and then, comparing (61) and (83), it is easy to verify that for tsunami waves we obtain an equation similar to Equation (71)

$$\tilde{\delta}_1 \exp\left(-\delta_1 H\right)\left[\frac{\tilde{\gamma}_2\tilde{\delta}_2 g}{\rho_{01}\omega^2}+\frac{\tilde{\delta}_2}{\rho_{02}}-\frac{\tilde{\gamma}_2}{\rho_{01}}\right]-\tilde{\delta}_2 \exp\left(-\delta_2 H\right)\left[\frac{\tilde{\gamma}_2\tilde{\delta}_1 g}{\rho_{01}\omega^2}+\frac{\tilde{\delta}_1}{\rho_{02}}-\frac{\tilde{\gamma}_2}{\rho_{01}}\right]=0. \quad (88)$$

For waves, the length of which is $\lambda \geq 500$ km, the value $\underline{\theta_{s2}}$ changes in the interval $\theta_{s2} \leq 5.5$ and, therefore, it cannot be neglected in comparison with the unit, in contrast to θ_{s2}^2. Then we will have:

$$\begin{cases} \tilde{\gamma}_2 = \dfrac{2}{\theta_{p1}}+\dfrac{1}{\theta_{s1}}\left[1+\sqrt{1+\theta_{s1}^2\left(1+\dfrac{\bar{\Omega}_1}{k^2}-x^2\right)}\right] \\ \delta_1 = -\tilde{\delta}_2 = -k\left(1+\dfrac{1}{\theta_{s2}}\right), \delta_2 = -\tilde{\delta}_1 = k\left(1-\dfrac{1}{\theta_{s2}}\right) \end{cases} \quad (89)$$

after which, Equation (88) takes the form

$$\begin{aligned} &\left\{\frac{2}{\theta_{p1}}+\frac{1}{\vartheta_{s1}}\left[1+\sqrt{1+\theta_{s1}^2\left(1+\frac{\bar{\Omega}_1}{k^2}-x\right)}\right]\right\} \\ &\times\left[\left(\frac{kg}{\omega^2}+\frac{1}{\theta_{s2}}\right)\tanh\left(kH\right)-1\right]-k\frac{\rho_{01}}{\rho_{02}}\tanh\left(kH\right)=0 \end{aligned} \quad (90)$$

Ignoring the last term, Equation (90) again splits into two equations:

$$\frac{2}{\theta_{p1}}+\frac{1}{\theta_{s1}}\left[1+\sqrt{1+\theta_{s1}^2\left(1+\frac{\bar{\Omega}_1}{k^2}-x^2\right)}\right]=0, \quad (91)$$

$$\left(\frac{kg}{\omega^2}+\frac{1}{\theta_{s2}}\right)\tanh\left(kH\right)-1=0. \quad (92)$$

From condition (56) it follows that Equation (91) has no solution and this means that during a tsunami, wave processes in the atmosphere are not generated. As for Equation (92), its solution is

$$U_p=\sqrt{\frac{\left(g/k\right)\tanh\left(kH\right)}{1-\tanh\left(kH\right)/\theta_{s2}}}. \quad (93)$$

Considering that $kH<1 \Rightarrow \tanh\left(kH\right)\cong kH$, we will have:

$$U_p=\sqrt{\frac{gH}{1-\dfrac{gH}{2C_2^2}}}. \quad (94)$$

For great depths, for example, at $H=10^4$ m, the value $gH/2C_2^2=0.02$ and it can be ignored. Thus, the phase velocity of the tsunami wave is $U_p=\sqrt{gH}$.

5. Conclusions

This article does not claim to be highly accurate or to be the ultimate truth. The

upper boundary of the troposphere changes depending on geographic parameters, and this concludes, that the average values of the problem parameters calculated here and, therefore, all the numerical data given in the article are rather conditional. However, undoubtedly the proposed method for solving the problem is new and makes it possible to trace the correlation between the ocean and the atmosphere during wave processes. In particular, it became clear why the atmospheric pressure in the ocean drops before the storm, as well as differentiating between the wavelengths of wind and tsunami became possible.

Its apparent advantage is also that at the level of a highly plausible hypothesis, it reveals the greatest mystery of nature called the "Killer Wave". Now it is clear why this wave is solitary. This is due to the fact that the flat topography of the ocean floor is disturbed at the distances of the order of the wavelengths that we calculated. It is also clear why a cavity, is formed before the wave. This is attributed to the fact that there is a region in front of the wave, where the pressure in the water sharply drops and in the atmosphere sharply increases.

All this became possible after the discovery of the isobaric speed of sound and the dependence of the true value of the speed of sound in the atmosphere on the vertical coordinate. This led to a radical change in many established dogmas and ideas in aero and hydrodynamics, which are recognized by the international scientific community (work [8] is posted on several sites on the Internet (see, for example, [16]) and work [14] is posted in the NASA database [17]). Despite this, the Internet [18] to this day gives the values of the speed of sound at different heights of the atmosphere, calculated by the formula (15). Obviously, this can be explained by the fact that our theoretical results have not been experimentally confirmed. We hope that this article will be able to help popularize this problem and in the relevant scientific community will show desire in conducting the necessary experiments. The importance of such experiments also lies in the fact that if it is possible in laboratories to create conditions corresponding to the upper boundary of the troposphere and to confirm the fact of heat release, we will get an alternative source of energy.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Kirtskhalia, V.G. (2016) *Journal of Fluids*, **2016**, Article ID: 4519201. https://doi.org/10.1155/2016/4519201

[2] Landau, L.D. and Lifchitz, E.N. (1988) Theoretical Physics, Hydrodynamics. Vol. 6, Nauka, Moscow.

[3] Kowalik, Z. (2012) Introduction to Numerical Modeling of Tsunami Waves. Institute of Marine Science University of Alaska, Fairbank. https://www.sfos.uaf.edu/directory/faculty/kowalik/Tsunami_Book

[4] Stoker, J.J. (1957) Water Waves. Interscience, New York.

[5] Whitham, G. (1974) Linear and Nonlinear Waves. John Wiley & Sons (Wiley-Interscience), New York.

[6] Gossard, E.E. and Hooke, W.H. (1975) Waves in the Atmosphere.

[7] Wang, G.S.K. (1986) *The Journal of the Acoustical Society of America*, **79**, 1359-1366.

[8] Kirtskhalia, V.G. (2012) *Open Journal of Acoustics*, **2**, 80-85. https://doi.org/10.4236/oja.2012.22009

[9] Kirtskhalia, V. (2012) *Open Journal of Acoustics*, **2**, 115-120. https://doi.org/10.4236/oja.2012.23013

[10] National Aeronautics and Space Administration (1976) U.S. Standard Atmosphere.

[11] Kirtskhalia, V.G. (2021) *IOP Conference Series: Materials Science and Engineering*, **1024**, Article ID: 012037. https://doi.org/10.1088/1757-899X/1024/1/012037

[12] Sorokhtin, O.G. (2009) The Process of Absorption of Ultra-Violet Radiation of the Sun Terrestrial Atmosphere. Bulletin of the Russian Academy of Natural Sciences, 2009/3.

[13] Kirtskhalia, V.G. (2015) *Journal of Modern Physics*, **7**, 948-954. https://doi.org/10.4236/jmp.2015.67099

[14] Kirtskhalia, V.G. (2013) *Journal of Modern Physics*, **4**, 1075-1079. https://doi.org/10.4236/jmp.2013.48144

[15] Kirtskhalia, V.G. (2019) *Journal of Modern Physics*, **10**, 452-458. https://doi.org/10.4236/jmp.2019.104030

[16] Kirtskhalia, V.G. (2012) *Open Journal of Acoustics*, **2**, 80-85. https://www.researchgate.net/publication/274750587_Speed_of_SoundinAtmospereof_the_Earth

[17] The Smithsonian/NASA Astrophysics Date System.

[18] http://www.sengpielaudio.com/calculator-speedsound.htm

Journal of Modern Physics, 2021, 12, 1366-1399
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Quantum Mysteries for No One

Frank Lad

Department of Mathematics and Statistics, University of Canterbury, Otautahi/Christchurch, New Zealand
Email: frank.lad@canterbury.ac.nz

How to cite this paper: Lad, F. (2021) Quantum Mysteries for No One. *Journal of Modern Physics*, **12**, 1366-1399.
https://doi.org/10.4236/jmp.2021.129082

Received: May 14, 2021
Accepted: July 23, 2021
Published: July 26, 2021

Abstract

I provide a critical reassessment of David Mermin's influential and misleading parable, "Quantum Mysteries for Anyone", identifying its errors and resolving them with a complete analysis of the quantum experiment it is meant to portray. Accessible to popular readership and requiring no knowledge of quantum physics at all, his exposition describes the curious behaviour of a machine that is designed to parody the empirical results of quantum experiments monitoring the spins of a pair of electrons under various conditions. The mysteries are said to unfold from contradictory results produced by a signal process that is proposed to explain them. I find that these results derive from a mathematical error of neglect, coupled with a confusion of two distinct types of experiments under consideration. One of these, a gedankenexperiment, provides the context in which the fabled defiance of Bell's inequality is thought to emerge. The errors are corrected by the recognition of functional relations embedded within the experimental conditions that have been long unnoticed. A Monte Carlo simulation of results in accord with the actual abstemious claims of quantum theory supports probability values that Mermin decries as unwarranted. However, the distribution it suggests is not definitive, in accord with the expressed agnostic position of quantum theory regarding measurements that cannot be executed. Bounding quantum probabilities are computed for the results of the gedankenexperiment relevant to Bell's inequality which inspired the parable. The problem is embedded in a 3 × 3 design of Stern-Gerlach magnet orientations at two observation stations. Computational resolution on the basis of Bruno de Finetti's fundamental theorem of probability requires the evaluation of a battery of three paired linear programming problems. Though technicalities are ornate, the message is clear. There are no mysteries of quantum mechanics that derive from mistaken understandings of Bell's inequality… for anyone.

Keywords

Bell Inequality Violation, Fundamental Theorem of Probability, Local Realism, Probability Bounds

1. Reassessing the Quantum Mysteries of David Mermin

Detailed mathematical formalities of theoretical quantum mechanics preclude their understanding by even the technically sophisticated among the generally educated public. Replete with measurement operator matrices on a Hilbert space of quantum states and a peculiar style of notation that is unique to them, engagement with their prescriptions is forbidding. Aware of the widespread public interest in the inscrutable content of the theory, David Mermin [1] devised an engagingly simple parable to provide an exhibition of touted features of mysterious quantum behaviour as they have been long understood. Requiring no knowledge of any aspect of quantum physics at all, the exposition merely describes a machine that sends a pair of balls in opposite directions from a central station C to detectors at stations A and B. The balls can address each detector in three different ways, represented by three numbered settings of a dial on its face. Thus, there are nine different conditions under which an experimental run of the machine can be conducted. Coloured lights, either red or green, at the two detectors, provide signals as to what occurs in the encounters of the balls at the two stations. Statistical properties of the signal performance in a sequence of operations of the device are reported and explained in such a way as to exhibit one of the defining puzzling mysteries of quantum theory: the purported defiance of Bell's inequality by the probabilistic behaviour of entangled particles.

Questions arise concerning the physical process of production of the machine's output. This evidently involves entangled probabilities of light signals at the stations A and B, each of which depends on both the dial setting at its own station and the setting at the other station. This is despite the fact that there is no physical connection between the stations which might convey information between them regarding their respective dial settings. An information transmission scheme is envisioned by which the pair of balls may carry within themselves unobserved encoded messages to stimulate the observed entangled behaviour of the light signals. Although this is shown capable of accounting for regularly matching signals when the dial settings are identical, an enigma arises when the settings are different. Any such scheme appears to instigate matching light signals at the two stations with a frequency exceeding 1/3 in situations for which the machine is known to exhibit such signalling with probability of only 1/4. The machine behaviour is touted as mysterious, defying explanation by encoded messages and portraying one of the great mysteries of quantum analysis.

Upon completion of the exhibition, it is explained to any QM-enlightened readership that the parable of the mysteries actually mimics the situation of a real quantum experiment. This would involve the transmission of a pair of electrons in opposite directions over long distances toward two observation stations at which Stern-Gerlach magnets identify the electromagnetic spin of each electron as directed up or down. The magnets at the two stations can each be set up in any of three differently angled directions perpendicular to the direction of the incoming electrons. These alternative directions are represented in the parable by the three different settings of the dials at stations A and B. The statistics re-

ported in the parable summarizing signal behaviour of the machine over sequences of experiments at each dial pairing correspond to what is expected of the spin observations according to the principles of quantum theory.

