Issues 1-4 2016 | Volume 12

The Journal on Advanced Studies in Theoretical and Experimental Physics, including Related Themes from Mathematics

PROGRESS IN PHYSICS

"All scientists shall have the right to present their scientific research results, in whole or in part, at relevant scientific conferences, and to publish the same in printed scientific journals, electronic archives, and any other media." — Declaration of Academic Freedom, Article 8

ISSN 1555-5534

The Journal on Advanced Studies in Theoretical and Experimental Physics, including Related Themes from Mathematics

PROGRESS IN PHYSICS

A quarterly issue scientific journal, registered with the Library of Congress (DC, USA). This journal is peer reviewed and included in the abstracting and indexing coverage of: Mathematical Reviews and MathSciNet (AMS, USA), DOAJ of Lund University (Sweden), Scientific Commons of the University of St. Gallen (Switzerland), Open-J-Gate (India), Referativnyi Zhurnal VINITI (Russia), etc.

Electronic version of this journal:
http://www.ptep-online.com

Postal Address

Department of Mathematics and Science,
University of New Mexico,
705 Gurley Ave., Gallup, NM 87301, USA

ISSN: 1555-5534 (print)
ISSN: 1555-5615 (online)

Standard Address Number: 297-5092
Printed in the United States of America

January 2016 **Vol. 12, Issue 1**

CONTENTS

Rybicki M. Type Ia Supernovae Progenitor Problem and the Variation of Fundamental Constants . . . 3

Hafeez H. Y., Chifu E. N., Musa I. M. Schrödinger Equation for a Half Spin Electron in a Time Dependent Magnetic Field . . . 15

Consiglio J. On the Absorber in Gravitation . . . 20

Scholkmann F. Power-Law Scaling of the Impact Crater Size-Frequency Distribution on Pluto: A Preliminary Analysis Based on First Images from New Horizons' Flyby . . 26

Gogberashvili M. Standard Model Particles from Split Octonions . . . 30

Zaninetti L. and Scholkmann F. The Impact Crater Size-Frequency Distribution on Pluto Follows a Truncated Pareto Distribution: Results from a First Data Set Based on the Recent New Horizons' Flyby . . . 34

Belyakov A. Determination of the Neutrino Mass . . . 36

Danilov V. On the Nature of the Magnetic Field of the Earth and Other Planets . . . 41

Weng S. A Classical Model of the Photon . . . 49

Potter F. Update on Pluto and Its 5 Moons Obeying the Quantization of Angular Momentum per Unit Mass Constraint of Quantum Celestial Mechanics . . . 56

Belyakov A. On the Deviation of the Standard Model Predictions in the Large Hadron Collider Experiments (*Letters to Progress in Physics*) . . . 59

Zhang T. X., Wilson C., Schamschula M. P. X-Ray Flares from Sagittarius A* and Black Hole Universe . . . 61

Crothers S. J. On an Apparent Resolution of the Catt Question . . . 68

Catania J. The Roland De Witte Experiment, R. T. Cahill, and the One-Way Speed of Light (*Letters to Progress in Physics*) . . . 70

Spivey R. J. A Non-anthropic Solution to the Cosmological Constant Problem . . . 72

Robitaille P.-M. Further Insight Relative to Cavity Radiation III: Gedanken Experiments, Irreversibility, and Kirchhoff's Law . . . 85

Rybicki M. Mansouri-Sexl Test Theory: The Question of Equivalence between Special Relativity and Ether Theories . . . 89

Information for Authors

Progress in Physics has been created for rapid publications on advanced studies in theoretical and experimental physics, including related themes from mathematics and astronomy. All submitted papers should be professional, in good English, containing a brief review of a problem and obtained results.

All submissions should be designed in LaTeX format using *Progress in Physics* template. This template can be downloaded from *Progress in Physics* home page http://www.ptep-online.com

Preliminary, authors may submit papers in PDF format. If the paper is accepted, authors can manage LaTeXtyping. Do not send MS Word documents, please: we do not use this software, so unable to read this file format. Incorrectly formatted papers (i.e. not LaTeXwith the template) will not be accepted for publication. Those authors who are unable to prepare their submissions in LaTeXformat can apply to a third-party payable service for LaTeX typing. Our personnel work voluntarily. Authors must assist by conforming to this policy, to make the publication process as easy and fast as possible.

Abstract and the necessary information about author(s) should be included into the papers. To submit a paper, mail the file(s) to the Editor-in-Chief.

All submitted papers should be as brief as possible. Short articles are preferable. Large papers can also be considered. Letters related to the publications in the journal or to the events among the science community can be applied to the section *Letters to Progress in Physics*.

All that has been accepted for the online issue of *Progress in Physics* is printed in the paper version of the journal. To order printed issues, contact the Editors.

Authors retain their rights to use their papers published in *Progress in Physics* as a whole or any part of it in any other publications and in any way they see fit. This copyright agreement shall remain valid even if the authors transfer copyright of their published papers to another party.

Electronic copies of all papers published in *Progress in Physics* are available for free download, copying, and re-distribution, according to the copyright agreement printed on the titlepage of each issue of the journal. This copyright agreement follows the *Budapest Open Initiative* and the *Creative Commons Attribution-Noncommercial-No Derivative Works 2.5 License* declaring that electronic copies of such books and journals should always be accessed for reading, download, and copying for any person, and free of charge.

Consideration and review process does not require any payment from the side of the submitters. Nevertheless the authors of accepted papers are requested to pay the page charges. *Progress in Physics* is a non-profit/academic journal: money collected from the authors cover the cost of printing and distribution of the annual volumes of the journal along the major academic/university libraries of the world. (Look for the current author fee in the online version of *Progress in Physics*.)

Type Ia Supernovae Progenitor Problem and the Variation of Fundamental Constants

Maciej Rybicki

Sas-Zubrzyckiego 8/27, 30-611 Krakow, Poland
E-mail: maciej.rybicki@icloud.com

Cosmological observations strongly suggest our universe is the interior of an expanding black hole. If the constant mass of the universe is assumed then from the equation for Schwarzschild radius: $r_S = 2Gmc^{-2}$ it follows that proportionality constant Gc^{-2} depends linearly on the universe's radius R_u, identified with r_S, i.e. $Gc^{-2} \sim R_u$, $M_u = const$. Because the Chandrasekhar limit M_{Ch} relates to the speed of light and to the Newton's constant as $M_{Ch} \sim (c/G)^{3/2}$ so expansion involves gradual decrease of M_{Ch}. In result, a single white dwarf can alone become the Type Ia supernova progenitor, which provides a complementary solution to single-degenerate and double-degenerate models for SNe Ia. Both alternative scenarios: $G \sim R_u$ and $c \sim R_u^{-1/2}$ are analyzed in regard of their consistence with observations, and their consequences to cosmology.

1 Introduction

On account of the supposed uniformity of their absolute magnitude, the Type Ia supernovae (SNe Ia) play an important role of "standard candles" in cosmology. A tight correlation between the peak light output and the light-curve width (width-luminosity relation) results from the way SNe Ia originate from white dwarfs (WDs) — the final remnants for low and medium mass stars. According to the current understanding, the carbon-oxygen (CO) thermonuclear fusion triggering the supernova explosion takes place in compact binary systems in either of two principal progenitor channels. A single-degenerate (SD) model (Whelan & Iben [80]) predicts that CO WD accretes matter from the companion, usually the red giant or the main sequence star. Just before approaching the Chandrasekhar mass-limit $M_{Ch} \approx 1.44 M_\odot$ for which electron degeneracy pressure becomes insufficient to prevent the gravitational collapse, the WD's core reaches the ignition temperature for the runaway carbon and oxygen fusion into heavier elements. In a preceding time lasting usually $\sim 10^6$ yr WD processes the transferred matter falling onto its surface through the accretion disc. In this phase, called "nuclear-burning white dwarf" (NBWD) the hydrogen-helium fusion releases energy in a form of copious X-radiation, observed as "super-soft X-ray source" (Di Stefano [17]).

Instead, the double-degenerate (DD) model (Webbink [79], Iben & Tutukov [37]) predicts that two WDs of the combined mass $\geqslant M_{Ch}$ form a compact binary system and subsequently spiral towards each other in a common envelope. Eventually, they collide and merge and, after exceeding the Chandrasekhar limit, explode as SN Ia. Unlike in accrete scenario the merging WDs are not expected to be the source of X-radiation until a short time preceding the supernova explosion. The X-ray signatures of SD and DD channels differ significantly, which makes them easy to distinguish. The DD model admits a broader range of progenitor mass and SNe Ia luminosity; thus is thought to be responsible for the non-standard SNe Ia explosions.

These two basic models (hereinafter collectively referred to as "SNe Ia binary paradigm") do not however provide a fair explanation to the diversity in the observed characteristics of SNe Ia and the paucity of their potential progenitors. The relevant SNe Ia progenitor problem amounts to the following two items. First is the problem of SNe Ia rate: the total number of potential progenitors seems to be inadequate to the number of observed SNe Ia events. Second is the problem of SNe Ia properties: the observed light-curves and remnants spectra do not match satisfactorily the detailed predictions of SD and DD models.

Our goal here is to provide a solution to the progenitor problem based on assumption of the varying Chandrasekhar mass, a consequence of varying constant Gc^{-2}. It's not been a century yet since one realized our universe has a turbulent history behind and some kind of final fate ahead. Compared with the prior model of eternal and basically invariable universe, this forms quite different ground for thinking about physical fundamental constants. One cannot ascribe logical necessity to any of fundamental constants (class C "universal" constants, according to Uzan's nomenclature(Uzan [74]) as e.g. in the case of mathematical constant π or the Euler's number e. Likewise, one cannot obtain them by pure deduction in a way similar to that Eddington tried (ineffectively) to do with the fine structure constant alpha. For the time being, they work as "free parameters". Hence, still valid is Dirac's opinion: "It is usually assumed that the laws of nature have always been the same as they are now. There is no justification for this. The laws may be changing, and in particular quantities which are considered to be constants of nature may be changing with cosmological time" (Dirac [16]). Let us complement this opinion with another one: "Ignoring the

possibility of varying constants could lead to a distorted view of our universe and if such a variation is established corrections would have to be applied" (Uzan [74]).

2 The SNe Ia progenitor problem: a brief overview

The question of identity of Type Ia supernovae progenitors is widely considered as the "major unsolved problem in astrophysics" (Maoz & Mannucci [47]). The main problem is the discrepancy between the observed SNe Ia rate and the number of potential progenitors. Taking into account the estimated rate of SNe Ia ($\sim (10^{-3} - 10^{-2})\mathrm{yr}^{-1}$ events in a typical spiral or elliptical galaxy) and the mean/median delay time for the SNe Ia progenitors ($\sim (0.5 - 1)$ Gyr for DD channel and $\sim (2 - 3)$ Gyr for SD channel), X-ray sources should manifest in thousands in any such galaxy including the Milky Way. Meanwhile, the X-ray flux from the sample of six neighboring spiral galaxies obtained from Chandra X-ray Observatory is a factor of 30-50 times fainter than expected (Gilfanov & Bogdan [29]). In some of SNe Ia previously thought to originate in SD channel no remnants of red giant has been observed (Schaeffer & Pagnotta [70], Li et al. [42], Nugent et al. [56]). Generally, in most cases red giants have been excluded as possible ex-companions in binaries. The discrepancy between the observed amount of X-ray sources and the assessed numbers of SNe Ia led to conclusion that accrete scenario is not a primary route to supernovae, giving priority to the merger scenario. Gonzalez Hernandez et al. [30] estimate that fewer than 20% of SNe Ia is produced in SD channel. Gilfanov & Bogdan [29] opt for even more stringent limit $\leqslant 5\%$ of total population. Di Stefano [17] indicates the lack of $90\% - 99\%$ of the required number of X-ray sources. She argues (Di Stefano [18]) that companion stars forming the double degenerates do not age at the same rate and thus do not become WDs at the same time; for that reason the common envelope phase should be preceded by a symbiotic pre-double-degenerate phase with the hydrogen-helium fusion similar to NBWD. Thus, merger channel should also produce X-ray flux comparable to the accrete channel prior to the common envelope phase, which puts into doubt DD model as an effective explanation.

A vital problem is the paucity of the observed white dwarfs mergers. According to Gilfanov [28] "...too few double-white-dwarf systems appeared to exist". One expects the ESO Supernovae Type Ia Progenitor Survey (SPY) (Napiwotzki et al. [54, 55]) and the ongoing Sloan White dwArf Radial velocity data Mining Survey (SWARMS) (Badenes et al. [3], Mullally et al. [53]) to provide evidences for the merger channel (DD) as the main route to SNe Ia. Badenes & Maoz [4] using Doppler techniques isolated 15 WD binaries from a sample of $\approx 4,000$ WDs brought by Sloan Digital Sky Survey (SDSS). They compared the rate of WD binaries with the rate of SNe Ia in the Milky Way-like Sbc galaxies and found a "remarkable agreement" between them. However, a majority of these WD binaries appeared to be sub-Chandrasekhar, although usually with total mass relatively close to M_{Ch} ($1.1 - 1.2M_{\odot}$).

Some of researches (Hachisu et al. [34], Van Kerkwijk et al. [40], Zhu et al. [82], Maoz & Mannucci [47]) claim that the requirement as to the total mass of merging CO WDs (i.e $1.4M_{\odot}$) is too restrictive. This would match observations of super-Chandra WD progenitor stars with the combined mass reaching $2.4 - 2.8M_{\odot}$. According to the respective models, the observed number of SNe Ia can be explained provided the wider range of combined mass: smaller than M_{Ch} (sub-M_{Ch} merger channel) or bigger than M_{Ch} (super-M_{Ch} merger channel), dependently on detailed conditions such as rotation, magnetic fields, metallicity and the host galaxy population. This would account for better agreement with observations, both as to the rate of SNe Ia and to the differences in their properties. The controversial point of these models is that they require special fine-tuning to be effective. Maoz & Mannucci [48] attribute some of discrepancies as caused by "deadly sins", i.e. incorrect or inadequate methods in measuring and analyzing the SNe Ia rates. They admit however the "detailed models still falls short of the observed number (of SNe Ia) by at least factor of a few".

Di Stefano [18] suggests that, possibly, only a small fraction of accreting WDs can be detected and identified as X-ray sources. This may occur by two reasons: either the winds from a companion giant reprocess the supersoft X-ray radiation into the radiation of longer wavelengths, or the duty cycle of nuclear burning is to low to be detected. However, neither of these solutions has been properly recognized and confirmed as yet. Another proposal (Di Stefano et al. [19]) links the mass of progenitor with the angular momentum gained from the donor star together with matter. The angular momentum prevents the super-M_{Ch} WD from collapse, which widens the potential range of SNe Ia progenitors. The relevant "spin-up/spin-down" models predict the existence of numerous WD "ticking bombs" waiting to explode until their rotation slows down to a proper level.

There is a broad agreement (e.g. Totani et al. [73], Maoz et al. [48], Mennekens et al. [50], Hachisu et al. [33]) as to the key role of "delay time distribution" (DTD) — the number of SNe Ia events in unit time as a function of time elapsed since starburst, in predicting the SNe Ia rates. It seems that DTD (indicated as t^{-1} power law) favors the DD scenario. Hachisu et al. [33] found a good agreement of DTD with SD model either, provided the donor stars are both red giants and the main-sequence stars. Undoubtedly, DTD introduces an indispensible methodological order to the SNe Ia progenitor problem. In general however, regarding DTD did not bring a decisive breakthrough so far in the question of identity of SNe Ia progenitors.

It has gradually become evident that SNe Ia are not "standard candles" in the originally attributed sense. Their intrinsic luminosity is neither considered nor demanded to be ex-

actly uniform, which gives priority to the more "capacious" merger channel. Instead of standard candles, SNe Ia are currently interpreted as "standarizable candles", which means that utilizing them as the correct distance indicators requires due calibration. This in turn demands better recognizing of their origin and nature. The study by Linden, Virey & Tilquin [43] revealed a likely positive correlation between the SNe Ia absolute brightness and distance, which may put in question the actually determined cosmological parameters. The observed relationship between the intrinsic color and ejecta velocity may help in reducing systematic biases in the estimates of distance (Foley et al. [25]). Instead, Sullivan et al. [72] point to the relationship between the luminosity of SNe Ia and metallicity of their hosts, while metallicity is supposed to depend on redshift. Gallagher et al. [26] comparing the spectra of a sample of 29 early elliptical galaxies of the age exceeding 5 Gyr with the general sample from SDSS including younger galaxies, find a strong correlation between the absolute magnitude of SNe Ia and the age of host galaxies while, most likely, "...the observed trend with metallicity is merely an artifact brought about the evolutionary entanglement of age and metallicity". These findings may help in recognizing the properties of SNe Ia, which is particularly important for the question of dark energy and the relevant accelerating expansion of the universe (Riess et al. [66], Perlmutter et al. [60]). The supposed correlation between the absolute magnitude and distance suggests the presence of a time dependent factor in the effective SNe Ia progenitor model.

3 Varying Chandrasekhar limit as the postulated main route to SNe Ia

The mass-limit formula for white dwarfs based on the equation of state for ideal Fermi gas (Chandrasekhar [11]) reads

$$M_{Ch} = 4\pi \left(\frac{K_2}{\pi G}\right)^{3/2} \omega_3^0 , \tag{1}$$

where ω_3^0 is the numerical constant equal to 2.018, derived from the explicit solution of the Lane-Emden equation for the polytropic index $n = 3$. The constant K in the general case connects pressure and density: $P = K\rho^{(n+1)/n}$ while in the case including white dwarfs (i.e. for $n = 3$) becomes specified as $P = K_2\rho^{4/3}$. Since K_2 is defined as

$$K_2 = \frac{1}{8}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \frac{hc}{(\mu_e m_H)^{4/3}} \tag{2}$$

(μ_e-mean molecular mass per electron, m_H-mass of hydrogen atom), so substituting gives

$$M_{Ch} = 4\pi \left[\frac{1}{8}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \frac{hc}{(\mu_e m_H)^{4/3}\pi G}\right]^{3/2} \omega_3^0 . \tag{3}$$

Collecting the pure numbers, and considering that $\hbar = h/2\pi$, one gets

$$M_{Ch} \approx 1.11065 \times 10^{54} \mu_e^{-2} m_P^3 , \tag{4}$$

where $m_P = (\hbar c/G)^{1/2}$ is the Planck mass. Since CO WDs are mainly composed of carbon-12 and oxygen-16, and because in both cases atomic number equal to half the atomic weight so one has $\mu_e = 2$, leading to $M_{Ch} \approx 1.44 M_\odot$. It is important that Chandrasekhar mass it is proportional to the cube of Planck mass:

$$M_{Ch} \sim m_P^3 . \tag{5}$$

Assuming $\hbar = const$ it relates to the speed of light and to the Newton's gravitational constant as

$$M_{Ch} \sim (c/G)^{3/2} . \tag{6}$$

(We use tilde for linear dependence in the cases when the variability of a reference quantity [here: c and G] is hypothetical. Instead, the symbol of proportionality [exact or approximate] $\propto$ is used when variation of a reference quantity is obvious or certain, e.g. cosmic time t or radius of universe R_u).

From this relationship it follows that any cosmological model postulating varying G or/and c (except the case they change accordingly) implies the postulate of varying M_{Ch}. This fact has not been properly explored so far. What we propose here is the "varying Chandrasekhar mass-limit" model (VCM) in which M_{Ch} decreases in cosmic time. VCM postulates that the currently known value of Chandrasekhar limit refers solely to the present epoch while in general:

$$M_{Ch}(\text{past}) > 1.44 M_\odot > M_{Ch}(\text{future}). \tag{7}$$

This determines a scenario for the single WD progenitors of SNe Ia, which can be outlined as follows. Once an individual WD is formed, it keeps its mass approximately constant during the cooling process while the Chandrasekhar limit gradually decreases in time. Eventually, it equates or approaches a given WD's mass triggering the SN Ia explosion. From a logical point of view, an effect of SN Ia caused by decreasing M_{Ch} reminds bringing water to a boil by reducing the atmospheric pressure without supplying heat. Hence, single WDs are, along with binary WDs, the potential progenitors of SNe Ia.

4 Varying constants and the black-hole cosmology

Varying Chandrasekhar limit, as a hypothesis based on assumption of varying constants c or/and G is closely related to the black-hole cosmology. A constitutive observation of the respective models is the coincidence between the radius of observable universe and the Schwarzschild radius, supposed to be valid over the whole course of the universe's history. According to a hypothesis advanced by Pathria [58] and Good [31], the universe is the interior of a black hole existing, among many others, within a larger structure called multiverse.

The recent multiverse model by Popławski [64, 65] uses the Einstein-Cartan-Sciama-Kibble theory removing from General Relativity the constraint of symmetry in the affine

connection, and regarding the antisymmetric variable torsion tensor in the Friedmann equations. The relevant cosmological scenario takes an advantage of the fact that most stars have a non-zero angular momentum. When a massive rotating star collapses to a (Kerr) black hole, the torsion of extremely dense matter inside the horizon prevents from the point singularity (replaced by the ring singularity). As a result, the black hole becomes a wormhole to another universe thought to originate in "big bounce". As far as our own universe is the interior of a black hole existing in another universe, any black hole in our universe is thought to contain (produce) a separate universe. The new universe is interpreted as a "white hole" — a time reversal black hole whose expansion, e.g. such as observed in our universe is driven by the torsion, identified with dark energy. This model predicts the presence of traces of primordial torsion in a form of slight anisotropies in both cosmic and nanoscopic scales. Some reported evidences of the preferred handedness of spiral galaxies (dipole asymmetry of the value 0.0408 ± 0.011 based on SDSS data sample containing 15,158 spiral galaxies with the redshift < 0.085) seem to support the idea of cosmic parity violation (Longo [44]). However, the area covered by this sample is still too small to derive unambiguous conclusions. According to Neta Bahcall "The directional spin of spiral galaxies may be impacted by other local gravitational effects".

Besides, even if the filaments forming the cosmic web are uniformly distributed, anisotropy connected with rotation will break the homogeneity in a deeper sense. In the isotropic cosmic space, the "center" is a purely relative concept connected with the notion of observable universe. But it is no longer relative in the anisotropic space with the fixed axis of rotation. The spinning universe implies, besides anisotropy, the presence of preferred points. We may think about analogies between directional spin of spiral galaxies and the Coriolis effects on the Earth, e.g. manifesting itself in different spin of hurricanes in north and south hemispheres. Anyway, the question of spinning universe is, in the end, a matter of (further) observations.

The model here proposed (VCM) bases on formal resemblance of our universe with a black hole (and thus we shall use the Schwarzschild equation for radius) yet does not settle whether the universe is a black hole in the literal sense. It seems instead that crucial property of the universe conceived as the interior of a black hole is that its total energy amounts to zero. In this regard, the black-hole cosmologies are close to the "zero-energy universe" theories.

The legitimacy for interpreting the universe in terms of a black hole depends on its parameters, in particular size, density and mass. Recent estimations concerning the radius of observable universe point to the value $\geqslant 14\,\mathrm{Gpc}$ ($4.3 \times 10^{26}\,\mathrm{m}$) or 28 Gpc in diameter. Cornish et al. [12] analyzing the WMAP data in search of the matched back-to-back circles predicted by various nontrivial topologies, settled the low bound of diameter of the last scattering surface of fundamental domain for 24 Gpc. Bielewicz & Banday [6], using similar methods extended this value to 27.9 Gpc. This admittedly does not prejudge the question of size, yet, provided the multi-connected space of universe, constraints the topology scale from below. An additional (though partly linked) difficulty comes out from the potential difference between the notions of entire and observable universe. In principle, entire universe may significantly surpass the observable universe (as inflationary theory predicts), but it can be as well slightly smaller due to nontrivial topology. The respective ratio may also change in time. Presumably, the black hole parameters describe the entire universe, and not just the universe currently observed. However, this distinction becomes important only insofar as "entire", by virtue of convention, denotes the biggest physically connected object defined according to the horizon problem of the early universe. Assuming the approximately linear rate of expansion after the end of inflationary epoch (or from the beginning), the parameters of the so defined "entirety" should not significantly differ from the "observable" parameters. Bearing in mind the obvious uncertainties, we shall use in calculations the value $10^{27}m$ for the universe's radius.

The critical density for a flat universe derived from Friedmann equation for the Hubble constant obtained from Planck telescope: $H_0 = 67.15\,\mathrm{kms^{-1}Mpc^{-1}}$ is $\rho_c = 3H^2/8\pi G \approx 0.85 \times 10^{-26}\,\mathrm{kgm^{-3}}$. The resultant total mass for $R_u = 10^{27}\,\mathrm{m}$ amounts to $M_u \approx 1.44 \times 10^{54}\mathrm{kg}$ (we shall use $10^{54}\mathrm{kg}$ in calculations). Considering the approximated values of gravitational constant: $G \approx 6.7 \times 10^{-11}\,\mathrm{m^3kg^{-1}s^{-2}}$ and the speed of light: $c \approx 3 \times 10^8\,\mathrm{ms^{-1}}$ ($c^2 \approx 10^{17}\,\mathrm{m^2s^{-2}}$) one obtains the numerical relationship connecting radius and mass:

$$10^{27} = 10^{-10}10^{54}10^{-17}\,(\mathrm{m}), \tag{8}$$

which means that equation for the Schwarzschild radius:

$$r_S = 2Gmc^{-2} \tag{9}$$

apparently applies to the universe as

$$R_u \approx GM_uc^{-2}. \tag{10}$$

We postulate that universe constantly fulfills the "black-hole condition" (BHC), which means that it is always fulfilled:

$$R_u \equiv r_S\,. \tag{11}$$

Together with assumption $M_u = const$, and the general assumption of isotropy of cosmic space, BHC implies

$$Gc^{-2} \propto R_u\,. \tag{12}$$

5 Models with varying constants

In the intensive discussion on the variability of fundamental constants, variation of c is probably the leading topic. A majority of the "variable speed of light" (VSL) models conceived as a challenge to inflation restricts the variation of c

to the early superluminary universe (Moffat [51], Albrecht & Magueijo [2], Magueijo & Smolin [45]). These models do not match BHC since, after restoring the local Lorentz invariance, the light is thought to travel at the presently measured speed. Likewise, assuming the change of c refers totally to the time preceding the structure formation, they would not imply the variability of M_{Ch}.

In some VSL models, the change in c value has been considered as a continuous process spread over the whole lifespan of the universe. Dicke's theory of gravity (Dicke [13]), developing the earlier considerations by Einstein [22, 23] explains the cosmological redshift as a result of c decreasing with time, which somehow corresponds with the steady state theory. However, this model does not predict the change of c to be a measurable effect since it assumes the units of length and time to change accordingly.

In turn, variability of G has been proposed in some scalar-tensor models modifying the Einstein's General Relativity, in particular the Brans-Dicke theory [9] inspired by Mach's principle, with the time and space dependent scalar field ϕ modifying the Newton's constant. A similar as to the general structure and conclusions model by Hoyle & Narlikar [36] originates from considerations concerning the action on distance. Petit [61, 62] advanced a model with joint variation of G, c and h decoding the Hubble's law in a static universe. One of the first models postulating varying G, and likely the most influential one, is the Dirac's "large number hypothesis" (Dirac [15]). From the supposed coincidence between two ratios: radius of universe (expressed as ct) vs. radius of electron, and electrostatic force vs. gravitational force between proton and electron (both of them yielding $\approx 10^{40}$), Dirac derived a conclusion that G changes as the inverse of cosmic time: $G \propto t^{-1}$, while the mass of universe increases as $M_u \propto t^2$. Provided the approximately linear relationship between time and radius ($R_u \propto t$), LNH satisfies BHC. However, LNH also implies $M_{Ch} \propto t^{3/2}$, which compared with the standard assumption of constant G makes the SNe Ia progenitor problem even more puzzling. A model proposed by the present author (Rybicki [69]) has postulated $G \propto R_u$, $M_u = const$, yet then with no reference to BHC and the SNe Ia progenitor problem.

A question underlying the varying constants models is whether the postulated changes in dimensional constants are physically meaningful. A long-lasting controversy over this subject has not been concluded so far. Some physicists (e.g. Barrow [5], Duff [20]) claim that only the (potential) change in dimensionless constants matters, e.g. the coupling constants of fundamental forces such as fine structure constant α, gravitational coupling constant α_G , or the masses of elementary particles related to Planck mass contributing to standard model. Instead, dimensional constants such as $\hbar$, c, G, e, or k may change in value dependently on the (arbitral) choice of units, thus being merely the "human constructs" or "conversion factors". Others (Okun [57], Veneziano [76]) consider as indispensable in shaping the fundamental theories respectively three (G, c and $\hbar$) and two (c and string length λ_s) dimensional constants.

From the "dimensionless" point of view as applied to BHC, no matter whichever of dimensional constants is thought to vary; only what counts is the change of $\alpha_G = Gm_e^2/\hbar c$. Since we discriminate here between the change of G and c treated as different solutions of BHC, so this question demands a clarifying comment. Let's start with two remarks: 1) There is no doubt that $Gc^{-2} \propto R_u$ implies the variability of α_G; 2) The fact that dimensional constant changes its numerical value together with the change of unit is trivial, and as such contributes nothing to discussion.

Let the increase of α_G be observed, correlated with the increase of R_u. Assuming $m_e = const$, $\hbar = const$, we conclude that it is either $G \propto R_u$ or $c \propto R_u^{-1}$ which, according to the "dimensionless" paradigm, we treat as fully equivalent (i.e. physically indistinguishable) interpretations of $\alpha_G \propto R_u$. However, from $Gc^{-2} \propto R_u$ it follows: $G \propto R_u \Rightarrow \alpha_G \propto R_u$, and $c \propto R_u^{-1/2} \Rightarrow \alpha_G \propto R_u^{1/2}$, which obviously differs from $\alpha_G \propto R_u$. Thus, G and c cannot be considered as "conversion factors" within BHC.

As we show in next sections, the Planck units of length and time react differently depending on whether G or c is postulated to vary. Besides, each of respective solutions affects entropy in a different way. We thus agree with the anonymous referee cited in Duff's paper: "It is true that if the fundamental "constants" $\hbar$, c, G, k ... are truly constant, then they do indeed only act as conversion factors and can e.g. be set equal to unity. However, when they are postulated (or discovered experimentally to vary) in time, then we have to take into account that varying one or the other of these constants can have significant consequences for physics" (Duff [20]).

6 Basics of the VCM hypothesis

Expressed in the here proposed nomenclature, our main idea consists in postulating VCM as being the consequence of BHC. Any model satisfying BHC makes the Planck units variable, and thus determines new parameters of the Planck era.

Identifying the mass in the equation for Schwarzschild radius with Planck mass: $m \equiv m_P$ gives

$$r_S = Gm_Pc^{-2} = G(\hbar cG^{-1})^{1/2}c^{-2} = (\hbar Gc^{-3})^{1/2} = \ell_P\,. \quad (13)$$

Accordingly, the black hole becomes the Planck particle. Implementing the Planck mass to the reduced Compton wavelength $\lambda/2\pi = \hbar m^{-1}c^{-1}$ makes the Planck particle the only one black hole whose Schwarzschild radius equals the Compton wavelength

$$\lambda/2\pi = \hbar(G\hbar^{-1}c^{-1})^{1/2}c^{-1} \equiv (\hbar Gc^{-3})^{1/2} = \ell_P\,. \quad (14)$$

Rewriting the Schwarzschild equation for the Planck particle:

$\ell_p \approx G m_p c^{-2}$ gives the identity

$$(\hbar G c^{-3})^{1/2} \equiv G(\hbar c G^{-1})^{1/2} c^{-2}, \tag{15}$$

which means that Planck particle's property of being a black hole is insensible to the change of G or/and c.

From $Gc^{-2} \propto R_u$ it follows $m_P \propto R_u^{-1/2}$; hence for $R_u \to 0$ the Planck mass tends to infinity. However, to avoid singularities (and also taking into account that Planck mass should have "realistic" reference), we assume that in the newly defined Planck era (denoted P_0) the Planck mass coincidences with the mass of universe:

$$m_{P_0} \equiv M_u \,. \tag{16}$$

Thus, the initial value of Schwarzschild radius becomes

$$r_{S_0} \approx G M_u c^{-2}. \tag{17}$$

This can be also obtained by expressing the Newton's constant in the equation $R_u = G M_u c^{-2}$ in terms of Planck units, namely: $G = \ell_P m_P^{-1} c^2$. Then

$$R_u = \ell_P M_u m_P^{-1} \tag{18}$$

and so

$$R_u \ell_P^{-1} = M_u m_P^{-1} \tag{19}$$

meaning that identity $R_{u_0} \equiv \ell_{P_0}$ becomes a consequence of the conjecture $M_u m_{P_0}^{-1} = 1$.We have thus arrived at conclusion that the universe at its initial stage (here called "primordial Planck era" — PPE) had the form of a quantum mechanical black hole identified with a single one "primordial Planck particle" (PPP), described by equation:

$$\ell_{P_0} = G_0 m_{P_0} c_0^{-2} \,. \tag{20}$$

Accordingly, the notion of PPP becomes coherent with the concept of the universe emerging from "nothing" due to the Heisenberg uncertainty.

From $M_u \approx 10^{54} kg$, provided $m_{P_0} \equiv M_u$, it follows

$$m_{P_0} m_P^{-1} \approx 10^{62} \tag{21}$$

a factor hereinafter denoted by δ.

Because $M_{Ch} \sim m_P^3$ so

$$M_{Ch_0} M_{Ch}^{-1} = \delta^3. \tag{22}$$

Obviously, M_{Ch_0} as related to the early universe, is a formal entity only. To be a physically meaningful concept, Chandrasekhar limit demands a proper physical "enviroment" (atoms, elements, stars). It belongs then to the epoch of structure formation starting from Population III stars. Provided the universe expanded in a roughly uniform rate, BHC can be expressed as the approximate function of cosmic time: $Gc^{-2} \propto t$. From the whole range of possible BHC scenarios, the two deserve special attention, namely: 1) $G \propto R_u$ i.e. $G \propto t$, $c = const$, and 2) $c \propto R_u^{-1/2}$, $G = const$, both analyzed in the next sections.

7 Assumption $c \propto R_u^{-1/2}$, $G = const$: collision with the second law of thermodynamics

The initial value of speed of light derived from $m_{P_0} = (\hbar c_0/G)^{-1/2}$ and $m_{P_0} \equiv M_u \approx 10^{54}\,\mathrm{kg}$ becomes $c_0 = 10^{132}$ ms^{-1}, yielding $c_0/c \approx 10^{124} = \delta^2$. The respective Planck length is (hereinafter, SI units always when omitted)

$$\ell_{P_0} = (\hbar G/c_0^3)^{-1/2} \approx 10^{-220} \tag{23}$$

a value equal to the Schwarzschild radius

$$r_S = G m_{P_0}/c_0^2 \approx 10^{-220} \tag{24}$$

and to the Compton wavelength

$$\lambda_0 = \hbar M_u^{-1} c_0^{-1} \approx 10^{-220}. \tag{25}$$

The initial Planck time would amount to

$$t_{P_0} = (\hbar G/c^5)^{-1/2} \approx 10^{-352}. \tag{26}$$

From $E = m_{P_0} c_0^2$ it follows

$$\hbar = E t_{P_0} (10^{-34} = 10^{318} 10^{-352}). \tag{27}$$

As derived from $c \propto t^{-1/2}$, with the age of universe $\approx 13.8 \times 10^9$ yr the current rate of decrease in the speed of light becomes

$$\dot{c}/c \approx -2.7 \times 10^{-11} \mathrm{yr}^{-1}. \tag{28}$$

Let us compare this prediction with the results obtained from observations of gas clouds spectra intersecting the distant quasars, the Oklo natural uranium fission reactor, and atomic clocks. In agreement with the VSL paradigm, the supposed change of α is usually interpreted as the change of c. For the approximate emission time connected with the observational data samples concerning quasars: $t_{EM} \approx 0.25\, t_0 / 0.85\, t_0$ covering $\approx 8.3\,\mathrm{Gyr}$ (here t_0 stands for the present moment), the reported values suggesting the change are: $\Delta c(t)/c = (-0.57 \pm 0.10) \times 10^{-5}$ (Webb et al. [77]), and $\Delta c(t)/c = (-1.09 \pm 0.17) \times 10^{-5}$ (Webb et al. [78]). At the same time, other groups (e.g. Chand et al. [10]) reported no detectable change in α value over the last 10-12 billion years. In the case of Oklo, for the respective operating time $t_{prev}/t_0 \approx 0.87$, Petrov et al. [63] obtained $\dot{\alpha}/\alpha = (-4+3) \times 10^{-17} \mathrm{yr}^{-1}$, in fact signifying no detectable change. In turn, Lamoreaux & Torgerson [41] reported a decrease in alpha at the level -4.5×10^{-8} over the last 2 billion years, which consequently should be interpreted as the increase of the speed of light. Observations based on atomic clocks give a direct insight to the possible current rate of change. Peik et al. [59], using cesium atomic clock set the limit of annual change of the present variation of alpha for $\dot{\alpha}/\alpha = (-1.2 \pm 4.4) \times 10^{-15} \mathrm{yr}^{-1}$. In turn, Rosenband et al. [67], based on the frequency ratio of Al^+ and Hg^+ in a single ion atomic clocks obtained a bound: $\dot{\alpha}/\alpha = (-1.6 \pm 2.3) \times 10^{-17} \mathrm{yr}^{-1}$.

Except the data provided by Webb et al. suggesting decrease of c at the level $10^{-15}\mathrm{yr}^{-1}$, and the opposite one (as to general conclusion) provided by Lamoreaux & Torgerson, all other results seem to point the zero change. This suggests the failure of assumption $c \propto R_u^{-1/2}$. Besides, the question of entropy provides us with an additional argument against declining c. As is known, entropy is proportional to the horizon surface area, which normally (i.e. by assumption $G = const$, $c = const$) implies linear dependence on the squared mass. Let us apply the Bekenstein-Hawking formula for the entropy of black hole:

$$S_{BH} = Akc^3(4G\hbar)^{-1} \tag{29}$$

or, written in terms of Planck length,

$$S_{BH} = Ak(4\ell_P^2)^{-1}, \tag{30}$$

where A is the surface area for event horizon, and k the Boltzmann constant. For the spherically symmetrical black hole, the surface area is $A = m^2 8\pi G^2 c^{-2}$ so entropy becomes $S_{BH} = m^2 2\pi G k c \hbar^{-1}$. Thus, despite increasing surface area $A = m^2 8\pi G^2 c^{-2}$, at the assumption $m \equiv M = const$, $G = const$ and $c \propto R_u^{-1/2}$, the entropy decreases according to $S_{BH} = m^2 2\pi G k c \hbar^{-1}$ being dependent on the decreasing speed of light: $S_{BH} \sim c$. One obtains therefore

$$S_{BH(present)}/S_{BH(primordial)} = \delta^{-2}, \tag{31}$$

which violates the second law of thermodynamics applied to the universe as a whole. This does not exclude VSL models in general; in particular, does not exclude VSL applied to the very early universe. However, BHC is not agreeable with VSL conceived as a continuous process. Therefore, in the further considerations, we shall specify BHC as a model defined by the assumption $G \propto R_u$, $c = const$. We shall also treat this model as a right basis for the VCM hypothesis and the respective quantitative predictions.

8 Assumption $G \propto R_u$, $c = const$: parameters of the universe at Planck era

Provided $m_{P_0} \equiv M_u \approx 10^{54}\,\mathrm{kg}$, the initial value of Newton's constant derived from $m_{P_0} = (\hbar c/G_0)^{-1/2}$ is $G_0 \approx 10^{-134}$, yielding $G/G_0 = \delta^2$. The initial Planck length becomes

$$\ell_{P_0} = (\hbar G_0/c^3)^{-1/2} \approx 10^{-97} \tag{32}$$

equal to the Schwarzschild radius:

$$r_S = G_0 M_u c^{-2} \approx 10^{-97} \tag{33}$$

and to the (constant) value of Compton wavelength for the universe:

$$\lambda_0 = \hbar M_u^{-1} c^{-1} \approx 10^{-97}. \tag{34}$$

All three quantities apply to the initial size of universe R_{u_0}:

$$R_{u_0} \equiv \ell_{P_0} \equiv r_S \equiv \lambda_0\,. \tag{35}$$

The initial Planck time is

$$t_{P_0} = (\hbar G_0 c^{-5})^{1/2} \approx 10^{-105}. \tag{36}$$

Hence,

$$\hbar = E_{P_0} t_{P_0} \approx 10^{-34}, \tag{37}$$

where $E_{P_0} = M_u c^2 \approx 10^{71}$. The invariability of Planck constant is a consequence of the fact that, although individually Planck energy and Planck time change in time, their product remains constant:

$$E_{P(variable)} \times t_{P(variable)} = \hbar_{(\mathrm{constant})}\,. \tag{38}$$

In general, initial values of the base Planck units relate to their present equivalents as

$$m_{P_0}/m_P = \ell_P/\ell_{P_0} = t_P/t_{P_0} = \delta\,. \tag{39}$$

The horizon problem in PPE is solved so to speak by definition, since

$$ct_{P_0} = \ell_{P_0}\,, \tag{40}$$

which means that the whole primordial universe fits in a light cone.

The density in the primordial Planck era is

$$\rho_{(PPE)} = M_u \ell_{P_0}^{-3} \approx 10^{344} \tag{41}$$

equal to initial Planck density:

$$\rho_{P_0} = c^5 \hbar^{-1} G^{-2} \approx 10^{344}. \tag{42}$$

Let us compare this with the critical density derived from the Friedmann equation: $\rho_c = 3H^2(8\pi G)^{-1}$, as calculated for PPE. The current value of Hubble constant ($\approx 70\,\mathrm{kms}^{-1}/\mathrm{Mpc}$) expressed in SI units amounts to

$$H_{(now)} \approx 2.27 \times 10^{-18}\,\mathrm{s}^{-1} \tag{43}$$

yielding the respective value of the Hubble constant in PPE:

$$H_{(PPE)} = H_{(now)} \times \delta^2 \approx 10^{106}\,\mathrm{s}^{-1}. \tag{44}$$

Approximating $8\pi G_0 \approx 10^{-133}$, one obtains the PPE critical density:

$$\rho_{c(PPE)} \approx 10^{212} 10^{133} \approx 10^{345}. \tag{45}$$

Hence, it is likely that also in PPE

$$\rho_{(PPE)} \equiv \rho_c \tag{46}$$

which solves the flatness problem.

In contrast to the previously considered assumption $c \propto R_u^{-1/2}$, $G = const$, the thermodynamic arrow of time becomes well defined. Considering $G/G_0 = \delta^2$, from $S_{BH} = m^2 2\pi G k c \hbar^{-1}$ it follows

$$S_{BH(present)}/S_{BH(primodial)} = \delta^2. \tag{47}$$

In the cosmological scenario based on assumption $G \propto R_u$, $c = const$, the expansion is linear, or roughly linear, including the early epoch. This means that $G \propto R_u$ is tantamount to $G \propto t$; in particular, $G_0(10^{-134})$ coincides with $t_{P_0}(10^{-105})$. At some additional assumptions, this scenario could be modified so as to regard nonlinear expansion during early epochs. However, considering that basic motives for invoking inflation (horizon problem and flatness problem) are absent in BHC scenario, inflation appears to be basically redundant.

9 Assumption $G \propto R_u$, $c = const$: question of consistence with observational tests of G variability

Provided the approximately uniform rate of Hubble flow, the derived from $G \propto R_u$ current rate of increase of G becomes a simple inverse of the age of universe. In fact, the Hubble time does not significantly differ from estimations of the age of universe derived from Friedman equation equipped with definite values of k and Λ. Whereas these estimations range from ≈ 13.798 Gyr (Lambda-CDM concordance model based on data from Planck satellite and WMAP) to $\approx$ 13.82 Gyr (Planck mission), the Hubble time ranges between ≈ 13.7 Gyr and ≈ 14.26 Gyr according to the current extreme estimates of the Hubble constant: ≈ 72 and and $\approx 67\,\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$ respectively. Thus, on the average, the Hubble time only slightly exceeds the supposed age of universe. Interpreting $G \propto R_u$ as $G \propto t$ and estimating the age of the universe for $\approx 13.8\times10^9$yr gives the current rate of change:

$$\dot{G}/G \approx 7.25 \times 10^{-11}\mathrm{yr}^{-1}. \tag{48}$$

Let us compare this prediction with the constraints put upon G variation, derived from different sources (paleontology and geophysics, celestial mechanics, stellar physics, cosmology). A handful of representative results covering the whole range are:

— paleontological data connected with Earth temperature: $|\dot{G}/G| < 2.0\times10^{-11}\mathrm{yr}^{-1}$ (Eichendorf & Reinhardt [21]);

— increase of Earth radius: $\dot{G}/G = (-0.5 \pm 2)\times10^{-11}\mathrm{yr}^{-1}$ (Blake [8]);

— stability of the radii of Earth, Moon and Mars: $-\dot{G}/G \leqslant 8 \times 10^{-12}\mathrm{yr}^{-1}$ (McElhiny et al. [49]);

— stability of the orbit of Mars (Mariner 9 and Mars orbiter data): $\dot{G}/G = (-2\pm10)\times10^{-12}\mathrm{yr}^{-1}$ (Shapiro [71]);

— systematic deviations from the Keplerian orbital periods of Moon: $\dot{G}/G \leqslant (3.2 \pm 1.1)\times10^{-11}\mathrm{yr}^{-1}$ (Van Flandern [75]);

— lunar laser ranging (LLR): $|\dot{G}/G| < 6\times10^{-12}\mathrm{yr}^{-1}$ (Dickey et al. [14]); LLR: $\dot{G}/G \leqslant (4 \pm 9)\times10^{-13}\mathrm{yr}^{-1}$ (Williams et al. [81]);

— spin-down of pulsar JP1953: $-\dot{G}/G < 5.8 \pm 1 \times 10^{-11}$ yr^{-1} (Mansfield [46]);

— pulsar timing PSR B1913+16: $\dot{G}/G \leqslant (4 \pm 5) \times 10^{-12}$ yr^{-1} (Kaspi et al. [38]);

— luminosity function of white dwarfs (cooling age): $-\dot{G}/G \leqslant 3^{+1}_{-3} \times 10^{-11}\mathrm{yr}^{-1}$ (Garcia-Berro et al. [27]);

— pulsating white dwarf data G117-B15A: $|\dot{G}/G| \leqslant 4.10\times 10^{-10}\mathrm{yr}^{-1}$ (Biesiada & Malec [7]);

— SNe Ia luminosity vs. redshift: $\dot{G}/G = (-3, +7.3) \times 10^{-11}\mathrm{yr}^{-1}$ (Mould & Uddin [52]);

— helioseismology: $|\dot{G}/G| \leqslant 1.6\times10^{-12}\mathrm{yr}^{-1}$ (Guenther et al. [32]);

— big bang nucleosynthesis (BBN): $|\dot{G}/G| \leqslant 9. \times 10^{-13}$ yr^{-1} (Accetta et al. [1]); BBN: $|\dot{G}/G| \leqslant 1.7\times10^{-13}$ yr^{-1} (Rothman & Matzner [68]).

One can easily notice that BHC prediction hardly matches the minority of the above bounds. However, a closer insight into methodology reveals various circumstances hidden behind the digits. We shall discuss them now, one by one.

9.1 Accuracy of the constraints on G variation and accuracy in measurements of the value of G

Unlike in the case of other fundamental constants, the increasing precision of measurements of G value is accompanied by increasing discrepancy of the obtained results. This led the CODATA to widen the uncertainty range from 0.013% to 0.15%. We ask whether this uncertainty may impinge on the G variability tests. This question does not seem groundless taking into account the ratio between typical bound put on the annual rate of change of G ($\sim 10^{-11}$) and the uncertainty range of G value (1.5×10^{-3}), roughly ten-billionth! To better realize the scale, imagine we test the Wegener's continental drift theory (btw unaccepted for a long time) by settling a constraint on the annual rate of relative motion between two continents, say, America and Europe. Assume we determine two points (measuring devices) placed on each of these continents, and estimate the distance between them for 5 thousand kilometers. However, due to hypothetic imperfection of measuring techniques, this distance is only known with the relative uncertainty 0.15%, which translates into 7.5*km*. Assume next that, undeterred by this immense inaccuracy, we derive the constraint for the drift rate for $10^{-11}\mathrm{yr}^{-1}$, i.e. 0.05 mm/year, while the drift rate estimated by the theory amounts to $7.25 \times 10^{-11}\mathrm{yr}^{-1}$, i.e. 0.36 mm/year (in fact, Wegener estimated the speed of drift for 2.5 m/year, while the currently observed rate amounts to about 2.5 cm/year).

Obviously, measuring a given value and measuring a change in this value are, basically, two different things; yet the mentioned discrepancy is too significant to be ignored. This in particular happens when a constraint depends on assumptions that are themselves encumbered by sizeable uncertainty (see subsection 9.3). In the above fictional example, before drawing ultimate conclusions as to the correctness of Wegener's idea, one should certainly aim at eliminating the distance uncertainty or try to find its hidden sources. Otherwise, any ultimate conclusions as to the change of distance

could not be considered reliable. There is no reason to assume the question of variability of Newton's constant should subject to different rules.

9.2 Differences in notation and the question of autonomy of particular constraints

There is no unique notation for the constraints on G variation; for different reasons, particular constraints (or their groups) are expressed in different mathematical forms. Revealing their meaning provides us with a better insight into the question of autonomy. The (here called) canonical form, $\dot{G}/G \leqslant (a \pm b) \times 10^{-c}\mathrm{yr}^{-1}$ (a positive/negative, b, c positive) reads: "An annual rate of increase/decrease of G, not greater than $a \times 10^{-c}$ has been observed, with the uncertainty range equal to $\pm b \times 10^{-c}$". If $a = 0$, it means that no change has been observed, although b still describes the range of uncertainty of that finding. Expression $-\dot{G}/G$, instead of $\dot{G}/G$, means that given constraint concerns solely (is design to detect) the decrease rate of G. This takes place when a theory predicting the decrease of G (e.g. Dirac's LNH) is tested, and thus respective assumptions are the base of derivation. In turn, the form $|\dot{G}/G|$ reads: "The possibility of G variation (including increase and decrease in equal degree) fits in the range..." However, $|\dot{G}/G|$ is sometimes used as equivalent to $-\dot{G}/G$, in particular when aimed at testing Dirac's hypothesis (e.g. Eichendorf & Reinhardt [21]). This form implies a to be indistinguishable from b, i.e. treats expressions "(rate) not greater than" and "with the uncertainty range" as tantamount to each other. Another way to identify the range of possible change with the range of uncertainty is the form $\dot{G}/G = (-b_1, +b_2) \times 10^{-c}\,\mathrm{yr}^{-1}$, $b_1 \neq b_2$. Although apparently similar to $|\dot{G}/G|$, this form indicates the observed tendency (i.e. increase or decrease) and thus seems to be basically equivalent to the canonical form; e.g. the term $(-2, +4)$ could be expressed as (1 ± 3). An alternative use of the relation symbols $<$, $\leqslant$ and $=$ in each of the above forms can be interpreted (dependently on the context) as a gradable expression of conviction as to the observed tendency. In particular, symbols $<$ and $\leqslant$, when used in the canonical form, play the role of additional proviso (apart of b term) due to general uncertainty; for example, if $|a|$ is greater than b then using $=$ unambiguously points to the observed change of G. Instead, using $<$ or $\leqslant$ weakens this statement, suggesting the change to be only probable.

Let us assume that, generally, all observations meet the criteria of scientific rigor. Apart of proper methodology and precision, this would also mean the unbiased standpoint as to the principal question, i.e. whether the Newton's constant is a true constant. Provided that, the postulate of autonomy says that each constraint should be interpreted in accordance with the sense of its notation and with regard to the underlying assumptions (usually not reflected in notation). In particular, weaker constraints should not be treated as "worse" than the stronger ones but, for the most part, as speaking in favor of variability.

9.3 Dependence on the employed theory and assumptions

Many factors involved in determination of the bounds put on G variation are theory or assumption dependent. For example, stringent constraints derived from BBN (Accetta et al. [1], Rothman & Matzner [68]) are valid only for Brans-Dicke theory; likewise, the constraint derived by Guenthner et al. [32] bases on the Brans-Dicke type theory with varying G. Most of constraints, even when not visibly shown in their notation, base on observations testing Dirac's LNH, i.e. are focused on the possible decrease of G. This in particular concerns the results derived from geophysical and paleontological data: impact of the Earth surface temperature on ancient organisms, expansion of Earth and the relevant difference in paleolatitudes between two sites of known separation (allowing to deduce the paleoradius), spin-down of the Earth due to its expansion, recession of the Moon and its impact on tides reflected in fossils. The respective data depend on too many conditions to repose excessive trust in their precision, and thus to consider them as fully reliable assumptions. In his extensive review study, Uzan [74] pays attention on these other sources of uncertainty connected with particular constraints.

9.4 Variation of Newton's constant and the age of universe

Assuming that increase of G extends the age of universe, the rate of G variation would be smaller than the here quoted value $7.25 \times 10^{-11}\mathrm{yr}^{-1}$ thus better fitting observations. However, according to the Friedmann equation

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\,\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3}\,, \qquad (49)$$

variation of G has a negligible impact on the age of universe. For $k = 0$ (flat universe) and $\Lambda = 0$, density becomes critical ($\rho_c = 3H^2(8\pi G)^{-1}$, and thus Friedmann equation reduces itself to identity $H^2 \equiv H^2$ becoming insensible to the change of G. In such a case, the age of universe simply equals the inverse of Hubble's constant ($t = H^{-1}$). However, for $\Lambda \neq 0$, currently estimated for $\Lambda_{(const)} \approx 10^{-52}\,\mathrm{m}$, dark energy (in a form of cosmological constant) predominates from a certain moment, so that t and H^{-1} more and more diverge. In an accelerating universe driven by dark energy, the rate of increase of G determined by $G \propto R_u$ also accelerates, which means that its declining in the unit time gradually slows down. Hence, in the far future, $G \propto R_u$ will translate to $G \propto H^{-1}$ rather than to $G \propto t$.

9.5 Equivalence of gravity and inertia

As is known, there are (currently) four notions of mass: 1) active gravitational mass — measure of ability to create gravi-

tational field or curvature, 2) passive gravitational mass — measure of "sensing" the gravitational field by a body (in the Newtonian depiction, respectively: measure of the force exerted by a body and measure of the force experienced by a body), 3) inertial mass — measure of resistance against the force accelerating a body (including the force of gravity), and 4) mass as a measure of energy according to $E = mc^2$. Numerous experiments performed over a long time up to present days have shown with increasing precision that inertial mass and passive gravitational mass are proportional to each other: $m_{(inert)} \sim m_{(pass)}$ (week equivalence principle). In turn, since active and passive gravitational masses are interchangeable according to the Newton's third law, so also active gravitational mass and inertial mass are proportional to each other:

$$m_{(act)} \sim m_{(inert)} \,. \tag{50}$$

The active gravitational mass is proportional to the Newton's constant: $m_{(act)} \sim G$. In fact $m_{(act)}$ is inseparable from G, which means that any change in the active mass should be interpreted as the change in the Newton's constant. Consequently, inertial mass is thought to follow the putative variation of G:

$$\Delta G \to \Delta m_{(inert)} \,. \tag{51}$$

This would make the so-called "inertial reaction force" always (i.e. also in the time-slice experiments) equivalent to the gravitational force. At the same time, variation of the Newton's constant would not affect the mass interpreted as the source of positive energy. Accordingly, the tests on G variation derived from celestial mechanics (e.g. LLR) would be basically ineffective, while the other ones (e.g. based on stellar physics) would still remain valid.

10 Quantitative predictions of the varying Chandrasekhar limit hypothesis, based on $G \propto R_u$, $c = const$

We shall now consider the VCM hypothesis in the form related to the BHC specified as $G \propto R_u$ i.e. $G \propto t$. On the assumption that the rate of Hubble expansion is approximately uniform, the Chandrasekhar limit depends on cosmic time as $M_{Ch} \propto t^{-3/2}$. This determines characteristic "delay time" for a single white dwarf, defined as the time needed to reach the WD's mass by the decreasing M_{Ch}. It makes thereby a basis for the quantitative predictions of VCM as to the rate of supernovae events, interpreted as a function of cosmic time. While, in general, the anticipated by VCM ability of a single WD to become the supernova meets the problem of the paucity of SNe Ia progenitors, the detailed predictions obviously demand more circumstantial investigation. One has to regard: 1) the number of single WDs within a given area (in particular, the number of their representative sample); 2) the mean/median mass of this sample; 3) the respective "delay time" for the median mass, determined by $M_{Ch} \propto t^{-3/2}$. Besides, in predicting the rate of distant SNe Ia one should also regard the related to distance intrinsic time of the observed events, and a corresponding value of Chandrasekhar limit. Once a distance is well defined, the respective limit should be treated as constant, considering the negligible (compared with the assumed rate of change in M_{Ch}) time devoted to observation. Instead, for the nearby SNe Ia one may fairly assume $M_{Ch} \approx 1.4 M_\odot$.

Let us apply the above to our Galaxy. For the sake of simplicity (an also taking into account the uncertainty in all data), we shall not regard the contribution of SNe Ia originated in binaries. We aim to estimate the present rate of SNe Ia, deriving it from accessible data, according to the above quoted three points. As is known, the Galaxy contains roughly 100-400 billion stars, above 97% of them supposed to end as white dwarfs, which however includes both actual WDs and the potential ones. According to the estimations based on SPY project, the space density of WDs within the radius of 20 pc is $(4.8 \pm 0.5) \times 10^{-3}\,\mathrm{pc}^{-3}$ while the corresponding mass density amounts to $(3.2 \pm 0.3) \times 10^{-3} M_\odot\,\mathrm{pc}^{-3}$, which gives the overall mean mass $(M)_{WD} \approx 0.665 M_\odot$ (Holberg et al. [35]). Instead Kepler et al. [39], basing on catalog elaborated by Eisenstein et al. [24] from the SDSS Data Release 4, found significant difference in the WD's mean mass between DA and DB stars (hydrogen and helium layers, respectively); namely $(M)_{DA} \approx 0.593 M_\odot$ and $(M)_{DB} \approx 0.711 M_\odot$. Considering the number of DA and DB in the sample (7167 and 507, respectively), one gets the $(M)_{WD} \approx 0.6 M_\odot$. We shall us this value in the further calculations.

In order to estimate the total number of white dwarfs in the Milky Way, we have to multiply the WD's space density by the Galaxy volume. Certainly, such an extrapolation is encumbered by significant uncertainty, as it is doubtful whether the sample obtained from the relatively close neighborhood (thin disc, in general) is typical for the whole Galaxy including thick disc, halo and the galactic bulge. Different parts of Galaxy vary in age, so WD's population is likely inhomogeneous in age and density. Evaluating the radius for 15,000 pc and the mean thickness for 5,000 pc and multiplying this by WDs' local density, one obtains: $(3.5 \times 10^{12}\,\mathrm{pc}^3) \times (5 \times 10^{-3}) \approx 1.7 \times 10^{10}$. This gives an insight into the actual number of WDs, consistent with a list brought by the Research Consortium on Nearby Stars (RECONS). According to the latter, 8 of the nearest 100 stars are the white dwarfs, which, provided this to be the representative ratio, gives the total number between 0.8×10^{10} to 3.2×10^{10} , dependently on the assumed total number of stars (100-400 billion).

The next step is to derive the "mean delay time" $(T)_{del}$ for the WD's mean mass $(M)_{WD}$. The respective algorithm reads

$$(T)_{del} = \left(\frac{t}{T_u} + 1 \right)^{2/3} \times T_u - T_u \tag{52}$$

T_u-age of universe, t – an auxiliary delay time not regarding

the power index, yielding

$$t = \frac{T_u}{(M_{Ch}/\Delta M_{WD}) - 1}\,, \tag{53}$$

where $\Delta M_{WD} = M_{Ch} - (M)_{WD}$. After conversion, one has

$$(T)_{del} = \left(\frac{M_{Ch}}{(M)_{WD}}\right)^{2/3} \times T_u - T_u\,. \tag{54}$$

Inserting $M_{Ch} = 1.4M_\odot$, $(M)_{WD} = 0.6M_\odot$ and $T_u = 13.8\,\mathrm{Gyr}$, one obtains $(T)_{del} \approx 10\,\mathrm{Gyr}$. Dividing the number of white dwarfs in Galaxy by that time gives the rate of roughly 1-3 events per year, a frequency exceeding the observed rate by a factor $> 10^2$. However, this prediction does not concern the present rate but a hypothetic rate averaged over the above calculated $(T)_{del}$. One should not identify (or confuse) "averaged" with "uniform" mainly because WD's masses subject, in general, to the Gaussian distribution:

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\sigma)^2/2\sigma^2} \tag{55}$$

(σ-standard deviation, μ-mean of the distribution) and the respective probability function:

$$Prob[a \leqslant x \leqslant b] = \int_a^b f(x)\,dx\,. \tag{56}$$

The observed standard deviation is significantly smaller than one ($\sigma^2 \ll 1$) yielding substantial peak around the median mass $0.6M_\odot$. Obviously, only WDs of the mass close to $1.4M_\odot$, corresponding with the relatively short mean delay time, contribute to the present rate of SNe Ia. We assume that any single white dwarf of the mass close to M_{Ch} is, dependently on specific conditions (rotation, chemical composition), a potential SN Ia at any moment during the slated delay time. Admittedly, the most massive known WD only slightly exceeds $1.3M_\odot$; this however should be associated with the fact that less than one-millionth of the whole population of WDs in Galaxy are identified so far. A similar difficulty concerns specifying the expression "close to $1.4M_\odot$". Bearing in mind an inevitable uncertainty, let us determine the respective range for $[b-a] \approx 0.1M_\odot$, assuming that, dependently on detailed conditions, any WD of the mass between $1.3 - 1.4M_\odot$ may become the SNe-Ia. For that mass range, the unit normal distribution yields less than 0.1% of the entire population, say, $\approx 10^7$. The mean mass of this "representative sample" is $1.35M_\odot$. It follows:

$$T_{del} \approx (1.4/1.35)^{2/3} \times 13.8 - 13.8 \approx 0.34\,(\mathrm{Gyr})\,. \tag{57}$$

The respective rate is then

$$\frac{10^7}{3 \times 10^8} = 3 \times 10^{-2}\,(\mathrm{yr}^{-1}). \tag{58}$$

This still slightly exceeds the observed rate, provided the latter is $\leqslant 1$ events per 100 years. However, considering the mentioned above reservations, it would not be reasonable to attach excessive importance to this or that particular number. The real number of single WDs from the representative sample may prove to be much smaller than 10^7 . The mass-range of potential progenitors may appear slightly narrower or wider. In general, more accurate data may support or falsify our hypothesis.

11 Conclusion

We have considered the SNe Ia progenitor problem in the context of general problem of the constancy of fundamental constants. Basing on arguments derived from the black-hole cosmology, we have singled out the Newton's constant as the most probable candidate for "inconstant constant". Since the increase of G involves the decrease in the value of Chandrasekhar limit M_{Ch}, both questions meet together yielding a hypothesis according to which a single white dwarf can alone become the progenitor of SN Ia.

Admittedly, the ongoing progress in observational techniques together with an improvement in stellar physics may bring solution to the progenitor problem dispensed with violating the constancy of Chandrasekhar limit. A tacit heuristic strategy connected with searching for the SNe Ia progenitors consists in attempts of making the SD and DD models flexible enough to eliminate the observed discrepancies. For the time being however the problem still exists, which makes solutions going beyond the binary paradigm justifiable and noteworthy.

The unbiased estimations seem to support the main thesis of this article, i.e. that M_{Ch} decreasing according to $G \propto R_u$ may explain the paucity of SNe Ia progenitors. It is to be noted that, predicted by $G \propto R_u$ immense growth of the Newton's constant from the initial to present value ($G/G_0 = \delta^2 \approx 10^{124}$) almost completely applies to the very early and early universe, preceding structure formation. Since the oldest SNe Ia detected so far: SN UDS1 0Wil (Wilson) and SN 1997ff reach about 11 Gyr the part of increase of the Newton's constant shaping the Chandrasekhar limit does not exceed the one order of magnitude, being much smaller in the case of overwhelming majority of the observed events.

Submitted on June 11, 2014 / Accepted on October 22, 2015

References

1. Accetta F.S., Krauss L.M., & Romanelli P. *Phys. Lett. B*, 1990, v.248, 146.
2. Albrecht J. & Magueijo J. *Phys. Rev. D*, 1999, v.59, 043516.
3. Badenes C. et al. *ApJ*, 2009, v.707, 971.
4. Badenes C. & Maoz D. *ApJ*, 2012, v.749, L11.
5. Barrow J.D. The Constants of Nature; from Alpha to Omega. Pantheon Books, New York, 2002.
6. Bielewicz P., & Banday A.J. *MNRAS*, 2011, 412(3), 2104.
7. Biesiada M., & Malec B. *MNRAS*, 2004, v.520(2), 644.

8. Blake G.M. *MNRAS*, 1977, v.181, 41.
9. Brans C.H., Dicke R.H. *Phys. Rev.*, 1961, v.123, issue 3, 925.
10. Chand H., et al. *A&A*, 2004, v.417, 853.
11. Chandrasekhar S. Nobel Prize lecture, 1983.
12. Cornish N.J. et al. *Phys. Rev. Lett.*, 2003, v.92, 201302.
13. Dicke R. *Rev. Mod. Phys.*, 1957, v.29, 363.
14. Dickey J.O. et al. *Science*, 1994, 265, 482.
15. Dirac P.A.M. *Nature*, 1937, 139, 323.
16. Dirac P.A.M. Lecture on the International Symposium on Contemporary Physics, Trieste, June 1968.
17. Di Stefano R. *ApJ*, 2010a, v.712, 728.
18. Di Stefano R. *ApJ*, 2010b, v.719, 474.
19. Di Stefano R., Voss R., Claeys J.S.W. *ApJL*, 2011, v.738(1).
20. Duff M.J. 2004, arXiv: hep-th/020809v3.
21. Eichendorf W. & Reinhardt M. *Mitt. Astron. Ges.*, 1977, v.42, 89.
22. Einstein A. *Annalen der Physik*, 1911, v.35, 898.
23. Einstein A. *Annalen der Physik*, 1912, v.38, 355.
24. Eisenstein D.J. et al. *ApJS*, 2006, v.167, 40.
25. Foley R.J. et al. *AJ*, 2012, v.143(5), 113.
26. Gallagher J.S. et al. *ApJ*, 2008, v.685(2), 752.
27. Garcia-Berro E. et al. *MNRAS*, 1995, v.277, 801.
28. Gilfanov M. Press conference, Chandra Press Room, NASA, Feb. 17, 2010.
29. Gilfanov M. & Bogdan A. *Nature*, 2010, v.463, 924.
30. Gonzalez Hernandez J.I. et al. *Nature*, 2012, v.489, 533.
31. Good I.J. *Physics Today*, 1972, v.25(7), 15.
32. Guenther D.B., Krauss L.M., & Demarque P. *ApJ*, 1998, v.498, 871.
33. Hachisu I., Kato M., Namoto K. *ApJ*, 2008, v.683, L127.
34. Hachisu I. et al. *ApJ*, 2012, v.744, 69.
35. Holberg J.B. et al. *AJ*, 2008, v.135, 1225.
36. Hoyle F. & Narlikar J.V. *Proceedings of the Royal Society of London A*, 1964, v.282, issue 1389, 19.
37. Iben I. Jr. & Tutukov A.V. *ApJS*, 1984, v.54, 335.
38. Kaspi V.M., Taylor J.H., & Riba M.F. *ApJ*, 1998, v.498, 871.
39. Kepler S.O. et al. *MNRAS*, 2007, v.375(4), 1315.
40. Kerkwijk Van M.H. et al. *ApJ*, 2010, v.722, L157.
41. Lamoreaux S.K. & Torgerson J.R. *Phys. Rev. D*, 2004, v.69, 121701.
42. Li W. et al. *Nature*, 2011, v.480, 348.
43. Linden S., Virey J.-M., Tilquin A. *A&A*, 2009, v.506(3), 1095.
44. Longo, M.J. 2011, Phys. Lett. B, 699 (4), 224
45. Magueijo, J., & Smolin, L. 2002, Phys. Rev. Lett. 88, 190403
46. Mansfield, V.N. 1976, Nature, 261, 560
47. Maoz D. & Mannucci F. *Publications of the Astron. Soc. of Australia*, 2012, v.29(4), 44.
48. Maoz D., Mannucci F., & Brandt T.D. *MNRAS*, 2012, v.426, 3282.
49. McElhiny N.W., Taylor S.R., & Stevenson D.J. *Nature*, 1978, v.271, 316.
50. Mennekens N. et al. *A&A*, 2010, v.515, A89.
51. Moffat J. *Int. J. Mod. Phys. D2*, 1993, 351.
52. Mould J. & Uddin S.A. To appear in PASA, 2014, arXiv: 1402.1534.
53. Mullally F. et al. *ApJL*, 2009, v.707, L51.
54. Napiwotzki R. et al. *Messenger*, 2003, v.112, 25.
55. Napiwotzki R. et al. *ESO Astrophysics Symposia*, 2003, v.134.
56. Nugent P. et al. *Nature*, 2011, v.480, 344.
57. Okun L.B. *Usp. Fiz. Nauk*, 1991, v.161, 177 (*Sov. Phys. Usp.*, 1991, v.34, 818).
58. Pathria R.K. *Nature*, 1972, v.240, 298.
59. Peik E. et al. *Phys. Rev. Lett.*, 2004, v.93(17), 170801.
60. Perlmutter S. et al. *ApJ*, 1999, v.517(2), 565.
61. Petit J.P. *Mod. Phys. Lett.*, 1988, v.A3, issue 16, 1527.
62. Petit J.P. *Mod. Phys. Lett.*, 1988, v.A3, issue 18, 1733.
63. Petrov Y.V. et al. *Phys. Rev. C*, 2006, v.74(6), 064610.
64. Popławski N.J. *Physics Letters B*, 2010, v.687(2), 110.
65. Popławski N.J. *Physics Letters B*, 2010, v.694(3), 181.
66. Riess A.G. et al. *AJ*, 1998, v.116(3), 1009.
67. Rosenband T. et al. *Science*, 2008, v.319, issue 5871, 1808.
68. Rothman T. & Matzner R. *ApJ*, 1982, v.257, 450.
69. Rybicki M. *Apeiron*, 2008, v.15(4), 465.
70. Schaefer B.E. & Pagnotta A. *Nature*, 2012, v.481, 7380.
71. Shapiro I.I. In *General Relativity and Gravitation*, ed. by N. Ashby, D.F. Bartlett and W. Wyss, Cambridge University Press, 1990.
72. Sullivan M. et al. *MNRAS*, 2010, v.406(2), 782.
73. Totani T. et al. *PASJ*, 2008, v.60, 1327.
74. Uzan J.-P. *Rev. Mod. Phys.*, 2003, v.75, 403.
75. Van Flandern T.C. In: *Precision Measurements and Fundamental Constants II*, *NBS circular*, 1981, v.617, 625.
76. Veneziano G. *Europhysics Lett.*, 1986, issue 2, 199.
77. Webb J.K. et al. *Phys. Rev. Lett.*, 1999, v.82(5), 884.
78. Webb J.K. et al. *Phys. Rev. Lett.*, 2001, v.87(9), 091301.
79. Webbink R.F. *ApJ*, 1984, v.277, 355.
80. Whelan J. & Iben I. *ApJ*, 1973, v.186, 1007.
81. Williams P.J., Newhall X.X., & Dickey J.O. *Phys. Rev. D*, 1996, v.53, 6730.
82. Zhu C. et al. *ApJ*, 2013, v.767, 164.

Schrödinger Equation for a Half Spin Electron in a Time Dependent Magnetic Field

Hafeez Y. Hafeez[1], E. N. Chifu[2], and Ibrahim M. Musa[3]

Physics Department, Federal University Dutse, P.M.B 7156, Jigawa State, Nigeria.
[1]E-mail: hafeezyusufhafeez@gmail.com
[2]E-mail: ebenechifu@yahoo.com
[3]E-mail: baffapet@yahoo.com

The behaviour of an electron with mass m_e and half spin when passing through a magnetic field with fixed strength B_0 is studied. The motion of the particle is restricted to a ring with radius R, thus assuming periodic boundary conditions. We also focused on magnetic field evolving adiabatically in time, the magnetic field is expressed as a function of angle ϕ and θ i.e only the direction of the magnetic field vectors change while the strength B_0 is kept fixed. Expression for eigenenergies were drawn for a fixed energy and sample values of α, ω, θ and $x = mR^2/\hbar^2$.

1 Introduction

An intriguing example emerging from asymmetric spin-interactions are skyrmion lattices. In 1989 Alexey Bogdanov predicted that for anisotropic chiral magnets there is a new magnetic order consisting of topologically stable spin whirls, named skyrmions after the English particle physicist Tony Skyrme, who showed that localized solutions to non-linear quantum field theories may be interpreted as elementary particles. Briefly speaking, skyrmions are topologically stable whirls in fields.

In 2009, a new magnetic order was observed in Manganese Silicide (MnSi) for specific temperatures and magnetic fields by Mühlbauer et al [1]. The physics of an electron moving through the magnetic field can be analyzed from two different points of view:

From the point of view of the electron, i.e. considering the problem in terms of emergent electric and magnetic fields, the change in spin orientation is equal to an effective Lorentz force acting on the electron, which is perpendicular to its motion [2]. As a result, the magnetic field induces a deflection of the electron, which can be measured by making use of the topological Hall-effect [3]. Because of the electron carrying an electric charge, a potential may be measured perpendicular to the direction of the current. Since the magnetic structure of the skyrmion lattice is very smooth, the adjustment of the spin of the electron to the magnetization of the skyrmion lattice can be considered an adiabatic process.

On the other hand, there must be a corresponding counterforce acting on the skyrmion. This force, arising from the transfer of angular momentum from the conduction electrons to the local magnetic structure (cf. [4]), can for example result in a drift of the domains of the lattice. A 1-D model of an electron passing over a static magnetic field has previously been investigated in the Bachelor's thesis of M. Baedorf [5]. Berry phase physics and spin-scattering in time-dependent magnetic fields has been studied by Sarah Maria Schroeter [6].

In this work, the behaviour of an electron with mass m_e, when passing through a magnetic field with a fixed strength B_0 is studied.

2 Formulation of the problem

The behaviour of a half spin particle, more specifically an electron, when passing through a magnetic field with a fixed strength B_0 is considered. The parameter ϕ sets the position where the particular magnetic field is measured. At every position ϕ on border of the circle, we attach an imaginary 3D-sphere which determines the direction of the field vector. In effect, the magnetic field is constituted by mere spherical coordinates. In addition, we allow variation of both angle ϕ and θ in time with frequency of ω_1 and ω_2 respectively:

$$B(r,t) = B_0\,\hat{n}(\phi,\theta,t) \tag{1}$$

$$B(r,t) = B_0 \begin{bmatrix} \sin(\theta-\omega_2 t)\cos(\phi-\omega_1 t) \\ \sin(\theta-\omega_2 t)\sin(\phi-\omega_1 t) \\ \cos(\theta-\omega_2 t) \end{bmatrix} \tag{2}$$

$$B(r,t) = B_0 \begin{bmatrix} \sin(\tilde{\theta})\cos(\tilde{\phi}) \\ \sin(\tilde{\theta})\sin(\tilde{\phi}) \\ \cos(\tilde{\theta}) \end{bmatrix} \tag{3}$$

where $\tilde{\phi} = \phi - \omega_1 t$ and $\tilde{\theta} = \theta - \omega_2 t$. The Hamiltonian is made up of a kinetic part and a part arising from the interaction of particle with the magnetic field:

$$H_0(r,t) = \frac{\hat{p}^2}{2m_e} + B(r,t)\frac{g_s|\mu_B|}{h}\,S \tag{4}$$

where S is the electron spin, g_s is the spin g-factor and μ_B is the Bohr magneton

$$|\mu_B| = \frac{|e|\hbar}{2m_e}\,.$$

We confine ourselves to the xy-plane, with the real space parameter $\theta = \pi/2$ and radius R kept fixed. The nabla-operator

is simplified as:

$$\nabla = \hat{e}_r \frac{\partial}{\partial r} + \hat{e}_{\tilde{\theta}} \frac{1}{R} \frac{\partial}{\partial \tilde{\theta}} + \hat{e}_{\tilde{\phi}} \frac{1}{R \sin\tilde{\theta}} \frac{\partial}{\partial \tilde{\phi}} \tag{5}$$

which becomes

$$\nabla^2 = \left(\frac{1}{R} \frac{\partial}{\partial \tilde{\phi}} \right)^2 . \tag{6}$$

Thus, we can now rewrite the Hamiltonian H_0 as

$$H_0 = -\frac{\hbar^2}{2mR^2} \left(\frac{\partial}{\partial \tilde{\phi}} \right)^2 + |\mu_B| \, B_0(r,t) \, \sigma \tag{7}$$

where σ is a vector of Pauli matrices and for any unit vector $\hat{n}$, we find a rotation matrix $\mathfrak{R}$ such that $\mathfrak{R}\hat{\tilde{\phi}} = \hat{n}$ so that (7) can be rewritten as

$$H_0 = \frac{\hbar^2}{mR^2} \left(-\frac{1}{2} \left(\frac{\partial}{\partial \tilde{\phi}} \right)^2 + \frac{|\mu_B| B_0}{\hbar^2/mR^2} \, \hat{n} \, \sigma \right) \tag{8}$$

$$H_0 = \frac{\hbar^2}{mR^2} \left(-\frac{1}{2} \left(\frac{\partial}{\partial \tilde{\phi}} \right)^2 + \alpha \hat{n} \sigma \right) = \frac{\hbar^2}{mR^2} \, \tilde{H}_0 \tag{9}$$

where

$$\alpha = \frac{|\mu_B| B_0}{\hbar^2/mR^2} \, ; S = \frac{\hbar}{2} \, \sigma \, ; g_s = 2.$$

Combining the operators generating the translation and rotation gives

$$g = -i\hbar \frac{\partial}{\partial S} \, 1 + \frac{\hbar}{2R} \, \sigma_z = -\frac{i\hbar}{R} \frac{\partial}{\partial \tilde{\phi}} \, 1 + \frac{\hbar}{2R} \, \sigma_z \tag{10}$$

$$\tilde{g} = -i \frac{\partial}{\partial \tilde{\phi}} 1 + \frac{\sigma_z}{2} \tag{11}$$

where $\tilde{g}$ is a rescaled version of g. By careful construction of g, $\tilde{H}_0$ and $\tilde{g}$ commute, consequently H_0 and g indeed commute.

$$[\tilde{H}_0, \tilde{g}] = \left[-\frac{1}{2} \left(\frac{\partial}{\partial \tilde{\phi}} \right)^2 + \alpha \hat{n} \sigma, -i \frac{\partial}{\partial \tilde{\phi}} 1 + \frac{\sigma_z}{2} \right] \tag{12}$$

$$\begin{aligned} [\tilde{H}_0, \tilde{g}] = &\left[-\frac{1}{2} \left(\frac{\partial}{\partial \tilde{\phi}} \right)^2 , -i \frac{\partial}{\partial \tilde{\phi}} \right] + \left[\alpha \hat{n} \sigma, -i \frac{\partial}{\partial \tilde{\phi}} \right] + \\ &+ \left[-\frac{1}{2} \left(\frac{\partial}{\partial \tilde{\phi}} \right)^2 , \frac{\sigma_z}{2} \right] + \left[\alpha \hat{n} \sigma, \frac{\sigma_z}{2} \right] \end{aligned} \tag{13}$$

$$\begin{aligned} [\tilde{H}_0, \tilde{g}] = &i\alpha \left(\frac{\partial}{\partial \tilde{\phi}}, \hat{n} \sigma \right) + \\ &+ \frac{\hat{n}\sigma}{2} \left([\sigma_x, \sigma_z], [\sigma_y, \sigma_z], [\sigma_z, \sigma_z] \right) \end{aligned} \tag{14}$$

with

$$[\sigma_i, \sigma_j] = 2i \in_{ijk} \sigma_k .$$

$$\begin{aligned} [\tilde{H}_0, \tilde{g}] = i\alpha \Big[&\frac{\partial}{\partial \tilde{\phi}} (\sin\tilde{\theta} \cos\tilde{\phi} \sigma_x + \sin\tilde{\theta} \sin\tilde{\phi} \sigma_y + \\ &+ \cos\tilde{\theta} \sigma_z) \Big] + i\alpha \hat{n} (-\sigma_y, \sigma_x, 0) \end{aligned} \tag{15}$$

$$\begin{aligned} [\tilde{H}_0, \tilde{g}] = &i\alpha \left(-\sin\tilde{\theta} \sin\tilde{\phi} \sigma_x + \sin\tilde{\theta} \cos\tilde{\phi} \sigma_y \right) + \\ &+ i\alpha \hat{n} (-\sigma_y, \sigma_x, 0) \end{aligned} \tag{16}$$

$$[\tilde{H}_0, \tilde{g}] = 0. \tag{17}$$

We have shown that $\tilde{H}_0$ and $\tilde{g}$ possess the same system of eigenfunctions, with that, we regard $\tilde{g}$ as a generalized momentum operator.