The reported results are both simple and stunning. The professor teased that he could actually create this machine using the results of paired quantum experiments as the generators of the random outcome sequences. Requiring an effort which he assured would be somewhat less than the order of the Manhattan project, he proclaimed that "the conundrum posed by the behavior of the device is no mere analogy, but the atomic world itself, acting at its most perverse."

So engaging, simple, and startling is Mermin's exhibition that the piece has become standard fare for the exposition of Bell's inequality to students ever since, both students of physics and of philosophy, even at graduate levels. Moreover, it is included in a welcome and popular collection of his essays on matters of theoretical physics meant for the generally educated public, *Boojums all the way through: communicating science in a prosaic age*. Immensely successful and influential, it has been reprinted by now in nine hardcover and five paperback editions. The exposition of "Quantum Mysteries" was lauded by Richard Feynman as "one of the most beautiful papers in physics that I know of" according to the preface to the volume [2].

However, the lionization of Mermin's article by another leading figure of twentieth century physics does not make it correct. I make bold here to display that his amusing allegorical presentation of the situation is both mistaken and misleading. A recognition of my assessment here, in tandem with my arguments in [3] and [4] will suggest a revision of physicists' attitudes towards the interpretation of quantum theory, and the mistaken supposed defiance of Bell's inequality in particular.

David Mermin is one of the most accomplished physicists of our era, a cherished professor in the Department of Physics at Cornell University for many years. It is edifying to view his curriculum vitae at the public website at cornell.edu. Along with the publication of his own extensive research results, he has been seriously committed to the exposition of contemporary findings of theoretical and applied physics to the general public. In addition to 138 technical publications, his curriculum vitae includes 20 pedagogical articles and 29 general writings. This does not make him immune to mistakes. We all make mistakes. It is with due respect for his accomplishments and appreciation of his personal style that I explain in this article the serious consequences of his mistaken understanding of Bell's theorem and its implications. Feeling gauche to refer to him regularly throughout the essay as "Mermin", I refer to him alternately as "the professor". I intend this with respect.

Of course I invite you to read or to reread the parable of "Quantum Mysteries for Anyone" for yourself, along with the preface to *Boojums* as a prelude to studying my exposition. Both are available online. However, to make this presentation self-contained I will begin Section 2 with a faithful outline of the mysteries as they are portrayed, firstly with a description of the properties of the ma-

chine's operation, and then with a display of the mysterious behaviour attributed to its conduct. Reflection on the structure of the argument allows us to recognize a sleight of hand in its application to the description of the machine's activity. This involves consideration of a gedankenexperiment which underlies the supposed defiance of Bell's inequality. Section 3 is devoted to the structure of the material problem of quantum physics that the professor would have us ignore in deference to thinking about his wondrous machinery. We shall find that the actual problem under consideration involves a system of restrictive functional restrictions that are ignored in his proposed assessment of its behaviour.

Section 4 presents a Monte-Carlo simulation of the quantum gedankenexperiment which recognizes these functional relations, displaying a frequency of matching lights on the order of 0.375 in situations for which Mermin proclaims his machine to provide only 0.25. The simulation subscribes completely to the probabilities specified by quantum theory in all appropriate instances. Section 5 then completes the computational analysis of the quantum gedankenexperiment, relying on the application of Bruno de Finetti's fundamental theorem of probability to identify the bounds on the relevant probabilities that quantum theory actually motivates. Quantum theory is quite explicit in refraining from asserting joint probabilities for the outcomes of measurement operators that do not commute. Nonetheless the prescriptions it motivates do imply precise bounds on such probabilities which cohere with the explicit assertions it does provide. My discussion concludes in Section 6 with an overview of what is to be learned from this exercise, and a recognition of precursings of the analysis here already aired in the technical physics literature.

2. Mermin's Machine and Its Puzzling Properties

From a box, labeled *c* in Figure 1, two apparently indistinguishable balls are ejected in opposite directions toward identical receivers at stations labeled *a* and *b*. There are no discernible connections between these components of the machine, *a*, *b*, and *c*. Each receiver has a dial on its face that can be positioned to any one of three settings, numbered 1, 2, and 3. Neither receiver is advised of the dial setting on the other receiver. In whatever way this pair of dials are set, when the balls enter the receivers at the stations, each station will flash one of two lights, coloured Green and Red. The results of a sequence of machine experiments are recorded using ordered notation such as 12*GR*. This would designate that an observation was made with the dial at *a* set to 1 and that at *b* set to 2, and that the coloured light observed at *a* is Green while that at *b* is Red. Thus, a

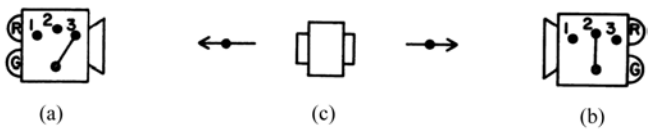


Figure 1. "The complete device. (a) and (b) are the two detectors. (c) is the box from which the two particles emerge." The original caricature of Mermin's mysterious machine, reprinted with permission from *The Journal of Philosophy* 78(7): 397-408, 1981.

sequence of such observations at various dial settings might look something like 12*GG*, 31*RG*, 13*GR*, 22*RR*, 32*RR*, 12*RG*, 21*GR*, and so on.

Here is how the machine works. A pair of apparently identical balls is ejected from box *C* in opposite directions toward the detectors *A* and *B* at whatever dial settings are arranged there, and the signal lights are observed. Once the result is recorded at these settings, another pair of balls is sent out to the detectors in whatever numbered dial settings are then arranged for them. Such experimentation continues sequentially with new pairs of indistinguishable balls. When the dials at *A* and *B* happen to be set to the same number for any run, then the colours of the flashing lights are always observed to be either both Red or both Green, with equal frequencies of 1/2 as *GG* and as *RR*. On the other hand, when the dials are set at different numbers, the signal lights flash the same colour 1/4 of the time, and flash different colours 3/4 of the time. In the former cases, half of the time the identical colours show *GG*, and half the time they show *RR*. In the latter cases, half of the time the flashing colours show *GR*, and half the time they show *RG*. There is no apparent regularity in the orders of their appearance.

The puzzling question and the source of the mystery involved concerns the determination of what could account for such observable results. It seems odd that the signal behaviour at each detector depends on the dial setting at the other detector, yet there is no obvious way for the two receivers to communicate with one another as to the positions of their dial settings. Proposed as a solution is that while the two balls sent to *A* and *B* are apparently identical to one another in every way, the character of each pair may be different in successive runs in a way that is not noticeable to the eye. On any given run, the two balls may be somehow encoded each with the same one of eight possible labels: *GGG*, *GGR*, *GRG*, *RGG*, *GRR*, *RGR*, *RRG*, or *RRR*. During a long sequence of runs, the source bin of the pairs of balls provides equal numbers of balls encoded with each of these eight configurations, in a random order. When either ball from an identically encoded pair such as *RGR*, for example, enters a detector station, the signal light would flash Red if the dial at that station were set at 1, would flash Green if the dial were set at 2, and would flash Red if it were at 3. That is, whatever the encoded message on the pair of balls may be, the colour flashed at each detector would match its dial setting with the associated colour designated at that position on the ball's encoding string.

Such a scheme would easily account for the fact that when the identically encoded pair of balls enter the two detectors, the signal lights would always flash the same colour if the detector dials were set to the same number. The balls would be coloured either both Green or both Red, depending on the specific identical encoding of the pair of balls and the setting of the dials. But what if the dials at *A* and *B* point to different number settings?

2.1. Mysterious Behaviour: Can You Believe It?

The professor pronounces that if such a scheme were in vogue, the proportion of runs in which the lights signal the same colour would exceed 1/3 whenever the

dials are set differently at the two stations. This would obviously defy the known result that the frequency with which matching lights are observed in such situations equals 1/4. Here is his reasoning. The dial settings at *A* and *B* are different from one another in six of the nine paired dial settings: 12, 13, 21, 23, 31, and 32. If the encoded message were *RRR* or *GGG*, the signal lights would always shine the same colour as the encoded ball enters the detectors at these settings. Now consider any encoded message that involves two designations of one colour and one of the other, for examples *RRG* or *GRG*. In response to such an encoding, say *RRG*, the colour signals at the two stations would be the same for two of the six possible dial settings of the receivers (settings 12 and 21). In the other four settings (13, 23, 31 and 32) the colour signals would be different. Thus, when responding to a long sequence of balls encoded in any of the eight ways, the lights will signal the same colour in at least 1/3 of the runs when the dial settings differ. Responding to two of the codes they always would match, while to any one of the other six codes the coloured lights would match at two of the six differing dial pairings. This argument is said to display the mysterious character of this machine. The encoding of the balls would suffice to explain why the flashing lights show the same colour when the dials are set identically at *A* and *B*. However, when the dials are set to different numbers, the encoding scheme would seem to imply that at least 1/3 of the observations should exhibit matching colours. Mysteriously, the machine is known to produce matching colours in only 1/4 of the runs with such settings.

An invisible encoding of the balls seems to contradict the facts of the empirical observations of the machine performance. It appears that the proposal of the hidden encoding cannot account for the facts. No other proposal has been offered that can account for the assured matching light signals whenever the dials happen to be set the same. There must be some mysterious connection between the machine components and the observation process itself to account for the facts.

Would you like to dwell on this puzzle yourself for a while if you have not already done so? Literally thousands upon thousands of people have done so, and have been taken in by a sleight of hand in the argument. Without warning or fuss, Professor Mermin has switched the setting of the game on us! Rather than counting the spin products as each pair of balls enters the machine at a dial setting, he is counting the spin products for each pair of balls as it would pass all six of the mixed dial settings. His reported lighting statistics pertains to one game, and his counting of the matching colours pertains to another, two completely different games. As we shall see, it makes sense to consider both games, but they are different for different reasons. We require some more thinking.

2.2. The Sleight of Hand

We could use a random number generator both to simulate both the behaviour of the mysterious machine and to simulate the emission of colour-coded balls, with quite different results. In the first case, it is easy enough to generate colour

signals that always appear randomly as *GG* or *RR* when the dials are set identically, but that appear in each of these ways only 1/8 of the time when the dials are set differently, while appearing then as *GR* and *RG* each 3/8 of the time. In the second case we would randomly pick an encoding design for the emitted pair of balls and a paired dial setting for the receivers, determining the light signals from the coding rule. These latter results would not match the operations of Mermin's machine. Examine firstly the columns of Table 1. These each display the light signal responses of the machines at one of the nine dial settings to individual pairs of the eight ball encodings listed along the rows, as described in Mermin's proposal. In columns for which the dials are set differently, the light signals are matching for 1/2 of the ball codings, not 1/4. The proportion of matching lights would reduce to 1/3 only if balls coded *GGG* or *RRR* were never introduced to the detectors, while the six mixed encodings were introduced at random. The proportion could reduce to 1/4 only if the distribution of emitted ball encodings varied according to the paired dial settings.

Table 1. Encoded messages and their induced responses. The minus symbol (−) designates the display of matching light colour signals, and the plus symbol (+) designates different coloured lights.

Dials	11	12	13	21	22	23	31	32	33
Setting	1	2	3	4	5	6	7	8	9
GGG	−	−	−	−	−	−	−	−	−
GGR	−	−	+	−	−	+	+	+	−
GRG	−	+	−	+	−	+	−	+	−
RGG	−	+	+	+	−	−	+	−	−
GRR	−	+	+	+	−	−	+	−	−
RGR	−	+	−	+	−	+	−	+	−
RRG	−	−	+	−	−	+	+	+	−
RRR	−	−	−	−	−	−	−	−	−

However, the professor motivated his claims regarding matching-colour frequencies *among encoded balls* by an argument based on a different situation: experimental results from sending each single pair of encoded balls to detectors at all nine dial setting pairs. His count of two matching lights among six observations arose from observing each pair of mixed-encoded balls such as RRG as it enters all six distinct station pairings with differing dial settings: 12, 13, 21, 23, 31, 32. These counts are exemplified in the *rows* of Table 1. To propose that counts from this experiment can represent counts from the original experiment (each pair of balls addresses only a single setting of the dials) amounts to a sleight of hand.

Now who said anything about subjecting a pair of encoded balls to all nine of the paired dial settings? In the operations of the machine which exhibit 1/4 matching lights at different dial settings, each pair of balls is ejected toward a single setting of the dials at stations *A* and *B*, and the result might be recorded as

something such as 32*GR* or 13*RG*. But not both!… not to speak of results of this pair of balls sent to the receivers at the seven other dial pairings. If we would like to study the observed light signals when any single pair of encoded balls is sent to the detectors at all nine paired dial settings, we would require a recording structure more elaborate. We shall have reason to make such a study if we are to examine the relevance of the machine behaviour to the touted violation of Bell's inequality by the probabilities of quantum physics, and we shall.