2.1 Solution to momentum operator

We now establish the eigenfunctions of $\tilde{g}$ solving the eigensystem

$$\left(-i \frac{\partial}{\partial \tilde{\phi}} 1 + \frac{\sigma_z}{2} \right) |\psi\rangle = K |\psi\rangle \tag{18}$$

$$-i \frac{\partial}{\partial \tilde{\phi}} 1 \, |\psi\rangle = \left(K - \frac{\sigma_z}{2} \right) |\psi\rangle = \begin{pmatrix} \left(K - \frac{1}{2} \right) & 0 \\ 0 & \left(K + \frac{1}{2} \right) \end{pmatrix} |\psi\rangle \tag{19}$$

with eigenvalues

$$\lambda_{\frac{1}{2}} = \left(K \mp \frac{1}{2} \right) \tag{20}$$

and the respective eigenfunctions

$$|\psi_1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{i\left(K - \frac{1}{2}\right)\phi} = \begin{pmatrix} \psi_1 \\ 0 \end{pmatrix} \tag{21}$$

$$|\psi_2\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{i\left(K + \frac{1}{2}\right)\phi} = \begin{pmatrix} 0 \\ \psi_2 \end{pmatrix}. \tag{22}$$

As we study the motion of a particle on a ring, we require $|\psi(\phi)\rangle$ to fill periodic boundary condition

$$|\psi(\phi)\rangle = |\psi(\phi + 2\pi)\rangle e^{i\left(K \mp \frac{1}{2}\right)2\pi} = 1K = n + \frac{1}{2}; n \in Z \tag{23}$$

This means that the momentum is quantized. The general solution to (18) is linear combination of both eigenfunctions

$$|\psi(\phi)\rangle = C_1(t)|\psi_1(\phi)\rangle + C_2(t)|\psi_2(\phi)\rangle = \begin{pmatrix} C_1(t)|\psi_1(\phi)\rangle \\ C_2(t)|\psi_2(\phi)\rangle \end{pmatrix} \tag{24}$$

where $C_1(t)$ and $C_2(t)$ do not depend on ϕ.

2.2 Solution to the time-dependent Hamiltonian

Ultimately, we are interested in computing the time-dependent coefficients $C_1(t)$ and $C_2(t)$ in order to receive full solution of the Schrödinger equation when solving the time-dependent Schrödinger equation, we employ the solution to

the momentum operator in order to simplify the eigensystem associated with $\tilde{g}$ as follows:

$$i\hbar\partial_t|\psi\rangle = H_0|\psi\rangle = \frac{\hbar^2}{mR^2}\left(-\frac{1}{2}\left(\frac{\partial}{\partial\tilde{\phi}}\right)^2 + \alpha\hat{n}\sigma\right)|\psi\rangle. \quad (25)$$

See the last page for intermediate equations (26) and (27)

$$i\hbar\partial_t|\psi\rangle = H_{0,K,\tilde{\phi}}(t)|\psi\rangle \quad (28)$$

where $H_{0,K,\tilde{\phi}}(t)$ is defined by equation (27).

2.2.1 Setting up the Schrödinger equation for the time-dependent coefficients

To set up the Schrödinger equation for the time-dependent coefficients $C_1(t)$ and $C_2(t)$ is by transforming the Schrödinger equation for $|\psi\rangle$:

$$\begin{aligned} i\hbar\partial_t|\psi\rangle\begin{pmatrix} C_1(t)\psi_1 \\ C_2(t)\psi_2 \end{pmatrix} &= H_0(t)\begin{pmatrix} C_1(t)\psi_1 \\ C_2(t)\psi_2 \end{pmatrix} \\ i\hbar\partial_t|\psi\rangle\begin{pmatrix} C_1(t) \\ C_2(t)\frac{\psi_2}{\psi_1} \end{pmatrix} &= H_0(t)\begin{pmatrix} C_1(t) \\ C_2(t)\frac{\psi_2}{\psi_1} \end{pmatrix}. \end{aligned} \quad (29)$$

Employing the equation (27) computed solution to the momentum operator, we know that $\frac{\psi_2}{\psi_1} = e^{i\phi}$ and may write (see the last page for intermediate equations (30) and (31)):

$$i\hbar\partial_t\begin{pmatrix} C_1(t) \\ C_2(t) \end{pmatrix} = H_{0,K,\omega}\begin{pmatrix} C_1(t) \\ C_2(t) \end{pmatrix} \quad (32)$$

where $H_{0,K,\omega}$ is defined by equation (31).

2.3 Moving into a rotating coordinate system

To solve the eigensystem, we transform $H_{0,K,\omega}(t)$ by changing into a coordinate system rotating clockwise with a frequency $\omega = \omega_1$

$$\begin{pmatrix} \tilde{C_1}(t) \\ \tilde{C_2}(t) \end{pmatrix} = e^{-\frac{i}{\hbar}S_z\omega t}\begin{pmatrix} C_1(t) \\ C_2(t) \end{pmatrix} = e^{-\frac{i}{2}\sigma_z\omega t}\begin{pmatrix} C_1(t) \\ C_2(t) \end{pmatrix}. \quad (33)$$

In another way, (33) becomes

$$\begin{pmatrix} C_1(t) \\ C_2(t) \end{pmatrix} = e^{\frac{i}{2}\sigma_z\omega t}\begin{pmatrix} \tilde{C_1}(t) \\ \tilde{C_2}(t) \end{pmatrix} \quad (34)$$

where

$$\begin{aligned} e^{\frac{i}{2}\sigma_z\omega t} &= \Sigma_n\frac{\left(\frac{i}{2}\sigma_z\omega t\right)^n}{n!} = \Sigma_n\frac{\left(\frac{i}{2}\omega t\right)^n}{n!}\begin{pmatrix} 1^n & 0 \\ 0 & (-1)^n \end{pmatrix} \\ &= \begin{pmatrix} e^{\frac{i}{2}\omega t} & 0 \\ 0 & e^{-\frac{i}{2}\omega t} \end{pmatrix}. \end{aligned}$$

Substituting of (32) in (34) gives:

$$\begin{aligned} &i\hbar\partial_t\begin{pmatrix} e^{\frac{i}{2}\omega t} & 0 \\ 0 & e^{-\frac{i}{2}\omega t} \end{pmatrix}\begin{pmatrix} \tilde{C_1}(t) \\ \tilde{C_2}(t) \end{pmatrix} \\ &= H_{0,K,\omega}\begin{pmatrix} e^{\frac{i}{2}\omega t} & 0 \\ 0 & e^{-\frac{i}{2}\omega t} \end{pmatrix}\begin{pmatrix} \tilde{C_1}(t) \\ \tilde{C_2}(t) \end{pmatrix}. \end{aligned} \quad (35)$$

Multiplying L.H.S of (35) by $e^{\frac{i}{2}\sigma_z\omega t}$, we obtain

$$\begin{pmatrix} e^{-\frac{i}{2}\omega t} & 0 \\ 0 & e^{\frac{i}{2}\omega t} \end{pmatrix}\begin{pmatrix} e^{\frac{i}{2}\omega t}\left(\frac{-\hbar\omega}{2} + i\hbar\partial_t\right) & 0 \\ 0 & e^{-\frac{i}{2}\omega t}\left(\frac{\hbar\omega}{2} + i\hbar\partial_t\right) \end{pmatrix}$$

$$\begin{pmatrix} \tilde{C_1}(t) \\ \tilde{C_2}(t) \end{pmatrix} = \begin{pmatrix} -\frac{\hbar\omega}{2} & 0 \\ 0 & -\frac{\hbar\omega}{2} \end{pmatrix}\begin{pmatrix} \tilde{C_1}(t) \\ \tilde{C_2}(t) \end{pmatrix} + i\hbar\partial_t\begin{pmatrix} \tilde{C_1}(t) \\ \tilde{C_2}(t) \end{pmatrix}.$$

Also multiplying R.H.S of (35) by $e^{\frac{i}{2}\sigma_z\omega t}$ we have:

$$\begin{pmatrix} e^{-\frac{i}{2}\omega t} & 0 \\ 0 & e^{\frac{i}{2}\omega t} \end{pmatrix}H_K\begin{pmatrix} e^{\frac{i}{2}\omega t} & 0 \\ 0 & e^{-\frac{i}{2}\omega t} \end{pmatrix}\begin{pmatrix} \tilde{C_1}(t) \\ \tilde{C_2}(t) \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{2}\left(K - \frac{1}{2}\right)^2 + \alpha\cos\tilde{\theta} & \alpha\sin\tilde{\theta} \\ \alpha\sin\tilde{\theta} & \frac{1}{2}\left(K + \frac{1}{2}\right)^2 - \alpha\cos\tilde{\theta} \end{pmatrix}\begin{pmatrix} \tilde{C_1}(t) \\ \tilde{C_2}(t) \end{pmatrix}.$$

As a consequence, (35) yields (see the last page for intermediate equation (36))

$$i\hbar\partial_t\begin{pmatrix} \tilde{C_1}(t) \\ \tilde{C_2}(t) \end{pmatrix} = C\begin{pmatrix} \tilde{C_1}(t) \\ \tilde{C_2}(t) \end{pmatrix}. \quad (37)$$

Comparing (37) with the corresponding static Schrödinger equation for time-independent coefficients, one observes that C is the Hamiltonian one receives when considering static magnetic field (cf. [5]) combined with an additional matrix

$$\begin{pmatrix} \frac{\omega mR^2}{2\hbar} & 0 \\ 0 & \frac{-\omega mR^2}{2\hbar} \end{pmatrix}.$$

We now deal with time-independent $\tilde{\theta}$ and time-dependent $\tilde{\phi}$, so that $\tilde{\theta} = \theta = constant$. As eigenvalues of the operator C we get (see the last page for equation (38)), which correspond to the energies of the lower and upper band. E_- corresponds to a magnetic moment which is parallel to the magnetic field.

2.4 Determining the rotated time-dependent coefficients

To determine the solution to (36) i.e find a representation of the rotated time-dependent coefficients $\tilde{C(t)}_1$ and $\tilde{C(t)}_2$, an equation of the form

$$i\hbar\partial_t\begin{pmatrix} \tilde{C}_1(t) \\ \tilde{C}_2(t) \end{pmatrix} = C\begin{pmatrix} \tilde{C}_1(t) \\ \tilde{C}_2(t) \end{pmatrix}$$

can immediately be found to have the solution

$$\begin{pmatrix} \tilde{C}_{1+}(t) \\ \tilde{C}_{2+}(t) \end{pmatrix} = e^{-iE_+t}X_+ \quad (39)$$

$$\begin{pmatrix} \tilde{C}_{1-}(t) \\ \tilde{C}_{2-}(t) \end{pmatrix} = e^{-iE_-t}X_+ \quad (40)$$

where E_+, E_- and X_+, X_- are the eigenvalues and corresponding normalized eigenvectors of the matrix C respectively.

More precisely, the later are found to be given by equation (41) and normalization factor (42) given on the last page.

2.5 Establishing the solution to the initial Schrödinger equation

Combining (39) and (40) with already computed static parts of the wave function (21) and (22) as well as multiplying the respective components with the e factor which sets the wave function back into a non-rotating coordinate system (see (34)), we receive the exact solutions to the initial Schrödinger equation (27)

$$|\psi\rangle_{K,+} = e^{-iE_+t}\begin{pmatrix} x_{1,+}e^{i\left(K-\frac{1}{2}\right)\phi}e^{i\frac{\omega}{2}t} \\ x_{2,+}e^{i\left(K+\frac{1}{2}\right)\phi}e^{-i\frac{\omega}{2}t} \end{pmatrix}; \quad (43)$$

$$_{+,K}\langle\psi|\psi\rangle_{K,+} = 1$$

$$|\psi\rangle_{K,-} = e^{-iE_+t}\begin{pmatrix} x_{1,-}e^{i\left(K-\frac{1}{2}\right)\phi}e^{i\frac{\omega}{2}t} \\ x_{2,-}e^{i\left(K+\frac{1}{2}\right)\phi}e^{-i\frac{\omega}{2}t} \end{pmatrix}; \quad (44)$$

$$_{-,K}\langle\psi|\psi\rangle_{K,-} = 1$$

The solutions (43) and (44) specific to energies E_- and E_+ (and respective bands + and -) corresponding to the solution to one K, hence the indices.

3 Numerical solution to the eigenenergies

First, let us turn back to the exact eigenenergies we computed in section 2.2, equation (38). We consider an incoming wave function with a fixed energy $\in$ (given on the last page). For a fixed energy $\in_n = \in_o + n\,\omega$ there are maximal four real solutions for $K(n, \sigma, \delta)$, which correspond to the propagation directions $\delta = l, r$ and the two possible eigenenergies of the respective wave functions, i.e. the alignment of the spin $\sigma = +, -$ with respect to the magnetic field, (see Fig. 1).

4 Discussion

The Schrödinger equation for a half spin particle in a time dependent magnetic field is presented. Depending on the energy, there are up to four real solutions for K. The energy function $E_+(K)$ lies below the function $E_-(K)$ for all specific K, (see Fig. 1). For a fixed energy below the minimum of E_- there are no real solutions. For a fixed energy between both minima there are two real solutions which correspond to a spin aligned in the direction of the magnetic field and waves propagating towards the left or the right. For an energy above two minima there are four real solutions. In this case, both directions of propagation and both spin orientations occur.

5 Conclusion

In this paper, the exact wave function of a particle moving through a non-colinear time-dependent magnetic field is computed. Also, it is confirmed that the motion of a half spin of an electron through the chosen magnetic field is an adiabatic problem evolving with time. We found that for a time-dependence of the position of the electron, there are no emergent electric fields since the undisturbed Hamiltonian can be mapped onto a time- independent one by unitary transformations.

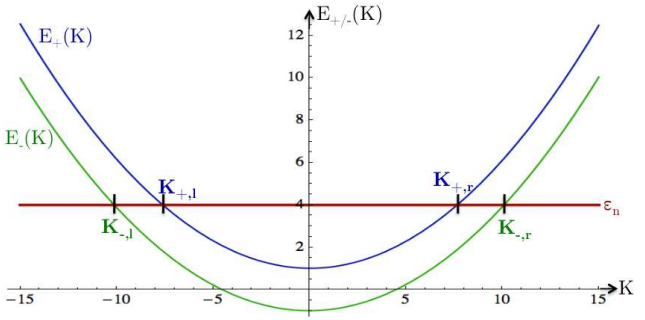

Fig. 1: Eigenenergies $E_\pm(K)$ plotted versus the momentum eigenvalue K for sample values of α, ω, θ and $x = mR^2/\hbar^2$. The points of intersection K_i with a fixed energy $\in$ determine the propagation direction and the spin alignment of the wave function. We set $\alpha = 10$, $\omega = 0.1$, $\theta = \pi$ and $x = mR^2/\hbar^2 = 10$.

Acknowledgements

The authors wish to thank Sarah Maria Schroeter for her time, useful discussions to this problem.

Submitted on October 10, 2015 / Accepted on October 17, 2015

References

1. Mühlbauer S., Binz B., Jonietz F., Pfleider C., Rosch A., Neubauer A., Georgii R., Boni P. Skyrmion Lattice in a Chiral Magnet. *Science*, 2009, v. 323, 915–919.
2. Pfleiderer C. and Rosch A. Single skyrmions spotted. *Nature*, 2010, v. 465 (7300), 880–881.
3. Neubauer A., Pfleiderer C., Binz B., Rosch A., Ritz R., Niklowitz P. G., and Boni P. Topological Hall Effect in the A Phase of MnSi. *Phys. Rev. Lett.*, 2009, v. 102, 186602.
4. Everschor K. Current-Induced Dynamics of Chiral Magnetic Structures. Ph.D. thesis, University of Cologne, 2012.
5. Baedorf M. Berry-Phase und Spinstreuung. Bachelor's thesis, University of Cologne, 2011.
6. Schroeter S. M. Berry Phase Physics and Spin-Scattering in Time-Dependent Magnetic Fields. Bachelor's thesis, University of Cologne, 2012.

$$i\hbar\partial_t|\psi\rangle = \frac{\hbar^2}{mR^2}\begin{pmatrix} -\frac{1}{2}\left(\frac{\partial}{\partial\tilde{\phi}}\right)^2 + \alpha\cos\tilde{\theta} & \alpha\sin\tilde{\theta}e^{-i\tilde{\phi}} \\ \alpha\sin\tilde{\theta}e^{i\tilde{\phi}} & -\frac{1}{2}\left(\frac{\partial}{\partial\tilde{\phi}}\right)^2 - \alpha\cos\tilde{\theta}\end{pmatrix}|\psi\rangle \tag{26}$$

$$i\hbar\partial_t|\psi\rangle = \frac{\hbar^2}{mR^2}\begin{pmatrix} \frac{1}{2}\left(K-\frac{1}{2}\right)^2 + \alpha\cos\tilde{\theta} & \alpha\sin\tilde{\theta}e^{-i\tilde{\phi}} \\ \alpha\sin\tilde{\theta}e^{i\tilde{\phi}} & \frac{1}{2}\left(K+\frac{1}{2}\right)^2 - \alpha\cos\tilde{\theta}\end{pmatrix}|\psi\rangle \tag{27}$$

$$i\hbar\partial_t\begin{pmatrix} C_1(t) \\ C_2(t)\frac{\psi_2}{\psi_1}\end{pmatrix} = \frac{\hbar^2}{mR^2}\begin{pmatrix} \left(\frac{1}{2}\left(K-\frac{1}{2}\right)^2 + \alpha\cos\tilde{\theta}\right)C_1(t) + \alpha\sin\tilde{\theta}e^{-i\tilde{\phi}}C_2(t)e^{i\phi} \\ \alpha\sin\tilde{\theta}e^{i\tilde{\phi}}C_1(t) + \left(\frac{1}{2}\left(K-\frac{1}{2}\right)^2 - \alpha\cos\tilde{\theta}\right)C_2(t)e^{i\phi}\end{pmatrix} \tag{30}$$

$$i\hbar\partial_t\begin{pmatrix} C_1(t) \\ C_2(t)\end{pmatrix} = \begin{pmatrix} \frac{1}{2}\left(K-\frac{1}{2}\right)^2 + \alpha\cos\tilde{\theta} & \alpha\sin\tilde{\theta}e^{i\omega_1 t} \\ \alpha\sin\tilde{\theta}e^{-i\omega_1 t} & \frac{1}{2}\left(K+\frac{1}{2}\right)^2 - \alpha\cos\tilde{\theta}\end{pmatrix}\begin{pmatrix} C_1(t) \\ C_2(t)\end{pmatrix} \tag{31}$$

$$i\hbar\partial_t\begin{pmatrix} \tilde{C}_1(t) \\ \tilde{C}_2(t)\end{pmatrix} = \frac{h^2}{mR^2}\begin{pmatrix} \frac{1}{2}\left(K-\frac{1}{2}\right)^2 + \alpha\cos\tilde{\theta} + \frac{\omega mR^2}{2\hbar} & \alpha\sin\tilde{\theta} \\ \alpha\sin\tilde{\theta} & \frac{1}{2}\left(K+\frac{1}{2}\right)^2 - \alpha\cos\tilde{\theta} - \frac{\omega mR^2}{2\hbar}\end{pmatrix}\begin{pmatrix} \tilde{C}_1(t) \\ \tilde{C}_2(t)\end{pmatrix} \tag{36}$$

$$E_\pm = \frac{\hbar^2}{mR^2}\left(\frac{K^2+\frac{1}{4}}{2} \pm \sqrt{\frac{\left(K-\frac{\omega mR^2}{2\hbar}\right)^2}{4} - \alpha\left(K-\frac{\omega mR^2}{2\hbar}\right)^2\cos\theta + \alpha^2}\right) \tag{38}$$

$$X_\pm = \begin{pmatrix} x_{1,\pm} \\ x_{2,\pm}\end{pmatrix} = \frac{1}{N_\pm}\begin{pmatrix} \frac{\hbar^2}{mR^2}\left(-\frac{1}{2}\left(K+\frac{1}{2}\right)^2 + \alpha\cos\theta\right) + \frac{\hbar\omega}{2} + E_\pm \\ \frac{\hbar^2}{mR^2}\alpha\sin\theta\end{pmatrix} \tag{41}$$

$$N_\pm^2 = \left(\frac{\hbar^2}{mR^2}\left(-\frac{1}{2}\left(K+\frac{1}{2}\right)^2 + \alpha\cos\theta\right) + \frac{\hbar\omega}{2} + E_\pm\right)^2 + (\alpha\sin\theta)^2 \tag{42}$$

$$E_\pm = \frac{\hbar^2}{mR^2}\left(\frac{K^2+\frac{1}{4}}{2} \pm \sqrt{\frac{\left(K-\frac{\omega mR^2}{2\hbar}\right)^2}{4} - \alpha\left(K-\frac{\omega mR^2}{2\hbar}\right)^2\cos\theta + \alpha^2}\right) = const = \epsilon_o \tag{45}$$

On the Absorber in Gravitation

Jacques Consiglio

52 Chemin de Labarthe, 31600 Labastidette, France
E-mail: Jacques.Consiglio@gmail.com

Assuming the big-bang is a periodic 4-dimensional event, we show that the main parameters of the ΛCDM model, namely matter, dark energy and total density, can be computed straightforwardly from Mach's principle and that the existence of dark matter is not necessary. As a result, we find that the cosmos expansion is the origin of mass and energy — but not the big-bang as a singular event.

1 Introduction

The object of this note is to show that once assumed that the big bang is a periodic event, and using absorber theory, the dark matter field is un-necessary in cosmology, and the dark energy is the natural free field of the absorber.

2 The Absorber and cosmology

Mach's principle states in a very general manner: *local physical laws are determined by the large scale structure of the universe*. This principle is the basis of the Wheeler-Feynman absorber theory [1, 2]. They suppose that the energy of particles is given by a time-symmetrical field; this interpretation was made by Tetrode and assumes that particles are not self-interacting. The main equations go as follows:

$$E_{total}(x_j,t) = \frac{\Sigma_{n\neq j}\left(E_n^{ret}(x_j,t)+E_n^{adv}(x_j,t)\right)}{2}, \qquad (2.1)$$

$$E_{damping}(x_j,t) = \frac{E_j^{ret}(x_j,t)-E_j^{adv}(x_j,t)}{2}, \qquad (2.2)$$

$$E_{total}(x_j,t) = E_{damping}(x_j,t)+\Sigma_{n\neq j}E_n^{ret}(x_j,t). \qquad (2.3)$$

They define the energies of the damping (2.2) and the total field (2.1–2.3), from advanced and retarded components for each particle (index j). The central idea is that the advanced field not being causal, it can only have damping effects while energetic interactions are causal. The theory was designed in electrodynamics but here we assume the field at the origin of gravitation and energy (including mass-energy and inertia), and propagating on the light cone.

The standard model of cosmology is based on general relativity theory (GRT). The idea is that the cosmos is self-contained (no outer realm), and internal metric expansion. However, it requires a unique event at its beginning, the so called big bang, resulting in the conceptual problem of its cause. Here we use Mach's principle on a larger scale: we assume the observed cosmos part of a wider 4-dimensional area. A 4-space denoted universe which we assume Euclidean with its own time and evolves as follows:

- A central location exists at the origin of the cosmos; we shall call it the emitter;
- A new cosmos or membrane is emitted periodically; the membranes separation is constant;
- The membrane progression is radial; the emitter produces more membranes and so on.

This structure is reminiscent of a wave; it is a manner to solve the problems of origin (the system is permanent); the membranes separation, if large enough, avoids the problem of instantaneous inflation. It also has the elegance of simplicity and the expansion is immediately linear. The idea at the basis of this concept was triggered-off by the recent observation of cosmological oscillations by Ringermacher and Mead [3].

3 Gravitation and energy

Now evidently, we have to build a theory from scratch; that is to say from experimental evidences. We shall use the following: we know from experimental gravitation physics that fixed clocks at different heights in the field have different rates; and the pulsation of photons and material system are constant in free fall. Equivalently, it is said that gravitation defines the context in which the rest of physics lives. According to Mach's principle it implies only a local variation of density which depends on the structure of the universe.

Denoting the density g, it varies according to $1/r$ as it addresses energy. This is classically written with:

$$g(r) = g^{\infty}\left(1-\frac{f(M)}{r}\right), \qquad (3.1)$$

where $f()$ is an undefined function of mass. The Newton potential reads:

$$\Gamma = \Gamma^0 - \frac{GM}{r}. \qquad (3.2)$$

Then G depends on $f()$, and Γ^0 is usually an arbitrary constant and the rest energy of a mass m is $E^0 = mc^2$. But now energy is given by the absorber mechanism and then the constant is $\Gamma^0 = c^2$. Then we write:

$$E = m\left(c^2 - \frac{GM}{r}\right). \qquad (3.3)$$

Therefore the density g defined in (3.1) is linked to mass-energy and to the velocity of light.

In a relativistic manner we can for instance define a variable c^*, use invariant masses and write:

$$c^{*2} = c^2 - \frac{2\,G\,M}{r}\,. \tag{3.4}$$

Since frequencies and wavelengths evolve conversely in the gravitational field, we write:

$$c^2\,d\tau^2 = c^{*2}\,dt^2 - \frac{c^2}{c^{*2}}\,dr^2. \tag{3.5}$$

Substituting from (3.4) this is the Schwarzschild metric. The pulsations of photons and of material systems in free fall are constant and then this equation applies identically to any form of energy. The concept is different from general relativity (GRT) but the equation is experimentally verified exactly in the same manner — that is to say uniquely in the solar system since all other verifications lead to suppose the existence of dark matter.

4 Dark energy and matter density

The absorber is time-symmetrical with causal effects; it concerns the total currents within the event horizon, say $M_A\,c^2$ the absorber "free" mass/energy. Equilibrium exists in the absorber process, and then the currents interfering with a mass m depend on m/M_A. We assume linear expansion; the visible cosmos radius is then $R_U = c/H = c\,T$ where H is the Hubble parameter and T the age of the membrane. Then by symmetry, we write:

$$\frac{E_m}{M_A\,c^2} = \frac{m}{M_A} \times \left(1 - \frac{M\,R_U}{M_A\,r}\right). \tag{4.1}$$

This is the Newton potential but the standard cosmological model is based on GRT which gives a factor $2GM$ like from (3.4–3.5), then in the standard theory the absorber free energy will be estimated from:

$$\frac{R_U}{2\,M_A} = \frac{G}{c^2}\,. \tag{4.2}$$

Using c, G and H we can now compute the absorber free energy; we find:

$$M_A = \frac{R_U\,c^2}{2\,G} = 8.790 \times 10^{52}\,\text{Kg}. \tag{4.3}$$

Considering visible energies $M_V c^2$, the ratio M_V/M_A is geometrical as it corresponds to the surface of a 4-sphere ; it is then $1/2\pi^2$. Then the factor 2 in (4.2) becomes $4\pi^2$ in 3+1D where masses interact. It gives:

$$2M_A = 4\pi^2\,M_V \to M_V = 4.453 \times 10^{51}\,\text{Kg}. \tag{4.4}$$

Summing (4.3–4.4), we get the total energy of the cosmos:

$$M_{total} = M_A + M_V = 9.236 \times 10^{52}\,\text{Kg}. \tag{4.5}$$

It corresponds to a density $\rho = 9.91 \times 10^{-27}\text{Kg/m}^3$ and the visible part (4.2) is 4.82% of the total. The benchmark at this time is the Plank mission results [4] which is $\rho = 9.90 \times 10^{-27}\text{Kg/m}^3$ and 4.9% of visible energy.

Hence according to the most favored model in cosmology we get three valid quantities in (4.3, 4.4, 4.5) which are *deduced* from the absorber symmetry and depend on geometry, c, G and $H = 1/T$. We do not get any dark matter, and assuming those results are significant we cannot afford any — though one could think that it may hide in M_A. But here the concept is different; the field is time-symmetrical and it cannot be an independent field as its relative amplitude is given by geometry.

With the results in this section we face two possibilities:

- The ΛCDM model parameters are tuned to match a linear expansion and it results in (4.3, 4.4, 4.5); which is a little surprising.
- A simple coincidence for M_A, but maybe a relevant result for M_V.

One way to make our mind is to develop the theory and check if the field needs dark matter.

5 The short range gravitational field

In (4.1) it appears that either G or M_V is variable; if we consider M_V constant, then G is a scale factor in proportions of R_U, but it is scale-independent on cosmological scales where R_U/r is constant.

In standard physics, one uses G, c and masses constant; we can then use the same constant quantities and it should give the differences between the Newton theory and the gravitational field given by our equations, at least a short range. In this section we consider that only t evolves and $T \gg t > 0$; it is linked to the Hubble factor H or R_U since the scenario of emission gives:

$$H(T)\,R_U(T) = c \to H(T) = \frac{c}{(R_0 + c\,T)} \approx \frac{1}{T}\,, \tag{5.1}$$

where $R_0 = R_U(T = 0)$ and T is the elapsed time since the separation of our membrane. Then from (4.2–5.1), denoting $R_U(T) \to R_U$ we can also write:

$$\frac{G\,M_A}{(R_U - c\,t)} = c^2. \tag{5.2}$$

Now all is constant except t and we can take a second order limited development; then denoting $H(T) \to H$, and using (5.1–5.2) we get:

$$\frac{G\,M_A\,H}{c} \times \left(1 + \frac{H\,r}{c} - \frac{H^2\,r^2}{c^2}\right) = c^2. \tag{5.3}$$

Multiplying G in the Newton potential by the terms of the limited development in (5.3) we introduce retarded interaction and then causality in the field (which is not in Newton's

theory). The potential is extended as:

$$\Gamma = \Gamma^0 - \frac{G\,M}{r} - \frac{G\,M\,H}{c} + \frac{G\,M\,H^2\,r}{c^2}\,. \tag{5.4}$$

Let us analyze how this potential works:

It first adds a constant negative energy term ($-G\,M\,H/c$) with no gravitational impact. It is then the contribution of the mass M to the constant c^2; it is the free absorber field and M must be summed to $2\,M_A$. Using (4.2) it leaves a negative constant $-c^2$ on the right-hand side. We get:

$$\Gamma = \Gamma^0 - c^2 - \frac{G\,M}{r} + \frac{G\,M\,H^2\,r}{c^2}\,.$$

Then $\Gamma^0 = c^2$ is immediate and the physical origin of energy is the expansion, not the big-bang.

The next term is then of identical nature and we sum again M to $2\,M_A$. Using (4.2) again yields $G\,M_A\,H^2\,r/c^2 = H\,c\,r$ (giving an acceleration $H\,c$). We now get:

$$\Gamma = \Gamma^0 - c^2 - \frac{G\,M}{r} + H\,c\,r\,. \tag{5.5}$$

But $\Gamma^0 = c^2$ and $\Gamma < 0$; then rescaling notations with $\Gamma + c^2 \to \Gamma$ and using (4.2) we choose to write:

$$\frac{\Gamma}{c^2} = 1 - \frac{M\,R_U}{2\,M_A\,r} + \frac{r}{R_U}\,. \tag{5.6}$$

It is well-known that stars at galaxies borders experience an anomalous centripetal acceleration in the range $H\,c$. This acceleration is the origin of the dark matter hypothesis by Oort in 1932.

Here the potential c^2 and the acceleration Hc are the effects of expansion and retarded interaction; it must be seen as the origin of energy and the known problem of conservation related to this acceleration is inexistent.

A second classical objection is that this anomaly is not observed in the solar system; however, we assume the absorber at the origin of mass/energy and the immediate consequence is that it transform in acceleration. We can directly transform the density g; that is, with acceleration $H\,c$ in any direction, a transformation L exists verifying:

$$L\left(Hc,\ g\left(T - \frac{r}{c}\right)\right) = g(T)\,. \tag{5.7}$$

The following transformation holds:

$$g\left(T - \frac{r}{c}\right) \times \left(1 + \frac{H\,r}{c}\right) = g(T)\,. \tag{5.8}$$

Because once extended to any acceleration A in place of $H\,c$, and replacing $r \to ct$, the non relativistic case gives:

$$g\,(T-t)\,\frac{A}{c} = \frac{g\,(T) - g\,(T-t)}{t}\,.$$

The right-hand of this equation is a time derivative, hence:

$$\frac{g\,A}{c} = \frac{dg}{dt} \to \frac{g}{c} = \frac{dg}{dv}\,. \tag{5.9}$$

It shows that a density obeying (5.8) creates resistance to acceleration and that mass increases with velocity. Hence the field is not Galilean, it is then a-priori relativistic. The equation (5.8) is equivalent to (and also justified by) the equation (5.2), but symmetrical where the field transforms in acceleration. This calculus shows, by symmetry, that a cosmological acceleration of the sun and its satellites in the direction of the galaxy core rescales the density and eliminates the term $H\,c\,r$; hence no second cosmological acceleration of its satellites can exist directed to the sun (and so on with planet's satellites).

6 Energy and the quantum world

6.1 Correspondence with the classical field

In this section, we shall continue using G constant and masses variable with time. The non-reduced Plank units and the Schwarzschild radius will be useful to the discussion. Recall:

$$M_P = \sqrt{\frac{h\,c}{G}}\,,\ \ l_P = \sqrt{\frac{h\,G}{c^3}}\,,\ \ t_P = \sqrt{\frac{h\,G}{c^5}}\,,\ \ R_S = \frac{2\,G\,m}{c^2}\,.$$

The equation (4.2) is equivalent to saying that the visible cosmos is defined by the Schwarzschild radius of M_A. The unique property of the Plank mass is that its Schwarzschild radius and wavelength are equal; it is then pivotal and using (4.2), we first write:

$$\frac{2\,M_A}{M_P^2} = 4\pi^2\frac{M_V}{M_P^2} = \frac{R_U\,c}{h}\,. \tag{6.1}$$

A similar equation can be written for any material system of mass m using its Schwarzschild radius:

$$\frac{2\,m}{M_P^2} = \frac{R_S\,c}{h}\,.$$

Hence, one could think that (6.1) is nothing new, but this is interesting firstly because this equation uses M_A and R_U, and not M_{total} as we may classically expect. It shows that any mass m and M_A come from the same mechanism, but in a reciprocal manner since the two quantities define opposite limit radius and obey the same equation. A complimentary equation gives unit-less ratios:

$$\frac{2\,M_A}{M_P} = \frac{R_U\,c^2}{G} \times \sqrt{\frac{G}{h\,c}} = \frac{R_U}{l_P} = \frac{T}{t_P}\,. \tag{6.2}$$

It expresses the same link with quantum physics; the system of units $[2M_A, R_U, T]$ is the time integral of the Plank system $[M_P, l_P, t_P]$. Again, it can be written with any mass m, but

not with M_V or M_{total}. Now using $h = c = G = 1$ we have $M_P = t_P = l_P = 1$, and the only evolving quantities are:

$$T = R_U = 2\,M_A = 4\pi^2\,M_V\,. \tag{6.3}$$

In the most natural system of units the cosmos energy is trivial and it appears to evolve. This is due to the choice of G constant. In facts, the cosmos expands exactly of one Compton wavelength of any massive system during one period of its pulsation (this is just $\lambda = h\,c/E$). The system $[2\,M_A, R_U, T]$ is just a time integral, and a system of units its differential. Consequently, the physical link with the quantum world is also trivial: *The cosmos expansion gives an action h at each period of any system pulsation.* It gives a very natural origin to the basics of quantum physic where energy is a time differential, $E = h\,\nu$.

We find identity of expansion, wave and energy, in perfect agreement with the results of the previous section.

6.2 The field

The Plank mass is pivotal in (6.1–6.2) then we model the absorber with an evolving field ϕ given by:

$$2\,M_A\,M_\phi = M_P^2 \to E_\phi = \frac{h\,c}{R_U} \approx 1.52\times 10^{-51}\,\mathrm{J}\,. \tag{6.4}$$

This is the energy of a field of wavelength R_U ($\approx 10^{-32}$ eV). Its energy is proportional to $1/R_U$ and decreases with time. But the laws of nature do not change; hence (6.4) is scale dependent but valid at any epoch and it is legitimate to write:

$$E_\phi(r) = \frac{h\,c}{r}\,, \qquad P_\phi(r) = \frac{h}{r}\,, \tag{6.5}$$

which addresses identically a hypothetical cosmos of radius r, and the field at a distant r of any mass.

A spherically inflating membrane defines a frame which is moving at velocity $v = c\,r/R_U$ at distance r from the attractive body M; then notice:

$$\frac{h}{M_\phi\,v} = r, \qquad \frac{h}{M_\phi(r)\,v} = R_U\,, \tag{6.6.1}$$

$$P_\phi(r) = \frac{h}{r} = M_\phi\frac{c^2}{v}\,. \tag{6.6.2}$$

The two expressions in (6.6.1) are equivalent to a de Broglie wavelength and in (6.6.2) momentum transfers on the light cone but in proportions of the phase velocity of the de Broglie wave. Now on top of the potential c^2, gravitation can be seen as a negative energy field. The equation (6.6.2) then corresponds to negative momentum on the light cone where the exchanged quantum is given by the de Broglie wave phase velocity $V = c^2/v$, and its emission rate is the Compton frequency of its source. In this way, we can write the field equations in an interesting semi-classical manner where all quantities depend on pulsation and momentum:

$$\frac{G}{c^2} = \frac{1}{P_\phi(R_U)}\times\frac{1}{\nu_A(T)} = const, \tag{6.7}$$

$$F = -\frac{P_\phi(r)^2}{P_\phi(R_U)}\times\frac{\nu_M(T)\,\nu_m(T)}{\nu_A(T)} = -\frac{GMm}{r^2}\,, \tag{6.8}$$

$$\frac{\Gamma}{c^2} = 1 - \frac{P_\phi(r)}{P_\phi(R_U)}\times\frac{\nu_M(T)}{\nu_A(T)} = 1 - \frac{GM}{rc^2}\,, \tag{6.9}$$

where notations are trivial for the Compton frequencies of the masses m, M, and $2\,M_A$ at the epoch T. From (6.4), the denominator is time independent, and then the choice of G constant is legitimate. (Though the alternate choice M_V constant where G is a scale factor also holds.)

6.3 Advanced and retarded components

Now let us show that the equations (6.8–6.9) are approximate *and come from causality*. Using constant masses, G is a scale factor and we can use the same limited development as before but with little interest; instead we shall use the absorber equations in section 2. In (6.8–6.9) the denominator is constant but the masses at the numerator evolve in proportion of time. Then using first (6.9) *without the potential* c^2, consider the distance $r = ct$ constant; at the time T the retarded and advanced momentum from M will be felt by m respectively like $P_\phi(r)\,\nu_M(T-t)$ and $P_\phi(r)\,\nu_M(T+t)$ in proportion of m. Recall also $\nu_M(T) = kT$, then we first write the damping potential; it gives the participation of M to the potential c^2 which we sum to the absorber mass:

$$\frac{\Gamma_{damping}}{c^2} = -\frac{P_\phi(r)\,\nu_M(T-t) - P_\phi(r)\,\nu_M(T+t)}{2\,P_\phi(R_U)\,\nu_A(T)}$$

$$= +\frac{P_\phi(r)\,\nu_M(t)}{P_\phi(R_U)\,\nu_A(T)} = +\frac{\nu_M(T)}{\nu_A(T)} \to +1. \tag{6.10.1}$$

Now the retarded potential:

$$\frac{\Gamma_{retarded}}{c^2} = -\frac{P_\phi(r)\,\nu_M(T-t) + P_\phi(r)\,\nu_M(T+t)}{2\,P_\phi(R_U)\,\nu_A(T)}$$

$$= -\frac{P_\phi(r)}{P_\phi(R_U)}\times\frac{\nu_M(T)}{\nu_A(T)}\,. \tag{6.10.2}$$

Of course their sum is causal and it gives:

$$\frac{\Gamma_{retarded}}{c^2} + \frac{\Gamma_{damping}}{c^2} = 1 - \frac{P_\phi(r)}{P_\phi(R_U)}\times\frac{\nu_M(T)}{\nu_A(T)}\,, \tag{6.10.3}$$

which is causal, agrees with (6.9), and now includes the potential c^2 from integration; but it misses the acceleration $H\,c$. A similar exercise is then needed on energy but we shall use forces as it will give the orientation of the acceleration; here we have to evaluate these on the full system (m plus M) exerted by all masses of the cosmos at the instant T. We shall do as if M and m were in a circular orbit at equal distance r of a third object (or their center of mass) as it is a representative test case. The retarded force on the system corresponds to the force from $M(T-t)$ to $m(T+t)$, summed with the force from $m(T-t)$ to $M(T+t)$; using again $r = ct$ we get:

$$\frac{F_{ret}}{c^2} = -P_\phi(2r)^2\frac{\nu_M(T-t)\,\nu_m(T+t) + \nu_M(T+t)\,\nu_m(T-t)}{P_\phi(R_U)\,\nu_A(T)}\,.$$

The advanced forces are identical and exerted at $T-t$; we get:

$$\frac{F_{adv}}{c^2} = -P_\phi(2r)^2 \frac{\nu_M(T+t)\,\nu_m(T-t) + \nu_M(T-t)\,\nu_m(T+t)}{P_\phi(R_U)\,\nu_A(T)}\,.$$

The damping force is null as it is the difference between those two expression; the retarded force is their sum and we extract the part related to $\nu_M(t)\,\nu_m(t)$ as the rest of the expression is identical to the potential; we get:

$$\Delta F = +P_\phi(r)^2 \frac{\nu_M(t)\,\nu_m(t)}{P_\phi(R_U)\,\nu_A(T)} > 0\,.$$

To simplify this expression we replace each momentum by its value (6.5) and use the linearity of $m(t) = m(T) \times t/T$:

$$\Delta F = +\frac{\nu_M(T) \times h\,\nu_m(T)}{\nu_A(T)\,R_U} \to -H\,c\,m(T) < 0\,. \qquad (6.10.4)$$

This expression depends only on T and we sum (for instance) M to get the effect on m of all masses of the cosmos; the sum is valid since the expression is independent of r. The sign of the force is negative since the masses in the sum are geometrically external to the system (except for the system itself which is negligible).

6.4 The Plank scale potential

At the Plank scale, (6.5) yields:

$$E_\phi(l_P) = \frac{h\,c}{l_P} = M_P\,c^2\,. \qquad (6.11)$$

This is the expected result in particles physics. But here the field is dependent on its source and this energy level does not pervade all space, the potential is c^2 and just multiplied by mass. Then, and more subtly, from (6.9), the main terms of the field potential cancel exactly at the Plank scale.

This section show the coherence of the classical field discussed in section 5 with the quantum world because the only equation introduced here is $Et = h$ or equivalently $P = h/r$.

7 Oscillations, expansion, black holes

A membrane of this kind has large thickness and the emission of the next membranes can be imprinted in the observable cosmos geometry; this imprint must be damped in proportions of the number of membranes existing between the emitter and ours. The oscillation recently observed by Ringermacher and Mead [3] corresponds to 7 minima and 6.5 ± 0.5 phases. The amplitude of the oscillations increases with distance. We interpret the minima as successive membrane emissions, and ≈6.5 visible oscillation phases for 7 membranes correspond to ≈ 50% of our membrane emission logically invisible, as a descending phase preceding its emission.

Now we want to understand the observation of 1A supernova since it leads to accelerated expansion and dark energy. The Chandrasekhar limit gives the mass of the type 1A supernova on which luminosity depends:

$$M_{limit} = \frac{k\,M_P^3}{(\mu_e\,m_h)^2}\,,$$

where k is a constant factor, M_P the Plank mass, μ_e the average molecular weight per electron, and m_h the mass of a hydrogen atom. Hence, with variable masses, M_{limit} evolves like $1/T^2$, which is in contradiction with observation (constant chemistry and atomic physics). Therefore, as it should in a gravitation theory, the field defines the context in which the rest of physics lives. It means that the same field is also at the origin of all charges interaction; not only of mass but of all forms of energy.

Consider then M_{limit} constant; the expression is epoch-independent and then also the emission luminosity. Now assume all measured red-shift are given by linear expansion (neglecting oscillations). Then photons will disperse more than with a decelerating expansion. A linear expansion is almost in perfect agreement with observation as shown by Perlmutter & al [5] and more recent works.

The ΛCDM model also uses baryonic acoustic oscillations to evaluate the size of the large structures of the cosmos; it requires dark matter and our equivalent is the acceleration $H\,c$ which becomes infinite when $T \to 0$. Then, large anisotropies of matter density should already be present at a very early epoch and primordial black-holes are also possible.

At the Schwarzschild radius the field potential reads:

$$\frac{\Gamma}{c^2} = 1 - 1 + \frac{R_S}{R_U}\,. \qquad (7.1)$$

The field is then compatible with the existence of black holes, which is obvious, but also with their stability since R_S/R_U is epoch-independent. Since the exchanges are time-symmetrical it creates neither black holes inflation (a known problem of pushing gravity) nor deflation.

8 Conclusion

We showed that the theory holds with no dark matter. It comes as a pressure field given by the very first quantum equation $P = h/r$; the gravitational field agrees with GRT results on a short scale and cosmology is straightforward. The field is coherent with Mach's principle; the emitter creates dissymmetry and the differential between the advanced and retarded field components create energy, gravitation, and the acceleration $H\,c$.

Interestingly, this field necessarily defines the context in which the rest of physics lives; hence it is also the origin of particles interaction and therefore it interacts with charges. Firstly the potential c^2 comes as a pressure field and can be interpreted as the Poincaré stress [6] and secondly it implies bottom-up unification.

9 Addendum

Still considering G constant, and since $H = 1/T$ and $m(T) = kT$, the force in (6.10.4) is time-invariant. In this equation, summing $m(T)$ to the absorber mass M_A gives:

$$\Delta F_{MA} = H\, c\, M_A(T) = \frac{c^4}{2\,G} = \frac{M_p\, c^2}{2\, l_p}\,, \tag{9.1}$$

which is half the Plank force; it also reads:

$$2\,\Delta F_{MA}\, l_p = M_p\, c^2\,. \tag{9.2}$$

This is the work of a force $2\Delta F_{MA}$ over the Plank length. We also have $M_p\, c^2 = h\, c/l_p$, and then:

$$2\,\Delta F_{MA}\, l_p^2 = h\, c\,, \tag{9.3}$$

which is the natural constant of the Yukawa interaction of the SM Higgs field. It also gives:

$$2\,\Delta F_{MA}\, l_p\, t_p = h\,. \tag{9.4}$$

This is the action of a force $2\,\Delta F_{MA}$ over the Plank length in the Plank time. Those equations read as if in a cosmos which radius is expanding at light speed (of length l_p in time t_p), a scalar field of constant hc is creating an additional dark energy $M_p c^2/2$ with an action h; then the total energy created by $\Delta\, F_{MA}$ since the big bang is:

$$M = \frac{M_p\, R_u}{2\, l_p} \to M = M_A\,, \tag{9.5}$$

which, of course, is identical to (4.2). Finally, we have just separated the forces of energy creation from the usual gravitation and then energy conservation.

This reasoning is circular as we introduce M_A at the beginning of the calculus; but there is no naturalness problem in the cosmology outlined here with respect to the constants of quantum physics (the cosmological constant and the so called "why now" problems are nonexistent). The novelty is the immediate significance of the Plank units and the permanence of energy creation; its power is constant and can easily be computed, it is half the Plank power which is then a constant of nature, and corresponds roughly to 2.4 W/Kg of dark or visible energy at the present epoch.

Submitted on October 30, 2015 / Accepted on November 1, 2015
Corrected on December 22, 2015

References

1. Wheeler J.A., Feynman R.P. Interaction with the Absorber as the Mechanism of Radiation. *Reviews of Modern Physics*, 1945, v. 17(2–3), 157–161.
2. Wheeler J.A., Feynman R.P. Classical Electrodynamics in Terms of Direct Interparticle Action. *Reviews of Modern Physics*, 1949, v. 21(3), 425–433.
3. Ringermacher H.I. and Mead L.R. Observation of discrete oscillations in a model-independent plot of the cosmological scale factor versus loopback time and scalar field. *Astronomical Journal*, 2015, v. 149, no. 4.
4. The Plank Collaboration. Planck 2015 results. I. Overview of products and scientific results. arXiv: 1502.01582
5. Perlmutter S. et al. Measurements of Omega and Lambda from 42 high redshift supernovae. *Astrophysical Journal*, 1999, v. 517(2), 565–86; arXiv: astro-ph/9812133.
6. Poincaré H. Sur la dynamique de l'électron. *Rendiconti del Circolo Matematico di Palermo*, 1903, v. 21, 129-176.

Power-Law Scaling of the Impact Crater Size-Frequency Distribution on Pluto: A Preliminary Analysis Based on First Images from New Horizons' Flyby

Felix Scholkmann

Research Office of Complex Physical and Biological Systems (ROCoS), Bellariarain 10, 8038 Zürich, Switzerland
E-mail: felix.scholkmann@gmail.com

The recent (14th July 2015) flyby of NASA's New Horizons spacecraft of the dwarf planet Pluto resulted in the first high-resolution images of the geological surface-features of Pluto. Since previous studies showed that the impact crater size-frequency distribution (SFD) of different celestial objects of our solar system follows power-laws, the aim of the present analysis was to determine, for the first time, the power-law scaling behavior for Pluto's crater SFD based on the first images available in mid-September 2015. The analysis was based on a high-resolution image covering parts of Pluto's regions *Sputnik Planum*, *Al-Idrisi Montes* and *Voyager Terra*. 83 impact craters could be identified in these regions and their diameter (D) was determined. The analysis revealed that the crater diameter SFD shows a statistically significant power-law scaling ($\alpha = 2.4926 \pm 0.3309$) in the interval of D values ranging from 3.75 ± 1.14 km to the largest determined D value in this data set of 37.77 km. The value obtained for the scaling coefficient α is similar to the coefficient determined for the power-law scaling of the crater SFDs from the other celestial objects in our solar system. Further analysis of Pluto's crater SFD is warranted as soon as new images are received from the spacecraft.

1 Introduction

The first close-up images of Pluto from NASA's New Horizons spacecraft, received in mid-September 2015, show a complex surface structure of Pluto never seen before in this detail. During the spacecraft's flyby of Pluto on 14th July 2015, images were taken by New Horizons' Long Range Reconnaissance Imager (LORRI) with a cooled 1024×1024 pixel CCD camera from a distance of approx. 12,500 km making it possible to obtain high-resolution images of Pluto's surface. Due to the slow transmission (about 1–2 Kbps), it will take around 16 months for all flyby images of Pluto to be received in full [1].

Many phenomena in astrophysics follow a power-law, i.e. the relationship between features exhibits a scale-invariance. Examples are the characteristics of the channel network on Mars [2], the relationship between solar flare occurrence and total flare energy [3], the correlation between a supermassive black hole mass and the host-galaxy bulge velocity dispersion ("M-sigma relation") [4], the distribution of initial masses for a population of stars ("initial mass function") [5], Kepler's third law, or the distribution of galaxies in the universe [6–8].

Size-frequency distributions (SFD) of natural objects also follow in general a power-law. Examples are the SFD of fragment sizes due to a fragmentation process [9], the SFD of landslides [10], the particle SFD of volcanic ash [11], the mass distribution objects of the Kuiper belt [12] — or the SFD of impact crater diameters on celestial objects.

Already in 1940 Young showed that the impact crater SFD on the Earth's Moon can be described by two power-laws with different scaling exponents. Further studies extended the analysis to other celestial objects, e.g. Earth [13], Mars [14–16] and Mercury [17].

Due to the lack of high-resolution images available, it has not been possible until now to analyze the impact crater SFD of Pluto. With the first images now available from NASA's New Horizons mission, the aim of the present paper was to conduct such a first, preliminary, analysis.

2 Materials and methods

2.1 Data

For the present analysis the raw images* obtained by the New Horizons' LORRI as of 14th September 2015 were visually inspected in order to find an image showing impact craters of Pluto with the highest resolution possible. An additional selection criterion was that the image had to be taken by LORRI at an angle capturing the region of interest maximally parallel to the camera, minimizing geometrical distortions of the features in the image.

The image lor_0299174809_0x630_sci_4 (in the following denoted as LOR-0299174809) was selected as fulfilling these criteria (see Figure 1(b)). LOR-0299174809 displays a particular area covering parts of Pluto's regions *Sputnik Planum*, *Al-Idrisi Montes* and *Voyager Terra*. The image was taken by LORRI on 14th July 2015, 10:14:50 UTC, with an exposure time of 150 ms.

2.2 Determination of crater diameter values

The image LOR-0299174809 was analyzed in Adobe Illustrator (version CS5; Adobe Systems, CA, USA) by first visually

*LORRI Images from the Pluto Encounter, http://pluto.jhuapl.edu/soc/Pluto-Encounter

identifying the craters on the image and measuring their diameters (D). The obtained values were than rescaled to give the final values in the unit km. To do so, the information given on NASA's website* was used. Information on the website indicated that image number three (from top), which covers the region displayed in LOR-0299174809, is 470 km in width.

2.3 Statistical analysis

For the statistical analysis we used the mathematical framework provided by the Santa Fe Institute [18, 19]. The data were analyzed in Matlab (version 2010a; Mathworks Inc., Natick, MA, USA).

2.3.1 Estimation of the lower-bound and exponent of the power-law model

A quantity x shows power-law scaling if it stems from a probability distribution $p(x) \sim x^{-\alpha}$, with the exponent α defining the characteristics of the scaling. To test if an empirically obtained probability distribution follows a power-law, classically a histogram is calculated and the distribution is analyzed on a doubly logarithmic plot. Since $p(x) \sim \alpha \ln(x) + \text{const.}$, a power-law distributed quantity x follows a straight line in the plot. Besides the fact that this method was (and is still) used to investigate power-law scaling of different quantities this approach can generate significant systematic errors [18]. Therefore, for the present analysis we used a framework presented by Clauset et al. [18] that circumvents these errors and also offers the possibility of estimating the lower bound of the power-law ($x_{\min}$). The determination of $x_{\min}$ is crucial when analyzing empirical data for power-law scaling since often the power-law behavior applies only for the tail region of the distribution, i.e. for values greater than the threshold value $x_{\min}$.

For the obtained crater diameter (D) data ($= x$) the power-law threshold value $D_{\min}$ ($= x_{\min}$) was determined based on the method described by Clauset et al. [18] which uses the Kolmogorow-Smirnov (KS) statistics. The scaling exponent α was then calculated with a maximum likelihood fitting method also described by Clauset et al. [18].

2.3.2 Determination of the statistical significance of the power-law fit

In order to determine if the fitted power-law can be considered statistically significant, a goodness-of-fit test described by Clauset et al. [18] was employed. To this end, power-law distributed synthetic data was generated with values of α and $x_{\min}$ that are equal to the values obtained by fitting the empirical data to the power-law model. Each synthetic data set is than fitted to the model and the KS statistics determined. Based on the occurrences of times the KS statistic is larger than for the empirical values, a p-value is calculated. For $p < 0.1$ the fit of the power-law model to the empirical data is considered to be statistically not significant, i.e. it can be ruled out that the empirical distribution obeys a power-law scaling. Thus, the p-value in this case represents a measure of the hypothesis that is tested for validity. A high value of p corresponds to a good fit.

3 Results, discussion and conclusion

83 impact craters could be identified and their diameter values were determined, ranging from 0.84 km to 37.77 km. Using the obtained 83 D values and the methods described in Section 2.3.1, the scaling coefficient α was determined to be $\alpha = -2.4926 \pm 0.3309$ and the scaling threshold value to be $D_{\min} = 3.75 \pm 1.14$ km. Thus, for D in the range $[3.75 \pm 1.14 \text{ km}, 37.77 \text{ km}]$ the values follow a power-law. The log-likelihood (L) of the data $D \geqslant D_{\min}$ under the fitted power law was determined to be $L = 104.5688$.

The statistical test, as described in Section 2.3.2, employing 10,000 synthetic data sets revealed a p value of 0.2241. Thus, according the test the hypothesis cannot be refuted that the data follows a power-law, i.e. Pluto's crater SFD shows a power-law scaling. Figure 1(c) visualizes the power-law behavior of the crater SFD.

How do the results of the presented analysis relate to the findings about characteristics of the crater SFD of other celestial objects? As mentioned in the introduction, it well established that the crater SFD of all investigated celestial bodies in our solar system exhibit a power-law scaling.

For example, according to an analysis performed by Robertson and Grieve from 1975 the crater SFD of the Earth is characterized by $\alpha \approx -2$ (for $D \geqslant 8$ km) [13]. An own analysis using the updated data of impact craters on Earth ($n = 188$) based on the Earth Impact Database† revealed $\alpha = 2.0286$ (for $D \geqslant 7$ km). The Earth Moon's crater SFD has been intensively investigated since the 1940's when Young [20] initially showed that for large D the crater SFD follows a power-law with $\alpha = -3$, and for small D the scaling is described by $\alpha = -1.5$. Subsequent studies described the scaling with laws governed by $\alpha = 2$ (for $D = [\sim 2 \text{ km}, 70 \text{ km}]$) [21], as well as $\alpha = -2$ (for $D < 100$ m) and $\alpha = -2.93$ (for $D >$ a few 100 m) [22], for example. Further studies showed that the scaling-relations of the lunar crater SFD need to include the observation that multiple power-laws are necessary to describe the whole SFD spectrum, i.e. α depends on D [23, 24]. A solution for optimally fitting the crater SFD was described based on the idea of using a polynomial function to fit the SFD data in the log-log space, i.e. it could be shown that a polynomial function of 7^{th} degree fit the data well for $D = [300 \text{ m}, 20 \text{ km}]$. The polynomial function included an extra term accounting for the fact that the scaling function also depends on the geological charac-

*http://tinyurl.com/n9k5mmc

†http://www.passc.net/EarthImpactDatabase

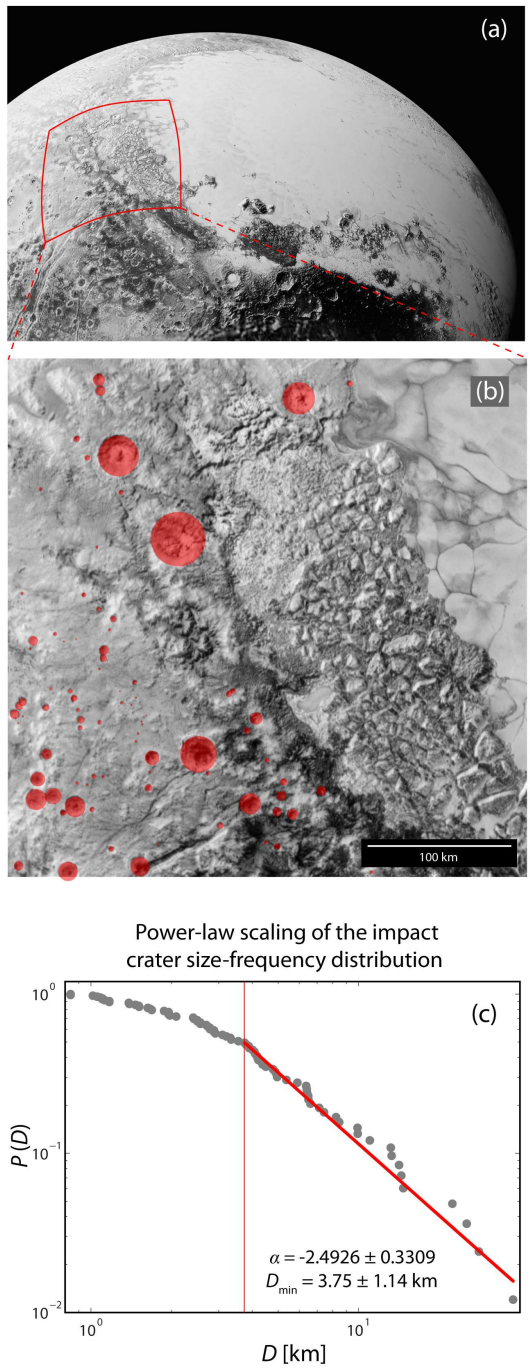

Fig. 1: (a) View of Pluto taken in July 2015 by LORRI on board NASA's New Horizons spacecraft. In the field of view the western lobe of the Tombaugh region is depicted. (b) LORI image lor0299174809_0x630_sci_4 showing a particular area covering parts of Pluto's regions *Sputnik Planum*, *Al-Idrisi Montes* and *Voyager Terra*. (c) Visualization of the power-law scaling of the impact crater size distribution. $P(D)$: complementary cumulative distribution function; D: crater diameter. Images (a) and (b) were obtained from NASA, Johns Hopkins University Applied Phsics Laboratory, Southwest Research Institute.

teristics of the region investigated — a finding also made by other studies (e.g., [24–26]). For an extended range of D, in later work a 11^{th} degree polynomial function was published by Neukum [27] valid for $D = [10\ \text{m}, 300\ \text{km}]$ and covering scaling exponents in the range of $\alpha = [-1, 4]$. For the Martian satellites Phobos and Deimos, the crater SFD was determined as being described by a power-law with $\alpha \approx 1.9$ for $D = [44\ \text{m}, 10\ \text{km}]$ [16].

Thus, the finding of the present analysis concerning the power-law characteristics (i.e., $\alpha = 2.4926 \pm 0.3309$ for $D = [3.75 \pm 1.14\ \text{km}, 37.77\ \text{km}]$) of the crater SFD of Pluto is comparable to the power-laws observed for the other celestial bodies. That Pluto's diameter scaling for $D < 3.75 \pm 1.14$km does not follow the $\alpha = -2.4926$ scaling relies most probably on the fact that small craters are much faster deteriorated due to erosion and that counting of craters with small D was not perfectly possible due to the limited resolution of the available image. The lowest D value (3.75 ± 1.14 km) for which the power law holds might interpreted as related to a transition from simple to complex craters. Interestingly, such a "transition diameter" was predicted for Pluto to be in the range of 4–5 km [28–30].

This analysis, of course, should be regarded only as a preliminary study for further follow-up as soon as the full set of images from Pluto is available and the images have been processed to deliver a high-resolution picture of Pluto's surface morphology. A limitation of the present analysis is that only one high-resolution image with sufficient craters was available. It was therefore only possible to obtain a relatively low number of crater diameter values ($n = 83$).

Knowledge of Pluto's crater SPF will not only give insights in the universality of the crater SFD scaling relations but necessarily will also help in the understanding of the history of Pluto and the characteristics of the Kuiper belt which Pluto is part of.

Acknowledgements

I thank Rachel Scholkmann for proofreading the manuscript, and the New Horizons project team at the Johns Hopkins University Applied Physics Laboratory for discussion concerning the copyright status of the publicly released LORRI images.

Submitted on October 15, 2015 / Accepted on November 6, 2015

References

1. Rienzi G. How exactly does New Horizons send all that data back from Pluto? Available from: `http://hub.jhu.edu/2015/07/17/new-horizons-data-transmission`
2. Caldarelli G., De Los Rios P., Montuori M., Servedio V.D.P. Statistical features of drainage basins in mars channel networks. *The European Physical Journal B*, 2004, v. 38 (2), 387–391.
3. Hudson H.S. Solar flares, microflares, nanoflares, and coronal heating. *Solar Physics*, 1991, v. 133 (2), 357–369.
4. Ferrarese L., Merritt D. A fundamental relation between supermassive black holes and their host galaxies. *The Astrophysical Journal*, 2000, v. 539 (1), L9–L12.

5. Kroupa P., Weidner C., Pflamm-Altenburg J., Thies I., Dabringhausen J., Marks M., Maschberger T. The stellar and sub-stellar initial mass function of simple and composite populations. In *Planets, Stars and Stellar Systems*, Oswalt T.D., Gilmore G. (Editors), 2013, Springer Netherlands, 115–242.
6. Watson D.F., Berlind A.A., Zentner A.R. A cosmic coincidence: The power-law galaxy correlation function. *The Astrophysical Journal*, 2011, v. 738 (1), 1–17.
7. Baryshev Y.V., Labini F.S., Montuori L., Pietronero L., Teerikorpi P. On the fractal structure of galaxy distribution and its implications for cosmology. *Fractals*, 1998 v. 6 (3), 231–243.
8. Joyce M., Sylos Labini F., Gabrielli A., Montuori M., Pietronero L. Basic properties of galaxy clustering in the light of recent results from the Sloan Digital Sky Survey. *Astronomy & Astrophysics*, 2005, v. 443 (1), 11–16.
9. Turcotte D.L. Fractals and fragmentation. *Journal of Geophysical Research*, 1986, v. 91 (B2), 1921–1926.
10. Stark C.P., Hovius N. The characterization of landslide size distributions. *Geophysical Research Letters*, 2001, v. 28 (6), 1091–1094.
11. Wohletz K.H., Sheridan F., Brown W.K. Particle size distributions and the sequential fragmentation/transport theory applied to volcanic ash. *Journal of Geophysical Research*, 1989, v. 94 (B11), 15703–15721.
12. Bernstein G.M., Trilling D. E., Allen R. L., Brown K. E., Holman M., Malhotra R. The size distribution of transneptunian bodies. *The Astrophysical Journal*, 2004, v. 128 (3), 1364–1390.
13. Robertson P.B., Grieve R.A.F. Impact structures in Canada — Their recognition and characteristics. *The Journal of the Royal Astronomical Society of Canada*, 1975, v. 69, 1–21.
14. Bruckman W., Ruiez A., Ramos E. Earth and Mars crater size frequency distribution and impact rates: Theoretical and observational analysis. *arXiv:1212.3273*, 2013.
15. Barlow N.G. Crater size-frequency distributions and a revised martian relative chronology. *Icarus*, 1988, 75, 285–305.
16. Thomas, P. Veverka, J. Crater densities on the satellites of Mars. *Icarus*, 1980, v. 41, 365–380.
17. Strom R.G., Chapman C.R., Merline W.J., Solomon S.C., Head J.W. Mercury cratering record viewed from Messergner's first flyby. *Science*, 2008, v. 321 (5885), 79–81.
18. Clauset A., Shalizi C.R., Newman M.E.J. Power-law distributions in empirical data. *SIAM Review*, 2009, v. 51 (4), 661–703.
19. Virkar Y., Clauset A. Power-law distributions in binned empirical data. *Annals of Applied Statistics*, 2014, v. 8 (1), 89–119.
20. Young, J., A Statistical Investigation of Diameter and Distribution of Lunar Craters. *Journal of the British Astronomical Association*, 1940, v. 50, 309–326.
21. Cross C.A. The size distribution of lunar craters. *Monthly Notices of the Royal Astronomical Society*, 1966, v. 134, 245–252.
22. Shoemaker E.M., Hait M.H., Swann G.A., Schleicher D.L., Dahlem D.H., Schaber G.G., Sutton R.L. Lunar regolith at tranquillity base. *Science*, 1970, v. 167 (3918), 452–455.
23. Chapman C.R., Haefner, R.R. A critique of methods for analysis of the diameter-frequency relation for craters with special application to the Moon. *Journal of Geophysical Research*, 1967, v. 72 (2), 549–557.
24. Neukum G., König B., Arkani-Hamed J. A study of lunar impact crater size-distributions. *The Moon*, 1975, v. 12 (2), 201–229.
25. Baldwin R.B. On the history of lunar impact cratering: The absolute time scale and the origin of planetesimals. *Icarus*, 1971, v. 14, 36–52.
26. Head J.W., Fassett C.I., Kadish S.J., Smith D.E., Zuber M.T., Neumann G.A., Mazarico E. Global distribution of large lunar craters: implications for resurfacing and impactor populations. *Science*, 2010, v. 329 (5998), 1504–1507.
27. Neukum, G. Meteoritenbombardement und Datierung planetarer Oberflächen. *Habilitationsschrift*. 1983, München, Germany: Universität München.
28. Moore J.M., Howad A.D., Schenk P.M., McKinnon W.B., Pappalardo R., Ryan E., Bierhaus E.B., et al. Geology before Pluto: Pre-encounter considerations. *Ikarus*, 2014, v. 246, 65–81.
29. Bray V.S., Schenck P.M. Pristine impact crater morphology on Pluto — Expectations for New Horizons. *Icarus*, 2015, v. 246, 156–164.
30. Zahnle K., Schenk P., Levison H., Donnes L. Cratering rates in the outer Solar System. *Icarus*, 2003, v. 163, 263–289.

Standard Model Particles from Split Octonions

Merab Gogberashvili

Javakhishvili Tbilisi State University, 3 Chavchavadze Avenue, Tbilisi 0179, Georgia
Andronikashvili Institute of Physics, 6 Tamarashvili Street, Tbilisi 0177, Georgia
E-mail: gogber@gmail.com

We model physical signals using elements of the algebra of split octonions over the field of real numbers. Elementary particles are corresponded to the special elements of the algebra that nullify octonionic norms (zero divisors). It is shown that the standard model particle spectrum naturally follows from the classification of the independent primitive zero divisors of split octonions.

The algebra of octonions [1–3] is interesting mathematical structure for physical applications (see reviews [4–7]). In this paper we suggest that split octonions over the reals form proper mathematical framework to describe elementary particles and show that some physical properties, like the variety of their spices, naturally follows from the structure of the algebra.

In [8–10] different aspects of geometrical applications of split octonions over the reals were considered. It is suggested to use split octonions as universal mathematical structure in physics, instead of vectors, tensors, spinors, etc. In this approach world-lines (paths) of particles are parameterized by the elements of split octonions,

$$s = \omega + \lambda^n J_n + x^n j_n + ctI \ . \quad (n = 1, 2, 3) \tag{1}$$

Here a pair of repeated upper and lower indices implies a summation, i.e. $x^n j_n = \delta_{nm} x^n j^m$, where δ^{nm} is Kronecker's delta.

Four of the eight real parameters in (1), t and x^n, denote the ordinary space-time coordinates, and ω and λ^n are interpreted as the phase (classical action) and the wavelengths associated with the octonionic signals.

The eight octonionic basis units in (1) are represented by one scalar (denoted by 1), the three vector-like objects J_n, the three pseudo vector-like elements j_n and one pseudo scalar-like unit I. The squares (inner products) of seven of the hypercomplex basis elements of split octonions give the unit element with the different signs,

$$J_n^2 = 1 \ , \quad j_n^2 = -1 \ , \quad I^2 = 1 \ . \tag{2}$$

It is known that to generate a complete basis of split octonions the multiplication and distribution laws of only three vector-like elements J_n are enough [1–3]. The three pseudo vector-like basis units, j_n, in (1) can be defined as the binary products,

$$j_n = \frac{1}{2} \varepsilon_{nmk} J^m J^k \ , \quad (n, m, k = 1, 2, 3) \tag{3}$$

where ε_{nmk} is the totally antisymmetric unit tensor, and thus describe orthogonal planes spanned by vector-like elements J_n. The seventh basis unit I (the oriented volume) is defined as the triple product of all three vector-like elements and has three equivalent representation in terms of J^n and j^n,

$$I = J_1 j_1 = J_2 j_2 = J_3 j_3 \ . \tag{4}$$

So the complete algebra of all non-commuting hypercomplex basis units has the form:

$$\left.\begin{aligned} &J_n J_m = -J_m J_n = \varepsilon_{nmk} j^k \\ &j_n j_m = -j_m j_n = \varepsilon_{nmk} j^k \\ &j_m J_n = -J_n j_m = \varepsilon_{nmk} J^k \\ &J_n I = -I J_n = j_n \\ &j_n I = -I j_n = J_n \end{aligned}\right\} . \tag{5}$$

The conjugation of vector-like octonionic basis units,

$$J_n^\dagger = -J_n \ , \tag{6}$$

can be understand as reflections. Conjugation reverses the order of J_n in products, i.e.

$$\left.\begin{aligned} &j_n^\dagger = \frac{1}{2} \left(\varepsilon_{nmk} J^m J^k \right)^\dagger = \frac{1}{2} \varepsilon_{nmk} J^{k\dagger} J^{m\dagger} = -j_n \\ &I^\dagger = (J_1 J_2 J_3)^\dagger = J_3^\dagger J_2^\dagger J_1^\dagger = -I \end{aligned}\right\} . \tag{7}$$

So the conjugation of the pass function (1) gives

$$s^\dagger = \omega - \lambda_n J^n - x_n j^n - ctI \ . \tag{8}$$

Using (2), (5) and (8) one can find the norm (interval) of the pass function (1),

$$N^2 = s s^\dagger = s^\dagger s = \omega^2 - \lambda^2 + x^2 - c^2 t^2 \ , \tag{9}$$

which is assumed to be non-negative. A second condition is that for physical events the vector part of (1) should be time-like [10],

$$c^2 t^2 + \lambda_n \lambda^n > x_n x^n \ . \tag{10}$$

One can represent rotations in the space of the split octonions (1) by the maps,

$$s' = e^{\epsilon\theta/2} s e^{-\epsilon\theta/2} \ , \tag{11}$$

where θ is some real angle and ϵ denotes the (3+4)-vector defined by the seven basis units J_n, j_n and I [1–3, 10]. The set of transformations (11), which satisfy the conditions (9) and (10), form the group $SO(3,4)$ of passive transformations of the coordinates x^n, λ^n and t [11]. However, to represent the active rotations in the space of s, which preserves the multiplicative structure (5) as well, we would need the transformations to be automorphisms. It means not all tensorial transformations of the coordinates λ_n, x_n and t, represent real rotations, only the transformations that have a realization as associative multiplications should be considered. Automorphisms of split-octonions form subgroup of $SO(3,4)$, the noncompact form of Cartan's smallest exceptional Lie group G_2^{NC} [12, 13].

Infinitesimal transformations of coordinates, which correspond to the action of the main geometrical group of the model, G_2^{NC}, can be written as [10]:

$$\left.\begin{aligned} x'_n &= x_n - \varepsilon_{nmk}\alpha^m x^k - \theta_n ct + \\ &\quad + \frac{1}{2}\left(|\varepsilon_{nmk}|\phi^m + \varepsilon_{nmk}\theta^m\right)\lambda^k + \left(\varphi_n - \frac{1}{3}\sum_m \varphi_m\right)\lambda_n \\ ct' &= ct - \beta_n\lambda^n - \theta_n x^n \\ \lambda'_n &= \lambda_n - \varepsilon_{nmk}\left(\alpha^m - \beta^m\right)\lambda^k + \beta_n ct + \\ &\quad + \frac{1}{2}\left(|\varepsilon_{nmk}|\phi^m - \varepsilon_{nmk}\theta^m\right)x^k + \left(\varphi_n - \frac{1}{3}\sum_m \varphi_m\right)x_n \end{aligned}\right\}, \quad (12)$$

with no summing over n in the last terms of x'_n and λ'_n. From the 15 parameters (five 3-angles: α^m, β^m, ϕ^m, θ^m and φ^m) in (12), due to the condition

$$\sum_n \left(\varphi_n - \frac{1}{3}\sum_m \varphi_m\right) = 0\,, \quad (13)$$

only 14 are independent.