Suppose we order the detector dial settings as 11, 12, 13, 21, 22, 23, 31, 32, 33, and send each pair of identical uncoded balls to all of them. Designating matching-light-colour observations by a −1 and mixed-light-colour observations by +1, the experimental results would be recorded not merely by something like 13*RG*, but rather something like $(-1,-1,+1,+1,-1,+1,-1,+1,-1)$. Recognizing that the vector components 1, 5, and 9 must all equal −1 because the light colours surely match at these settings, it would appear there would be scope for a sizeable number of distinct observation vectors to arise from such a 9-ply experimental run, perhaps even $2^6 = 64$ of them if the other six might each be either −1 or +1. We shall learn about this in time when we address the actual gedankenexperiment of quantum physics that the mysterious machine is proposed to emulate.

In order to make a Monte Carlo experiment of the imagined scenario, sending each pair of balls to all nine dial-pair settings at the stations, we would need to observe the light signal responses at every one of them. The accumulating data matrix would have size $N \times 9$ rather than merely $N \times 1$. Of course we shall want to use appropriate quantum probabilities when generating such a sequence, and we shall. (These, remember, involve matching coloured light probabilities of 1/4 when the two dials are set differently.) When we do this in the context of a real quantum experiment, surprisingly we shall find frequencies of matching lights exceeding 1/3 among the different-dial-setting runs, just as Mermin has tendered in his consideration of the encoded balls. The same quantum probabilities that generate his results of 1/4 matching light signals in runs on a sequence of balls each sent to a single dial setting also generate results of matching signal frequencies exceeding 1/3 when each pair of balls addresses all nine settings. There is nothing mysterious about it. What has been missed in Mermin's accounting are the same type of functional relations among the spin-products in a gedankenexperiment that Aspect/Bell missed in their simpler polarization experiment with paired photons. It will take a while to explain the situation.

However there is another peculiarity to be noticed in the presentation of Table 1. While you follow the professor across a row for any mixed-colour encoded ball in noting the two of six matching light colour results when the two dials are set differently, notice also that there are only *four* distinct rows of nine-vectors that can possibly result from the scheme using the eight types of ball encodings. There are only eight rows to the Table, and the final four rows duplicate the first four rows, listed in reverse order. Just a few paragraphs ago we were imagining the possibility of several possible nine-vector observational results in the second

scenario of machine operation, as many as 64. Now it is evident that the vector $(-1,-1,+1,+1,-1,+1,-1,+1,-1)$ which we suggested as an exemplary possibility would not be a possibility at all under the encoded-ball scheme, despite it identifying matching lights in two of the six paired dial settings that differ. In fact, there are only four distinct nine-vectors of result possibilities arising from the eight encoding designs. The bottom line for now is that this scheme of sending encoded balls to detectors at all nine dial settings is a proposition completely different from that which yields the proclaimed results of Mermin's machine. We shall sort this all out forthwith.

To clarify the situation requires a diversion into the real quantum experiment that the professor would have us ignore while we are enticed to marvel at his mysterious machine. My plan is to begin with a presentation of the relevant practical quantum experiment that can be and has been conducted many times. Then we shall embellish the context to a gedankenexperiment designed to assess the implications of Einstein's principle of local realism and his challenge to the completeness of quantum theory. This is the context in which the specification of Bell's inequality is entertained, and the context for which Mermin's second version of the game is appropriate as an emulation. It is only once we recognise the structure of this matter that we will be able to identify prospective quantum probabilities when a single pair of balls visits all nine designs of dial settings.

Once we have studied the structure of the real quantum experiment and its associated gedankenexperiment, we shall design and conduct a Monte Carlo experiment as a prelude to a complete analysis of the entire situation based strictly on the limited claims of quantum theory. Surprisingly, the simulation also exhibits matching light frequencies exceeding 1/3 under conditions that Mermin proposes as mysterious. But the Monte Carlo simulation will not constitute the end of our analysis. As with my solution to the simpler experimental context of Aspect/Bell in [3], we will find in the complete analysis that quantum theory does *not* propose a joint probability distribution over the complete space of possible gedanken observations. Rather, in its current incomplete form it specifies a multi-dimensional polytope of such distributions, and explicitly renounces any prospect for refining it. The simulation design, while natural, is not the only design that QM theory would allow, and we shall see why. This is all sounding fairly complicated. However, it is merely a matter of plodding on to sort things out.

3. What Are We Really Talking about?

Professor Mermin understands the mystery to convey that while there are no obvious physical connections between the three pieces of the experimental device, *A*, *B*, and *C*, the attempt to explain its experimental features by unobservable instructions encoded within the balls is futile. Such an explanation might account for the specified observable outcomes of the machine when the dials are at *A* and *B* set identically as 11, 22, or 33, but it seems to provoke specific observations of flashing lights that do not match what we experience when running the

machine at other dial pairings. The alternative he proposes is to recognise that indeed the operation of the recorders actually is connected in some mysterious way, suggesting "connections of no known description, that serve no purpose other than relieving us of the task of accounting for the behavior of the device in their absence." This is the purportedly mysterious behaviour of quantum mechanics as is currently widely promoted. However Mermin engages such speculations no further, as the task proposed for his exposition was merely to state the conundrum, not to resolve it. The parable is concluded.

After completing his description of the mystery, the professor presents an insightful discussion of the relevance of the parable to issues raised by Einstein, Podolsky, and Rosen [5] in their proposition that the theory of quantum mechanics must be incomplete. While they had presented arguments that may appear telling regarding the activity of the machinery when the dials are set *identically* at A and B, their arguments appear to fail in situations in which the dials are set differently. This was a situation they did not assess, consumed as they were in their article with claims about the reality of quantum states and their observations that could be predicted with certainty. The implications for quantum behavior portrayed in the parable by different dial settings at A and B did not become evident until the startling research results of John Bell. These have been understood to display that if one presumes Einstein's principle of local realism and the relevance of hidden variables, then the specifications of quantum theory defy some standard inequalities of probability theory.

Mermin's exposition concludes with a description of the contextual quantum experiment that the parable is meant to portray, emphasizing that such detail can be conveniently ignored while the significance of the mystery is absorbed in awe. This is a well-known quantum experiment involving a pair of electrons that are propelled in opposite directions toward identical detecting devices of Stern-Gerlach magnets at stations A and B, each of the magnets oriented at one of three specific angles within the plane perpendicular to the incoming electrons. The detectors identify the magnetic spins of the electron pair, each in either the direction "up" or "down", denoted by $A=+1$ or $A=-1$, and similarly for the value of B. Rather than ignoring the experimental physics as suggested, the remainder of my exposition now is oriented to a detailed assessment of the exact specification of this experiment and the proclamations of quantum theory that concern it. We shall find that the parable fails to represent the situation adequately, for the same reason that Aspect's assessment of Bell's inequality fails in the simpler case of a pair of photons presented to two paired polarization angles. It is a mathematical error of neglect.

3.1. The Quantum Experiment in Its Gedanken Extension

Remember that the simple quantum gedankenexperiment of Aspect/Bell concerned a cerebral assessment of possibilities for the combined result of four practical experiments, each of which can be engaged, but for which the engagement of all four simultaneously is recognized as impossible. The same style of

investigation will pertain to our considerations now. We shall examine the (im)possible imagined results of a nine-ply Stern-Gerlach thought experiment conducted on a single pair of electrons. This is the context to which Bell's inequality pertains and in which the principle of local realism is relevant. Now it is a pair of electrons that are propelled in opposite directions toward the stations of Alice and Bob. In a real experiment, each is charged with observing one of the pair as it passes a magnet oriented in one of three different directions relative to vertical up and down. The vertical position is designated as the zero position, and the other two are directed in twists of negative and positive angles relative to this zero. The possible pairings of these magnet orientations at the two stations specify 3 × 3 possibilities for a paired choice of them at the two stations during any experimental run. The vector outcome of spin-products occurring in a run of a thought experiment, sending a single electron pair to *all nine* paired magnet orientations, will be denoted $\boldsymbol{G}_9$, the "$\boldsymbol{G}$" standing for "gedankenvector".

Initially we shall designate the three possible magnet orientations of each spin observation variable by the subscripts *n*, *z*, or *p*, so to represent its alignment relative to vertical as negative, zero, or positive. We may write A_p or B_z, for examples. When referring to a spin observation at a generic magnet orientation we may write the quantities A and B without subscript, or we may use a subscript letter "d" considered as a variable. Eventually we shall assess the specific setup in which the two chosen magnet orientations *differ* by the angles −120˚, 0˚, and +120˚. This is the setup relevant to the probability assessments prescribed in Mermin's parable. Deliberations of quantum theory specify probabilities for the possible paired observations of spins as up or down at the two stations, casually denoted as P_{++}, P_{+-}, P_{-+}, and P_{--} appropriate to any such angle pairing. Equivalently, they specify the expectation of the spin product, $E(AB)$.

In the gedankenexperiment, Alice and Bob will observe the spins of a pair of electrons in *every* paired directional setting of their magnets. Their respective observations named A and B would be recorded as either +1 or −1 to designate an observation of spin "up" or "down". We shall denote the possible results of their nine paired observations by product event designations such as $(A_n=+1)(B_z=-1)$, $(A_n=+1)(B_z=+1)$, $(A_p=+1)(B_n=+1)$, or $(A_z=+1)(B_z=-1)$, and so on, results exhibiting spin-products of −1, +1, +1 and −1 respectively. My use of arithmetic notation means that each of these *product events* indicates whether the *joint* observation of spin values at site A and site B arises in a particular configuration or not. There will be nine of them. The number of *prospective* nine-tuples of observation products could be as large as $2^9=512$. This number of possibilities will be reduced shortly on account of theoretical speculation and on account of the particular directional angles we employ in the experiment's design.

We shall begin by considering a list of all the possible results of the paired observations at A and B that could be entertained according to Einstein's contentious (and currently widely rejected) "locality" condition. This involves a proposition that lies outside the technical domain of quantum theory. It accepts that

the electron spin observation made by Alice in any specific magnet orientation is a result assessed with a quantum probability entangled with that of Bob's magnet direction in that instance. However it proclaims that no matter what her magnet orientation may be in any particular experimental run, Alice's spin observation in this instance would be the same (either up or down) no matter what be the corresponding orientation of Bob's magnet and his spin observations in imagined companion experiments *on the same pair* of electrons.

The reason such a claim lies outside the scope of quantum theory is that the distinct operator matrices for observation of a single pair of electrons addressing two different designs of magnet orientations do not commute. Thus, quantum theory itself says nothing about their *joint* product results at the two designs. The complete experiment, a thought experiment, presumes that a single pair of electrons passes by the two magnets in all nine of their paired orientations. The principle of local realism stipulates that if Alice's spin observation is, say, +1 in a specific magnet orientation when Bob's relative orientation angle is +120˚, then Alice's would also equal +1 in this instance if Bob's magnet were oriented relatively at 0˚ and/or at −120˚ as well. Bob's spin observations might equal either +1 or −1 in either case. Although Alice's actual measurement for a particular electron spin is proposed to be invariant with respect to the setting of Bob's detection angle, this principle respects nonetheless an assertion of entanglement of the electrons. This understanding derives from the specification of quantum theory that $P\left[\left(A=+1\right)\left(B=-1\right)|\,\theta\right]=\frac{1}{2}\cos^2\left(\theta/2\right)$ at any single relative angle setting, where θ is the relative angle between Alice's and Bob's magnet orientations. Equivalently the specification is $E\left(AB|\,\theta\right)=1-2\cos^2\left(\theta/2\right)$. This is the relevant prescription of quantum theory.

The principle of local realism implies that in measuring the spins at all nine angle orientation pairs for the gedankenexperiment, *each* of the observers would register *only three distinct* spin values. According to this premise, each of the observed values of A_n, A_z, and A_p in the nine-ply experiment would be the same no matter which of the station *B* magnet orientations it were paired with, B_n, B_z, or B_p. The same would hold for the observation values of the *B*'s. These six observation values would display themselves among nine specific observation pairs of the form $\left(A_d, B_d\right)$. There appear to be only $2^6=64$ conceivable instantiations of these six spin observations that would respect the principle of local realism in any run of the gedankenexperiment. However, the probabilistic assertions of quantum theory reduce the number of these possibilities still further. Consider the prescription of quantum theory pertinent to a any experimental setup in which Alice's and Bob's magnet orientations are identical, and we measure the spin values $\left(A_n, B_n\right)$, or $\left(A_z, B_z\right)$, or $\left(A_p, B_p\right)$. Quantum theory stipulates that in any such experiment we must observe opposite spin values at stations *A* and *B*. For the quantum prognostication stipulates that $P\left(A_dB_d=-1|\,\theta=0\right)=1$, and equivalently $E\left[A_dB_d\,|\left(\theta=0\right)\right]=-1$ whenever the two magnet orientations are identical. It is impossible for the spin observa-

tions at Alice's and Bob's stations to be the same when their magnet orientations are the same. An implication of these quantum probabilities along with the principle of local realism is that there are not 64 possible results of the gedankenexperiment, but rather only 8. Let's examine them.

3.2. Specifying the Possible Results of the Stern-Gerlach Gedankenexperiment

Let's cut to the quick, and present for comment a list of all possible results of the 3×3 gedankenexperiment that might be observed for a single pair of electrons passing the magnets in all of their possible relative orientations. The possibilities constituting the experimental ensemble are presumed to be limited by the principle of local realism and the assertions of quantum theory, particularly as they are relevant to component experiments in which the orientations of Alice's and Bob's magnets are identical. That is, we shall examine the "realm matrix" of all possibilities for the unknown observation vector $\boldsymbol{G}_6 \equiv \left(A_n, A_z, A_p, B_n, B_z, B_p\right)^{\mathrm{T}}$ that could result from the running of such a nine-ply-thought-experiment. These are followed by the vector of spin-products

$$\boldsymbol{G}_9 = \left(A_nB_n, A_nB_z, A_nB_p, A_zB_n, A_zB_z, A_zB_p, A_pB_n, A_pB_z, A_pB_p\right)^{\mathrm{T}}$$

that such an array of spin observations would imply for the nine paired magnet orientations. Partitioned (for reasons to be seen) vertically into three sections and horizontally into two for examination, the realm matrix appears as

$$\boldsymbol{R}\begin{pmatrix} A_n \\ A_z \\ A_p \\ B_n \\ B_z \\ B_p \\ *\,*\,* \\ A_nB_n \\ A_nB_z \\ A_nB_p \\ A_zB_n \\ A_zB_z \\ A_zB_p \\ A_pB_n \\ A_pB_z \\ A_pB_p \\ *\,*\,* \\ {}^1A_nB_n \\ {}^2A_nB_z \\ {}^3A_nB_p \\ {}^6A_zB_p \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & * & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & * & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & * & 1 & -1 & 1 & -1 \\ -1 & -1 & -1 & -1 & * & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & * & -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 & * & -1 & 1 & -1 & 1 \\ * & * & * & * & * & * & * & * & * \\ -1 & -1 & -1 & -1 & * & -1 & -1 & -1 & -1 \\ -1 & -1 & 1 & 1 & * & 1 & 1 & -1 & -1 \\ -1 & 1 & -1 & 1 & * & 1 & -1 & 1 & -1 \\ -1 & -1 & 1 & 1 & * & 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & -1 & * & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & -1 & * & -1 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 & * & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & * & -1 & 1 & 1 & -1 \\ -1 & -1 & -1 & -1 & * & -1 & -1 & -1 & -1 \\ * & * & * & * & * & * & * & * & * \\ -1 & -1 & -1 & -1 & * & -1 & -1 & -1 & -1 \\ -1 & -1 & 1 & 1 & * & 1 & 1 & -1 & -1 \\ -1 & 1 & -1 & 1 & * & 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 & * & -1 & 1 & 1 & -1 \end{pmatrix}$$

Within the first vertical partition of the named vector appear the six observa-

tions of Alice and Bob in the performance of the nine experiments: $\left(A_n, A_z, A_p, B_n, B_z, B_p\right)^{\mathrm{T}}$. While each component of the 6 × 1 vector can equal either −1 or +1, the components of the second triple in any column, B_n, B_z, and B_p, must be the negative values of the first three components of that column. This is a prescription of quantum theory. The spin directions of Alice and Bob must oppose one another when their magnet orientations are the same. Each of the eight columns displaying an array of Alice's spin possibilities at her three magnet orientations is accompanied by a display of Bob's opposing spin values at his matching orientations. To the contrary, when the magnet orientations of Alice and Bob differ, then a spin observation at A in either direction might be accompanied by a spin observation at B in either direction as well. The spin-product might equal −1 or +1. Composing the columns of the middle partitioned section of this realm matrix are the arithmetical *products* of each of the three A's with each of the three B's appearing in the same column. There are nine such product quantities. The product $A_d B_d$ equals −1 whenever the magnet directions at stations A and B are identical. The third partitioned sections of the named quantity vector and of its realm matrix merely repeat rows 1, 2, 3, and 6 of the middle partition. We shall discuss them when it becomes appropriate.

3.3. A Substantive Recognition: Functional Relations among Spin-Products

A *substantive* matter to recognise about this realm matrix is that the four columns of *the right partition of the middle section* constitute a *folded replica of the columns of the left partition* of that section. The middle section of column 5 is identical to that of column 4. That of 6 is identical to column 3, and so on, until the midsection of column 8 is identical to that of column 1. Thus, the middle partition of the realm matrix has only 4 *distinct* columns, rather than 8 as the top partition appearing above it does. Moreover, the middle partition matrix has only four distinct rows rather than nine as one might naively suspect. This fact has important ramifications for the analysis of the hypothetical problem of nine distinct Stern-Gerlach experiments on the same pair of electrons!

In the first place, the realm matrix for the spin-product vector resulting from the gedankenexperiment on a single pair of electrons consists of only the left half of the middle partition matrix above, a 9 × 4 matrix. For reference in ensuing discussions, we shall refer to this realm matrix for the vector of nine spin-product observations in the gedankenexperiment as $\boldsymbol{R}_{9,4}$. What is more, some of the rows of this complete matrix are obtainable via specific functions of other rows in that section. As long as any two rows of $\boldsymbol{R}_{9,4}$ constitute the cartesian product $\{-1,+1\}^2$, the remaining seven rows are determined by a function of them, an "into" mapping with the structure $\{-1,+1\}^2 \to \{-1,+1\}^7$. Consider, for example, the rows 2 and 3 of $\boldsymbol{R}_{9,4}$. Their column pairs exhaust the cartesian product $\{-1,+1\}^2$, constituting the domain of a function. For each of these pairs in the domain, the remaining seven columns provide a unique vector within $\{-1,+1\}^7$. We shall designate this function with the notation $23 \to 1456789$.

In fact there are twelve such functional relations that inhere within the structure of the realm matrix $\boldsymbol{R}_{9,4}$. Using the same functional notation, we can list them as:

$$\begin{array}{lll} 23 \to 1456789 & 34 \to 1256789 & 47 \to 1235689 \\ 26 \to 1345789 & 36 \to 1245789 & 48 \to 1235679 \\ 27 \to 1345689 & 38 \to 1245679 & 67 \to 1234589 \\ 28 \to 1345679 & 46 \to 1235789 & 78 \to 1234569 \end{array}$$

The first of these arrows denotes the functional relation among the columns of spin-products we have just described. It denotes a mapping from $\{-1,+1\}^2$ into $\{-1,+1\}^7$. The subsequent arrow structures all describe functions as well. All that is required is that the two rows of $\boldsymbol{R}_{9,4}$ corresponding to the domain variables exhibit among their column pairs the component vectors of $\{-1,+1\}^2$. This list of functional relations embedded in the realm matrix is exhaustive. There are six paired spin products corresponding to experiments with different magnet orientations at the stations of Alice and Bob, these being the components 2, 3, 4, 6, 7, and 8 of any spin-product vector. There are ${}^6C_2 = 15$ possible choices of two spin-products to consider as elements of a possible function domain. However three such choices of two of them fail to provide a valid function domain: (2, 4), (3, 7), and (6, 8). That is, attempted mappings of 24 onto 1356789, of 37 onto 245689, and of 68 onto 1234579 all fail to identify a function. For example, the column pairs from rows 2 and 4 do not exhaust $\{-1,+1\}^2$. Furthermore, when they repeat they correspond with two distinct tentative objects in $\{-1,+1\}^7$. The relation this structure provides for consideration does not constitute a function.

In the second place worthy of note, the twelve embedded functional relations among the spin products we have recognized are not linear. If they were, the rank of the realm matrix $\boldsymbol{R}_{9,4}$ would be only two, but it is four! This is a feature crucial to the implications of quantum theory for assessing the prospective results of the gedankenexperiment. Quantum theory makes specific expectation (and probability) assertions regarding the spin product possibilities for any two domain variables among our listed functions, considered as distinct isolated experiments. If the functional relations we have enumerated were linear then these would imply precise expectations for the range variables. As it is, the assertions of quantum theory will stipulate only bounds on the expectations for the spin-products of the range variables in any such case. This situation inheres some intrigue. The conditional distribution for any seven products in a range vector given the results of the domain vector would be degenerate at their function value. Yet their joint probability distribution with the domain variables cannot be determined precisely. We shall see more of this.

A further remark of note concerns the paired repetitions found among rows 2 and 4, rows 3 and 7, and rows 6 and 8 in the central vertical partition of this realm matrix. These identities exhibit specific symmetries among components of

the spin-product vector of a gedankenexperiment run: the commutativity of spin-product observations with respect to the orientations of their detecting magnets. Noting the names of the spin-product quantities whose observation possibilities constitute the repeating rows, these repetitions specify that $A_n B_z = A_z B_n$, $A_n B_p = A_p B_n$, and $A_z B_p = A_p B_z$ in any imagined run of the experiment. This feature of symmetry will come to bear on our computations of probability bounds for gedankenresults deriving from the claims of quantum theory, restricted by its avowed uncertainty principle.

A final surprise can be seen among the *columns* of the partition matrix $\boldsymbol{R}_{9,4}$. Although we have mentioned nothing at all about encoded balls while constructing it, it is apparent that the columns of $\boldsymbol{R}_{9,4}$ match the row designations in Table 1 which repeat themselves. These exhibit the light signal responses of Mermin's mysterious machine to his explanatory suggestion of encoded balls. We shall come back to this recognition too in due time.

4. A Simulation Experiment, Using QM Motivated Probabilities

We shall now capitalize on our recognition of the functional relations embedded in the realm matrix $\boldsymbol{R}_{9,4}$, by conducting a simulation experiment appropriate to the restricted experiment. It represents the situation Professor Mermin thought he was assessing when he evaluated the behaviour of his mysterious machine in response to colour-encoded balls at all nine dial-pair combinations. Our simulation is meant to elucidate the structure that the theory of quantum mechanics designates as appropriate to the assessment of Bell's inequality in the context of its associated gedankenexperiment. It will generate a sequence of twelve million gedankenvectors $\boldsymbol{G}_9$, these being the simulated spin-products of electron pairs that each pass all nine paired magnet orientations, obeying the probabilistic prescriptions of quantum theory.

The experiment will be conducted in twelve configurations, corresponding to the twelve functional relations enumerated in Section 3.3. Data are generated that correspond to a gedankenexperiment of Stern-Gerlach apparatus with the angled magnet direction possibilities for both Alice and Bob set at all three of their orientations relative to vertical in the (x, y) plane, −120˚, 0˚, and +120˚. This plane of directions is perpendicular to the direction of the incident electrons, exactly as described in the conclusion to the "Quantum Mysteries for Anyone". We shall denote the relative angle between Alice's and Bob's magnet orientations in any paired direction setup by $\theta = d_B - d_A$. The nine paired direction possibilities yield orientations that differ by angles of $\theta = 0^\circ$ when the A and B observations are (A_n, B_n), (A_z, B_z), and (A_p, B_p); of $\theta = +120^\circ = -240^\circ$ when the A and B observations are (A_n, B_z), (A_z, B_p), and (A_p, B_n); and of $\theta = -120^\circ = +240^\circ$ when the A and B observations are (A_n, B_p), (A_z, B_n), and (A_p, B_z). Rows 1, 5, and 9 of our realm matrix $\boldsymbol{R}_{9,4}$ correspond to an orientation difference angle $\theta = 0^\circ$; row numbers 4, 8, and 3 pertain to the difference angle $\theta = -120^\circ$; and rows 7, 2, and 6 pertain to the difference angle $\theta = +120^\circ$.

The simulation begins with a routine pertinent to the spin-product function $23 \to 1456789$. It first generates observation values for components 2 and 3 of the spin-product vector independently according to standard QM specifications of probabilities for differing spin observations at these magnet orientation pairings. These prescribe the quantum probabilities

$$P\left[A_n B_z = -1 \,|\, \theta = -120^\circ\right] = 1/4 = P\left[A_n B_p = -1 \,|\, \theta = +120^\circ\right],$$

which correspond to the frequencies reported by Mermin in the observations of his machine. The coloured lights match only 1/4 of the time when the stations' magnet orientations differ.

Appropriately then, from each eventuality of the two outcome values so generated for components 2 and 3, the associated spin-product values for components 1, 4, 5, 6, 7, 8, 9 are computed according to the functional rule we have designated with the notation $23 \to 1456789$. In each of these cases, the distribution of these latter component values given the spin-product pair simulated for the components of the domain is degenerate on their function value. This derives from the structural features of the possible spin-product observations. From one million such generations the number of occurrences of −1 are counted and recorded in the spin-product columns 1 through 9.