The transformations (12) can be divided into several distinct classes [10]. For instance, the G_2^{NC}-rotations by the angles α^n, β^n and θ^n of the space-time coordinates only, imitate the ordinary infinitesimal Poincaré transformations of (3+1)-Minkowski space,

$$\left.\begin{aligned} ct' &= ct - \theta_n x^n + a_0 \\ x'_n &= x_n - \varepsilon_{nmk}\alpha^m x^k - \theta_n ct + a_n \end{aligned}\right\}, \quad (14)$$

where the space-time translations are defined as:

$$\left.\begin{aligned} a_0 &= -\beta_n\lambda^n \\ a_n &= \frac{1}{2}\,\varepsilon_{nmk}\theta^m\lambda^k \end{aligned}\right\}. \quad (15)$$

Time translations a_0 are smooth, since β_n are compact angles, but θ^m are hyperbolic and for the spatial translations a_n we have the Rindler-like horizons.

Note that Poincaré-like transformations (14) do not form subgroup of G_2^{NC} (the subgroup structure of G_2^{NC} one can be find, for example, in [13]), since we had neglected rotations of the extra time-like parameters λ_n. Complete G_2^{NC}-transformations reveal some new features in compare to the Minkowski case, like parity violations [10].

Another class of automorphisms,

$$\left.\begin{aligned} x'_n &= x_n + \left(\varphi_n - \frac{1}{3}\sum_m \varphi_m\right)\lambda_n \\ t' &= t \\ \lambda'_n &= \lambda_n + \left(\varphi_n - \frac{1}{3}\sum_m \varphi_m\right)x_n \end{aligned}\right\}, \quad (16)$$

represent rotations through hyperbolic angles, φ_1, φ_2 and φ_3 (of the three, due to (13), only two are independent) of the pairs of space-like and time-like coordinates x_n and λ_n, into the orthogonal planes $(x_1 - \lambda_1)$, $(x_2 - \lambda_2)$ and $(x_3 - \lambda_3)$. It is convenient to define 2-parameter Abelian subalgebra of G_2^{NC} by the generators of two independent rotations in these planes. It is known that the rank of G_2^{NC} is two, as of the group $SU(3)$ [13, 14]. In terms of the two parameters, K_1 and K_2, which are related to the angles φ_n as

$$\left.\begin{aligned} K_1 &= \frac{1}{3}\,(2\varphi_1 - \varphi_2 - \varphi_3) \\ K_2 &= -\frac{1}{2\sqrt{3}}\,(2\varphi_3 - \varphi_1 - \varphi_2) \end{aligned}\right\}, \quad (17)$$

the transformations (16) can be written more concisely,

$$\begin{pmatrix} \lambda'_1 + Ix'_1 \\ \lambda'_2 + Ix'_2 \\ \lambda'_3 + Ix'_3 \end{pmatrix} = e^{(K_1\Lambda_3 + K_2\Lambda_8)I}\begin{pmatrix} \lambda_1 + Ix_1 \\ \lambda_2 + Ix_2 \\ \lambda_3 + Ix_3 \end{pmatrix}, \quad (18)$$

where I is the vector-like octonionic basis unit ($I^2 = 1$) and Λ_3 and Λ_8 are the standard 3×3 Gell-Mann matrices [10]. Then one can classify irreducible representations of G_2^{NC} by two fundamental simple roots of the algebra (K_1 and K_2) and using analogies with $SU(3)$ interpret them as corresponding to the spin and hypercharge of particles. It is known that all quarks, antiquarks, and mesons can be imbedded in the adjoint representation of G_2^{NC} [14].

In the approach [8–10] the norm (9) can be viewed as some kind of space-time interval with four time-like dimensions. The ordinary time parameter, t, corresponds to the distinguished octonionic basis unit, I, while the other three time-like parameters, λ_n, have a natural interpretation as wavelengths, i.e. do not relate to time as conventionally understood. Within this picture, in front of time-like coordinates in the expression of pseudo-Euclidean octonionic intervals there naturally appear two fundamental physical parameters, the light speed and Planck's constant. Then from the requirement of positive definiteness of norms under G_2^{NC}-transformations,

together with the introduction of the maximal velocity, there follow conditions which are equivalent to uncertainty relations [9, 10]. Also it is known that a unique physical system in multi-time formalism generates a large variety of "shadows" (different dynamical systems) in (3+1)-subspace [15–19]. One can speculate that information of multi-dimensional structures, which is retained by these images of the initial system, might takes the form of hidden symmetries in the octonionic particle Lagrangians [10].

Split algebras contain special elements with zero norms (zero divisors) [1], which are important structures in physical applications [20]. For the coordinate function (1) we can define the deferential zero divisor,

$$\frac{d}{ds} = \frac{1}{2}\left[\frac{d}{d\omega} - J_n\frac{d}{d\lambda_n} - j_n\frac{d}{dx_n} - I\frac{d}{cdt}\right], \qquad (19)$$

such that its action upon s is:

$$\frac{ds}{ds} = 1\,. \qquad (20)$$

The operator (19) annihilates $s^\dagger$, while the conjugated derivative operator,

$$\frac{d}{ds^\dagger} = \frac{1}{2}\left[\frac{d}{d\omega} + J_n\frac{d}{d\lambda_n} + j_n\frac{d}{dx_n} + I\frac{d}{cdt}\right], \qquad (21)$$

is zero divisor for s, i.e.

$$\frac{ds^\dagger}{ds} = \frac{ds}{ds^\dagger} = 0\,. \qquad (22)$$

From these relations it is clear that the interval (9) is a constant function for the restricted left octonionic gradient operators,

$$\left.\begin{aligned}\frac{d}{ds_L}\left(s^\dagger s\right) &= \left(\frac{ds^\dagger}{ds}\right)s = 0\\ \frac{d}{ds_L^\dagger}\left(ss^\dagger\right) &= \left(\frac{ds}{ds^\dagger}\right)s^\dagger = 0\end{aligned}\right\}, \qquad (23)$$

and the invariance of the intervals,

$$ds^2 = dsds^\dagger = ds^\dagger ds\,, \qquad (24)$$

in the space of split octonions can be viewed as an algebraic property.

The octonionic wavefunctions Ψ, in general, should depend on s and on $s^\dagger$ as well. Thus we need the condition of analyticity of functions of split octonionic variables,

$$\frac{d\Psi(s, s^\dagger)}{ds^\dagger} = 0\,, \qquad (25)$$

which is similar to the Cauchy-Riemann equations from complex analysis. It can be shown that the system of eight algebraic conditions (25), in certain cases [21], lead to the octonionic Maxwell and Dirac equations [8].

Now consider non-differential zero divisors. These type of quantities are distinct elements of the algebra and thus in physical applications could be corresponded to the unit signals (elementary particles). In the algebra of split octonions two types of primitive zero divisors, idempotent elements (projection operators) and nilpotent elements (Grassmann numbers), can be constructed [1, 10]. There exist four classes (totally eight) of primitive idempotents,

$$\left.\begin{aligned}D_n^\pm &= \frac{1}{2}\,(1 \pm J_n)\,, \quad (n = 1, 2, 3)\\ d^\pm &= \frac{1}{2}\,(1 \pm I)\end{aligned}\right\}, \qquad (26)$$

which obey the relations:

$$\left.\begin{aligned}D_n^\pm D_n^\pm &= D_n^\pm\\ d^\pm d^\pm &= d^\pm\end{aligned}\right\}. \qquad (27)$$

The pairs (D_n^+, D_n^-) and (d^+, d^-) are zero divisors for each other,

$$\left.\begin{aligned}D_n^\pm D_n^\mp &= 0\\ d^\pm d^\mp &= 0\end{aligned}\right\}, \qquad (28)$$

and thus commute,

$$[D_n^+, D_n^-] = [d^+, d^-] = 0\,. \qquad (29)$$

We have also twelve classes (twenty four in total) of primitive nilpotents,

$$\left.\begin{aligned}G_{nm}^\pm &= \frac{1}{2}\,(J_n \pm j_m)\,, \quad (n, m = 1, 2, 3)\\ g_n^\pm &= \frac{1}{2}\,(I \pm j_n)\end{aligned}\right\}, \qquad (30)$$

which are zero divisors for themselves,

$$\left.\begin{aligned}G_{nm}^\pm G_{nm}^\pm &= 0\\ g_n^\pm g_n^\pm &= 0\end{aligned}\right\}. \qquad (31)$$

We see that separately the quantities (30) can be considered as the Grassmann numbers, but do not commute with their conjugates,

$$\left.\begin{aligned}&G_{nn}^\pm G_{nn}^\mp = d^\mp\\ &G_{nm}^\pm G_{nm}^\mp = \epsilon_{nmk}D_k^\pm\,, \quad n \neq m\ \ (n, m, k = 1, 2, 3)\\ &g_n^\pm g_n^\mp = D_n^\pm\end{aligned}\right\}, \qquad (32)$$

in contrast to the case of projection operators (29). The quantities $G_{nm}^\pm$ and $g_n^\pm$ are the elements of so-called algebra of Fermi operators with the anti-commutators,

$$\left.\begin{aligned}G_{nm}^\pm G_{nm}^\mp + G_{nm}^\mp G_{nm}^\pm &= 1\\ g_n^\pm g_n^\mp + g_n^\mp g_n^\pm &= 1\end{aligned}\right\}, \qquad (33)$$

which is some syntheses of the Grassmann and Clifford algebras.

We want to emphasize that the number of distinct primitive idempotents (four) and nilpotents (twelve), and there conjugates, coincides with the number of particle/antiparticle spices (bosons and fermions, respectively) of the standard model. This justifies our assumption that primitive zero divisors, which describe unit signals in the space of split octonions, can be corresponded to the elementary particles. The properties that the product of two projection operators reduces to the same idempotent (27), while the product of two Grassmann numbers is zero (31), naturally explains the validity of the Bose and Fermi statistics for the corresponding particles. In this picture distinct statistics follows from the existence of the two types of "light-cones" in the octonionic (4+4)-space (9), what shows itself in the definitions of the primitive zero divisors (26) and (30). Also note that the number of the standard model particle generations and the amount of spatial dimensions, both follow from the structure of the algebra of split octonions and are connected with the exitance of the three fundamental vector-like elements J_n.

To conclude, in this paper geometrical applications of real split octonions are considered and elementary particles are connected with zero divisors, the special elements of the algebra which nullify octonionic intervals. It is shown that the standard model particle spectrum naturally follows from the classification of the independent primitive zero divisors of the algebra.

Acknowledgments: This research was supported by the Shota Rustaveli National Science Foundation grant ST09_798_4 – 100.

Submitted on November 18, 2015 / Accepted on November 19, 2015

References

1. Schafer R. Introduction to Non-Associative Algebras. Dover, NY 1995.
2. Springer T.A and Veldkamp F.D. Octonions, Jordan Algebras and Exceptional Groups. Springer Monographs in Mathematics, Springer, Berlin 2000.
3. Baez J.C. *Bull. Am. Math. Soc.*, 2002, v.39, 145, arXiv: math/0105155 [math.RA].
4. Okubo S. Introduction to Octonion and Other Non-Associative Algebras in Physics. Cambridge Univ. Press, Cambridge, 1995.
5. Finkelstein D. Quantum Relativity: A Synthesis of the Ideas of Einstein and Heisenberg. Springer, Berlin 1996.
6. Gürsey F. & Tze C. On the Role of Division, Jordan and Related Algebras in Particle Physics. World Scientific, Singapore 1996.
7. Lõhmus J., Paal P. & Sorgsepp L. Nonassociative Algebras in Physics, Hadronic Press, Palm Harbor 1994; *Acta Appl. Math.*, 1998, v.50, 3.
8. Gogberashvili M. *Int. J. Mod. Phys. A*, 2006, v.21, 3513, arXiv: hep-th/0505101; *J. Phys. A*, 2006, v.39, 7099, arXiv: hep-th/0512258.
9. Gogberashvili M. arXiv: hep-th/0212251; *Adv. Appl. Clif. Alg.*, 2005, v.15, 55, arXiv: hep-th/0409173; *Adv. Math. Phys.*, 2009, 483079, arXiv: 0808.2496 [math-ph].
10. Gogberashvili M. & Sakhelashvili O. *Adv. Math. Phys.*, 2015, 196708, arXiv: 1506.01012 [math-ph].
11. Manogue C.A. & Schray J. *J. Math. Phys.*, 1993, v.34, 3746.
12. Cartan É. Sur la structure des groupes de transformations finis et continus. Thèse, Paris 1894, p.146; *Ann. Sci. Ecole Nonn. Sup.*, 1914, v.31, 263; Reprinted in: *Oeuvres complètes*, Gauthier-Villars, Paris 1952.
13. Beckers J., Hussin V. & Winternitz P. *J. Math. Phys.*, 1986, v.27, 2217.
14. Günaydin M. & Gürsey F. *J. Math. Phys.*, 1973, v.14, 1651; *Lett. Nuovo Cim.*, 1973, v.6, 401; *Phys. Rev. D*, 1974, v.9, 3387.
15. Cole E.A.B. *Nuov. Cim. A*, 1977, v.40, 171; *J. Phys. A*, 1980, v.13, 109.
16. Gogberashvili M. *Phys. Lett. B*, 2000, v.484, 124, arXiv: hep-ph/0001109.
17. Gogberashvili M. & Midodashvili P. *Phys. Lett. B*, 2001, v.515, 447, arXiv: hep-ph/0005298; *Europhys. Lett.*, 2003, v.61, 308, arXiv: hep-th/0111132.
18. Christian J. *Int. J. Mod. Phys. D*, 2004, v.13, 1037, arXive: gr-qc/0308028; in *Relativity and the Dimensionality of the World*, ed. V. Petkov, Springer, NY 2007, arXiv: gr-qc/0610049.
19. Velev M.V. *Phys. Essays*, 2012, v.25, 3.
20. Sommerfeld A. Atombau und Spektrallinien, II Band. Vieweg, Braunschweig 1953.
21. Gogberashvili M. *Eur. Phys. J. C*, 2014, v.74, 3200, arXiv: 1410.4136 [physics.gen-ph].

The Impact Crater Size-Frequency Distribution on Pluto Follows a Truncated Pareto Distribution: Results from a First Data Set Based on the Recent New Horizons' Flyby

Lorenzo Zaninetti[1] and Felix Scholkmann[2]

[1] Dipartimento di Fisica, University of Turin, Via Pietro Giuria 1,10125 Turin, Italy. E-mail: zaninetti@ph.unito.it
[2] Research Office of Complex Physical and Biological Systems (ROCoS), Bellariarain 10, 8038 Zürich, Switzerland
E-mail: felix.scholkmann@gmail.com

Recently it could be shown (Scholkmann, *Prog. in Phys.*, 2016, v. 12(1), 26-29) that the impact crater size-frequency distribution of Pluto (based on an analysis of first images obtained by the recent New Horizons' flyby) follows a power law ($\alpha = 2.4926 \pm 0.3309$) in the interval of diameter (D) values ranging from 3.75 ± 1.14 km to the largest determined value of 37.77 km. A reanalysis of this data set revealed that the whole crater SFD (i.e., with values in the interval of 1.2–37.7 km) can be described by a truncated Pareto distribution.

1 Introduction

The recent flyby of NASA's New Horizons spacecraft allowed high-resolution images of Pluto's surface morphology to be obtained, thus enabling a first determination of the impact crater size-frequency distribution (SFD) of a specific region, i.e., covering parts of Pluto's regions *Sputnik Planum*, *Al-Idrisi Montes* and *Voyager Terra* [1].

The first analysis of the crater SFD used a power law of the type $p(x) \sim x^{-\alpha}$ to model the data. In the present paper we show the results of an extended analysis. The inverse power law scaling is known as the Pareto distribution $p(x) \sim x^{-(c+1)}$. In the present paper we tested the hypothesis that an upper truncated Pareto distribution (i.e., a Pareto distribution in which the probability range is limited rather than infinite) can improve the modelling of the empirical crater SFD presented in [1].

We review the properties of the Pareto and the truncated Pareto distributions in Section 2, and report in Section 3 the results of applying the truncated Pareto distribution to the novel Pluto crater SFD data set.

2 From the Pareto to the truncated Pareto distribution

In the follwing we report the definitions of the probability density function (PDF), the distribution function (DF), the survival function (S) and the maximum likelihood estimator (MLE) for the two distributions analyzed here. The sample is made by crater diameter (D) values ($n = 83$) denoted by x_i.

2.1 The Pareto distribution

The Pareto PDF is given by

$$f(x; a, c) = ca^c x^{-(c+1)}, \tag{1}$$

with $c > 0$; the Pareto DF is defined as

$$F(x; a, c) = 1 - a^c x^{-c}, \tag{2}$$

and the survival function is given by

$$S(x; a, c) = 1 - F(x; a, c). \tag{3}$$

The parameter values can be estimated by applying the MLE:

$$a = \min(x_i), \tag{4a}$$

$$\frac{1}{c} = \left(\frac{1}{n}\right) \sum_{i=1}^{n} \ln\left(\frac{x_i}{\tilde{a}}\right). \tag{4b}$$

More details can be found in [2].

2.2 The truncated Pareto distribution

An upper truncated Pareto random variable is defined in the interval $[a, b]$, and the PDF is given by

$$f_T(x; a, b, c) = \frac{ca^c x^{-(c+1)}}{1 - \left(\frac{a}{b}\right)^c}; \tag{5}$$

and the truncated DF is defined as

$$F_T(x; a, b, c) = \frac{1 - \left(\frac{a}{x}\right)^c}{1 - \left(\frac{a}{b}\right)^c}. \tag{6}$$

The MLE determines the parameters according to

$$a = \min(x_i), \tag{7a}$$

$$b = \max(x_i), \tag{7b}$$

$$0 = \frac{n}{\tilde{c}} + \frac{n\left(\frac{a}{b}\right)^{\tilde{c}} \ln\left(\frac{a}{b}\right)}{1 - \left(\frac{a}{b}\right)^{\tilde{c}}} - \sum_{i=1}^{n} [\ln x_i - \ln a], \tag{7c}$$

where the value of $\tilde{c}$ can be found using Brent's method to find a root of a nonlinear function, i.e., by applying the FORTRAN subroutine ZBRENT [3]. More details can be found in [4].

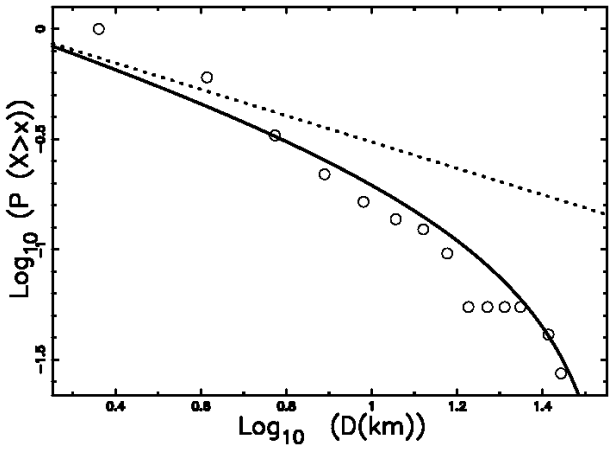

Fig. 1: Survival function (S) in a log-log plot for crater size in D = [1.38 km, 37.77 km]. Empty circles: empirical data, full line: S of the truncated Pareto PDF, dotted line: S of the Pareto PDF. The K-S test for the truncated Pareto gave $P_{KS} = 0.128$ and $d_{max} = 0.134$. The K-S test for the Pareto gave $P_{KS} = 0.0075$ and $d_{max} = 0.192$.

3 Data analysis and results

For statistical testing the Kolmogorov–Smirnov (K-S) test [5–7] was employed which does not require data binning. The K-S test, as implemented by the FORTRAN subroutine KSONE [3], finds the maximum distance (d_{max}) between the theoretical and the empirical DF as well the significance level P_{KS} (see equations 14.3.5 and 14.3.9 in [3]). A value of $P_{KS} \geqslant 0.1$ assures that the fit is acceptable.

When using the impact crater SFD data of Pluto with D = [3.75 km, 37.77 km] the Pareto PDF gave $c = 1.5299$ and thus $\alpha = 2.5299$ (similar to the value $\alpha = 2.4926$ reported by [1]), and the K-S test gave $P_{KS} = 0.866$ and $d_{max} = 0.091$. Figure 1 shows the empirical and and the two fitted distributions when the interval of crater size values is enlarged so that all D values are included in the fitting, i.e., D = [1.2 km, 37.77 km]. The truncated Pareto distribution describes the empirical crater SFD quite accurately over the whole interval of D values available.

4 Conclusions

The distribution of crater diameters of planets is commonly modeled by a power law. A small modification of the "simple" PDF by a truncated Pareto PDF (as given by equation (5)) allows the dichotomy of the infinite rather than finite range of existence to be avoided and provides better K-S test statistics with respect to the Pareto PDF (i.e., a "simple" power law), see captions of Figure 1.

In conclusion, we were able to show that the empirical impact crater SFD of Pluto (using a first data set based on recent New Horizons flyby) closely agrees with a truncated Pareto distribution. Applying the same modelling approach to an extended data set of Pluto's crater values is warranted to confirm our results – a task to be done as soon as new images of the New Horizon spacecraft are available.

Submitted on November 30, 2015 / Accepted on December 2, 2015

References

1. Scholkmann F. Power-law scaling of the impact crater size-frequency distribution on pluto: A preliminary analysis based on first images from New Horizons flyby. *Progress in Physics*, 2016, v. 12(1), 26–29.
2. Evans M., Hastings N. and Peacock B. Statistical Distributions (third edition). John Wiley & Sons Inc, New York, 2000.
3. Press W. H., Teukolsky S. A., Vetterling W. T. and Flannery B. P. Numerical Recipes in FORTRAN. The art of scientific computing, Cambridge University Press, Cambridge, UK, 1992.
4. Zaninetti L. and Ferraro M. On the truncated Pareto distribution with applications. *Central European Journal of Physics*, 2008, v. 61, 1–6.
5. Kolmogoroff A. Confidence limits for an unknown distribution function. *The Annals of Mathematical Statistics*, 1941, v. 12(4), 461–463.
6. Smirnov N. Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, 1948, v 19(2), 279–281.
7. Massey F. J. Jr. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 1951, v. 46 (253), 68–78.

Determination of the Neutrino Mass

Anatoly V. Belyakov
E-mail: belyakov.lih@gmail.com

The neutrino mass in four different independent formulations have been successfully calculated on the basis of the mechanistic interpretation of J. Wheeler's geometrodynamics concept. Mechanical analogue of the weak interaction is presented. Its adequacy is confirmed by the various variants for calculating the neutrino mass. The calculated mass agrees well with the indirect estimation of the neutrino mass obtained on the basis of cosmological data. It has been established that neutrinos can change its structure and properties, in particular, a magnetic moment, that leads to changes in the power of detected neutrinos flow (neutrino oscillations). The time constant of neutrino oscillations is calculated.

1 Introduction

The geometrodynamics of the famous scientist John Archibald Wheeler, who passed away in 2008, does not seem to find favor among modern physicists.

According to J. Wheeler's geometrodynamic concept charged microparticles are considered therein as singular points located in a non-unitary coherent two-dimensional surface and connected to each other through "wormholes", current tubes, or current force lines of the input-output (source-drain) kind in an additional dimension, thus forming a closed contour. However, "wormholes" in space, if they are not considered as purely mathematical constructions, in its physical embodiment can only be vortex formations of some kind substance that has the properties of an ideal fluid.

Assuming their existence, consistently developing and complicating the concept, one has managed to develop the mechanistic model, in which the properties of objects in both the microcosm and space scales are grounded and defined by using only the most general physical laws [1–4]. The determination of the neutrino mass and the calculation of other characteristic parameters provided out later in this article are the final confirmation of the correctness of the chosen model.

Experiments on the direct measurement of the neutrino mass, based on the kinematics of weak decays, to date do not give the exact value of neutrino masses, but only set the upper limit for it (the limit is permanently decreasing). The lowest limit is obtained indirectly by studying cosmological data on the relict radiation, the galaxies recession and other. According Adam Moss and Richard Battye's analysis of the data of Planck Space Telescope and their comparison with gravitational lensing observations on distant galaxies gives an upper limit for the total amount of neutrino masses of about 0.320 ± 0.081 eV [5].

2 Initial conditions

Recall that, in the proposed model, from a purely mechanistic viewpoint the *charge* only manifests the degree of the nonequilibrium state of physical vacuum; it is proportional to the momentum of physical vacuum in its motion along the contour of the vortical current tube. Respectively, the *spin* is proportional to the angular momentum of the physical vacuum with respect to the longitudinal axis of the contour, while the *magnetic interaction* of the conductors is analogous to the forces acting among the current tubes [1].

In such a formulation the electric constant ε_0 makes sense the linear density of the vortex current tube

$$\varepsilon_0 = \frac{m_e}{r_e} = 3.233 \times 10^{-16}\ \text{kg/m}, \tag{1}$$

and the value of *inverse magnetic constant* makes sense of the centrifugal force

$$\frac{1}{\mu_0} = c^2 \varepsilon_0 = 29.06\, n \tag{2}$$

appearing due to rotation of an element of the vortex tube having the mass m_e and the classical radius r_e with the light velocity c; this force is equivalent to the force acting between two elementary charges at the given radius.

Elementary particles are like vortex structures in an ideal fluid which can stay in two extreme forms: the vortex *at the surface* along the X-axis (let it be the analogue of a fermion of the mass m_x), and the vortex thread or a *sub-surface* vortical current tube having of the peripheral velocity v, the radius r and the length l_y along the Y-axis (let it be the analogue of a boson of the mass m_y). These structures oscillate inside a real medium, passing through one another (forming an oscillation of oscillations) showing that a mass (an energy) can have two states and pass from one form to another.

In paper [2], proceeding from the conditions of conservation of charge and constancy parameters μ_0 and ε_0, the parameters of the vortex thread m_y, v, r for an arbitrary p^+–e^--contour were defined as

$$m_y = (an)^2\, m_e\,, \tag{3}$$

$$v = \frac{c_0^{1/3}\, c}{(an)^2}\,, \tag{4}$$

$$r = \frac{c_0^{2/3}\, r_e}{(an)^4}\,, \tag{5}$$

where n is the quantum number, a is the inverse fine structure constant, c_0 is the dimensionless light velocity $c/[\text{m/sec}]$.

Wherein, referring to the constancy ε_0 (linear density), it is clear that the relative length of the tube current in the units of r_e is equal the boson mass m_y in the units of m_e, i.e.

$$l_y = m_y = (an)^2 . \tag{6}$$

In the framework of the model, the particles themselves are a kind of a contour of a subsequent order, formed by the intersection of the X-surface with the current tube, and they have their own quantum numbers defining the influence zone of these microparticles.

In [2] we determined that

$$n_p = \left(\frac{2c_0}{a^5}\right)^{1/4} = 0.3338 \tag{7}$$

for a proton, while for the electron we have $n_e = (n_p)^{1/2} = 0.5777$.

To calculate the mass of an arbitrary fermion m_i a formula was obtained

$$m_i = m_e \left(\frac{n_e}{n_i}\right)^{14} . \tag{8}$$

Hereinafter all the numerical values of the mass, size and speed are given in dimensionless units: as the respective proportions of the electron mass m_e, its radius r_e and the speed of light c.

It is important to note that the vortex tube contour (which the vortex thread fills helically) can be regarded as completely "stretched", i.e. elongated proportionally to $1/r$ or, contrary, extremely "compressed" i.e. shortened proportionally to $1/r$ and filling all the vortex tube of the radius r_e. In the latter case its compressed length $L_p = l_y\, r$ is numerically equal to the energy of the contour boson mass in the mass-energy units $m_e c^2$.

Indeed, because $r = v^2$, the numerical values of the aforementioned quantities (expressed in dimensionless units) are in all cases identical, and for an arbitrary axis are

$$L_{p_i} = l_i\, r_i = m_i\, r_i = m_i\, v^2 = \frac{c_0^{2/3}}{(an_i)^2} . \tag{9}$$

It is obvious that the mass of an arbitrary boson in the mass-energy units matches its own numerical value m_y only in the case of ultimate excitation of the vortex tube wherein we have $r \to r_e$ and $v \to c$.

When considering a closed contour having contradirectional currents from the balance of the magnetic and gravitational forces recorded in a "Coulombless" form, the characteristic size of the contour comes as a *geometric mean* of two linear values [2], which in the r_e units has the form:

$$l_k = (l_i\, r_i)^{1/2} = \left(\frac{z_{g_1} z_{g_2}}{z_{e_1} z_{e_2}}\right)^{1/2} (2\pi\gamma\rho_e)^{1/2} \times [\text{sec}], \tag{10}$$

where z_{g_1}, z_{g_2}, z_{e_1}, z_{e_2}, r_i, l_i are gravitational masses and charges expressed through the mass and charge of the electron, the distance between the current tubes and theirs length, γ is the gravitational constant, while ρ_e is the electron density $m_e/r_e^3 = 4.071 \times 10^{13}$ kg/m^3.

In the $p^+ - e^-$-contour, proton quarks become an active part of the proton mass, and are involved in the circulation. Their mass as z_g enters into the equation (10). When a proton and an electron are approaching, for example, in the case of e-capture, the contour becomes deformed and reduced.

3 Determination of the neutrino mass from the conditions of weak interaction

Let the neutrino is a particle having fermion and boson parts; the latter is separated in the weak interaction process (for example, when electron-proton absorption occurs) from the proton-electron X-contour into the region Y; see Figure 1. Let us find the neutrino mass *on the basis on the parameters of the neutrino vortex tube*.

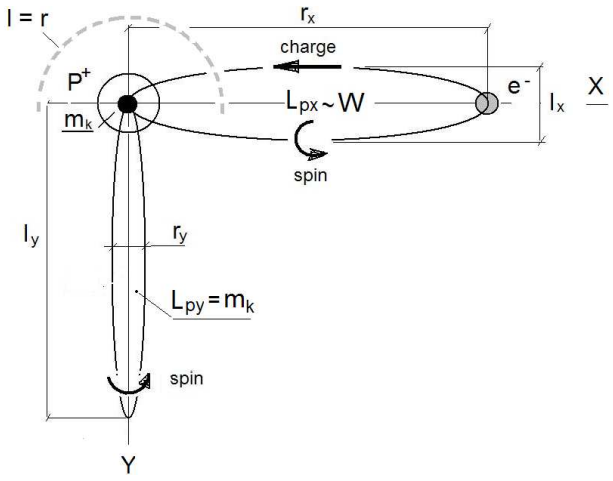

Fig. 1: Scheme of formation of the neutrino.

For the X-contour, referring to (9), its energy-mass in units of $m_e c^2$ is

$$L_{p_x} = \frac{c_0^{2/3}}{(an_x)^2} . \tag{11}$$

It is necessary to define the same parameters n_y and L_{p_y} for the neutrino. Because of the special stability of the neutrino, one can assume that its structure is characterized by all possible balances and symmetries.

Proceeding from energy balance, we assume that the active part of the proton, i.e. the quark energy-mass, is equal to the neutrino boson vortex tube energy-mass

$$m_k = L_{p_y} = \frac{c_0^{2/3}}{(an_y)^2} . \tag{12}$$

For the $p^+ - e^-$-X-contour it is accepted: $z_{g_1}/z_{e_1} = 1$, $z_{e_2} = 1$ and $z_{g_2} = m_k = L_{p_y}$. Then, using (9), from (10) we get:

$$\frac{L_{p_x}}{L_{p_y}} = 2\pi\gamma\rho_e \times [\sec^2]. \quad (13)$$

Assume, due to symmetry, that the contour large axis along X-axis and the neutrino vortex tube along Y-axis are equal, i.e., $r_x = l_y$. Then, referring to (5) and (6), the relation between the quantum parameters X and Y-contours is

$$n_y n_x^2 = \frac{c_0^{1/3}}{a^3}. \quad (14)$$

Proceeding from the formulas (11–14), as a result, we have

$$L_{p_x} = c_0^{4/9}\left(2\pi\gamma\rho_e \times [\sec^2]\right)^{1/3} = 1.51 \times 10^5\ (77\,\text{GeV}), \quad (15)$$

$$L_{p_y} = m_k = \frac{c_0^{4/3}}{L_{p_x}^2} = 8.83\ (4.51\text{MeV}), \quad (16)$$

as well as the quantum parameter of the neutrino vortex tube

$$n_y = \frac{c_0^{1/3}}{aL_{p_y}^{1/2}} = 1.643. \quad (17)$$

Now, according to the equation (8) the neutrino fermion mass is found:

$$m_v = \left(\frac{n_e}{n_y}\right)^{14} = \left(\frac{0.5777}{1.643}\right)^{14} = 4.39 \times 10^{-7}\ (0.225\ \text{eV}). \quad (18)$$

Additional sequels appear: the X-contour energy-mass is very close to a W-boson mass (80 GeV), and the estimated mass of a quark agrees well with that of the d-quark (4.8 MeV).

In more detailed consideration of the weak interaction, process the possibility of finding the neutrino mass *from the conservation of energy and symmetry* is detected. In the process of *e*-capture, the proton-electron X-contour is reduced and deformed in the Y-region. Being already the neutrino Y-contour, it contains the neutrino mass instead of the electron mass. Let us assume that at some intermediate state, before the allocation in the vortex tube form, Y-contour still maintains its momentum (the unit charge). In this case the formula (15) includes a neutrino mass m_v (in the m_e units), and, at $z_{e_2} = 1$, applies to the neutrino contour. It has the form:

$$(L_{p_x})_v = c_0^{4/9} m_v^{1/3}\left(2\pi\gamma\rho_e \times [\sec^2]\right)^{1/3}. \quad (19)$$

At the same time the X-contour initial energy-mass L_{p_x} has been transformed into the proton active part energy-mass (i.e., the quark mass $(L_{p_y})_v$). Then, referring to (16), we can write

$$L_{p_x} = (L_{p_y})_v = \frac{c_0^{4/3}}{(L_{p_x})_v^2}. \quad (20)$$

As a result, considering (15) and (19) from (20) we obtain:

$$m_v = \left(2\pi\gamma\rho_e \times [\sec^2]\right)^{-3/2} \quad (21)$$

that gives 4.5×10^{-7} (0.23 eV), the amount actually coincided with the result of the formula (18). With making the similar actions under the condition of the short axes equality $r_y = l_x$, then the same result has been got. In this case, contrary, $L_{p_y} = (L_{p_x})_v$ that apparently corresponds to the inverse process of the neutron in proton transformation.

Finally, the neutrino mass can be derived from the *conditions of complete symmetry*, i.e. from the state that is intermediate between the neutron and the proton when the X and Y-contours merge into one symmetrical contour at the zero point coordinates. This state apparently occurs only under some distinctive amount of the neutrino contour charge, namely — it is the charge value per one structure unit of the standard contour (per one photon) or e_0/a [1].

Indeed, since for a symmetrical contour $n_x = n_y$, $l = r = c_0^{2/9}$ and $L_{p_x} = L_{p_y} = c_0^{4/9}$, then by introducing into the initial formula (10) $z_{e_2} = 1/a$ from (19) we obtain

$$m_v = a^{-1}\left(2\pi\gamma\rho_e \times [\sec^2]\right)^{-1}, \quad (22)$$

that gives 4.28×10^{-7} (0.219 eV), the same amount as the resulting from the formulas (18) and (21).

Note that if a single photon has a linear size of $1/a$ of the standard contour length, i.e. the value of $c_0^{2/3}/a$, then the neutrino has a similar size of $c_0^{2/3} m_v$ or $0.192\, r_e$. This value is about 1/3 of the proton diameter; it is the linear quark dimension along the axis X. Indeed, since for the quark we have $n = 0.48$, then $r_x = c_0^{2/3}/(an)^{3.5} = 0.194\, r_e$ [2]. This coincidence additionally points to the correctness of the proton quark model, as set out earlier.

Full symmetry and the combining of the $p^+ - e^-$-contour and the neutrino contour are possible only in a special excited state of the nucleon. In reality, the electromagnetic interaction (nominal axis X) and weak interaction (nominal axis Y) are realized separately, and then only in a certain scale range, forming three generations of elementary particles [2]. That is, here is a *mechanical analogue of spontaneous electroweak symmetry breaking in the SM*.

Thus, the proposed model clearly describes the process of the weak interaction (how a proton absorbs an electron). The proton-electronic contour is reduced until the energy-mass becomes equal to the energy of W-particles. Then it transmits this energy and momentum (charge) to the proton, transforming it into an excited state (the neutron); further the contour is allocated into Y-region as the neutrino vortex tube with the parameter $n_y = 1.643$, keeping its spin and having the value of energy-mass equal to that of the light d-quark.

4 Determination of the neutrino mass from the limit conditions

At last, the neutrino mass is possible to be found directly from the magnetic-gravitational equilibrium conditions; from the equation (10), by substituting the limit conditions.

A vortex thread or tube in a non-viscous medium can be either closed or having an output to the surface of X, that is having a charge. The neutrino has no detectable charge and, therefore, it represents a closed structure or a contour.

Assume that Planck's size $r_h = (\hbar\gamma/c^3)^{1/2}$ has a physical meaning and it is the minimum size of the elementary neutrino vortex contour, i.e., $r_i = r_h = 1.616 \times 10^{-35}$ m or $5.735 \times 10^{-21}\, r_e$. Then, taking into account (5) and (6), a geometric mean is obtained from (10) as

$$l_k = c_0^{1/6} r_h^{1/4}. \tag{23}$$

In [2] it is shown that any electron vortex tube includes three vortex zones. But as one of the zones needs to be double, there should in general be four vortex threads each containing one-quarter of the electron total momentum (charge). Therefore, the elementary neutrino should be viewed as a pair of the closed vortex threads. Accordingly, two types of neutrinos are possible there: a pair of left-right rotation and, conversely, a pair of right-left rotation, obviously, as a neutrino and an antineutrino.

For a pair of the vortex threads at $z_{e_1} = z_{e_2} = \frac{1}{4}\, e_0$ and at $z_{g_1} = z_{g_2}$, having in mind (23), from (10) it should be:

$$z_g = m_v = \frac{c_0^{1/6} r_h^{1/4}}{(32\pi\gamma\rho_e \times [\mathrm{sec}^2])^{1/2}}, \tag{24}$$

that gives 4.31×10^{-7} (0.220 eV), the amount actually coinciding with the results of the formulas (18), (21), and (22). It should be noted that these results are only the ones of its kind since these formulas include only the fundamental constants.

Thus, the two states of the neutrino are obtained — *at the moment of birth in the form of a vortex tube and in its ultimate state in the form of a closed structure*, and the fermion neutrino mass in the initial state turned out equal to the gravitational mass of the neutrino vortex threads in the ultimate state. Is it possible to reconcile these very different states? Perhaps, it must be admitted that since the neutrino's vortex tubes initially contain all four single vortex threads then further the neutrino transforms into two potentially possible final forms (neutrino and antineutrino) maybe passing some intermediate states.