The results of the simulation yield surprising counts which can be displayed as

23	1000000	250191	250332	250191	1000000	625225	250332	625225	1000000	1456789

The number at the far left edge of this row of such counts designates the row numbers of function *domain* observations that were selected by the pseudo random numbers of MATLAB, while the number at the far right edge identify the corresponding rows in the *range* that were computed via appropriate function rules. Understanding this, you should read this report line as identifying 1000000 counts of spin-product observations −1 in experiment columns 1, 5, and 9. (Mermin's lights always flash the same colour when the dials at A and B read the same.) Counts of −1 amounted to 250191, 250332, 250191, and 250332 as recorded in experiments 2, 3, 4, and 7, respectively, two pairs of which involve repetitions. Finally, identical counts of 625225 were recorded in both columns 6 and 8. (The identical counts in three of the column pairs result from the commuting symmetries which we found to insist that $A_n B_z = A_z B_n$, that $A_z B_p = A_p B_z$, and that $A_p B_n = A_n B_p$ when the experiments run gedankenly in tandem. This feature arises naturally in the function-based generations of the simulation.)

Notice particularly, that the repeated count of 625225 corresponding to spin products $A_z B_p$ and $A_p B_z$ differs markedly from the claims of Professor Mermin that the experiment should yield a number near to 250000 in *every* spin-product column for which the difference in orientation angle is $\theta = -120^\circ$ or $\theta = +120^\circ$. We can now understand why this has occurred! On account of the functional relations required among the spin-product observations, only a select two of the spin-product values can be generated freely according to QM

probability relations as specified by the well-known cosine squared equations.

$$P\left[\left(A=+1\right)\left(B=-1\right)|\,\theta\right] \;=\; \frac{1}{2}\cos^2\left(\theta/2\right)$$

The remaining seven must be determined from the values of these two according to the functional relation we are considering, which binds them all.

Now there is nothing special about the experimental components 2 and 3 which we allowed to be chosen freely by their quantum probabilities. We have seen that there are twelve such domain choices of experimental pairs that can be used to generate the gedanken spin vector. We shall now examine a *full array* of simulation results deriving the other eleven choices of the function domain as well. They will be found to allay misdirected concerns aired in the parable regarding proportions of matching light colours proclaimed to differ from 1/4. Such concerns *are* appropriate in real physical experiments in which sequences of electron pairs engage any single pair of differing magnet orientations. But they *are not* appropriate to a gedankenexperiment motivated by quantum theory and local realism in which each pair passes all nine magnet orientation pairings. Let's look at the complete results.

4.1. An Array of Simulation Results Generated by Twelve Functional Relations

Displayed in Table 2 are results of twelve simulation runs that are structurally identical in their generation to those I have already described for the function $23\rightarrow 1456789$. Distinct runs were based on QM probability assertions applied to the two domain variables of each of the twelve functional relations we have recognized. In each set of runs, spin-products were generated for two appropriate function domain variables using QM probability simulations. Then the remaining seven were computed from these using the relevant function specifications we have identified. Nonetheless the output of each 9-vector generated respects the injunctions of all twelve function rules. Without further ado, Table 2 below presents the results of these simulations, presented according to the same reporting structure used in the previous Subsection. In fact, the first row of this Table repeats the results that were reported there.

Each row of these Simulation Count results is based upon quantum probabilities relevant to a distinct function whose domain rows are displayed in its left edge column. Columns numbered 1, 5, and 9 (not counting the edge columns) show that the generated spin-product is negative at all three configurations in which magnet orientations at the sites A and B would be identical. The products of the simulated spin observations yield counts of 1000000 in these columns, exactly as expected in quantum theory. The negative spin-product counts at the other gedanken configurations vary. For example, row 3 of the Table, which identifies in its left edge column the relevant functional relation as $27\rightarrow 1345689$, exhibits counts of 250096 and 250274 in matrix columns 2 and 7, as expected according the probability specifications of quantum theory applied to a single

Table 2. The letter D at top left stands for "Domain" of a function. Twelve simulation counts of negative spin-products.

D	$^{1}A_nB_n$	$^{2}A_nB_z$	$^{3}A_nB_p$	$^{4}A_zB_n$	$^{5}A_zB_z$	$^{6}A_zB_p$	$^{7}A_pB_n$	$^{8}A_pB_z$	$^{9}A_pB_p$	*Range*
23	1000000	250191	250332	250191	1000000	625225	250332	625225	1000000	1456789
26	1000000	249641	625501	249641	1000000	249912	625501	249912	1000000	1345789
27	1000000	250096	250274	250096	1000000	624192	250274	624192	1000000	1345689
28	1000000	250188	625260	250188	1000000	250060	625260	250060	1000000	1345679
34	1000000	250777	250397	250777	1000000	624338	250397	624338	1000000	1256789
36	1000000	625459	249849	625459	1000000	249814	249849	249814	1000000	1245789
38	1000000	624890	250619	624890	1000000	250277	250619	250277	1000000	1245679
46	1000000	250093	624872	250093	1000000	249855	624872	249855	1000000	1235789
47	1000000	249640	249716	249640	1000000	625256	249716	625256	1000000	1235689
48	1000000	249483	625658	249483	1000000	249411	625658	249411	1000000	1235679
67	1000000	625506	249710	625506	1000000	249974	249710	249974	1000000	1234589
78	1000000	625491	249736	625491	1000000	249681	249736	249681	1000000	1234569
	Sum Simulation Counts by Product Column:									
	12000000	4501455	4501924	4501455	12000000	4497995	4501924	4497995	12000000	
	Proportions of Negative Spin-Products by Product Column:									
	1.0000	0.3751	0.3752	0.3751	1.0000	0.3748	0.3752	0.3748	1.0000	

experimental magnet settings. However, in columns 6 and 8 of row 3, among the range variables of the constraining function both counts are found to be 624192, not near to 250000 at all!

Similar structures govern the simulation counts in all rows of the count matrix shown in Table 2: two of the column elements of each row exhibit identical counts in the vicinity of 625000 while four column elements are in the vicinity of 250000, arising as two identical count pairs. No matter which pair of function domains is used to generate the nine columns of results, the counts are always identical in columns 2 and 4, in columns 6 and 8, and in columns 3 and 7. The requirements of commuting spin observations such as $A_nB_z = A_zB_n$ are satisfied simply by recognition of the functional relations among spin products.

Summing the twelve columns of these simulation counts (which each arise from 1 million simulated Stern-Gerlach experiments) yields further results of interest. Quite striking in fact are the implied *proportions* of differing spin-products exhibited at the several paired angle orientations. These proportions are *not* displayed as three 1's and six 0.25's as proclaimed by Professor Mermin. The three 1's surely appear in the expected places, but of the remaining six columns we find *all* the proportions near to 0.375, defying his claim to the proportion arising as 1/4. Indeed, the proportions we have generated in the quantum gedankensimulation are reminiscent of the frequency behaviour of encoded balls which had worried him in his parable, motivating him (with many others) to decry the sensibility of Einstein's suggestion of hidden variables in quantum behaviour. Nevertheless, such ball coding was not employed at all in the simulated generation of these results.

4.2. Comments, a Qualification, and a Query

Don't get me wrong. If you would do a sequence of simulation experiments at a specific pairing of differing magnet orientations using quantum probabilities, you would find the spin-product values to equal −1 in close to 1/4 of these cases. However, if you do a long sequence of simulated experiments that gedankenly subjects the electrons to all nine paired magnet angle directions in the way local realism restricts them, you would find the proportion of spin-products equal to −1 at about 0.375 whenever the relative angle between the magnets equals −120˚ or +120˚. This happenstance governs the counts displayed in columns 2, 3, 4, 6, 7, and 8. The result has nothing to do with Mermin's proposed explanation of "the mystery" involving colour-encoded balls. It derives from a recognition of the functional relations embedded into spin-product possibility vectors in the gedankenexperiment. A situation clearly distinct, when the rows are produced by gedankenly submitting each pair of electrons to all nine of the relative angle settings, many elements of the cartesian product $\{-1,+1\}^6$ for the unequal magnet angle designs would constitute impossible outcomes of the spin-product functions that govern the experiment. Each of the allowable result vectors respects twelve functional mappings of $\{-1,+1\}^2$ onto $\{-1,+1\}^7$.

Professor Mermin's fabulous machinery produces no mysterious results at all. The character of the matrix of results would be different, depending on which of the two different ways that the results are generated. This is not surprising.

One way to simulate the gedankenexperiment as supported by the probabilities of quantum theory would be to pick sequentially (randomly, uniformly) one of the functional relations that bind the spin-products $\{-1,+1\}^2 \to \{-1,+1\}^7$, to generate a vector of nine-tuple observations as we have described. Then pick a functional relation again to generate the next 9-vector of results. Continuing with this process we would generate a sequence of such 9-vectors, and accumulate the counts of negative spin-products at the nine angle pairings across the sequential generations. This process would result in proportions of negative spin-products as appear in the final line of Table 2. However, this result could hardly be claimed to be a definitive prognostication of quantum theory relevant to the gedankenexperiment. Be aware that the nine component vector results of the experimental runs involve simultaneous results of observations at angle pairings whose operator matrices do not commute! Quantum theory explicitly says nothing specific about such impossible experimental results.

So what *does* quantum theory say, and how can we present it in a complete and concise way? Consider again a single functional relation such as $23 \to 1456789$. Well, quantum theory is quite specific in identifying a probabilistic structure governing the possible results at either of the two relative angles between the Stern-Gerlach magnets that appear in the function domain. However, it is also quite explicit in denying any motivation for making claims about *simultaneous* spin-product results at any two relative angle settings for which the observation matrix operators do not commute, these two in particular. The results of this simulation activity I have just proposed *could be* admissible according to the

logic of quantum theory, but there is no reasoning that would make the joint distribution they imply definitive. There is no requirement that function domains be picked randomly and uniformly at all, as I have done here. We could generate other distributions of results if we picked them according to some other random scheme.

The way to *characterize* the *complete* space of joint probability distributions over gedanken results that cohere with the positive claims of quantum theory is to assess a battery of linear programming computations. These can identify the *bounds* on probabilities for the range settings of the constraining functions that would cohere with the specifications that quantum theory does provide for results of function domain settings. To produce such an assessment is the burden of our next Section. This involves investigating the implications of Bruno de Finetti's "fundamental theorem of probability" for the nine proposed Stern-Gerlach gedankenexperiments, all performed on the same pair of electrons. We turn to this investigation now.

5. Characterizing a QM Probability Polytope for Stern-Gerlach Gedankenexperiments via de Finetti's FTP

If you are not familiar with Bruno de Finetti's fundamental theorem of probability [6], Chapter 3.10, then you should read a brief commentary and description of the theorem in linear programming form which I attach to this essay as an Appendix. In a word, the theorem identifies a linear programming problem that characterizes the bounds on the probability for any event that are required by its coherency with other probabilities that are taken as given. An application of this theorem to the assessment of the Aspect/Bell error was made in Section 7 of my article [3] about supposed inequality violations entitled, "Quantum violations of Bell's inequality: a misunderstanding based on a mathematical error of neglect". If you would like a pedagogical introduction to the theorem, explaining both its construction and an explanation of what is so fundamental about it, either before or after you continue reading the present text, I suggest an examination of my book [7], Chapter 2.10, pages 99-113.

The remainder of this Section presents an analysis of the implications of theoretical quantum probabilities for the imagined results of the now classic gedankenexperiment on a pair of electrons propelled toward the Stern-Gerlach magnets of Alice and Bob. It covers a description and a formalization of QM-motivated probability assertions appropriate to the several problems posed, and displays computational results that identify the bounding implications for other probabilities about which the theory is silent. Although quantum theory says nothing precise about the gedanken results, the FTP places definitive restrictions on what would cohere with what quantum theory does say.

Section 5.1 is simply discursive, describing conversationally the setup of the linear programming problems relevant to our discussion. It does not dwell on explicit formal definitions of all notation, but rather proceeds straightforwardly with discussion using vocabulary that is standard in LP methods. Section 5.2

presents formal algebraic detail of the quantities and constraints involved in these LP problems. Section 5.3 then displays numerical results that portray the polytope of probability vectors representing the coherent content of recognized quantum theory as it is relevant to these matters.

5.1. A Linear Programming Problem Identifying QM Probabilities That Recognize Functional Restrictions: A Discursive Introduction

To begin, we shall presume standard expectations (and probabilities) of quantum theory for the spin-products in the domain experiments 23, and use a linear programming format to specify bounds on implied probabilities for spin-product observations in the range experiments 1456789 imagined to be concurrent in the thought experiment. Once we are clear on how this works we shall describe how such a set of min/max computations would be replicated and concatenated for all twelve functional restriction structures.

Familiar by now with the quantum probability distributions for the outcomes of any real experiment, you should recall that the probabilities for the four possible outcomes of the spin observations, ++, +−, −+, and −−, can all be identified from the probability for any one of them, say P_{++}. For the four quantum probabilities resolve to $P_{++} = P_{--}$ along with $P_{+-} = P_{-+} = \frac{1}{2}\left[1-2P_{++}\right]$. Furthermore, this probability is uniquely related to the expected value of the spin product via the equation $E(AB) = P_{++} - P_{+-} - P_{-+} + P_{--}$, which then equals $4P_{++} - 1$. The probability for a spin-product of −1 resolves to $P[AB=-1] = \left[1-E(AB)\right]/2$.

In the context of any component experiment for which the relative angle between the two magnet orientations is equal to θ, these relations specify $P_{++}(\theta) = \frac{1}{2}\left[1-\cos^2(\theta/2)\right]$ and $E\left[AB(\theta)\right] = 1-2\cos^2(\theta/2)$. For reference in our computations relevant to Mermin's problem, these spin-product expectations resolve to $E\left[AB \mid \theta = -120^\circ\right] = E\left[AB \mid \theta = +120^\circ\right] = 0.5$ and $E\left[AB \mid \theta = 0^\circ\right] = -1$.

Correspondingly,

$P\left[AB=-1 \mid \theta = -120^\circ\right] = P\left[AB=-1 \mid \theta = +120^\circ\right] = 0.25$, and

$P\left[AB=-1 \mid \theta = 0^\circ\right] = 1$.

Here is the problem, stated directly in conversational English, presuming familiarity with all algebraic notation and detail. We are to investigate the QM-motivated probability specifications for the 4 possible observation vectors of nine spin-products observed by Alice and Bob in a 3 × 3 paired-angle-thought-experiment on the same pair of electrons. Each of these vectors specifies an array of values for all nine components of the spin-product gedanken vector we shall call

$$\boldsymbol{G}_9 \equiv \left(A_nB_n, A_nB_z, A_nB_p, A_zB_n, A_zB_z, A_zB_p, A_pB_n, A_pB_z, A_pB_p\right)^{\mathrm{T}}$$

These are the quantities that are crucial to the specifications of quantum theory. A complete list of their possible gedanken observations constitutes the four columns of the left half of the middle partition of the realm matrix we created in Section 3.2, and designated as $\boldsymbol{R}_{9,4}$. Only four of the nine rows of this

matrix are distinct, the other five being repetitions.

Among these possible 9-dimensional vectors of spin-products, there are only two dimensions of free observations, on account of the functional restrictions embedded within them. For example, the observations at the magnet configurations numbered 2, 3 would functionally identify the results at configurations 1, 4, 5, 6, 7, 8, 9 according to quantum theoretic specifications enhanced by Einstein's presumed principle of local realism. However, this functional relation is evidently non-linear, for the rank of the realm matrix of nine spin-product possibilities is 4. This rank corresponds to any four distinct rows of $\boldsymbol{R}_{9,4}$. Thus, the specification of quantum expectations pertinent to the domain configurations of rows 2 and 3 would place only polytopic bounds on the cohering probabilities for the other spin-products they imply.

If we were to specify a *complete* distribution of probabilities over the four possible spin-product outcome components of the function domain, $\{-1,+1\}^2$, our problem would be over. However, quantum theory *explicitly disavows* an assertion of a *complete* distribution vector $\boldsymbol{q}_4$ over these possibilities, for this would entail a specification of joint probabilities for the results of non-commuting Hermitian operators on the state space of the electron pair. While quantum theory specifies precise probabilities for the two possible values of the spin-product observation $A_d B_d$ at any experimental paired angle setting of Stern-Gerlach magnets, it explicitly says nothing about the *joint outcomes* of the spin-products observed at *both* settings 2 *and* 3, for example. Nonetheless, quantum theory *does* specify explicitly cohering probabilities regarding the outcomes of the spin-product experiment at each of the configurations 2 and 3 separately. Observations at this pair of settings constitute the domain of the function we have designated as $23 \to 1456789$.

An aside of detail should clarify the preceding disavowal. While QM theory would clearly specify probabilities such as $P\left[\left(A_n=1\right)\left(B_z=-1\right)\right]$ and $P\left[\left(A_n=1\right)\left(B_p=+1\right)\right]$ for example, deriving from the expectations $E\left(A_n B_z\right)$ and $E\left(A_n B_p\right)$, it explicitly disavows assertions of the form $P\left[\left(A_n=1\right)\left(B_z=-1\right)\left(A_n=1\right)\left(B_p=+1\right)\right]$. The former two assertions each specify a standard probabilistic assertion of quantum theory relative to a pair of spin observations at a specific paired angle setting; whereas the latter assertion would entail a claim about joint observations of spins at both B_z *and* B_p, observations represented by non-commuting Hermitian operators. However, a joint probability of this sort would be required in order to specify precise values for components of the vector $\boldsymbol{q}_4$, flaunting this abstemious honesty. Quantum theory does *not* provide for a complete distribution of probabilities for the four possible spin-product components of $\{-1,+1\}^2$, neither at the domain pairing 23, nor any other pair of domain variables among the twelve functions embedded in the observation realm.

Suppose then that we entertain the cohering expectations of quantum theory pertinent to the isolated domain products 2 and 3. What would these imply for the cohering expectations of spin-products 1, 4, 5, 6, 7, 8, and 9 determined in

their function range? Based on the associated rows of the realm matrix for the spin products, these expectations would place two linear restrictions on the components of any prospective gedankenvector $\boldsymbol{q}_4$. Along with the constraint that their components are all non-negative and sum to 1 (unity), a linear programming routine would identify a pair of solution vectors $\boldsymbol{q}_{4\min}$ and $\boldsymbol{q}_{4\max}$ that produce the extreme feasible values of the objective functions $E\left(A_d B_d\right)$ for any range orientation pairing, subject to the quantum theoretical linear constraints on spin-products 2 and 3. (These are in addition to the requirement that $E\left(AB\right)=-1$ for the spin-products of orientation pairs 1, 5, and 9, a condition specified by quantum theory which underlies the entire problem.) The linear coefficients of the objective function can be identified from appropriate rows of the realm matrix.

As it turns out, with expectations specified for any pair of two domain variables, there is only one range variable whose extreme cohering expectations we shall need to investigate. Remember that considerations of symmetry in the problem imply the equality of the spin products $A_n B_z = A_z B_n$, $A_n B_p = A_p B_n$, and $A_z B_p = A_p B_z$. These correspond to the identity of rows 2 and 4, 3 and 7, and 6 and 8 in the realm matrix $\boldsymbol{R}_{9,4}$. Thus, with expectations settled at −1 for the negative spin-products at orientations 1, 5, and 9, asserting quantum theoretic expectations at orientations 2 and 3 would imply the same expectations at orientations 4 and 7 as well. This would leave only spin-product expectations for orientations 6 and 8 to be investigated, and these must be identical. As a result, a single pair of min/max linear programming problems would identify the bounds on the entire cohering expectation vector. This is what we shall formalize now.

The paired solution vectors $\boldsymbol{q}_{4\min}$ and $\boldsymbol{q}_{4\max}$ contain the information that puts bounds on the general problem solution we seek: to identify the extreme vectors $\boldsymbol{q}_4$ that satisfy all QM-motivated probability specifications relevant to the gedankenexperiment and also support appropriate cohering range expectations. We will need to determine this pair of solution vectors for the min/max LP problems appropriate to each of the twelve function domains we have identified. It would be the entire resulting space of vectors that represents the implications of quantum theory pertinent to the magnetic spin-product gedankenexperiment on an electron pair.

Before presenting the numerical results of these computations, let's examine a formal identification of the quantity vectors that play the central roles in the linear restrictions of the several linear programming problems we have delineated.

5.2. Algebraic Representation of the LP Constraints

Our goal is identify the prognostications of quantum theory regarding a gedankenexperiment: Mermin's physics problem of two electrons engaging the nine component 3 × 3 design of Stern-Gerlach magnets at the stations of Alice and Bob.

Let $\boldsymbol{G}_9 = \left(A_n B_n, A_n B_z, A_n B_p, A_z B_n, A_z B_z, A_z B_p, A_p B_n, A_p B_z, A_p B_p\right)^{\mathrm{T}}$ denote the

column vector of imagined spin-product outcomes of the nine experiments on a single pair of electrons.

Let $\boldsymbol{R}_{9,4} \equiv \boldsymbol{R}\left(\boldsymbol{G}_9\right)$ designate the realm matrix of possibilities for the gedankenobservation vector $\boldsymbol{G}_9$. We have already displayed it as the left half of the middle matrix partition of the large realm matrix we constructed in Section 3.2. We shall refer to individual *rows* of this matrix by using the denotation $\boldsymbol{r}_{i\cdot}$ for values of the row numbers $i = 1, \cdots, 9$, and to individual columns of this matrix by $\boldsymbol{r}_{\cdot j}$ for values of the column numbers $j = 1, 2, 3, 4$.

The bold vector $\mathbf{1}_4$ denotes a *row* vector of four 1's.

Furthermore, let the column vector $\boldsymbol{b}_9$ denote the numerical values of the quantum theoretic expectations for the components of $\boldsymbol{G}_9$ when they each designate the outcome of a single real experiment on a pair of electrons at a specific angle pairing of the S-G magnets.

Finally, let $\boldsymbol{Q}_4$ denote the column vector of events whose components identify whether the observation vector $\boldsymbol{G}_9$ would happen to equal the various columns of $\boldsymbol{R}_{9,4}$. That is, the component $Q_j = \left(\boldsymbol{G}_9 = \boldsymbol{r}_{\cdot j}\right)$. Each of these four events equals 1 or 0, and only one of them equals 1.

Notice that whereas quantum theory specifies expectations for the components of $\boldsymbol{G}_9$ when each is entertained as the spin-product of a unique pair of electrons in individual experiments standing alone, it does not specify expectations for the components of $\boldsymbol{Q}_4$. For these events identify the joint outcomes of several incompatible experimental observations. However, the linear programming problems we now address codify restrictions on the space of such expectation vectors that would cohere with what quantum theory does say about the domain experiments individually.

To begin, we should formalize the linear programming investigations required by the spin-product function $23 \to 1456789$ which we have discussed informally in the details of the preceding subsection:

Find the vectors $\boldsymbol{q}_{4\min}$ and $\boldsymbol{q}_{4\max}$ that yield the minimum and the maximum values of $\boldsymbol{r}_{6\cdot}\boldsymbol{q}_4$ subject to the conditions that

$$\begin{pmatrix} \boldsymbol{r}_{2\cdot} \\ \boldsymbol{r}_{3\cdot} \\ \mathbf{1}_4 \end{pmatrix} \boldsymbol{q}_4 = \begin{pmatrix} b_2 \\ b_3 \\ 1 \end{pmatrix},$$

where each component $q_i \geq 0$.

We shall denote these two solution vectors by $\boldsymbol{q}_{4\min 23\cdot 6}$ and $\boldsymbol{q}_{4\max 23\cdot 6}$. For once we determine them, we shall need to repeat such computational LP searches so to determine extreme vectors appropriate to the other eleven spin-product functions that govern possibilities for the spin-product vector $\boldsymbol{G}_9$ as well. Formally, this would amount merely to changing the coefficient vectors $\boldsymbol{r}_{2\cdot}$ and $\boldsymbol{r}_{3\cdot}$ in the LP domain constraints, and changing the objective function coefficients $\boldsymbol{r}_{6\cdot}$ accordingly so to represent the range variable whose expectation bounds we search. In principle, there could be 24 such extreme solution vectors. However, on account of duplications among the row vectors of $\boldsymbol{R}_{9,4}$,

there are only three distinct LP problem pairs among these, and six distinct solution vectors. It should be evident, for example, that $\boldsymbol{q}_{4\min 23\cdot 6} = \boldsymbol{q}_{4\min 47\cdot 8}$ and $\boldsymbol{q}_{4\max 23\cdot 6} = \boldsymbol{q}_{4\max 47\cdot 8}$. The LP problems that they resolve are numerically identical.

It is a feature of coherent probability structures that any convex combination of coherent expectation assertions is also coherent. It is the convex hull of all the six extreme expectation vectors that represents the quantum theoretic prognostications for the gedankenexperiment.

5.3. Computational Results

The six solution vectors that resolve the three distinct pairs of LP problems involved in this investigation are displayed as 4 × 1 columns in the top section of Table 3. The notational subscripts in their column headings specify the form of the LP problems from which they derived. Although they were designed to identify extreme values of expected spin-products at specific magnet orientations, the probabilities underlying each of these extreme solution vectors would specify cohering expectation values for every one of the spin-products that result from a "run" of the gedankenexperiment at all nine relative angle configurations. These implied nine-vectors of expected spin-products are printed immediately below the solution vectors in the Table to which they correspond. The rank of the matrix of solution vector columns is 4.