As for the muon and the tau-neutrinos, the electron mass in the formula (10) can be formally replaced by the masses of the muon and the tau-particles, provided that the linear density of the contour tube remains unchanged (that is not obvious). Then, as follows from the above formulas, the contours' parameters are changed, and the contours are deformed "stretching" along their axes; the X-contour energy-mass increases as the cube root of the relative weight of the microparticle. For the muon contour $L_{p_x} = 456\,\mathrm{GeV}$, which is equal to twice the value of the total energy-mass of the standard $p^+ - e^-$-contour (229 GeV) [1]. For the τ-contour $L_{p_x} = 1170\,\mathrm{GeV}$. This value is the sum of the neutrino energy and that of the expected boson energy-mass of the third generation, the heaviest one, which is not yet registered in experiment; that is, having the value of about 1000 GeV, which matches to the value defined earlier in [2]. As follows from the above formulas masses of the muon and the tau-neutrinos must be much less than that of the electron neutrino, and the resulting formulas give different results that may indicate instability of these neutrinos, like other particles of the second and third generations.

The fact of the neutrino transformation is derived from the model and confirmed by the experimentally detected neutrino oscillations.

5 Neutrino magnetic properties and its oscillations

The neutrino boson vortex tube retains the electron spin, and has a magnetic moment μ. The magnetic moment is determined relative to the axis Y. By definition, the μ is the product of the charge × the velocity × the path. Suppose that for the vortex thread the peripheral speed is v, while the path is πr. Revealing v and r through (4) and (5), as a result we obtain

$$\mu = \frac{\pi c_0\, c\, e_0\, r_e}{(an)^6}\ \mathrm{Am}^2. \tag{25}$$

(Ampere at a "Coulombless" system is equivalent to the acting force.)

The neutrino magnetic moment in the moment of its allocation μ_{v_0} according to formula (25) at $n_y = 1.643$ is $9.81 \times 10^{-31}\ \mathrm{Am}^2$. Moreover, it appears that this value with high accuracy is equal to the geometric mean of the proton magnetic moment μ_p and the vortex tube magnetic moment with average parameter l_k (Compton wavelength), which complies to $n_y = 8.07$ [2]. Its magnetic moment $\mu_k = 6.99 \times 10^{-35}\ \mathrm{Am}^2$, which corresponds to 0.75×10^{-11} Bohr's magneton. That is,

$$\mu_{v_0} = \left(\mu_p \mu_k\right)^{1/2}. \tag{26}$$

Such a large magnetic moment of neutrinos are not detected, but what is significant, it is the magnetic moment μ_k close to 10^{-11} Bohr's magneton that requires the neutrino to explain the anticorrelation of the registered neutrino flow with the magnetic flow near the sun surface. It is assumed that the neutrino magnetic moment interacts with the magnetic field in the outer convective layers of the sun, which leads to the spin precession of neutrinos changing its helicity from left to right; and the right neutrinos are not registered by detectors [6, 7]. The same neutrino magnetic moment is required because of some astrophysical limitations regarding the dynamics of stars [7].

So it is logical to assume that the neutrino magnetic moment, an originally very large magnitude, rapidly decreases

to the value of about 10^{-11} Bohr's magneton at the intersection of the Sun's surface, and in the neutrino ultimate state it becomes absolutely negligible. The reason for this is the transformation of the neutrino contour, which is analogous to the process of the transformation of a neutron into a proton.

Indeed, if the counter comprises several vortex threads with co-directed currents, they must be rotated relative to the longitudinal axis. At the same time, since by definition an elementary unit of the model medium (vortex thread) is absolutely inelastic and at the same time is absolutely deformed, the closed counter must be deformed ("turned out") in different structures by changing its parameters.

From the equality of the magnetic and inertial (centrifugal) forces for the vortex threads the peripheral rotation speed relative to the longitudinal axis of the contour is obtained

$$v_0 = \frac{(z_{e_1} z_{e_2})^{1/2} r_e}{(2\pi)^{1/2} \times [\text{sec}]} . \qquad (27)$$

This speed does not depend on the length of the vortex threads and distances between them and for the unit charges is 1.124×10^{-15} m /sec.

Earlier [2], it was found that the *time constant* of the transformation process (the ratio of the counter size to the peripheral speed) has appeared equal to the neutron lifetime.

Similarly, the time constant for the neutrino can be expressed in the forms $\tau_v = r_y/v_0$. Then, referring to (5), with $n = 1.643$, we obtain $\tau_v = 4.37 \times 10^{-4}$ sec (the time constant should be increased with the decrease of the residual charge of the neutrino). During this time the neutrinos having the speed of light move away from the source at a distance of 1.31×10^5 m. If they would transform to another form, a decrease in their number would be registered when the detector would be displaced from the source at a distance not less than the calculated value.

It is the distance the largest neutrino detector KamLAND (Kamioka Liquid Scintillator Anti-Neutrino Detector, Honshu island, Japan) has registered a decrease of the neutrino flow in the nuclear reactor antineutrino experiments [8]); see Figure 2 (the data are taken from [8]).

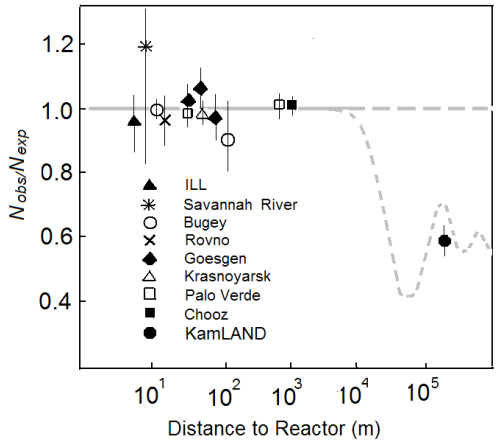

Fig. 2: The ratio of the measured neutrinos flows in the expected ones if there is no oscillations for experiments with reactor neutrinos.

6 Conclusion

Thus, one value of the neutrino mass has been derived by theoretical methods. Moreover, the same result was then obtained in four different formulas and three of them on the basis of the classical mechanistic model (actually through the analogue of spontaneous electroweak symmetry breaking in the SM). The results coincided with the indirect estimate of the neutrino mass derived from cosmological data. It was established that neutrinos may exist in various forms. It arises in the form of the electron neutrino with the mass of about 0.22 eV and further during the transition to its final state with the same mass may possibly change its parameters like the mass and magnetic moment, which results in the changes of a detectable power neutrino flow (oscillations). It is possible that the muon-neutrinos and tau-neutrinos are not stable. Apparently, they are the intermediate states of the totally stable electron neutrino.

The fact that the same neutrino mass is obtained in several ways may indicate that the values of other fundamental constants can also be obtained through the neutrino mass, which apparently is a key element of matter.

Submitted on October 14, 2015 / Accepted on November 17, 2015

References

1. Belyakov A.V. Charge of the electron, and the constants of radiation according to J. A. Wheeler's geometrodynamic model. *Progress in Physics*, 2010, v. 6, issue 4, 90–94.
2. Belyakov A.V. Macro-analogies and gravitation in the micro-world: further elaboration of Wheeler's model of geometrodynamics. *Progress in Physics*, 2012, v. 8, issue 2, 47–57.
3. Belyakov A.V. Evolution of stellar objects according to J. Wheeler's geometrodynamic concept. *Progress in Physics*, 2013, v. 9, issue 1, 25–40.
4. Belyakov A.V. Nuclear power and the structure of a nucleus according to J.Wheeler's geometrodynamic concept. *Progress in Physics*, 2015, v. 11, issue 1, 89–98.
5. Battye R.A., Adam M. Evidence for massive neutrinos from CMB and lensing observations. arXiv: 1308.5870, August 27, 2013.
6. Wolfenstein L., Bayer Yu.U. Neutrino oscillations and solar neutrinos. *Physics Today*, July 1989, v. 42, no. 7, 28–36.
7. Derbin A.V. The search for the magnetic moment of the neutron. *Particle Physics and Nuclear Physics*, Petersburg Nuclear Physics Institute at Gatchina, Russia, 2001, v. 32, v. 3, 734–749.
8. Orekhov D.I. Toolkit for neutrino physics. Dept. of Physics, Moscow State University, August 27, 2006.

On the Nature of the Magnetic Field of the Earth and Other Planets

Vladimir Danilov
E-mail: danvlad@bk.ru

This study presents a hypothesis of the origin and maintain of the magnetic field of the Earth and the planets. The mechanism of the tides on the opposite side of the Earth from the Moon is considered. The possible causes that enforce the continents to displace are discussed in couple with the causes that distort the shape of the Earth, and the causes of the jumps of the astronomical time. A mechanism of earthquakes is proposed, as well as a version of the appearance of the "magnetic tubes" in the Sun. The source of the forces causing the equatorial current and wind is shown.

Contents

1. Introduction
2. Tides
3. Currents
4. Earthquakes
5. Time jumps and killer waves
6. Causes of a dip appearing in the gravity graph during the Sun's passage across the sky
7. On motion of the magnetic poles
8. Conclusion

Appendix. A short comparison of the planets' magnetic fields depending on the number of their moons and other properties

---

> Books of physics are full of complicated mathematical formulae. But thought and ideas, not formulae, are the beginning of every physical theory.
>
> Albert Einstein

> That hypothesis which explains the current world with the fewest assumptions and means should have an advantage, because it is less arbitrary.
>
> Empedocles, *On Nature, the Law of Economy*

> The form of development of natural science, in so far as it thinks, is the hypothesis... If one should wait until the material for a law was in a pure form, it would mean suspending the process of thought in investigation until then and, if only for this, reason, the law would never come into being.
>
> Friedrich Engels, *Dialectics of Nature*

1 Introduction

The Earth's magnetic field makes our planet habitable — there would be no life on the planet without it. It protects the Earth's biological envelope from the hostile lifeless space and devastating effects of cosmic-ray particles. The habitability-determining need for a magnetic field reduces the number of potentially habitable planets. It is hard to enumerate all the effects of the field on inhabitants of the planet. Its properties are used by both humans and animals, while the scientific community has no unambiguous approach to understanding the mechanism of the field's creation and maintenance, as well as on the factors affecting its behavior.

One of the most popular hypotheses trying to explain the nature of the field is the dynamo theory. It proposes that convective and/or turbulent motions of conductive fluid in the core trigger self-excitation of a magnetic field and maintain the field stable.

However, it is hard to imagine the core steadily moving up to the surface in the same direction due to temperature — if it is convective motion; or the turbulence created by rotation being so stable that it could maintain self-excitation, and even in the same direction. Though, the nature of turbulence is not clear either. Over time, in the absence of external forces, the inner substance of the Earth will also rotate together with the shell due to its viscosity. The origin of the potentials in the core is also unclear. Why are they not compensated, if the substance is conductive? The authors of this hypothesis themselves thought it was a far cry from being proven. Although the hydrodynamic dynamo hypothesis explains many well-known facts, it is clear that the power triggering the "dynamo" has been defined incorrectly.

Another hypothesis proposes that the magnetic field is created in the ionosphere by the solar wind.

The third one says about salt-water flows in the oceans.

None of these theories can be applied to all the planets of the Solar System free of contradictions. For example, Jupiter spins in the same direction as the Earth does, but Jupiter's magnetic field is directed opposite to the Earth's one. Venus and Mars have no strong fields.

Anyway, it is not fair to believe that the Earth owns some unique features that no other planet has. After all, it is not the only planet that has a magnetic field, and it is not quite the thing to do to come up with its own mechanism for creating a magnetic field for each planet either. So what could be right? There should be a single physics of this phenomenon.

It just manifests itself somewhat differently because of different conditions of existence of different planets.

I would like to note here that the modern model of the Earth (with a hard core inside, surrounded by liquid alloy) is based on the study of behavior of acoustic (seismic) waves and their ability to pass in solid and liquid media differently. High-temperature plasma with close-packed nuclei will conduct seismic waves as a solid (crystalline) material, which is consistent with the measured data, and the adopted boundary of the solid core may be a boundary of transition to the plasma state. Generally it is hard to imagine — without inventing new forms and states — that some substance would "float" in a hard form in the same melted substance without melting itself.

This article presents a hypothesis of emergence and maintenance of the planet's magnetic field taking into account its own travel (axial inclination) in the solar ecliptic and the properties of the planet itself and its moons, if any. It shows that the outer shell of the planet is "independent" from the processes occurring in the planet's interaction with other bodies, thus allowing the magnetic poles to move, up to their inversion.

Attempts to find the answers to the following questions

1. What is the origin of the Earth's and other planets' magnetic fields?
2. Why does the far side of the Earth furthest from the Moon has tides too?
3. Why do the Moon and most moons keep the same side turned to their planets?
4. What force causes the continents to move?
5. What causes earthquakes?
6. Why is the Earth not round?
7. What are the reasons for sharp changes in astronomical time?
8. How do "killer-waves" occur?
9. Why is there a dip in the gravitation graph during the Sun's passage across the sky?
10. What are the reasons for periodic variations of geophysical fields and seismic activity?
11. What gives rise to and maintains major ocean currents and equatorial winds?

have given rise to the following hypothesis:

> The main reason for all of the above phenomena is the gravitational interaction of the Sun and moon(s) with a moving core of the planet.

The main proof of the hypothesis is the clear connection in the chain "**planet — satellite(s) — planet's magnetic field**" for various planets of the Solar System, bearing in mind that each planet is a moon of the Sun in its turn.

Thus, it can be noted that:

1. The magnetic field is effective if a planet has a moon or more. The field is small if the planet has no moons (e.g., Venus and Mercury have no moons, and their magnetic fields are very small);
2. If the planet cooled down and does not have a liquid core, it does not have a magnetic field either (e.g., the Moon);
3. Direction and shape of a planet's magnetic field depends on both the direction of rotation of the planet itself in the ecliptic plane and the orbit of the moon revolving around the planet (e.g., Mars and Uranus have reverse rotation of moons and reverse magnetic fields);
4. In the presence of multiple moons, the field becomes complex, and priority in the field's direction is determined by the more closely spaced or the more massive moon (for example, Uranus or Neptune);
5. Direction of the main winds and location of dust clouds on most of the planets in the Solar System coincides with the direction of their moons' motion.

In addition, the fact that the most moons revolve around their planets turning one side on them, and the rotation of planets such as Venus and Mercury is synchronized with the motion of the Earth (the two planets turn the same hemisphere to the Earth when approaching it), shows that cosmic bodies interact with each other not as uniform bodies, but as bodies with misplaced centers of mass. At the same time, in the case of a liquid core, the center can move within the hard shell of the planet.

Let's consider the mechanism of occurrence of a magnetic field (MF) in the example of the Earth. It will be the same for any Earth-like planet.

Imagine the Earth as a fixed sphere filled with substances of various densities and various specific gravity, and the Sun as a source of gravity affecting these substances. It is obvious that the heavier structures will gravitate to the shell of the sphere that is closest to the source of gravity, and distribution of density and mass within the Earth will be uneven not only in depth, but also towards the Sun (see Fig. 1).

According to modern theories of the Earth structure, substances below the lower mantle are in a liquid state (metallic phase) — plasma — where electrons are separated from the

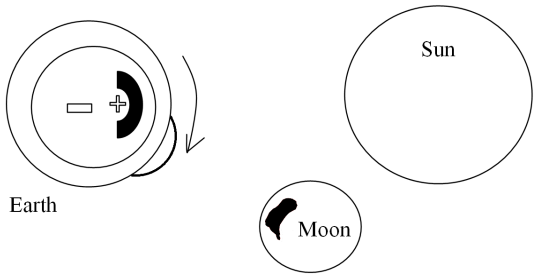

Fig. 1: Mass distribution.

nuclei. But, as the nuclei are much heavier than the electrons, it is clear that these are "precipitating" nuclei. Then a division inside the Earth's core occurs not only by mass but also by electric potential. The core of the Earth has become a dipole with the center of mass shifted significantly, where "+" and the bulk mass of the core are closer to the Sun.

While the Earth rotates, this part of the Earth's core follows the Sun and thereby create directed motion of electrically charged particles and circular, cyclic displacement of the center of mass of the Earth relative to its shell.

In 1878, Henry A. Rowland proved that charges moving on a moving conductor are identical in their magnetic effect to conduction current in a conductor at rest. Thus, in our case, the right-hand rule is generally appropriate, as evidenced by the direction of motion of the core part carrying a positive charge and the force lines of the Earth's magnetic field.

It certainly does not mean that one side of the sphere is pure "+" and the other is "−". Otherwise there would be no magnetic field formed in rotation of such a dipole because of the mutual compensation. There are just different motion radii, and different linear speeds respectively, and hence current potentials are different too. There may occur some compensation in motion of various charges, but "+" prevails.

More information on polarization of plasma in massive astronomical objects due to gravitational forces and their interaction with Coulomb forces is available in works by Igor Iosilevskiy (for example, in his publications [1, 2]).

By the way, if we accept the proposed hypothesis, the formation of the dipole inside the planet is a practical proof of the theoretical assumptions made by Iosilevskiy.

Of course, besides the Sun, the behavior of the charged core is also influenced by all the planets and the Moon in particular (see the section on tides).

Another proof of the hypothesis are daily and annual changes in the magnetic field direction, i.e., dependence of the field on the Earth's position relative to other objects affecting division by mass, charge, and trajectory of the core. (In the case of the now accepted hypothesis of a hydrodynamic dynamo, there should be no such influence.)

In fact, the heavy part of the core moves from East to West and in spirals from North to South and back with changes in axial inclination (change of season).

A very interesting measured data were provided by Yury P. Malyshkov and Sergey Yu. Malyshkov [3] on the basis of their research done in the Institute of Monitoring Climatic and Ecological Systems, Russsian Academy of Sciences.

Based on years of research on natural pulse electromagnetic fields of the Earth (NPEMFE) in seismically active areas of the Baikal Lakeside, they came to a conclusion on the motion of the planet's core and related natural phenomena — seismic activity, impact on the human body and so forth. The figures showing intensity of NPEMFE changes at different points in time exactly repeat the expected movement pattern of the dipole's heavy part.

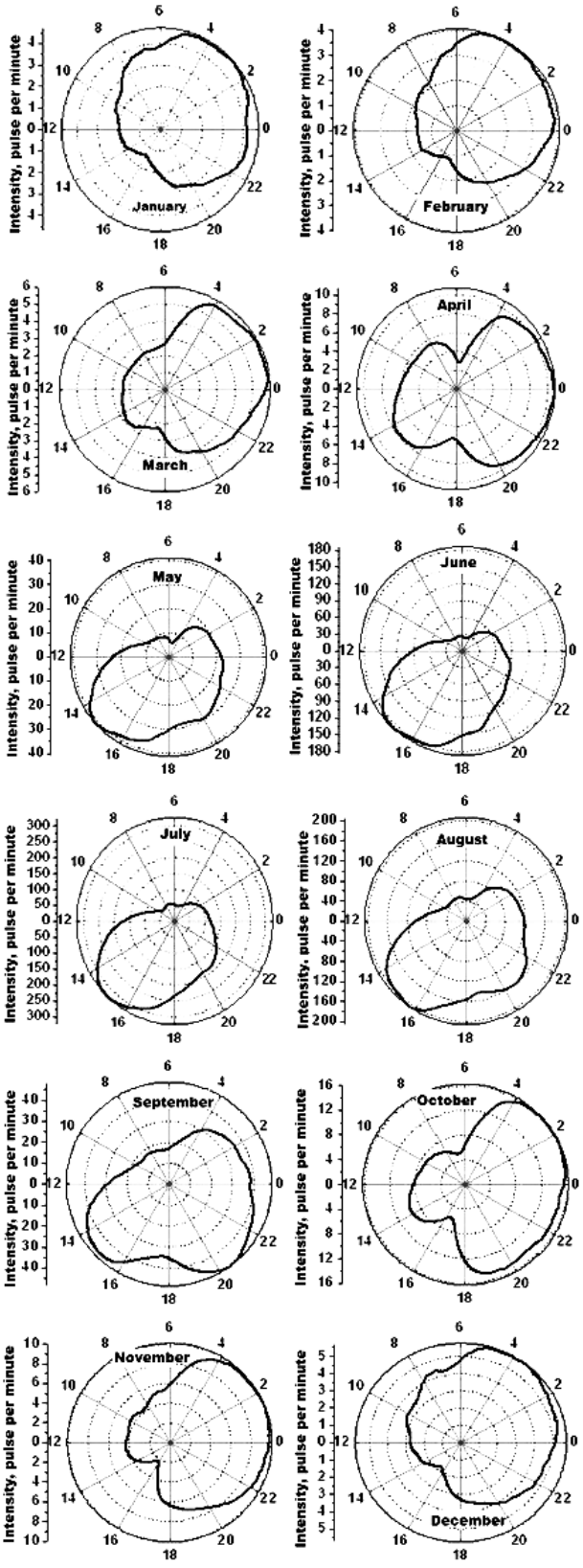

Fig. 2: Average rounded daily variations in NPEMFE in polar coordinates for the period from 1997 to 2004.

These figures show the way the intensity of the electromagnetic field disturbances is changing during the time of day depending on the season. We can see that the intensity is significantly reduced in winter months with its maximum at night, that is when it is day time and summer in the Southern Hemisphere, where the heavy part of the core is, and there are more storms.

It is very sad that such an enormous result obtained by Y.Ṗ. Malyshkov and S. Yu. Malyshkov [3] on these measurements, systematization, analysis and so on cannot be continued because of lack of funding.

It becomes clear how the Earth's magnetic field is formed and why other planets and the Sun have magnetic fields too, if they have moons, or no magnetic fields, if they don't (eg., Venus has a very slow spin — 243 Earth days — that is there are no gravitational forces to create a moving charge), or if the planet cooled down and has no liquid core (Moon), as well as reversal of polarity with reversed rotation of the moon(s) (Mars), and presence of a complex field due to the planet's complex relationship with moons (Uranus and Neptune). It is interesting that Mercury, while having no moons, has a field similar to the Earth's one, though much smaller. However, it itself is a moon of the Sun, and the closest one. It quickly orbits the Sun — in 89 Earth days. Mercury's field is symmetric and directed along the axis of rotation. Its equator is only 0.1 degree tilted to the orbit plane.

A good illustration of the influence of the planet-moon system on a magnetic field's form is a comparison of the fields of Jupiter and Earth. Jupiter's field is more like a flat disk — even most of its moons rotate in correct circular orbits in the equatorial plane — and the axis of rotation of the planet itself is negligibly tilted. There is no change of seasons. On the other hand, the form of the Earth's field resembles an apple, and the planet itself swings relative to the plane of the ecliptic. This can be compared as fields from two different electromagnetic coils — one loop-to-loop wound around the coil-tube and the other being similar to a cassette tape.

Thus, the charges forming the magnetic field of a planet having a liquid core are created and propelled by the total gravitational force from its moons, the Sun, and other planets moving nearby relative to the planet. The charges also influence on the field shape. Of course, MF depends on the distance between the planet and the Sun. Influence of the latter is paramount. For example, as shown by Alexander L. Chizhevsky, "Taking into account the diameter of the Sun equal to 1,390,891 km* and the tremendous power of physical and chemical processes occurring on the Sun, it must be recognized that the Globe is under its enormously intensive influence" [4].

A short comparison of the planets' magnetic fields depending on the number of their moons and other properties is given in Appendix.

The generated pulsating (for a point on the surface) — with a day-and-night period — magnetic field of the Earth is supported by the magnetic properties of the planet's body that smooths and stabilizes its behavior, and sometimes distorts, creating local anomalous areas.

According to the research conducted by Hrvoje Tkalčić, College of Physical and Mathematical Sciences, Australian National University [5], he found that spins of different layers of the Earth are not synchronic. The red-hot core of the Earth inexplicably begins to gain momentum and then slow down, and spins faster or slower than the Earth does. To detect the desynchronization phenomenon, the researchers used a very effort-consuming method of studying double earthquakes, i.e. the earthquakes that occur in the same place at intervals of two weeks to decades. Comparison of seismic waves made it possible to reveal changes in the deep layers of the Earth and learn about changing spin speed of the planet's core.

It is quite hard to measure the spin speed in discrete measurements as, in this case, we need some kind of a marker on the core's surface; all the more so as said that the speed is unstable and variable. We can only determine that there is a position change. If changing the model of the Earth's internal structure, the measured result change too. However, the fact that these changes take place also verifies the hypothesis, and it can broadly explain the physics of motion.

2 Tides

Let's consider the effect of gravitational force in the example of the Earth. The primary influence is caused by the Sun and the Moon. The Sun's influence is (according to various data) 30 to 200 times stronger than the Moon's. However, despite the fact that the Sun's gravitational force is almost 200 times greater for the Globe than the gravitational force of the Moon, the tidal forces generated by the Moon are almost twice as much as generated by the Sun. This is due to the fact that the tidal forces do not depend on the magnitude of the gravitational field and its degree of heterogeneity. With increasing distance from the source of the field, heterogeneity decreases more rapidly than the size of the field itself. Since the Sun is almost 400 times farther from Earth than the Moon, the tidal forces caused by the solar gravitation are weaker.

In other words we can say that the tidal force of the Moon is more "superficial", local, and more affecting the ocean and the upper mantle, whereas the solar gravity is more uniform, affecting the whole body of the planet. The solar gravity can be considered roughly equal anywhere on the Earth. It is the solar gravity that makes the core move and separate into charges. Naturally, this mechanism will slightly vary for other planets, but the physics of the phenomenon is the same.

With spin of the Earth, these two forces are added and the tidal wave, which has the shape of an ellipsoid, is a superposition of two *double-humped waves*, formed as a result

*According to recent data, the Sun's diameter is 1,392,000 kilometers, while the Earth is located at 107 Sun diameters from the Sun.

of gravitational interaction of the Earth-Moon planetary pair and the gravitational interaction of the pair with the central luminary — the Sun.

Thus, the words *lunar tide* hereinafter mean a tide caused by the cumulative influence of the Sun and Moon on the body of the planet.

In addition to the tides on the Earth's side facing the Moon, there are tides on the other side. They are about the same in magnitude. In literary sources, the existence of this phenomenon is explained by reduced gravity of the Moon and the centrifugal forces created by rotation of the Earth-Moon pair. But then there would be a tide on the other side of the Moon too, and this would happen there all the time, especially as the Moon moves at the larger distance from the center of mass than the other side of the Earth does. We know about the shifting center of mass and elongation of the Moon towards the Earth, but there are no tides on the far side. In addition, as it was said above, the tides are caused not only by the Moon, but by the Sun and the Moon together, so we have to find now the center of mass for three planets.

If we compare the forces affecting the Earth's surface in low-tide areas (Point 2) and high-tide areas of the *dark* side of the Earth (Point 1), the gravity forces in the *dark* should be stronger, as the gravity of the Earth's center is added (though weakened) the gravity of the Moon and the Sun. This means that the sea level in Point 1 should be lower than the sea level at low tide in Point 2, but it is actually almost the same as it is in Point 3. How else can it be explained?

Following the hypothesis, we can assume that the heaviest part of the Earth's core following the Moon and the Sun is displaced so far from the opposite edge of the Earth, that the square of the distance has its effect, and the gravity force of the core on the surface is weakened thus causing a tidal effect. In other words, the force of gravity at the point on the Earth depends not only on the position of the Moon and the Sun, but also the center of mass of the Earth (see Fig. 3 and Fig. 4).

Apparently, these processes occurred on the Moon too. When cooling, the heavy mass of the inner substance clustered mostly in the side of the planet facing the Earth, thus making the Moon a kind of Roly-Poly and forcing it to turn the same heavy side to us.

This is also confirmed by the fact that earlier, as it is known, it had a strong magnetic field which now exists only in residual form.

Thus, the force of the Earth's gravity (together with the Moon's gravity force) not only holds the Moon in the moon orbit, but also makes it spin thus requiring energy. Perhaps this interaction further heats the inner substance of the planet, preventing it from cooling down. This can refute the theory of a thermonuclear source maintaining the planet's core in a "warm" state. Otherwise, at least we would have long been bald.

The same core makes the Earth to "bulge" at the equator, giving it a form other than a sphere. The same bulging is a

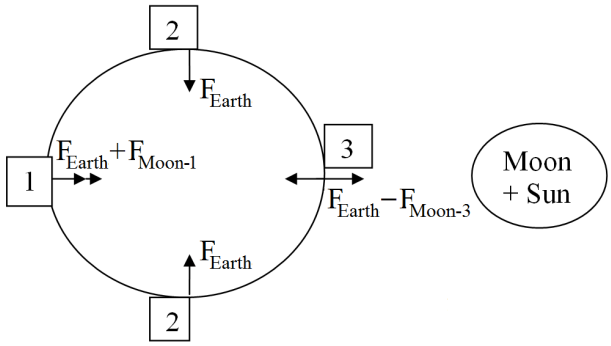

Fig. 3: The forces affecting the points on the Earth's surface with uniform mass distribution.

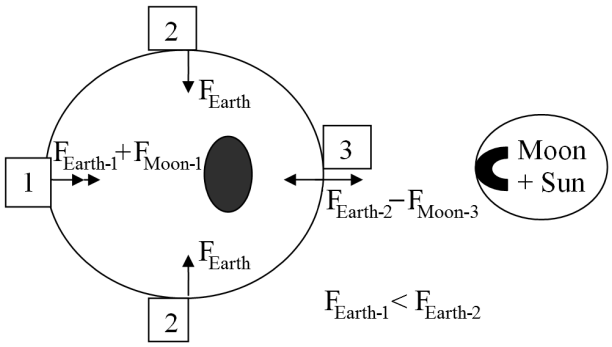

Fig. 4: The forces affecting the points on the Earth's surface with the shifted center.

characteristic of Jupiter with its high speed of spin, where this is further contributed by centrifugal forces.

A similar phenomenon seems to be happening with the Sun and its moon-planets.

If we imagine that the "heavy" center of the Sun following the moon-planets "floats" on the surface with a strong gravitational pull of planets, is charged with the electric potential, and is in motion, this may cause *magnetic flux tubes* on the surface, i.e. output points of the both poles of the magnetic field.

Over many years of research on the impact of solar activity on the biosphere, Chizhevsky has clearly shown a direct relationship of these processes, assuming that the perturbations observed as *sunspots* are causing radiation that reaches the Earth's surface and penetrates into it affecting all the living and non-living things [5]. The proposed hypothesis can explain the appearance of wide-frequency-range electromagnetic radiation as a result of abruptly changing fluxes of charged solar material.

3 Currents

Literature sources used to explain the nature of the equatorial currents by the winds constantly blowing in the same direction, while the nature of the winds was explained by surface heating and spinning of the Earth. Of course, this does affect

the ocean and the air masses too, but, in my opinion, they are primarily influenced by the gravity force from moving *Earth core — the Moon* and *Earth core — the Sun* pairs affecting everything that gets between them and that is carried from East to West by their gravity force. It should not be regarded as a process with tight fixing. It is more similar to stirring a teaspoon in a large pot in the same direction — not hard, but for a long time.

4 Earthquakes

There is still no clear definition of the nature of earthquakes. It is quite possible that it may look as follows: Employ your imagination — Where will a body located at the center of the planet gravitate at the slightest deviation from the center?

If a substance is distributed unevenly (assume that it is denser to the center), it is just like as written in textbooks. But what forces draw it in the center? It should be a substance having infinite density. It sounds more like fiction.

If the Earth had the form of an empty sphere, there would be no gravity force inside it. The point inside the Earth would be influenced by the gravity forces of external bodies — the Moon, the Sun, etc. This point would tend to follow the direction of the sum vector of the forces of these bodies.

If the Earth had uniform distribution of substance by density, then (if the substance is liquid) it would be the same.

In both cases, the substance inside the hard shell will gravitate to the shell from the inside toward the outside forces from other planets.

All the above was said without taking pressure into account, but let's consider the pressure's behavior upon submersion — naturally it increases in the beginning (as the mass "over the head" increases), but further on the gravity force decreases and the pressure gradually "stabilizes". In the end we have a closed space with approximately even pressure throughout volume, and its influence may be small compared to the gravity forces. It is just the same as in ordinary life — the atmospheric column presses down on all of us, but it still lets the gravity forces to drop an apple on the ground.

It turns out that the interior of the Earth can be similar in structure to a chicken egg and have the same distribution of substance by density as it is on the surface — solid-liquid — and all these at enormous pressure and temperature.

Now, if we imagine, the glowing mass exposed to various — addable or deductible — gravitational forces from various planets is moving in the "inner" surface of the earth, constantly blending and running into irregularities. At the same time, the interior of the Earth's shell is constantly exposed to momentum which is transmitted to the tectonic plates, forcing them to move gradually, thereby moving the continents. This is confirmed by the fact that the continents are moving in the latitudinal direction (East-West) and do not move in the longitudinal one (South-North).

Sometimes the forces are added in such a way that parts of the core get into the central zero-gravity zone and, after breaking away from the bulk mass, "fall" on the opposite side of the sphere, which might cause an earthquake. A very good illustration of such a case is behavior of water in a zero-gravity environment shot by US astronauts. Behavior of water balls in a "bubble" could well be similar to that of the inner core of the planet.

By the way, the zero-gravity zone is not fixed in a permanent place, but is following the main mass of the core in rough circles.

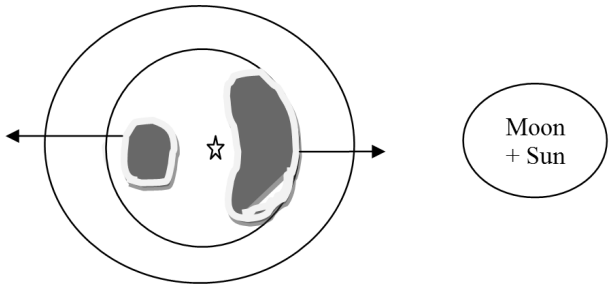

Fig. 5: A part of the core falls on the opposite side of the Earth's shell after it has separated and moved to a gravity-equilibrium zone.

There may also occur a sort of a wave with a crest when climbing an inner roughness, with a further collapse, which may also cause an earthquake.

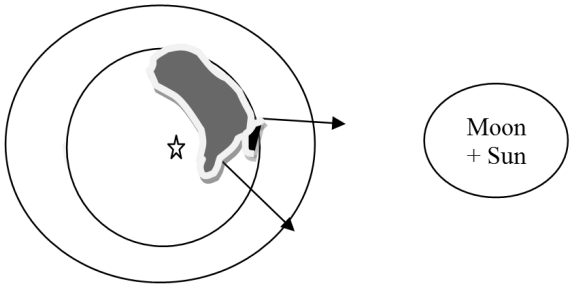

Fig. 6: Collapse of a core part.

This mechanism of earthquakes may be even more likely, since the majority of seismic focuses are located at the boundaries of tectonic plates or in areas of geological irregularities.

These two phenomena can cause shifts in the surface layers of the mantle triggering creation of additional seismic focuses and aftershocks.

It should be also noted that, as is known, magnetic storms on the Earth are accompanied with low-frequency vibrations of the Earth's body, and vice versa, earthquakes are accompanied by electromagnetic radiation, i.e. these two phenomena are interrelated. This can also serve as a verification of the suggested hypothesis, as there are surges of electric charge (current), and the transition process (as we know) has a wider range than direct current.

5 Time jumps and killer waves

With the advent of new, more precise time measuring means, it was observed that sometimes the celestial (stellar) time flows changing relative to the reference atomic one in jumps*. How can this be explained but through the Earth being exposed to forces, turning it at a certain angle? We see no external forces of such a power, so we have internal ones left.

It is quite possible that, when running into an internal "roughness", the core "pushes" the main body of the planet, altering astronomical time relative to the stable reference one.

Mariners now a natural phenomenon known as the "killer wave" (also known as periodic wave, monster wave, rogue wave, freak wave, onde scelerate, or galejade). Some ten to fifteen years ago, scientists believed that seafarers' stories about giant killer waves that emerged from nowhere and took down ships were nothing but maritime folklore.

The existence of sea waves twenty to thirty meters high contradicts the laws of physics and does not fit into any mathematical model of formation of waves. It should be noted that these waves appear on relatively calm water surface. They can be a crest or a trough, single one or coming in a set.

The proposed hypothesis can logically explain the mechanism of their occurrence through the same interactions between the moving core and the internal irregularities of the planet's body, which are carried over to the sea surface.

6 Causes of a dip appearing in the gravity graph during the Sun's passage across the sky

Following the work with a new directional gravimeter, Evgeny Orlov presented some interesting data. As shown in his article [6], round-the-clock registration of gravimeter readings made it possible to determine the original geometrical shape of the solar gravitational signal (see Fig. 7).

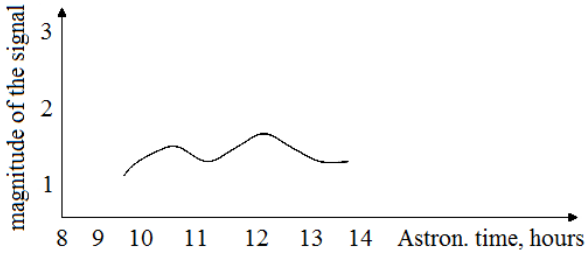

Fig. 7: The original geometrical shape of the solar gravitational signal as registered by Orlov [6].

It is registered in the daytime, in the form of double-humped curve with a dip in the range from 11 a.m. to 01 p.m., so the dip comes where the Sun would draw the load the hardest. The author of the article explains this by the fact that the volume of the gravitating mass of the planet facing the Sun on both sides of the planet exceeds the gravitating mass at its center. However, in my opinion, it is determined by the fact that the hardest part of the core comes closer to the Earth's surface and the distance to the measuring part of the gravimeter is reduced, thereby increasing the gravity to the Earth and compensating the gravity to the Sun.

*Please do not confuse it with a correction of calendar time.

7 On motion of the magnetic poles

It also turns out that the outer shell of the Earth is weakly related to the processes taking place between the planets causing appearance of a magnetic field, and therefore is "free" to move relative to the center of mass (it is similar to rotation of the outer rim of a bearing with internal one being fixed), while changing the position of the magnetic poles on the surface of the Earth, but without changing the position in space. At the same time, the position of the outer sphere of the Earth depends on the interaction strength of the core magnetic field and the magnetic properties of the sphere itself, which, among other things, may be affected by anthropogenic factors. A shift occurs before the mantle comes into one of the local stability points. It does not have to be a complete polarity reversal.

8 Conclusion

The suggested hypothesis is not loaded with mathematical calculations for yet for a number of reasons, including the following:

1. There are too many factors affecting the field;
2. One can always bring math under any theory by introducing correction factors and hiding the lack of physics of the phenomenon.

Of course, this hypothesis is presented in yet "unfledged" form and requires much to be done to verify and expand understanding of the physics of the processes.