Table 3. Extreme spin-product expectations for unique LP solutions.

LP solutions	$\boldsymbol{q}_{4\min 23\cdot 6}$	$\boldsymbol{q}_{4\min 26\cdot 3}$	$\boldsymbol{q}_{4\min 36\cdot 2}$	$\boldsymbol{q}_{4\max 23\cdot 6}$	$\boldsymbol{q}_{4\max 26\cdot 3}$	$\boldsymbol{q}_{4\max 36\cdot 2}$		
q_1	0.25	0.25	0.25	0	0	0		
q_2	0	0	0.75	0.25	0.25	0.5		
q_3	0	0.75	0	0.25	0.5	0.25		
q_4	0.75	0	0	0.5	0.25	0.25		
E(spinprod)							**QM**	**Sim**
$E(A_nB_n)_1$	−1	−1	−1	−1	−1	−1	−1	−1
$E(A_nB_z)_2$	0.5	0.5	−1	0.5	0.5	0	0.5	0.25
$E(A_nB_p)_3$	0.5	−1	0.5	0.5	0	0.5	0.5	0.25
$E(A_zB_n)_4$	0.5	0.5	−1	0.5	0.5	0	0.5	0.25
$E(A_zB_z)_5$	−1	−1	−1	−1	−1	−1	−1	−1
$E(A_zB_p)_6$	−1	0.5	0.5	0	0.5	0.5	0.5	0.25
$E(A_pB_n)_7$	0.5	−1	0.5	0.5	0	0.5	0.5	0.25
$E(A_pB_z)_8$	−1	0.5	0.5	0	0.5	0.5	0.5	0.25
$E(A_pB_p)_9$	−1	−1	−1	−1	−1	−1	−1	−1

To the right of the six *E*(spinprod) vectors in the Table appear two additional column vectors for purposes of comparison. The column headed by **QM** exhibits the standard expectations of quantum mechanics for the spin-product to be observed in an actual experiment at any single one of the various paired angle configurations. The final additional column headed by **Sim** exhibits expectations for spin-products as 0.25 corresponding to those proportions of differing spin observations (−+) or (+−) that were generated in our simulated gedankenexperiment in Section 4.1, with those proportions rounded to 0.375. The simulations, remember, relied upon quantum-theory-motivated probabilities to generate the emergence of spin-products at each of the domain configurations, and then relied on the functional relations to yield the other seven dimensional components accordingly. The simulated behaviour of each electron pair involved its engaging the magnet orientations at detection stations in all nine of their experimental paired orientations. This is the situation that provoked Mermin's rejection of the message-encoded balls explanation.

The glory of Table 3 is that its columns specify the extreme points of a bounding polytope of expectation vectors supported by quantum theory as it would be relevant to the spin-products of a gedankenexperiment on a single pair of electrons at all nine of the possible orientation pairings. The fundamental theorem of probability does not identify a specific vector of expectations for spin-products at every one of the nine paired magnet orientations. Neither do the prescriptions of quantum theory, which avoid precise assertions regarding the *joint* outcomes of non-commuting measurements. Rather, they specify a space of such allowable expectation vectors which cohere with the assertions about real experiments that quantum theory actually is endowed to assess. Based on the QM motivated probabilities that constrained the several LP computations, the conclusion to this exercise is that the sought-for vector of gedanken expectations needs merely sit somewhere within the convex hull of the E(spinprod) vectors appearing as the first six columns in the lower half of the Table.

The rank of the matrix of 9-D expectation vectors appearing in Table 3 is only four. These four dimensions are spanned by the rows 1, 2, 3, and 6 of the Table. The other rows are repeats of these, so we could have listed four different rows, but the result would be the same. In order to discuss these results in the same terms with which Professor Mermin assessed the behaviour of his machine, we shall transform these expectations into the probabilities they imply for negative spin values in Table 4. Expressed as a function of θ, this linear transform is $P\left[AB(\theta)=-1\right]=\left\{1-E\left[AB(\theta)\right]\right\}/2$ for each relative angle between the magnet orientations.

One of the features of these computational results is that we can now make sense of the simulation results we generated in Table 2. These, remember, defied Professor Mermin's claims regarding results that would obtain in such (impossible) experimentation on a single electron pair. The simulated counts of negative spin-products shown in that Table yielded proportions on the order of 0.375 for

six of the Stern-Gerlach orientation pairings. These differ markedly from the Mermin proclamation that proportions of negative spin-products would hardly differ from 0.25 at these settings, but they do exhibit the order of magnitude that he found upsetting. As a marginal statement in the context of the gedankenexperiment, quantum theory specifies only $P(A_nB_z=-1)\in[0.25,1]$, for example, which can be seen by reading along the rows of Table 4. But we can be more specific in understanding the three such marginal assessments of probability regarding A_nB_z, A_nB_p, and A_zB_p. Since the top row of the Table is constant at the value of 1, we can recognize that the convex hull enclosing the first six column vectors of this matrix constitutes a 3-D polytope within a hyperplane in the 4-D space. It is displayed in Figure 2 as a tetrahedron that has lost one of its tips.

Table 4. Probabilities for negative spin-products along extreme solution vectors.

LP solutions	$q_{4\min 23\text{-}6}$	$q_{4\min 26\text{-}3}$	$q_{4\min 36\text{-}2}$	$q_{4\max 23\text{-}6}$	$q_{4\max 26\text{-}3}$	$q_{4\max 36\text{-}2}$	QM	Sim
$P(A_nB_n=-1)_1$	1	1	1	1	1	1	1	1
$P(A_nB_z=-1)_2$	0.25	0.25	1	0.25	0.25	0.5	0.25	0.375
$P(A_nB_p=-1)_3$	0.25	1	0.25	0.25	0.5	0.25	0.25	0.375
$P(A_zB_p=-1)_6$	1	0.25	0.25	0.5	0.25	0.25	0.25	0.375

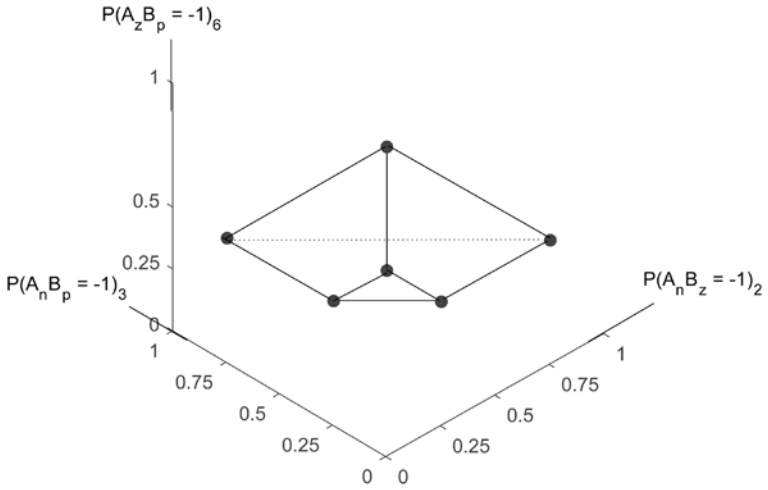


Figure 2. The three dimensional polygon constituting the convex hull of probabilities for negative spin products in a gedankenexperiment when the Stern-Gerlach magnet directions are not set identically. Professor Mermin's proclaimed point of probabilities in these three dimensions, (0.25, 0.25, 0.25), is exterior to this polytope, while the simulation vector of probabilities (0.375, 0.375, 0.375) is a point well within the hull as a convex combination of its vertices.

Regarding the first six columns of Table 4 as vertices of a quantum theoretical probability polytope for negative spin-products, we can determine that the appended final column vector headed by **Sim** is located within their convex hull. Algebraically, it is equal to the convex combination of those columns with con-

vexity coefficients $(1/6, 0, 1/12, 0, 1/2, 1/4)$. In contrast, the appended column headed by **QM** does *not* lie in this convex hull. It cannot be expressed as a convex combination of the vertex vectors. To proclaim it as representing the prescriptions of quantum mechanics for the outcome of the gedankenexperiment would be incoherent. The accompanying relegation of Einstein's principle of local realism amounts to nonsense. When the operation of Professor Mermin's machine is applied to the gedankenexperiment, the crude vector of quantum probabilities representing his provocative claims lies outside of the convex hull of probability vectors that are supported by the results of quantum theory. His assertions and his concerns derive from a mathematical error of neglect, similar in structure to the error we have found in the claims of Aspect/Bell.

6. A Mystery Exposed

In challenging the purported mysteries of quantum physics, we have now been through an exercise of tedious mathematics and computation. It is time to conclude with an overview of what we have learned. I will be brief and frank in concluding. An extensive discussion of issues in quantum theory that impinge on the foundations of probability will require a presentation of its own [8].

There is nothing wrong with Professor Mermin's machine, and furthermore there is nothing mysterious about it. The machine quite accurately portrays the probabilistic structure of the current propositions of quantum theory regarding the experimental observation of magnetic spins of paired electrons as they pass angled Stern-Gerlach magnets at two distantly separated detection stations. To be explicitly precise, the machine is designed to exemplify the structure of theoretical and empirical results from sequential observations of distinct electron pairs ejected to any one of nine different paired experimental settings, similar in design. Each of the paired dial settings on the two machines constitutes a different type of experiment. At three of the paired angle magnet orientations when the two settings are identical, every observation of the spin-product equals −1. The spins (A, B) observed at the two stations are recorded as either $(-1, +1)$ or $(+1, -1)$. At any one of the six other paired settings in which the magnet orientations differ, long sequences of observations yield spin-products equal to −1 in 1/4 of the runs, and +1 in 3/4 of the runs. For any individual unique experiment designed with such a structure, quantum theory asserts only the probability of 1/4 for the negative spin-product. The quantum probability for the electron spin behaviour at station A depends on the orientation of the magnet at station B as well as that at station A.

Many regard this result to be mysterious in itself, involving what Einstein had referred to as "spooky action at a distance". The source of this attitude derives from the imagination that the probabilistic behaviour of quantum activity corresponds to a feature of randomness inherent in the particle structure. Nature at its fundamental base is considered to be random, governed by recognizable probability distributions, derivable by theoretical acumen and tamed at classical

scales of mass phenomena by the laws of large numbers. It is this conception of the matter to which Einstein objected. He proposed with EPR that the stochastic aspect of quantum theoretic results derives rather from the incompleteness of the theory and from our uncertainty regarding the influence of unknown and unobserved "supplementary variables" pertinent to the conditions of any specific experimental run. This proposal would relegate "the mystery" of quantum results to the same category involved in common mysteries of activity at classical scales of magnitude, such as "where did I leave my keys?". Maybe here, maybe there. It is the codification of symmetries in our uncertainty regarding conditions of the experimental problem that yields probabilistic prescriptions regarding the quantum mechanics. Conditional probabilities for the outcome of one quantity that vary with the outcome of a conditioning event are a standard feature of distributions that represent exchangeable (symmetric) judgments. There is no implication that the conditioning event has actually been observed. A conditional probability will be specified conditioned on the negated event as well. *Both*, together, characterise an asserted quantum probability distribution for the possible paired results of the experiment.

When regarded as properties of the particles themselves, the quantum probabilities of Mermin's machine do seem to pose the mysterious question of how the probabilistic activity of the electron at station A can depend on the magnet angle (dial) setting at station B if there is no way for the status of the setting at B to be communicated to station A when an electron (a ball) arrives there. The resolution of this enigma proceeds from recognition that probabilities are not properties of particles at all, but rather formally assessed representations of our considered uncertainty about observable quantities. For now, I shall focus my conclusion here on what the professor proclaims as a challenge to this point of view. It underlies a mistaken attitude that is held virtually universally among quantum theorists today: the defiance of Bell's inequality in a gedankenexperiment on a single pair of electrons at all nine experimental settings defies not only the reasonability of local realism at the quantum scale and the proposition of supplementary variables pertinent to quantum behaviour, but the uncertainty interpretation of probabilities itself in accounting for the evidence of quantum experiments. Indeed it is a common misconception that the defiance of the inequality arises in quantum mechanical assessment of spin product expectations for any electron pairs whatsoever, gedanken or not.

Professor Mermin models the action of supplementary variables by encoded designations of colour schemes on the balls ejected toward the stations. He proposes this as a model of most any supplementary variables explanation of the quantum experiment, on the same metaphorical order as the machine with balls models the observation of electron spin behaviour. Such a proposal could be fair enough, though one might quibble with its adequacy for representing the substance of the supplementary variables viewpoint. Nonetheless, his analysis of the activity of an encoded pair of balls when they arrive at all nine dial settings generates what he considers to be an even deeper mystery. Although the scheme

surely ensures matching light signals when the dials are set identically, he motivates the proportion of matching lights (spin-products equal to −1) deriving from such a scenario as equal to more than 1/3. It is this result, which he proposes as an instance of the supposed defiance of Bell's inequality, that is seen to constitute the mystery of quantum behaviour: apparently no supplementary features of the experimental situation can account for the known behaviour of quantum experiments.

It is this result that is just plain wrong. Examining the real quantum experiment which we are coaxed to ignore, we have found embedded within the corresponding thought experiment a surprising feature that has long been unnoticed. Subjecting each pair of electrons to spin-detection at all nine of the paired angle settings would engender an array of restrictive functional relations among the nine observed spin results. The professor neglects these symmetric functional relations mapping $\{-1,+1\}^2$ into $\{-1,+1\}^7$ in his analysis of the situation. His claims regarding the machine behaviour yielding matching lights in 1/4 of such gedanken observations when the switches differ are blatantly false. They rely on the possibility that any string of nine-tuples deriving from the cartesian product $\{-1,+1\}^6$ of spin possibilities could designate the outcome of such a thought experiment. We have seen otherwise … that many such strings of spin-products are impossible. The space of possibilities derives rather from the cartesian structure $\{-1,+1\}^2$, replicated in several different pairs of angled magnet orientations. The four possible results of each domain pair of spin observations is mapped into restrictive completions within the space of $\{-1,+1\}^7$. Moreover, we have used computational procedures of linear programming to identify precisely the polytope of probabilistic assertions regarding the outcomes of the gedankenexperiment that represent the honest claims of quantum theory relevant to this matter.

The probabilities for matching lights proclaimed by Mermin are representable by a nine-tuple vector that *does not* lie within the convex hull of the coherent vectors supported by quantum theory. We have created a Monte-Carlo simulation of results of a scenario which is both wholly consistent with quantum theory and also respects the restrictive symmetric functional relations that govern the structure of the experiment. It generates proportions of matching lights on the order of 0.375, precisely on the order of magnitude that the professor would have us suspect on account of his error of neglect. This vector *does reside* within the convex hull of extreme gedanken probabilities required by coherency. But there is still more to say about this!

One of the most famous features of quantum theory, known widely by name to the general public, is the relevance of Heisenberg's uncertainty principle. Under the guise of that name, the principle concerns physical experiments with electrons that attempt to measure both the position and the velocity of an electron at a point in time. What the principle recognizes is that it is impossible to make such a joint measurement of both of these characteristics of the electron at the same time. We can make a measurement of one or the other, but not both.

Technically, the quantum theory identifies this impossibility by the characterization of the two measurements of the quantum state via Hermitian operators that do not commute. Quantum theory can specify probability distributions for possible values of either of these measurements on an electron. However, it cannot and does not provide an assessment of a joint probability for the observation of both measurements. For such an operation is impossible. It is the incommensurability of measurements characterised by matrix operators which do not commute that is the general form of the uncertainty principle of theoretical quantum mechanics.

The incommensurability of simultaneous observation of paired electron spins at several different paired magnet orientations in a gedankenexperiment belies the professor's claims about his machine in this context. Quantum theory does not identify a joint probability distribution for the results of all nine of them. This limitation has long been recognized since the early insights of Fine [9] [10] and subsequently in challenges by Hess and Phillip [11], among others. These have been regularly rebuffed in mainstream literature which has heralded the rejection of local realism on this account. Nonetheless the message relevant to the incompleteness of quantum theory is clear and is accentuated by the deliberations I have reported herein. A stirring review by Kupczynski [12] has recently supported a revived reconsideration of the widely acclaimed rejection of local realism.

Quantum theory does provide precise probabilities for the four possible observations of spin pairs at any single paired orientation, these being ++, +−, −+, or −− at the two detection stations. Moreover, it can stipulate probabilities for such outcomes from several distinct experiments on an electron pair at differing magnet angles, realizing albeit that only one of them can be engaged. These are the probabilities used in linear programming routines that identify the restrictions on range probabilities provided by individual assessments of domain probabilities. However, for the joint distribution of all nine measurements it leaves four dimensions of freedom remaining unascertained. The probability distributions of quantum theory for the results of a real experiment on an electron pair at any one of the nine design orientations *may not* be considered to be a marginal distribution from a joint distribution. There is no joint distribution over these imagined experimental results supported by quantum theory in its current formulation! Any vector within the polytope of feasible joint distributions can be transformed mechanically into a vector of marginal probabilities for the spin products at any magnet orientation pair, but no one of these constitutes a marginal distribution pertinent to all feasibilities.

In particular, the result of our simulated experiment using the probabilistic assertions of quantum theory cannot be presumed to provide a definitive proclamation of quantum mechanics. It is intriguing that it generates coherent probabilistic activity that emulates precisely behaviour of the sort that Professor Mermin had relegated. Yet the implications of the simulation are not decisive. Its generation had embedded into it a formulation of independence among the

outcomes of any two spin-products observed among the domain arguments of the functions. While feasible in the context of the agnostic stance of the theory relative to incommensurable measurements, this is surely not a requirement of the theory. Furthermore, in the context of the active claims of the theory regarding the entanglement of spin observations at distant detection stations, it ought well be considered suspect by those who might like to think about such things. We are left with the realization that quantum theory provides only a convex polytopic boundary of probabilities for the result of the gedankenexperiment to which Bell's inequality is relevant. Bell's inequality is defied by none of the distributions within the polytope that is entertained by quantum theory as we know it. While we are avowedly still to understand completely the physical details of quantum behaviour, they inhere no mysteries in themselves … for anyone. More hoojums than boojums.

Acknowledgements

The University of Canterbury provided computing and research facilities. Thanks to Larry Dennis for helpful discussions, to the managing editor and three reviewers for very helpful comments, and to Paul Brouwers, Steve Gourdie, and Allen Witt for IT service and consultation.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Mermin, N.D. (1981) *Journal of Philosophy*, **78**, 397-408. https://doi.org/10.2307/2026482

[2] Mermin, N.D. (1990) Boojums All the Way Through: Communicating Science in a Prosaic Age. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511608216

[3] Lad, F. (2021) *Journal of Modern Physics*, **12**, 1109-1144. https://doi.org/10.4236/jmp.2021.128067

[4] Lad, F. (2020) *Entropy*, **22**, 19. https://doi.org/10.3390/e22070759

[5] Einstein, A., Podolsky, B. and Rosen, N. (1935) *Physical Review*, **47**, 777-780. https://doi.org/10.1103/PhysRev.47.777

[6] de Finetti, B. (1970) Teoria delle probabilità. Ed. Einaudi, 2 voll., Torino. (English version: Theory of Probability. A. Machi and A. Smith (trs.), Chichester, Wiley, 1974, 1975).

[7] Lad, F. (1996) Operational Subjective Statistical Methods: A Mathematical, Philosophical, and Historical Introduction. John Wiley, New York.

[8] Lad, F. (2021) On Probability and Quantum Physics. Preprint on Researchgate. https://doi.org/10.13140/RG.2.2.21376.87048

[9] Fine, A. (1982) *Physics Review Letters*, **48**, 291-295. https://doi.org/10.1103/PhysRevLett.48.291

[10] Fine, A. (1989) Do Correlations Need to Be Explained? In: Cushing, J.T. and McMullin, E., Eds., *Philosophical Consequences of Quantum Theory* (pp. 175-194). University of Notre Dame Press, Notre Dame, Indiana.

[11] Hess, K. and Philipp, W. (2005) Bell's Theorem: Critique of Proofs with and without Inequalities. *AIP Conference Proceedings*, **750**, 150-157. https://doi.org/10.1063/1.1874568

[12] Kupczynski, M. (2020) *Frontiers in Physics*, **8**, 273. https://doi.org/10.3389/fphy.2020.00273

[13] Bruno, G. and Gilio, A. (1980) *Statistica*, **40**, 337-344.

[14] Lad, F., Dickey, J.M. and Rahman, M.A. (1990) *Statistica*, **50**, 19-38.

Appendix: The Fundamental Theorem of Probability

The fundamental theorem of probability (FTP) specifies that when probabilities or expectations for any N quantities whatsoever are assessed with the vector of values $\boldsymbol{p}_N$, then bounds on a cohering expectation for any further $(N+1)^{\text{st}}$ quantity can be computed via a linear programming routine. The theorem was first *named* in de Finetti [6] (Chapter 3.10), though it is as old as his famous lectures at the Institute Henri Poincaré in 1935. It was first presented in linear programming form in the article of Bruno and Gilio [13]. The theorem extends naturally to specify bounds on expectations for general quantities ("previsions" in de Finetti's nomenclature) as presented in the article of Lad, Dickey, and Rahman [14]. It is discussed pedagogically in Lad [7] (Chapter 2.10, pp. 99-113).

Theorem: Suppose the realm matrix $\boldsymbol{R}(\boldsymbol{X}_{N+1})$ for the vector of quantities X_1 through X_{N+1} has K columns. These columns exhaust all possibilities for prospective quantity observations under consideration. Define the vector $\boldsymbol{r}_{N+1}$ as the final *row* of this realm matrix corresponding to the possibilities for the quantity X_{N+1} as the last component of the observation vector, and the matrix $\boldsymbol{R}_{N,K}$ as the N initial rows of the realm matrix corresponding to the concomitant possibilities for the first N components of $\boldsymbol{X}_{N+1}$. The design of the linear programming routine is to find the column vectors $\boldsymbol{q}_K$ for which the linear combination $\boldsymbol{r}_{N+1}\boldsymbol{q}_K$ achieves minimum and maximum values subject to the N linear restrictions that $\boldsymbol{R}_{N,K}\boldsymbol{q}_K$ equals $\boldsymbol{p}_N$, along with the restrictions that the components of $\boldsymbol{q}_K$ are non-negative and that they sum to 1. If there is no feasible solution to these problems, then the assertion of the N expectations that have been presumed is incoherent.

Comment: The linear restrictions on $\boldsymbol{q}_K$ ensure that as long as $E(X_{N+1})$ is within the extremes of $\boldsymbol{r}_{N+1}\boldsymbol{q}_K$ determined by the theorem, the expectation of the full vector $E(\boldsymbol{X}_{N+1})$ would then lie within the convex hull of the columns of its realm matrix. This is the general condition of coherency.

Call for Papers

Journal of Modern Physics

ISSN: 2153-1196 (Print) ISSN: 2153-120X (Online)
https://www.scirp.org/journal/jmp

Journal of Modern Physics **(JMP)** is an international journal dedicated to the latest advancement of modern physics. The goal of this journal is to provide a platform for scientists and academicians all over the world to promote, share, and discuss various new issues and developments in different areas of modern physics.

Editor-in-Chief

Prof. Yang-Hui He City University, UK

Subject Coverage

Journal of Modern Physics publishes original papers including but not limited to the following fields:

Biophysics and Medical Physics
Complex Systems Physics
Computational Physics
Condensed Matter Physics
Cosmology and Early Universe
Earth and Planetary Sciences
General Relativity
High Energy Astrophysics
High Energy/Accelerator Physics
Instrumentation and Measurement
Interdisciplinary Physics
Materials Sciences and Technology
Mathematical Physics
Mechanical Response of Solids and Structures
New Materials: Micro and Nano-Mechanics and Homogeneization
Non-Equilibrium Thermodynamics and Statistical Mechanics
Nuclear Science and Engineering
Optics
Physics of Nanostructures
Plasma Physics
Quantum Mechanical Developments
Quantum Theory
Relativistic Astrophysics
String Theory
Superconducting Physics
Theoretical High Energy Physics
Thermology

We are also interested in: 1) Short Reports—2-5 page papers where an author can either present an idea with theoretical background but has not yet completed the research needed for a complete paper or preliminary data; 2) Book Reviews—Comments and critiques.

Notes for Intending Authors

Submitted papers should not have been previously published nor be currently under consideration for publication elsewhere. Paper submission will be handled electronically through the website. All papers are refereed through a peer review process. For more details about the submissions, please access the website.

Website and E-Mail

https://www.scirp.org/journal/jmp **E-mail: jmp@scirp.org**

What is SCIRP?

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

What is Open Access?

All original research papers published by SCIRP are made freely and permanently accessible online immediately upon publication. To be able to provide open access journals, SCIRP defrays operation costs from authors and subscription charges only for its printed version. Open access publishing allows an immediate, worldwide, barrier-free, open access to the full text of research papers, which is in the best interests of the scientific community.

- High visibility for maximum global exposure with open access publishing model
- Rigorous peer review of research papers
- Prompt faster publication with less cost
- Guaranteed targeted, multidisciplinary audience

Scientific Research Publishing

Website: https://www.scirp.org

Subscription: sub@scirp.org

Advertisement: service@scirp.org