Submitted on November 24, 2015 / Accepted on December 5, 2015

References

1. Iosilevskiy I.L. Plasma polarization in massive astrophysical objects. *Journal of Physics A: Mmathematical and Theoretical*, 2009, v. 42, 214008.
2. Iosilevskiy I.L. Plasma polarization in high gravity astrophysical objects. *Contrib. Plasma Phys.*, 2009, v. 49, no. 10, 755–759.
3. Malyshkov Yu.P., Malyshkov S.Yu. Periodic variations in geophysical fields and seismic activity, and their possible connection with motion of the Earth's core. *Geology & Geophysics*, 2009, v. 50, no. 2, 152–172 (in Russian).
4. Chizhevsky A. Physical Factors of the Historical Process. *Cycles*, Special Issue, January 1971, 11–27.
5. Tkalčić H., Young M.K., Bodin T., Ngo S., and Sambridge M., The shuffling rotation of the Earth's inner core. *Nature Geoscience*, 2013, v. 6, 497–502.
6. Orlov E.F. Experimental propagation speed of the gravitational interaction. Part 1. *Researches in Science*, May 2013, no. 5 (in Russian, accessed from http://science.snauka.ru)

Appendix. A short comparison of the planets' magnetic fields depending on the number of their moons and other properties

Planet	Moons	Magnetic field
Mercury	No	One percent of the Earth's field; of dipole-type, directed along the axis of rotation which is perpendicular to the orbit plane.
Comment: The intensity of Mercury's magnetic field is 100 times smaller than that of the Earth. Mercury's magnetic field has a dipole structure and is highly symmetrical. Its axis is only two degrees tilted from the spin axis of the planet.		
Venus	No	Almost absent: the planet's spin is very slow.
Comment: Since the planet's own magnetic field is absent, it should be assumed that there is no motion of charged particles — electric current — in its iron core that could cause a magnetic field. Therefore, the core substance does not move.		
Mars	2	The planet's magnetic field is 500 times weaker than the Earth's one. The field's polarity is reverse to that of the Earth. Phobos rises in the West and goes down in the East. Its size is very small. The influence of Deimos is weaker because of its remoteness.
Comment: Mars has a magnetic field, but it is weak and extremely unstable. In various parts of the planet, its intensity may vary from 1.5 to 2 times. Its magnetic poles do not coincide with physical ones.		
Jupiter	17 + ring	Twenty times as strong as the Earth's. The polarity is reverse to that of the Earth.
Comment: Jupiter's moon system consists of at least 67 moons, including four large moons. Jupiter has a strong magnetic field. The dipole axis is tilted to the axis of rotation at 10°. Its polarity is reverse to the polarity of Earth's magnetic field. All the major moons of Jupiter rotate synchronously and always keep the same face turned to Jupiter due to the influence of powerful tidal forces of the giant planet. Jupiter's rotation speed is so high that the planet bulges along the equator.		
Saturn	18 + ring	Almost equal to the Earth's and reverse in direction.
Comment: By its strength, Saturn's magnetic field is in the middle between the magnetic field of the Earth and the more powerful field of Jupiter. The magnetic field is nearly a dipole, similar to that of the Earth, with north and south magnetic poles. The north magnetic pole is located in the northern hemisphere, and the south one is in the South, unlike Earth, where the location of the geographic poles is reverse to that of magnetic ones. Saturn has 62 known moons. Most of the moons, except Hyperion and Phoebe, spin synchronously — they always keep the same side turned to Saturn.		
Uranus	21 + ring	Less than that of the Earth and has axial tilt at 60 degrees. The polarity is reverse to the Earth's. Uranus rotates reversely. The moons rotate reversely too. The moons' orbits are steeply tilted to the ecliptic.
Comment: The equatorial plane of Uranus is tilted to the plane of its orbit at an angle of 97.86° — that is, the planet rotates "lying on its side." This gives the season changing process completely different from the other planets of the Solar System. If other planets may be compared to a spinning top, Uranus is more like a rolling ball. Uranus has a very specific magnetic field that is not directed from the geometric center of the planet, but is tilted towards the axis of rotation by 59 degrees. In fact, the magnetic dipole is shifted from the center to the south pole of the planet about one third of the planet's radius. This unusual geometry results in a very asymmetric magnetic field.		
Neptune	8	A complex magnetic field
Comment: Neptune resembles Uranus in its magnetosphere, with a magnetic field strongly tilted relative to its rotational axis at 47°. Neptune has 13 known moons. Triton is the largest Neptunian moon, comprising more than 99.5% of the mass in orbit around Neptune, and it is the only one massive enough to be spheroidal. Unlike all other large planetary moons in the Solar System, Triton has a retrograde orbit. It is close enough to Neptune to be locked into a synchronous rotation, and it is slowly spiraling inward because of tidal acceleration.		

A Classical Model of the Photon

Shixing Weng

11 Metzak Dr., Brampton, Ontario L6Z 4N3, Canada
E-mail: wengs2015@gmail.com

A desired solution of the four-potential is presented for free-space photons, obtained with wave equations derived from the Maxwell equations and the Lorenz condition. The solution shows that an electromagnetic field in wave form propagating at the speed of light with a fixed internal phase may exist as a particle taking a limited space at a specific point of time. It reveals the existence of electric charge distributed as an electric capacitor on the parallel cylindrical surface of constant radius to the central axis of the solution. And the charge distribution has a phase change both in the azimuthal angle and along the direction of the wave propagation. The solution is applied to the case of a model photon to determine several parameter values of the solution, which in turn provides a view on the model photon.

1 Introduction

The year of 2015 has been the International Year of Light and Light-Based Technologies, designated by the United Nations Educational, Scientific, Cultural Organization (UNESCO). This designation further emphasizes the importance of light to people's life. As a part of the support for the designation, we present in this paper a theoretical model for the elements of light, photons, based on our knowledge of classical electrodynamics, classical mechanics and mathematical method for quantization rules.

In this paper we consider a single free photon in which photon-photon interactions [1] are neglected. A photon [2] is a quantum of light which is a wave form of the electromagnetic radiation and is characterized by its speed c and wavelength λ. It is known that a photon has both physical properties of wave and particle.

As a particle, the photon has a certain energy and momentum. In the study of the black body radiation [3], Max Planck proposed that the energy ϵ of a radiation oscillator was quantized and each energy was proportional to its vibrational frequency ν as

$$\epsilon = h\nu\,, \tag{1}$$

where h is the Planck constant. Then Einstein applied the idea to the light and proposed that light was made of quanta, inseparable entities, with the energy ϵ in terms of the frequency being given in Eq. (1), which successfully explained the photoelectric effect [4].

The Compton Scattering Experiment [5] further demonstrated that a photon had a certain energy as specified in Eq. (1) as well as a momentum in the direction of its motion. And the magnitude of the momentum p is given by

$$p = \frac{\epsilon}{c} = \frac{h\nu}{c} = \frac{h}{\lambda}\,, \tag{2}$$

where the relation $\nu = c/\lambda$ is used.

Furthermore it is known from quantum mechanics [6], that there is an angular momentum difference involved in the magnitude of integral $\hbar$ between the two transitional atomic states, where $\hbar$ is the reduced Planck constant which equals to the Planck constant h divided by 2π. In the case of light emission this angular momentum difference may be transfered to the photon.

On the other hand the Young's two slit experiment [7] shows the wave property of light. In a typical Young's experiment one observes the interference pattern of light from a monochromatic light source of wavelength λ passing through two small-spaced parallel slits, which demonstrates the wave property of light.

It is also known that light is a form of the electromagnetic wave. In the electromagnetism [8], the set of Maxwell equations for vacuum gives relationships among the electric field $\mathbf{E}$, magnetic field $\mathbf{B}$, electric charge density ρ, and electric current density $\mathbf{J}$ as following:

$$\nabla\cdot\mathbf{E} = \frac{\rho}{\epsilon_0}\,, \tag{3}$$

$$\nabla\cdot\mathbf{B} = 0\,, \tag{4}$$

$$\nabla\times\mathbf{E} + \frac{\partial\mathbf{B}}{\partial t} = 0\,, \tag{5}$$

$$\nabla\times\mathbf{B} - \frac{1}{c^2}\frac{\partial\mathbf{E}}{\partial t} = \mu_0\mathbf{J}\,, \tag{6}$$

where ϵ_0 is the permittivity of vacuum and μ_0 is the permeability of vacuum; ∇ represents the differential operator and $\nabla = \hat{\mathbf{i}}\frac{\partial}{\partial x} + \hat{\mathbf{j}}\frac{\partial}{\partial y} + \hat{\mathbf{k}}\frac{\partial}{\partial z}$ in Cartesian coordinates with $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, $\hat{\mathbf{k}}$ being unit vectors for the Cartesian coordinates; t represents the time and x, y, z are, respectively, the Cartesian components; the "$\times$" symbol represents the cross operation and the "$\cdot$" represents the dot operation. In this paper we use SI units. And for simplicity we shall consider in the following the medium to be vacuum. For vacuum where $\rho = 0$ and $\mathbf{J} = 0$, the following equations may be obtained for the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$ from Eqs. (3) to (6),

$$\frac{1}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} - \nabla^2\mathbf{E} = 0\,, \tag{7}$$

$$\frac{1}{c^2}\frac{\partial^2 \mathbf{B}}{\partial t^2} - \nabla^2 \mathbf{B} = 0\,, \tag{8}$$

where c is the speed of light, which is equal to $1/\sqrt{\epsilon_0\mu_0}$ for vacuum, and ∇^2 is the Laplacian operator. Eqs. (7) and (8) are wave equations with the propagation speed equal to the speed of light, which shows the light to be a form of the electromagnetic wave. But we believe that the achieved solution from Eqs. (7) and (8) so far for free-space photon is limited to one-dimension and our current view on the photon is limited.

As we know that an electric field or a magnetic field has energy. And the total energy density η is equal to the sum of the electric field energy density $\eta_{\mathbf{E}}$ and the magnetic field energy density $\eta_{\mathbf{B}}$ and is given by

$$\eta = \eta_{\mathbf{E}} + \eta_{\mathbf{B}} = \frac{1}{2}\,\epsilon_0 |\mathbf{E}|^2 + \frac{1}{2\mu_0}|\mathbf{B}|^2\,, \tag{9}$$

where $|\mathbf{E}|$ is the magnitude of the electric field and $|\mathbf{B}|$ the magnitude of the magnetic field.

The Poynting vector $\mathbf{S}$, which is the energy current density of the electromagnetic wave, is given by

$$\mathbf{S} = \frac{1}{\mu_0}\,\mathbf{E}\times\mathbf{B}\,. \tag{10}$$

The Poynting vector is perpendicular to both $\mathbf{E}$ and $\mathbf{B}$ vectors and is in the direction of the thumb while using the right-hand-rule turning fingers from $\mathbf{E}$ to $\mathbf{B}$.

Both the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$ can be expressed in terms of the four-potential, a scalar electric potential ψ plus a magnetic vector potential $\mathbf{A}$ as following,

$$\mathbf{B} = \nabla\times\mathbf{A}\,, \tag{11}$$

$$\mathbf{E} = -\nabla\psi - \frac{\partial \mathbf{A}}{\partial t}\,. \tag{12}$$

The Lorenz condition [9], named after the Danish mathematician and physicist, L. V. Lorenz, provides a covariant form of the four-potential and is given by

$$\nabla\cdot\mathbf{A} + \frac{1}{c^2}\frac{\partial\psi}{\partial t} = 0\,. \tag{13}$$

Eq. (13) appears similar to the continuity equation and may represent a "local form" of the conservation of electric potential energy for a point charge in the electromagnetic field. With the Lorenz condition, both the scalar potential ψ and the vector potential $\mathbf{A}$ satisfy the following equations, respectively,

$$\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} - \nabla^2\psi = \frac{\rho}{\epsilon_0}\,, \tag{14}$$

$$\frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla^2\mathbf{A} = \mu_0\mathbf{J}\,. \tag{15}$$

The purpose of the paper is to present a model view of the photon by obtaining a three-dimensional solution from Eqs. (14) and (15) for vacuum without external electric charge nor external electric current. The three-dimensional solution hence is theoretical analyzed to reveal its physics meaning. It is finally applied to the case of a model photon to gain a deep insight into the photon, which is new since we are not aware of such a report in the literatures.

This paper is organized as these: Introduction, Solution, Discussions, and Conclusion. The Introduction section provides a brief overview on our fundamental understandings of light and photon. In the Solution section, two expressions of the four-potential as a solution for three-dimensional space are presented, which are obtained from Eqs. (14) and (15) for vacuum without external electric charge nor external electric current. The characteristic of the solution shows that its quantities are in limited space at a specific point of time, which is desirable for photons. In the Discussions section, expressions for the electric field and the magnetic field are derived from the four-potential solution. An analysis of the electric field reveals the existence of electric charge distributed on the parallel cylindrical surface of constant radius to the central axis of the solution. The solution then is applied to the case of a model photon to determine the constant parameter values of the solution from physical quantities of the photon, which in turn provides a view on the model photon. The Conclusion section provides a brief summary of the paper together with some comments.

2 Solution

In vacuum where electric charge density $\rho = 0$ and electric current density $\mathbf{J} = 0$, Eqs. (14) and (15) are reduced respectively to

$$\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} - \nabla^2\psi = 0\,, \tag{16}$$

$$\frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla^2\mathbf{A} = 0\,. \tag{17}$$

Eqs. (16) and (17) are wave equations and their solutions for one-dimensional space are easily obtained and are known as a traveling wave,

$$\psi = \psi_0 \sin(kx - \omega t)\,, \tag{18}$$

$$A = A_0 \sin(kx - \omega t)\,, \tag{19}$$

where ψ_0 represents the amplitude of the scalar potential, A_0 the amplitude of the vector potential, ω is the angular frequency which equals to $2\pi\nu$ and ν is the wave frequency, and k is the wavenumber and $k = \omega/c = 2\pi/\lambda$. The reason to choose the sine function instead of the cosine function here is arbitrary, but with no difference, since the sine and cosine functions are different by a phase difference of $\pi/2$, they may represent the same physical wave. Also as we know that the electric potential is a measurable quantity which is real, we shall restrict the solution to the real number domain in this paper.

In the following, Eqs. (16) and (17) are solved for three-dimensional space to reveal more features of the solution. First we choose the circular cylindrical coordinates (or cylindrical polar coordinates) as in Fig. 1 for our coordinate system [10]. Here we use the r symbol to represent the polar axis since the ρ symbol is used for the electric charge density. And ϕ represents the azimuthal angle and z represents the central axis and is the same as the Cartesian z axis. Their respective unit vectors are $\hat{\mathbf{r}}$, $\hat{\phi}$, and $\hat{\mathbf{z}}$ as in Fig. 1.

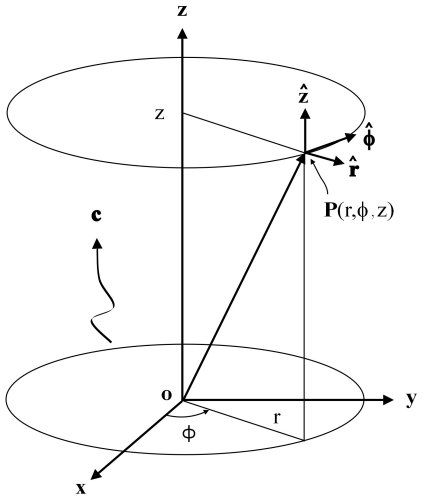

Fig. 1: A drawing of the circular cylindrical coordinate system with respect to the Cartesian coordinates, where $\hat{\mathbf{r}}$, $\hat{\phi}$, and $\hat{\mathbf{z}}$ are unit vectors for the coordinate system. The wave symbol represents a photon moving in the direction of the positive z axis at the speed of light c.

The Laplacian operator ∇^2 in the cylindrical coordinates is expressed as

$$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2}, \tag{20}$$

and hence we get a solution of the four-potential from Eqs. (16) and (17) as following

$$\psi = \psi_0 \sin(kz + m\phi - \omega t) \begin{cases} \left(\dfrac{r}{r_0}\right)^m & r < r_0\,, \\ \left(\dfrac{r_0}{r}\right)^m & r > r_0\,, \end{cases} \tag{21}$$

$$\mathbf{A} = \hat{\mathbf{z}} A_0 \sin(kz + m\phi - \omega t) \begin{cases} \left(\dfrac{r}{r_0}\right)^m & r < r_0\,, \\ \left(\dfrac{r_0}{r}\right)^m & r > r_0\,, \end{cases} \tag{22}$$

where we choose the wave to propagate along the positive z axis, ψ_0 is a strength constant for the scalar potential and A_0 is a strength constant for the vector potential whose direction is in that of the wave propagation, r_0 is a constant polar radius to be determined in the next section by the wavelength of the photon, m is a positive integer to satisfy the 2π periodic boundary condition of the azimuthal angle. Here m is a quantum number which may be associated with the angular momentum of the wave. Again the choice of the sine function instead of the cosine function here is arbitrary but has no physics difference. The solution at r_0 is not defined but has finite quantities. r_0 is a boundary of the solution and in the following treatment we shall let the boundary thickness to approach to zero so the solution is approximately defined at r_0.

Eqs. (21) and (22) represent a traveling wave propagating along the positive z axis. The solution by the two expressions is desirable since its quantities are limited in the polar axis. It is worthwhile to mention that this solution may be for individual photons free from interactions with each other. The study of photon interactions is out of the scope of this paper. In the following section we will analyze the solution to reveal its physics meaning.

3 Discussions

Applying the Lorenz condition, Eq. (13), to Eqs. (21) and (22), we have

$$A_0 = \frac{\psi_0}{c}\,. \tag{23}$$

Hence the vector potential and the scalar potential are related to each other, only one of them is independent.

Now applying Eqs. (11) and (12) to the solution Eqs. (21) and (22) and using Eq. (23), we may have for the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$ as following:

$$\mathbf{E} = -m\psi_0 \begin{cases} \dfrac{r^{m-1}}{{r_0}^m}\Big(\hat{\mathbf{r}}\sin(kz + m\phi - \omega t) + \\ + \hat{\phi}\cos(kz + m\phi - \omega t)\Big) & r < r_0\,, \\ \dfrac{{r_0}^m}{r^{m+1}}\Big(-\hat{\mathbf{r}}\sin(kz + m\phi - \omega t) + \\ + \hat{\phi}\cos(kz + m\phi - \omega t)\Big) & r > r_0\,, \end{cases} \tag{24}$$

$$\mathbf{B} = mA_0 \begin{cases} \dfrac{r^{m-1}}{{r_0}^m}\Big(\hat{\mathbf{r}}\cos(kz + m\phi - \omega t) - \\ - \hat{\phi}\sin(kz + m\phi - \omega t)\Big) & r < r_0\,, \\ \dfrac{{r_0}^m}{r^{m+1}}\Big(\hat{\mathbf{r}}\cos(kz + m\phi - \omega t) + \\ + \hat{\phi}\sin(kz + m\phi - \omega t)\Big) & r > r_0\,, \end{cases} \tag{25}$$

where $\hat{\mathbf{r}}$ is the unit vector for the polar axis, $\hat{\phi}$ is the unit vector for the azimuthal angle. From Eqs. (24) and (25) we know that both the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$ are traveling in the direction of the positive z axis and are perpendicular to the direction of the wave propagation. Furthermore we have $\mathbf{E} \cdot \mathbf{B} = 0$, meaning that the electric field and the magnetic field are perpendicular to each other, which is consistent with the basic electromagnetic theory for free-space.

For better understanding of the fields, in the following discussions we shall restrict ourself to the case of the angular momentum number $m = 1$, which may correspond to the case of the photon we know. For general case of $m > 1$, following treatments are similarly applicable. Hence Eqs. (24) and (25) become

$$\mathbf{E} = -\psi_0 \begin{cases} \frac{1}{r_0}\Big(\hat{\mathbf{r}}\sin(kz+\phi-\omega t)+ \\ +\hat{\phi}\cos(kz+\phi-\omega t)\Big) & r < r_0\,, \\ \frac{r_0}{r^2}\Big(-\hat{\mathbf{r}}\sin(kz+\phi-\omega t)+ \\ +\hat{\phi}\cos(kz+\phi-\omega t)\Big) & r > r_0\,, \end{cases} \tag{26}$$

$$\mathbf{B} = A_0 \begin{cases} \frac{1}{r_0}\Big(\hat{\mathbf{r}}\cos(kz+m\phi-\omega t)- \\ -\hat{\phi}\sin(kz+m\phi-\omega t)\Big) & r < r_0\,, \\ \frac{r_0}{r^2}\Big(\hat{\mathbf{r}}\cos(kz+m\phi-\omega t)+ \\ +\hat{\phi}\sin(kz+m\phi-\omega t)\Big) & r > r_0\,. \end{cases} \tag{27}$$

From Eqs. (26) and (27), for $r > r_0$ both field strengths are inversely proportional to r^2 and approach to zero as r goes to infinity, which is a desirable result because a photon takes a limited space at a specific point of time. The electric field $\mathbf{E}$ at r_0, or on the parallel cylindrical surface in a three-dimensional view, is not continue in the radial direction, meaning charge may exist on the surface. To derive an expression for the surface charge density σ, apply Eq. (3) to Eq. (26), we have

$$\sigma = 2\epsilon_0\psi_0\frac{1}{r_0}\sin(kz+\phi-\omega t)\,. \tag{28}$$

Hence the charge density is also in the form of a traveling wave, moving uniformly in the direction of the positive z axis with a fixed internal phase both in the azimuthal angle and along the z axis.

To get a precise sense of the fields and charge distribution, we simplify Eqs. (26), (27), and (28) by letting $z = 0$, and $t = 0$, which allows us to better understand the solution at the specific point of time and space. And hence we have

$$\mathbf{E} = \psi_0 \begin{cases} -\frac{1}{r_0}\hat{\mathbf{j}} & r < r_0\,, \\ \frac{r_0}{r^2}\Big(\hat{\mathbf{i}}\sin(2\phi)-\hat{\mathbf{j}}\cos(2\phi)\Big) & r > r_0\,, \end{cases} \tag{29}$$

$$\mathbf{B} = A_0 \begin{cases} \frac{1}{r_0}\hat{\mathbf{i}} & r < r_0\,, \\ \frac{r_0}{r^2}\Big(\hat{\mathbf{i}}\cos(2\phi)+\hat{\mathbf{j}}\sin(2\phi)\Big) & r > r_0\,, \end{cases} \tag{30}$$

and

$$\sigma = 2\epsilon_0\psi_0\frac{1}{r_0}\sin\phi\,, \tag{31}$$

where $\hat{\mathbf{i}}$ is the unit vector for the x axis and $\hat{\mathbf{j}}$ is the unit vector for the y axis. In deriving Eqs. (29) and (30), we use the following relations for unit vector transformations between the polar and Cartesian coordinates

$$\hat{\mathbf{r}} = \hat{\mathbf{i}}\cos\phi + \hat{\mathbf{j}}\sin\phi\,, \tag{32}$$

$$\hat{\phi} = -\hat{\mathbf{i}}\sin\phi + \hat{\mathbf{j}}\cos\phi\,. \tag{33}$$

The electric field $\mathbf{E}$, magnetic field $\mathbf{B}$, and the surface charge density σ at $z = 0$ and $t = 0$ are shown in Fig. 2.

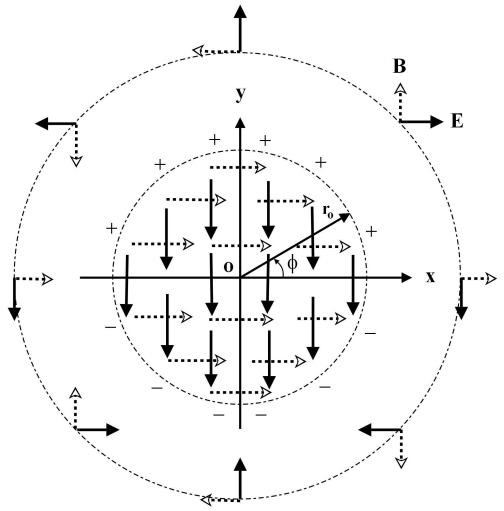

Fig. 2: A schematic diagram of the electric field $\mathbf{E}$ (solid lines), magnetic field $\mathbf{B}$ (dash lines), and charge distribution ("+" for positive charge and "−" for negative charge) on an imaging cylindrical surface ($r = r_0$) of the solution in the x-y plane, where $z = 0, t = 0$. The wave is propagating along the positive z axis (pointing out of the x-y plane). r_0 is the constant radius, and ϕ is the azimuthal angle.

As we know from Eqs. (29) and (30), both the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$ are constant inside of the circle r_0; For outside of the r_0 both fields decreases as the radius squared, r^2, increases, and the field direction changes two times as fast as the azimuthal angle ϕ (Fig. 2). The distribution of the surface charge density σ is described by the sine function of the azimuthal angle, and the total charge by the r_0 circle is zero. Referring to Fig. 2, the charge distribution is polarized, i.e., the positive charge on its corresponding half-circle at r_0 is distributed symmetrically to the negative charge on the other half-circle, or vice versa. The total charge distribution appears as an electric capacitor made of circularly distributed electric dipoles.

In the following discussions we apply the solution to a model photon and shall use the physical quantities of the photon to determine the values of the constants used in the solution.

For $z \neq 0$ and $t = 0$ the electric field $\mathbf{E}$, the magnetic field $\mathbf{B}$ and the surface electric charge density σ are distributed around the central axis z with a certain phase. And the phase

change depends on both the azimuthal angle ϕ and the z axis. We show the charge distribution for $z < 0$ and $t = 0$ in Fig. 3.

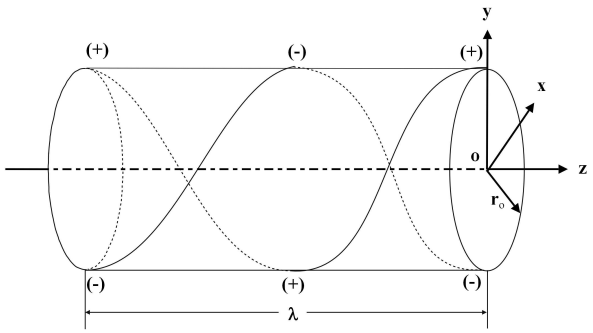

Fig. 3: A schematic diagram showing the surface charge distribution("+" for positive charge and "−" for negative charge) on the surface of $r = r_0$ in the z axis direction for one wavelength λ, where $t = 0$, the model photon is moving along the positive z axis at the speed of light c and r_0 is the constant radius. For clarity we only show two lines of charges here.

The charge distribution appears as a circularly distributed electric dipole "twisted" in the azimuthal angle and along the z axis. The twisting phase change is exactly the same as that of the photon (one cycle of the charge phase change by one wavelength λ). The model photon picture in Fig. 3 represents a "frozen" view at $t = 0$. For $t \neq 0$, by the phase analysis of the sine wave (Eq. (28)), the model photon is doing a displacement along the positive z without changing its internal phase. Now imaging that if we place an observer facing the incoming photon at a fixed z position, it may see the circularly distributed charge rotating counter-clockwise (in the direction of the azimuthal angle) around the photon's central axis. Since this rotation represents a certain angular momentum, the photon may carry an angular momentum in the phase of the charge distribution.

In the following we shall assume that the length of the model photon, l, equals to $n\lambda$, where n is a positive integer to satisfy the periodic condition in the propagation direction. Here n may be considered as a quantum number and its minimum value is one, which makes a minimum complete cycle.

Now applying Eq. (23) to Eqs. (26) and (27), we find that the electric field energy density $\eta_{\mathbf{E}}$ and the magnetic field energy density $\eta_{\mathbf{B}}$ (Eq. (9)) are equal to each other for the photon. And we have the total energy density η as following

$$\eta = \epsilon_0 |\mathbf{E}|^2 = \epsilon_0 {\psi_0}^2 \begin{cases} \dfrac{1}{{r_0}^2} & r < r_0 \,, \\ \dfrac{{r_0}^2}{r^4} & r > r_0 \,, \end{cases} \tag{34}$$

where $|\mathbf{E}|$ is the magnitude of the electric field. The energy density is constant for $r < r_0$ and is inversely proportional to r^4 for $r > r_0$. The photon energy (Eq. (1)) may be equal to the integration value of Eq. (34) in the photon space at time $t = 0$. The integration path for r is 0 to r_0 and r_0 to ∞, for z is -$n\lambda$ to 0, and for ϕ is 0 to 2π. And hence we find the ψ_0 to have the following relationship

$$\psi_0 = \sqrt{\frac{\hbar c}{\epsilon_0 n}} \frac{1}{\lambda} \,. \tag{35}$$

In deriving Eq. (35) we used Eq. (1). It is interesting to note that the potential strength constant, ψ_0, is inversely proportional to the wavelength λ.

By using Eqs. (10), (26), and (27), the Poynting vector is

$$\mathbf{S} = \hat{\mathbf{z}} \frac{\psi_0 A_0}{\mu_0} \begin{cases} \dfrac{1}{{r_0}^2} & r < r_0 \,, \\ \dfrac{{r_0}^2}{r^4} & r > r_0 \,. \end{cases} \tag{36}$$

According to Eq. (36), the photon energy flows in the direction of the positive z axis, which is consistent with the photon direction of motion. The total energy by the Poynting vector for the photon is $h\nu$, which may be calculated by integrating out the Poynting vector, Eq. (36), for the photon and using Eqs. (23) and (35). This is an expected result.

Since the charge is distributed in the r_0 cylindrical surface, which generates a surface electric current by the displacement of the photon at the speed of light, the density of the photon self energy may also be expressed in the following relationship,

$$\eta' = \frac{1}{2} \sigma\psi + \frac{1}{2} \mathbf{A} \cdot \mathbf{J}' \,, \tag{37}$$

where η' represents the surface energy density, σ the surface charge density, ψ the electric potential, $\mathbf{A}$ the vector potential, and $\mathbf{J}'$ represents the surface electric current density. For the photon, $\mathbf{A} \cdot \mathbf{J}' = AJ'$ and $J' = \sigma c$, the second term is equal to the first term on the right hand side of Eq. (37) and we have.

$$\eta' = \sigma\psi \,. \tag{38}$$

Using Eqs. (28), (21) for $m = 1$, and (35), we may calculate the photon energy ϵ by integrating out Eq. (38) on the r_0 cylindrical surface of length $n\lambda$,

$$\begin{aligned} \epsilon &= \int_{-n\lambda}^{0} \int_0^{2\pi} \eta' \, dS = \int_{-n\lambda}^{0} dz \int_0^{2\pi} \sigma\psi r_0 \, d\phi \\ &= \int_{-n\lambda}^{0} dz \int_0^{2\pi} 2\epsilon_0 \psi_0^2 \sin^2(kz + \phi) \, d\phi \\ &= \int_{-n\lambda}^{0} dz \int_{kz}^{kz+2\pi} 2\epsilon_0 \psi_0^2 \sin^2(\phi') \, d\phi' \\ &= n\lambda 2\epsilon_0 \psi_0^2 \pi = h\nu \,, \end{aligned} \tag{39}$$

where dS represents an infinite small area on the r_0 cylindrical surface, the time $t = 0$, and a variable change, $kz + \phi = \phi'$. Hence we get that the energy is $h\nu$. This result indicates that

it is equivalent to consider the photon energy being stored in the r_0 cylindrical surface.

Now we evaluate the value of the constant length of the polar radius, r_0, of the model photon. We first assume that r_0 is proportional to the wavelength λ as

$$r_0 = \frac{\lambda}{2\pi} . \tag{40}$$

Then we support it by two reasons. The first reason is that with this assumption the phase velocity of the charge distribution on the r_0 cylindrical surface is equal to the speed of light c, i.e., $\omega r_0 = 2\pi\nu\lambda/2\pi = \nu\lambda = c$. This is consistent with the nature of the photon. This velocity may be physically experienced by an electron in an atom as in light absorption.

The second reason is that the angular momentum carried by the photon is $\hbar$, which is consistent with the angular momentum number $m = 1$. To evaluate the angular momentum, we use following expression

$$d\mathbf{J} = \mathbf{r}_0 \times d\mathbf{P} , \tag{41}$$

where we consider the angular momentum to be generated in the r_0 cylindrical surface, $d\mathbf{J}$ represents an infinite small quantity of angular momentum, $d\mathbf{P}$ represents an infinite small quantity of momentum in the cylindrical surface, and $\mathbf{r}_0$ is the polar radius vector pointing to the cylindrical surface where the small momentum is considered. Referring to Fig. 3, an observer like an electron in an atom may experience a rotational force from the photon, which corresponds to a momentum in the direction of the azimuthal angle ϕ. This momentum may generate an angular momentum in the direction of the positive z axis.

Similar to Eq. (2), the magnitude of the infinite small quantity of momentum dP may be written as

$$dP = \frac{d\epsilon}{c} , \tag{42}$$

where $d\epsilon$ represents an infinite small amount of energy in the cylindrical surface and c is the speed of light. Using Eq. (38), we have for the $d\epsilon$,

$$d\epsilon = \eta' dS = \sigma\psi dS , \tag{43}$$

where dS represents an infinite small area on the r_0 cylindrical surface. And finally we have for the magnitude of the infinite small quantity of the angular momentum dJ as

$$dJ = \frac{r_0}{c}\,\sigma\psi dS , \tag{44}$$

where r_0 is given in Eq. (40). The direction of the angular momentum is in the positive z axis.

By integrating out Eq. (44) for the photon on the r_0 cylindrical surface at the time $t = 0$, as has been done in Eq. (39), we get that the total angular momentum of the photon is indeed $\hbar$. Hence from the second reasoning we prove that the constant radius r_0 of the photon cylindrical surface is $\lambda/2\pi$.

This angular momentum, derived from the classical mechanics, may be considered as the spin angular momentum of the photon since it is generated by the self-rotation around its central axis.

Now based on the solution of Eqs. (21) and (22), we have built a consistent three-dimensional model of the photon: a quantized electromagnetic wave of length $n\lambda$ with a charged cylindrical surface core of radius $\lambda/2\pi$. Such a model may be tested for it is expected that the photon is very hard to pass a pinhole of radius less than $\lambda/2\pi$.

4 Conclusion

Conclusion by summarizing what have been presented in the paper. First a desirable solution was shown in terms of the two expressions, Eqs. (21) and (22), for the four-potential, obtained from wave Eqs. (16) and (17) derived by using the Maxwell equations together with the Lorenz condition. Although we assumed the medium to be vacuum in the solution for simplicity, our solution may be extended to the case of a homogeneous medium by using the medium parameters of the permittivity, permeability, and the speed of light. Also for clarity we limited our consideration in the Discussions section to the case of $\phi \geq 0$ and $t \geq 0$, but the solution itself is equally applicable if we substitute ϕ by -ϕ or t by -t. In the case of ϕ, the ± signs respectively may represent the right or left spin state of the photon.

Then the solution was analyzed for understanding its characteristics, which showed that an electromagnetic field in isolated wave form at the speed of light might exist in a limited space at a specific point of time. The solution requires the existence on the r_0 cylindrical surface of electric charge distributed in certain phase with the azimuthal angle ϕ and along the direction of the light propagation. The solution was specifically studied for the case of the angular momentum number $m = 1$.

We then applied the solution to the case of a model photon and determined the constant values of the solution in terms of the photon quantities. By doing that, a detailed theoretical three-dimensional model of the photon was achieved. We showed that the angular momentum of the photon might be considered as coded in the r_0 cylindrical surface by the phase of the charge distribution.

Notice that we have solved a special case of Eqs. (16) and (17) by restricting the angular momentum of the photon in the direction of the light propagation. Furthermore, the length of the photon was assumed to be $n\lambda$, but the upper bound of n was not determined specifically.

Finally it is theoretically interesting to mention that by letting the angular momentum number $m > 1$ in the solution, which could correspond to a photon with spin larger than one, we may get similar results as the spin one photon in terms of the wave taking a limited space at a specific point of time.

Acknowledgements

The author would like to express his thanks to all teachers for his education up to the post-doctoral position. Helpful discussions with Dr Wing-Ki Liu are acknowledged.

Submitted on November 10, 2015 / Accepted on December 7, 2015

References

1. Glauber R. J. The quantum theory of optical coherence, *Physical Review*, 1963, v. 130(6), 2529.
2. Kragh H. Photon: New light on an old name. arXiv:1401.0293.
3. Planck M. *Verhandl, Dtsch. Phys. Ges.*, Berlin, 1900, v. 2, 237. (English translation: Planck M. On the theory of the energy distribution law of the normal spectrum. In: *The Old Quantum Theory*, ed. by D. ter Haar, Pergamon Press, 1967, 82.)
4. Einstein A. *Annalen der Physik* (in German), 1905, v. 17(6), 132. (English translation: Einstein A. On a heuristic point of view concerning the production and translation of light. In: The Swiss Years: Writings, 1900–1909, v. 2, 86.)
5. Compton A. H. A quantum theory of the scattering of X-rays by light elements. *Physical Review*, 1923, v. 21 (5), 483.
6. For example: Davydov A. S. Quantum Mechanics. 2rd ed., Translated by D. Ter Haar, Pergamon Press, New York, 1988, p. 22.
7. For example: Kenyon I. R. The Light Fantastic. Oxford University Press, New York, 2008.
8. For example: Jackson J. D. Classical Electrodynamics. 3rd ed., John Wiley & Sons Inc., New York, 1999.
9. Lorenz L. V. On the identity of the vibrations of light with electrical currents. *Philos. Mag.*, Ser 3, 1867, v. 34, 287.
10. Arfken G. Mathematical Methods for Physicists, 3rd Ed., Academic Press, Inc., New York, 1985, 95.

Update on Pluto and Its 5 Moons Obeying the Quantization of Angular Momentum per Unit Mass Constraint of Quantum Celestial Mechanics

Franklin Potter

Sciencegems.com, 8642 Marvale Drive, Huntington Beach, CA 92646 USA
E-mail: frank11hb@yahoo.com

In July, 2015, the New Horizons spacecraft passing by Pluto did not discover any more moons. Therefore, we know the Pluto system total angular momentum to within 2.4%, more accurately than any other system with more than two orbiting bodies. We therefore update our previous analysis to determine whether a definitive test of the quantum celestial mechanics (QCM) angular momentum constraint can now be achieved.

1 Introduction

In 2012 we analyzed the angular momentum properties of the Pluto system with its 5 moons [1] not knowing the total angular momentum in the system. The New Horizons spacecraft passing by Pluto and its large moon Charon in July, 2015, did not discover any more moons than its earlier discovery of 4 additional tiny moons. Therefore, the Pluto system that we know is the final configuration of orbiting bodies, so we now know its total angular momentum to within 3%. Consequently, we can consider this gravitationally bound system as a possible definitive test of the theory called quantum celestial mechanics (QCM) first proposed in 2003 by H. G. Preston and F. Potter [2].

They derived a new gravitational wave equation from the general relativistic Hamilton-Jacobi equation for a test particle of mass μ as given by Landau and Lifshitz:

$$g^{\alpha\beta}\,\frac{\partial S}{\partial x^{\alpha}}\,\frac{\partial S}{\partial x^{\beta}} - \mu^2\,c^2 = 0\,, \tag{1}$$

where $g^{\alpha\beta}$ is the metric of the general theory of relativity (GTR) and S is the action. This general relativistic Hamilton-Jacobi equation becomes a scalar wave equation via the transformation to eliminate the squared first derivative, i.e., by defining the wave function $\Psi(q, p, t)$ of position q, momentum p, and time t as

$$\Psi = e^{iS'/H} \tag{2}$$

with $S' = S/\mu c$. The H is defined as the Preston distance characterizing the specific gravitational system and is a function of *only two physical parameters* of the system

$$H = \frac{L_T}{M_T\,c}\,, \tag{3}$$

where M_T is the total mass of the system and L_T its total angular momentum. Only these two parameters of the system are required to define all the stable quantization states of the gravitationally bound system. We call the resulting theory quantum celestial mechanics or QCM.

The end result of the transformation is the new scalar "gravitational wave equation" (GWE)

$$g^{\alpha\beta}\,\frac{\partial^2\Psi}{\partial x^{\alpha}\,\partial x^{\beta}} + \frac{\Psi}{H^2} = 0. \tag{4}$$

One can now consider the behavior of the test particle in different gravitational metrics. In the Schwarzschild metric, we find good agreement with predictions for all systems to which the QCM constraints have been applied.

There have been numerous applications of QCM to gravitationally bound systems, including multi-planetary exosystems [3], the Solar System [2], the five moons of Pluto [1], the S-stars at the galactic center [4], and circumbinary systems [5, 6] with planets. All these systems have been shown to obey the quantization of angular momentum per unit mass constraint dictated by QCM in the Schwarzschild metric approximation for each orbiting body μ_i, i.e.,

$$\frac{L_i}{\mu_i} = m_i\,c\,H. \tag{5}$$

Of course, one assumes that the body in consideration has been in an equilibrium orbit for at least tens of millions of years. Then if one knows the semi-major axis r, the eccentricity e, and the period of orbit, the QCM value for L_i in the specific equilibrium orbit equals the Newtonian value $L = \mu\,\sqrt{GM_T\,r\,(1-e^2)}$. The value of M_T is nearly the central body mass for most cases.

Knowing the period of orbit is an additional constraint that allows one to determine a set of integers m for the QCM angular momentum per unit mass linear regression fit, with $R^2 > 0.999$, which we seek in all cases. Moreover, if one knows the total angular momentum for the gravitationally bound system, then a unique set of m values is possible. However, if the system total angular momentum is unknown, then several sets of integers could meet the liner regression fit, in which case we will accept the set beginning with the smallest integer.

From the slope of the resulting plot of $L/\mu c$ vs. m for all

the known orbiting bodies in the system, one can calculate the predicted QCM total system angular momentum L_T and therefore can predict whether additional mass orbiting the star is needed to account for this total angular momentum value. Many m values for the gravitationally bound system will be unoccupied, for the occupancy of the specific QCM orbits depends upon the history of formation and the subsequent evolution of the planetary system.

For simplicity, applications have concentrated on circular or near-circular orbits only. Whereas in GTR and its Newtonian approximation all allowed circular or nearly-circular orbits about a massive central object are equilibrium orbits, QCM dictates that only a subset of these equilibrium orbits are permitted by the quantization of angular momentum per unit mass constraint.

With any new theory, one needs a definitive test. Until now there has been no laboratory test of QCM. Finding a convincing, definitive test for QCM has not been successful. As of this date, the satellites of Pluto actually offer the best test of QCM and its quantization of angular momentum per unit mass prediction. Why? Because the total angular momentum of the Pluto-Charon system with its 4 tiny moons is well-known now to within 2.4%.

One would expect that the Solar System as a whole or the many satellites of the Jovian planets would be a better test. However, one does not know the total angular momentum to within 10% of either the Solar System or each of the Jovian planets. The Jovian planets themselves dominate the angular momentum contributions in their systems but their internal differential rotations lead to large uncertainties in their total angular momentum.

And, unfortunately, we do we not know the total angular momentum of the Solar System to within 10%. Why not? Because the Oort Cloud dominates the Solar System angular momentum [7], providing about 50 times the total angular momentum contribution from the Sun and the planets! The total mass of the Oort Cloud is unknown but can be estimated by assuming perhaps 100 Earth masses of ice chunks at more than 40,000 AU. The dominance of the Oort Cloud can be verified by estimating the Newtonian value of its angular momentum.

Although we have determined excellent linear regression fits to all planetary-like systems by the QCM angular momentum constraint, there remain two limitations of the fits: (1) they are not unique and (2) all integers are candidates for m, i.e., there being no upper limit. For example, even with a linear regression fit $R^2 = 1.000$ for the set of m values 3, 5, 8, 14, 17, for a 5 planet system, the set of double values 6, 10, 16, 28, 34, fits equally well. The slope of the graph of $L/\mu c$ versus m is used to predict the total angular momentum of the system, the former set predicting twice the angular momentum. However, if one knows the total system angular momentum value, such as we do now for the Pluto system, then the set of m values is unique.

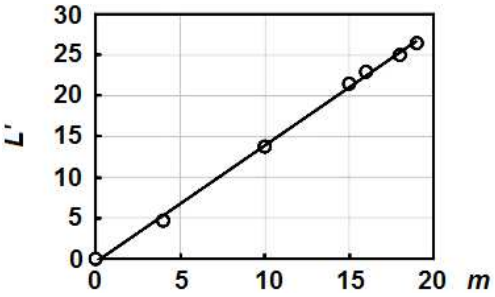

Fig. 1: The Pluto System fit to QCM.

	$r \times 10^6$ m	Period (d)	m	P_2/P_1	$(n_2/n_1)^3$
Pluto	2.035	6.38723	4		
Charon	17.536	6.38723	10	1	1
Styx	42.656	20.16155	15	3.156	3.077
Nix	48.694	24.85463	16	3.891	3.691
Kerberos	57.783	32.16756	18	5.036	5.153
Hydra	64.738	38.20177	19	5.981	6.011

Table 1: Pluto system orbital parameters and QCM m values.

2 Pluto and its 5 moons

Will a *random* set of orbital distances fit the QCM angular momentum quantization constraint? Yes, because there is no upper limit to the integers available for the m values. One can always fit the constraint using very large integers! This possibility is eliminated when the total angular momentum is known. If one uses this random set of orbital distances with a specific mass for the central star but the other masses are unknown, the system obeys Newton's law of universal gravitation and the angular momentum *per unit mass* is known but the unique set of integer values for m cannot be achieved.

The New Horizons spacecraft passing Pluto in July, 2015, did not discover any more moons. The Pluto satellite system [8] has five moons, Charon, Styx, Nix, Kerberos, and Hydra, which are nearly in a 1:3:4:5:6 resonance condition! The orbital behavior of the five moons is considered by using distances from the Pluto-Charon barycenter. The important physical parameters of the Pluto system satellites are given in Table 1. The orbits are very close to circular.

The system total mass is essentially the combined mass of Pluto (13.05×10^{21} kg) and Charon (1.52×10^{21} kg). The QCM values of m in the fourth column were determined by the linear regression fit ($R^2 = 0.998$) to the angular momentum quantization per mass equation as shown in Figure 1 with $L' = L/\mu c$ plotted against m with resulting slope $H = 1.43$ meters. The uncertainty bars are within the circles. Our previous fit [1] of these Pluto moons proposed the m values 2, 6, 9, 10, 11, 12, with $R^2 = 0.998$ also.

This new value of H produces a total angular momentum value $L_T = 6.28 \times 10^{30}$ kg m^2/s that is commensurate with the total angular momentum of $6.26(\pm 0.14) \times 10^{30}$ kg m^2/s for the

known Pluto system when both orbital and rotational angular momentum are included.

In QCM the predicted period ratios for the orbital resonance conditions in the last column of Table 1 are calculated from the m values using

$$\frac{P_2}{P_1} = \frac{(m_2 + 1)^3}{(m_1 + 1)^3} \,. \tag{6}$$

With Charon as the reference, this system of moons has nearly a 1:3:4:5:6 commensuration, with Kerberos having the largest discrepancy of about 5.2%.

These moons have distances from the barycenter that are within 2.4% of their QCM equilibrium orbital radii. If in the next few million years they adjust their orbital semi-major axes, their positions on the plot may improve to increase the R^2 value but their m values will remain the same. Dynamic analysis via the appropriate QCM equations could be done to predict their possible movements.

Note that some additional extremely tiny moons of Pluto may be found at some of the non-occupied m values, but their angular momentum contributions will be very small. The formation history of Pluto determines which m values are actually occupied by orbiting bodies.

3 Discussion

QCM predicts the quantization of angular momentum per unit mass for all orbiting bodies in gravitationally bound systems. Unfortunately, the total angular momentum of planetary-like systems is usually not known to within 10%. Fortunately, the New Horizons spacecraft passing by Pluto in 2015 did not discover any additional moons of Pluto, so we now know the extent of this system and its total angular momentum to within 2.4%.

We have determined the best set of m integers for a fit to the QCM angular momentum constraint, and the predicted resonances in its moon system are in agreement with the measured period ratios to within 5.2%.

Therefore, we claim to have a definitive test of QCM in the Schwarzschild metric in a planetary-like system because the best understood system, Pluto and its 5 moons, obeys the quantization of angular momentum per unit mass constraint. Consequently, we expect that all such systems obey QCM, and in the future we will search for systems that seem to violate the angular momentum constraint.

One would prefer the ability to vary the parameters in a gravitationally bound system, but we do not have that luxury in astronomical systems. A laboratory test would allow the variation of the system parameters in a controlled manner and should be undertaken with perhaps a pendulum in a vacuum chamber near to a rotating mass. In the ideal case one would expect the maximum repulsion of the pendulum to occur when the angular momentum constraint is met and its magnitude to be comparable to the Newtonian attraction. This type of additional definitive test of QCM might be able to achieve an reduced uncertainty down to about 0.1%.

Acknowledgements

We thank Sciencegems.com for continuing to encourage this research into the applications of QCM to gravitational systems.

Submitted on December 7, 2015 / Accepted on December 9, 2015

References

1. Potter F. Pluto moons exhibit orbital angular momentum quantization per mass. *Progress in Physics*, 2012, v.8(4), 3–4.
2. Preston H.G., Potter F. Exploring large-scale gravitational quantization without $\hbar$ in planetary systems, galaxies, and the Universe. arXiv: gr-qc/0303112.
3. Potter F. Multi-planet exosystems all obey orbital angular momentum quantization per unit mass predicted by Quantum Celestial Mechanics (QCM). *Progress in Physics*, 2013, v. 9(3), 60–61.
4. Potter F. Galaxy S-stars exhibit orbital angular momentum quantization per unit mass. *Progress in Physics*, 2012, v. 8(4), 29–30.
5. Potter F., Preston H.G. Kepler-16 circumbinary system validates Quantum Celestial Mechanics. *Progress in Physics*, 2012, v. 8(1), 52–53.
6. Potter F. Kepler-47 circumbinary planets obey angular momentum quantization per unit mass predicted by Quantum Celestial Mechanics (QCM). *Progress in Physics*, 2014, v. 10(1), 19–20.
7. Weissman P.R. The angular momentum of the Oort cloud. *Icarus*, 1991, v. 89, 190–193.
8. Youdin A.N., Kratter K.M., Kenyon S.J. Circumbinary chaos: using Pluto's newest moon to constrain the masses of Nix & Hydra. arXiv: 1205.5273v1.

LETTERS TO PROGRESS IN PHYSICS

On the Deviation of the Standard Model Predictions in the Large Hadron Collider Experiments

Anatoly V. Belyakov

E-mail: belyakov.lih@gmail.com

The newest Large Hadron Collider experiments targeting the search for New Physics manifested the possibility of new heavy particles. Such particles are not predicted in the framework of Standard Model, however their existence is lawful in the framework of another model based on J. A. Wheeler's geometrodynamcs.

The main task of the Large Hadron Collider is to look for true deviations from the Standard Model (SM) if any. The collider has done hundreds of such experiments already. Some experimental results of these really deviate from the theoretical results predicted in the framework of SM. The newest Large Hadron Collider experiments done in look for New Physics manifested the possibility of new heavy particles.

The ATLAS collaboration team and the CMS collaboration team reported on the experimental search for heavy particle-resonances [1–3]. So, the ATLAS team, while experimental search for heavy resonances of a mass in the scale from 1 to 3.5 TeV decaying into a pair of bosons (i.e., into WW-, WZ-, or ZZ-pairs), discovered an anomalous number of events having an invariant mass of ~ 2 TeV. While the CMS team looked for the events in which many hadrons and an electron-positron pair were born then scattered with high energies. In the scale of invariant masses of ~ 2 TeV, they registered an anomalous many events. The obtained picture is like the production and decay of new heavy particles.

Such particles are not predicted in the framework of SM. However their existence is lawful in the framework of a model based on J. A. Wheeler's geometrodynamic concept.

In this geometrodynamic model, any elementary particle is considered as a trace appeared due to that a vortical tube (Wheeler's wormhole) transits the surface of our world (i.e. as a fermion), and also as a contour or a vortical tube (i.e. as a boson). So there can be connected contours of the first and higher order, which give birth to a few generations of the elementary particles [4]. As a result, any particle corresponds to two quantum numbers depending on that the particle is considered either as a fermion (an analogy of a proton joined into the large contour of the next family of particles), or as the boson mass of the contour of the previous family of particles.

In this way, only three families of the elementary particles can exist.

The first generation of the particles is a proton contour (a proton itself) having the same fermionic and bosonic masses, the sum of which is approximately equal to the sum of all π-mesons and K-mesons (1899 MeV).

The second generation is the standard proton-electron contour (the μ-analogy of the proton) having a bosonic mass close to the summary mass of the W and Z-bozons (229 GeV; the fermionic masses of the contour and those of the following contour are neglected).

The third generation is the largest contour wherein the parameters of the vortical tube reach its critical numerical values (the τ-analogy of the proton). The mass of the vortical tube is 3.1 TeV. It is logically lawful to guess that, in analogy to the second generation, this mass consists as well of three bosons (the average mass of each is 1 TeV).

According to the formulae obrained in [4] on the basis of Wheeeler's geometrodynamics, the aforementioned mass can be expressed in the $m_e c^2$ units as

$$M_y = \frac{1}{3}\left(\frac{2a^3}{c_0^{1/3}}\right)^{7/4} = 2.1\times 10^6 \ \ (1.07\ \text{TeV}), \qquad (1)$$

where a is the reverse fine structure constant, while c_0 is the dimensionless light speed.

The characteristic mass close to 1 TeV can also be found proceeding from other consideration. As was found in [5, 6], the mass of the active part of the proton (the mass of its quark) enrolled into a circulation contour having a quantum contour parameter n_y answers the relation $m_k = c_0^{2/3}/(an_y)^2$. It is shown in [4, 6] that not only 1/3 but also 1/4 of this value can be the minimally possible charge (mass). Thus, in the ultimate small value can be $m_k = \frac{1}{4}m_e$. As a result, the ultimate heavy bosonic mass of the contour (in its excited state) is equal to

$$M_y = (an_y)^2 = 4c_0^{2/3} = 1.79\times 10^6 \ \ (0.916\ \text{TeV}). \qquad (2)$$

At last, it was found in [6] while considering the process of appearance of the neutrino that, if the mass-energy of a $p^+ - e^-$-contour is close to the mass of a W-boson, replacing the electron mass with the τ-particle mass we obtain

$$M_y = c_0^{4/9} m_\tau^{1/3}\left(2\pi\gamma\rho_e \times [\sec^2]\right)^{1/3} = \\ = 2.29\times 10^6 \ \ (1.17\ \text{TeV}) \qquad (3)$$

that corresponds to the mass of the guessed boson of the third generation. Herein, m_τ is the τ-particle mass in the m_e units, γ is the gravitational constant, ρ_e is the density inside the electron ($m_e/r_e^3 = 4.071 \times 10^{13}$ kg/m^3).

Thus, proceeding from the viewpoint of the suggested model, such heavy particles decaying into the boson pair having a summary mass of ~ 2 TeV are very possible.

Submitted on November 10, 2015 / Accepted on November 11, 2015

References

1. ATLAS Collaboration. Search for high-mass diboson resonances with boson-tagged jets in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. arXiv: 1506.00962; CERN Report: CERN-PH-EP-2015-115.
2. CMS Collaboration. Search for massive WH resonances decaying to $\ell\nu b\bar{b}$ final state in the boosted regime at $\sqrt{s} = 8$ TeV. CERN Report: CMS-PAS-EXO-14-010.
3. CMS Collaboration. Search for heavy neutrinos and W bosons with right-handed couplings in proton-proton collisions at $\sqrt{s} = 8$ TeV. arXiv: 1407.3683; CERN Reports: CMS-EXO-13-008, CERN-PH-EP-2014-161.
4. Belyakov A.V. Macro-analogies and gravitation in the micro-world: further elaboration of Wheeler's model of geometrodynamics. *Progress in Physics*, 2012, v. 8, issue 2, 47–57.
5. Belyakov A.V. Nuclear power and the structure of a nucleus according to J. Wheeler's geometrodynamic concept. *Progress in Physics*, 2015, v. 11, issue 1, 89–98.
6. Belyakov A.V. Determination of the neutrino mass. *Progress in Physics*, 2016, v. 12, issue 1, 34–38.

X-Ray Flares from Sagittarius A* and Black Hole Universe

T. X. Zhang, C. Wilson, and M. P. Schamschula
Department of Physics, Alabama A & M University, Normal, Alabama 35762, USA
E-mail: tianxi.zhang@aamu.edu

Sagittarius (Sgr) A* is a massive black hole at the Milky Way center with mass of about 4.5 million solar masses. It is usually quite faint, emiting steadily at all wavelengths including X-rays. Since the beginning of this century, rapid and intensive X-ray flares are regularly detected from Sgr A* at a rate of about once a day. Conventionally, these mysterious events daily occurred at the Milky Way center are believed to be caused by the falling of objects such as asteroids, comets, and planets onto the massive black hole. However, the physical process of how the falling objects to produce the observed X-ray flares is still poorly understood. It is unclear why the gases, formed by tearing the falling objects apart, can be heated up to 100 million degrees Celsius so suddenly on a regular basis. This study develops a new alternative mechanism and provides a possible explanation for the observations of X-ray flares from Sgr A*, in accordance with the black hole universe model that was recently proposed by Zhang. The results obtained from this study indicate that X-ray flares from the Milky Way center can be understood as emissions of the dynamic massive black hole (i.e. Sgr A*). A massive or supermassive black hole, when accreting matter or objects from the outside, becomes dynamic and breaks its event horizon, which leads to the inside hot (or high-frequency) blackbody radiation leaking and produces X-ray flares or bursts. The energies and spectra of X-ray flares that Sgr A* can produce when it accretes objects with various sizes including asteroids, comets, planets, and stars are theoretically analyzed and numerically calculated. In terms of results obtained from these analyses and calculations, we explain the current measurements of X-ray flares from Sgr A*, predict events that will possibly occur at our galactic center in future, and compare the extremely intensive events predicted with the strong X-ray flares measured from other normal and active galactic centers. This study develops a new physical mechanism for the origin of X-ray flares from galactic centers and deepens our understanding to the black hole dynamics, galactic activities, and cosmological evolutions.

1 Introduction

Sagittarius (Sgr) A* is a compact astronomical radio source that was first discovered by [1] at the center of the Milky Way, near the border of the constellations, Sagittarius and Scorpius. The orbital motions of stars around the Milky Way center indicate the presence of a massive black hole with about 4.5 million solar masses, which is spatially coincident with Sgr A* [2–3].

In general, Sgr A* is very faint and emits steadily at all wavelengths, especially in the range of soft X-rays (2-10 keV) with luminosity about 2×10^{33} erg/s [4]. Recently, NASA Chandra X-ray Observatory and other missions such as Swift, NuStar, XMM-Newton, and Roast have discovered intensive and rapid X-ray flares at a rate of about once a day from Sgr A*, with luminosity at the peak up to a few times 10^{35} erg/s [5–7]. The brightest X-ray flare ever observed so far emits in total $\sim 10^{39} - 10^{40}$ ergs of X-rays (2-10 keV) and last a few thousand seconds or hours [8–9]. The X-ray echoes recently discovered reveal that Sgr A* would have been a very violent past with luminosity of $\sim 10^{39}$ erg/s (i.e., a million times brighter than its present normal emission) during the X-ray outbursts of the past few hundred years [10]. X-ray outbursts from some other inactive galaxies can be even much more intensive with luminosity $\sim 10^{44}$ erg/s [11–12]. Luminosities of an active galactic nuclei or a quasar can be extremely high up to 10^{46} erg/s [13–15].

To explain the mysterious X-ray flares, astronomers have suggested that there exists a gas cloud around Sgr A* containing hundred-trillions of asteroids, comets, and planets that are stripped from their parent stars by the tidal forces of the massive black hole. When these objects rain down or are accreted onto the massive black hole, X-ray flares take place via physical processes such as the non-thermal synchrotron emission [16], the inverse-Compton scattering [17], and stochastic electron acceleration [18]. To emit the high-energy X-rays detected, an object that was striped from its parent star had to be torn apart into gases during its falling and the gases when arriving nearly at the massive black hole had to spike to hundreds of million degrees Celsius, which is ten or more times hotter than the center of the Sun. However, why the gases heats up so suddenly and efficiently on a regular ba-

sis is still poorly understood. One possible heating scenario recently guessed is based on the physics of solar flares by considering that the lines of magnetic energy contained in the gas flowing into Sgr A* got tangled and the reconnection of magnetic lines leads [19-20], but there still lacks of a quantitative study on this magnetic mechanism. Especially, Sgr A* may not be able to gravitationally tear an asteroid into parts as small as a human body, because the gravitational field difference between the head and feet of a 2-meter height person, who stands on Sgr A* surface is only 10^{-3} m/s^2. Up to the date, astronomical communities are still out on what really caused these giant X-ray flares from Sgr A*. The mechanism for the origin of X-ray flares from the galactic center is still a mystery and in pending for a physical explanation.

Recently, postulating the equivalence between a spactime and a black hole, Zhang [21–22] developed a new cosmological model called black hole universe, which is consistent with Mach's principle, governed by Einstein's general relativity with the cosmological principle of spactime isotropy and homogeneity, and able to explain the existing observations of the universe without encountering difficulties such as the flatness, horizon, inflation, dark matter, and dark energy problems. The studies that have been conducted so far have explained the origin, structure, evolution, expansion, cosmic microwave background radiation, quasar formation and emission, gamma ray bursts (GRBs), and acceleration of black hole universe [15, 22–27]. According to this new cosmological model, the universe originated from a star-like black hole with several solar masses, grew up through a supermassive black hole with billions of solar masses to the present state with hundred sextillions of solar masses by accreting ambient matter and merging with other black holes. More aspects about the black hole universe model have been presented in a sequence of American Astronomical Society (AAS) meetings [28–37]. The black hole universe model establishes a complete new understanding to the dynamics of black holes, so that offers a unique explanation to the observations of various events that relate to the activities of black holes such as quasars [15], gamma ray bursts [25], and X-ray flares from galactic centers (this paper).

This study will focus our investigations on the physical mechanism of X-ray flares from Sgr A*, a massive black hole at the Milky Way center, and provides an alternative explanation for the energy and spectrum measurements of X-ray flares according to the black hole universe model. The results indicate that X-ray flares from the galactic center can be understood as emissions of the dynamic massive black hole. As pointed out in Zhang's early studies, a black hole, when it accretes its ambient matter or objects, becomes dynamic. A dynamic black hole has a broken event horizon and thus cannot hold the inside hot (or high-frequency) blackbody radiation, which leaks out and produces a gamma ray burst if it is a star-like black hole or an X-ray flare if it is a massive or supermassive black hole. The energies and spectra of X-rays obtained by this study for the X-ray emissions from Sgr A* when it accretes appropriate size objects such as asteroids, comets, and planets can be consistent with the measurements.

2 X-ray emissions of dynamic massive black holes

In accordance with the black hole model of the universe developed recently by [21–22], a black hole constructs an individual spacetime (spatially singular and temporally noncausal to the outside) and a spacetime encloses a black hole. Black hole and spactime are equivalent. According to this equivalence, our four-dimensional (4D) spacetime universe is a fully grown extremely supermassive black hole. The observed star-like, massive, and supermassive black holes are subspacetimes of our black hole universe. Upon the view from the outside, a star-like or supermassive black hole is a singular sphere, from which no matter and radiation can escape. In general, a star-like (or larger) black hole can be considered as an ideal blackbody, with the following Mach-Schwarzschild mass-radius (M-R) relation

$$\frac{2GM}{c^2R} = 1, \tag{1}$$

where c is the light speed in the free space and G is the gravitational constant.

The temperature inside a star-like black hole, though it cannot be measured from outside, should be as high as that of a neutron star because both types of objects are comparably compact. At the moment of its birth via a supernova explosion, a neutron star can reach 10^{12} K and then quickly cools down to about 10^8 K because of its strong radiation and neutrino emission [38]. A black hole can hold the high temperature reached at the moment of its birth because it does not radiate significantly. When a star-like black hole accretes matter and radiation from outside, it expands and cools down. As a star-like black hole grows up as big as the present universe, the inside temperature decreases from 10^{12} K to about 3 K. In the black hole universe model, the observed 3 K cosmic microwave background radiation is the internal blackbody radiation of the black hole universe, an ideal blackbody [23, 29].

The spectral energy density of blackbody radiation within a black hole including the black hole universe can be determined according to Planck's law as

$$u(\nu, T) = \frac{8\pi h\nu^3}{c^3} \frac{1}{\exp\left(\frac{h\nu}{kT}\right) - 1}, \tag{2}$$

where ν is the radiation frequency, T is the temperature, h is the Planck constant, and k is the Boltzmann constant. In the SI unit system, the unit of $u(\nu, T)$ is J/m^3/Hz, which is equivalent to 2.41×10^{17} J/m^3/keV. Figure 1 plots the spectral energy density as a function of photon energy $\epsilon = h\nu$ at temperature equal to 10^6, 10^7, 10^8, and 10^9 K, respectively. It is seen that the spectral energy density significantly varies with the temperature and photon energy. Inside a black hole

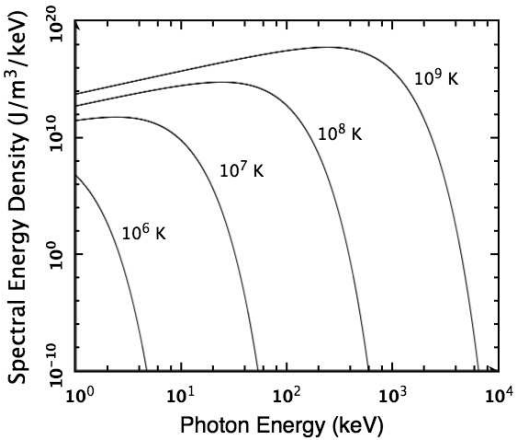

Fig. 1: The spectral energy density of blackbody radiation as a function of radiation energy at temperature equal to 10^6, 10^7, 10^8 K and 10^9 K, respectively.

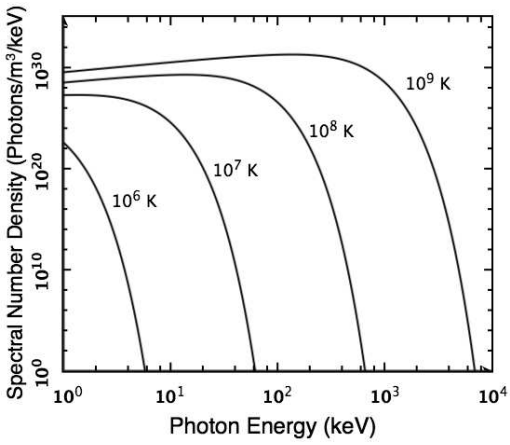

Fig. 2: The spectral number density of blackbody radiation as a function of radiation energy at temperature equal to 10^6, 10^7, 10^9 K and 10^{12} K, respectively.

with temperature of $10^7 - 10^8$ K (e.g. a massive black hole with millions of solar masses), the blackbody radiation dominates at the frequency of X-rays with photon energy in the range of 1 – 200 keV. The spectral photon number density $f(\nu, T) \equiv u(\nu, T)/\epsilon$ is plotted in Figure 2

Integrating the spectral energy density (Eq. 2) with respect to the frequency of radiation in the entire range, we have the energy density of the blackbody radiation inside a black hole including the black hole universe,

$$\rho_\gamma \equiv \int_0^\infty u(\nu, T)d\nu = \beta T^4, \tag{3}$$

where the constant β is given by $\beta \equiv 8\pi^5 k^4/(15h^3c^3) \simeq 7.54\times 10^{-16}$ J/m^3/K^4. Inside a black hole with temperature $\sim 10^7 - 10^8$ K, the energy densities of radiation are $\sim 10^{13} - 10^{17}$ J/m^3.

As a black hole including the black hole universe accretes its outside matter and radiation, it expands and cools down. Considering that the increase of the Planck radiation energy within the black hole equals to the radiation energy inhaled from the outside space, we have [23]

$$\frac{dT}{dR} = -\frac{3T}{4R}\left[1 - \left(\frac{T_p}{T}\right)^4\right]. \tag{4}$$

where T is the temperature inside the black hole and T_p is the temperature outside the black hole. This equation determines the temperature inside a black hole in accordance with its size. The solution of Eq. (4) for the dependence of T on R depends on T_p or on the relation between T and T_p. In the early studies [23, 29], Eq. (4) was solved for the present black hole universe that grew up from a hot star-like black hole through a supermassive black hole.

For star-like or supermassive black holes, the temperatures inside should be much greater than the temperatures outside, i.e., $T \gg T_p$. In this case, Eq. (4) can be simply solved as

$$R^3T^4 = C, \tag{5}$$

where C is a constant. Zhang [26] has assumed this constant to be the same for all size star-like or supermassive black holes and quantitatively explained the measurements of GRBs as emissions of dynamic star-like black holes. The value of the constant was determined according to the radius R_s and temperature T_s of a particular (or reference) black hole as $C = R_s^3T_s^4$. For a three-solar-mass black hole ($M_s = 3M_{\rm Sun}$) to be the reference black hole, its radius is about $R_s = 2GM_s/c^2 \sim 8.89$ km. Choosing its temperature to be $T_s = 10^{12}$ K, we have $C \sim 7\times 10^{59}$ m^3 K^4. The temperature of a star-like or supermassive black hole decreases as it expands in size according to $T \propto R^{-3/4}$.

Figure 3 plots the temperature of a black hole as a function of the radius or mass of the black hole. The the temperature of a three-solar mass black hole is chosen to be $T_s = 5\times 10^{11}$ K and 10^{12} K. For Sgr A* with mass of 4.5 million solar masses or radius of 1.33×10^{10} m, the temperature is $\sim 10^7 - 10^8$ K. The frequency of blackbody radiation at the peak to this temperature range is $\sim 10^{18} - 10^{19}$ Hz (or the energy of X-rays at the peak is $\sim 4 - 40$ keV).

From Eqs. (3) and (5), we obtain the total radiation energy U inside a black hole with volume V or radius R to be a constant and independent of its size or mass,

$$U \equiv \rho_\gamma V = \frac{4}{3}\pi\beta R^3T^4 = \text{Constant}. \tag{6}$$

It is seen that the total radiation energy inside a black hole (either a star-like or supermassive black hole) remains the same or is conserved. A black hole can grow its size by accreting mater from the outside space or merging with other black holes, but cannot increase its total radiation energy.

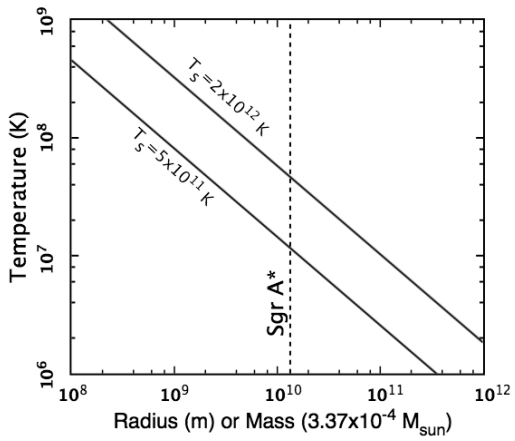

Fig. 3: The temperature of a massive black hole as a function of the radius or mass of the black hole with $T_s = 10^{12}$ K or 5×10^{11}. The vertical dashed line represents the radius, mass, and range of temperature.

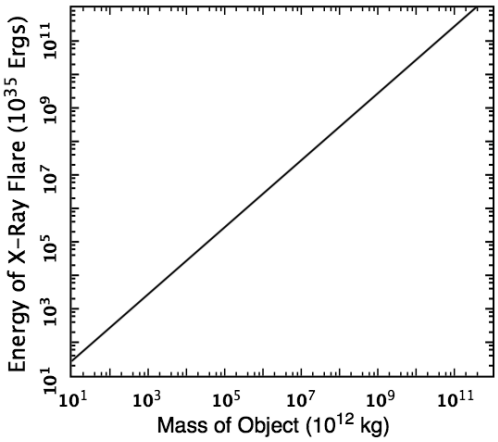

Fig. 4: The energy of X-ray flares from Sgr A* versus the mass of the object accreted.

A star-like black hole with several solar masses holds the same amount of radiation energy as a supermassive black hole with billions of solar masses does. The difference is only the temperature or frequency of the radiation. Dynamic star-like black holes with thousand billions of Kelvins radiate gamma rays [26], while dynamic massive or supermassive black holes with millions to billions of Kelvins radiate X-rays such as X-ray emissions from quasars [15] and X-ray flares from Sgr A*, a massive black hole at the Milky Way center as shown in this study.

3 Energy and energy spectrum of X-ray flares from Sgr A*

According to the black hole universe model, X-ray flares from the Milky Way center are the emissions of the dynamic massive black hole, Sgr A*, which is accreting objects that fail to orbit around Sgr A*.

The energy emitted by Sgr A* with mass M and radius R, after it has accreted an object with mass m and radius r, can be determined by the difference of gravitational potential energies subtracting all other losses or dissipations during the falling of the object towards Sgr A*

$$E = U_{\rm M} + U_{\rm m} + U_{\rm Mm} - U_{\rm M+m} - E_{\rm loss}\,, \qquad (7)$$

where $U_{\rm M}$ is the gravitational potential energy of Sgr A* before the object is accreted,

$$U_{\rm M} = -\frac{3GM^2}{5R} = -\frac{3}{10}\,Mc^2; \qquad (8)$$

$U_{\rm m}$ is the gravitational potential energy of the object (e.g. an asteroid),

$$U_{\rm m} = -\frac{3Gm^2}{5r} = -\frac{3}{10}\frac{r_g}{r}\,mc^2, \qquad (9)$$

with $r_g = 2Gm/c^2$ is the Schwarzschild radius of an object with mass m; $U_{\rm M+m}$ is the gravitational potential energy of Sgr A* after the object is accreted,

$$U_{\rm M+m} = -\frac{3G(M+m)^2}{5(R+\delta R)} = -\frac{3}{10}\,(M+m)\,c^2; \qquad (10)$$

and $U_{\rm Mm}$ is the gravitational potential energy between Sgr A* and the object when the object is initially on the orbit,

$$U_{\rm Mm} = -\frac{GMm}{R_{\rm orbit}} = -\frac{1}{2}\frac{R}{R_{\rm orbit}}\,mc^2, \qquad (11)$$

with $R_{\rm orbit}$ is the radius of asteroid's initial orbit around Sgr A*; and $E_{\rm loss}$ is the energy lost or dissipated during the object is falling into Sgr A*. Substituting Eq. (8) through Eq. (11) into Eq. (7), we have

$$E = \frac{3}{10}\left(1 - \frac{r_g}{r} - \frac{5R}{3R_{\rm orbit}}\right)mc^2 - E_{\rm loss}. \qquad (12)$$

Since $r_g \ll r$ and $R \ll R_{\rm orbit}$, Eq. (12) simply reduces to

$$E \sim \frac{3}{10}\,mc^2, \qquad (13)$$

if the loss or dissipation is negligible in comparison with the rest energy of the object. Therefore, the energy of X-ray flares from Sgr A* approximately depends on the mass of the object that Sgr A* has accreted from outside. Figure 4 plots the energy of X-rays emitted by the massive black hole Sgr A* when it accretes an object as a function of the object mass. It is seen that Sgr A* emit more X-rays if it accretes more massive object. For instance, Sgr A* can emit up to 10^{39} ergs of X-rays if it accretes an asteroid with mass of 10^{17} kg.

Table 1 lists the energies of X-ray flares from Sgr A* by accreting some particular objects. Hourly accreting some small-sized asteroids can explain the faint and steady emissions of Sgr A* ($\sim 10^{33}$ ergs/s). Daily accreting one medium-sized asteroid can explain the present observations of X-ray

Type of Object	Mass (kg)	Energy (erg)
Asteroid (small size)	10^{13}	3×10^{36}
Asteroid (medium size)	10^{16}	3×10^{39}
Asteroid (large size)	10^{20}	3×10^{43}
Planet (Pluto size)	1.3×10^{22}	4×10^{45}
Planet (Earth size)	6×10^{24}	2×10^{48}
Planet (Jupiter size)	2×10^{27}	6×10^{50}
Star (1 solar mass)	2×10^{30}	6×10^{53}
Star (100 solar mass)	2×10^{32}	6×10^{55}

Table 1: Mass of various objects and energy of X-ray flares from Sgr A* when it accretes these objects.

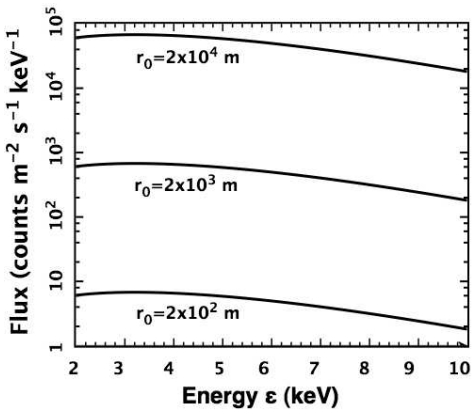

Fig. 5: The spectral flux of dynamic massive black hole Sgr A* as a function of radiation photon energy.

flares (about hundred times more luminous than the steady emission) from Sgr A* at a rate of once a day. Occasionally accreting of large-sized asteroids or pluto-sized planets can explain the X-rays outbursts of a million times brighter than the normal emission of Sgr A*, which occurred during the past few hundred years. For big size planets, this may also explain the X-ray outbursts from some other inactive galactic centers. In future, when Sgr A* accretes a star including neutron star daily (or yearly for a large star), an active galactic nucleus (AGN) or quasar will form or is born in our galaxy. It should be noted that the G2 cloud with 3 Earth masses, if it is accreted by Sgr A*, will produce a super X-ray flare, billions times brighter than the normal emissions.

The spectral energy flux $S(\nu, T)$ of the blackbody radiation from a dynamic black hole can be determined by,

$$S(\nu, T) = cu(\nu, T). \tag{14}$$

Dividing the spectral energy flux $S(\nu, T)$ by the energy of photon, we have the spectral photon flux as,

$$J(\nu, T) \equiv \frac{S(\nu, T)}{h\nu} = cf(\nu, T). \tag{15}$$

For the radiation observed at the Earth, the spectral flux of an X-ray flare produced by the dynamic massive black hole Sgr A*, when it accretes an object, is given by,

$$J(\nu, T) = cf(\nu, T)\left(\frac{r_0}{d_L}\right)^2, \tag{16}$$

where d_L is the luminosity distance and r_0 is the radius of radiation area, which is the area of the horizon broken. The temperature T of Sgr A* can be estimated, according to Eq. (5), by

$$T = T_s\left(\frac{R_s}{R}\right)^{3/4} = T_s\left(\frac{c^2 R_s}{2GM}\right)^{3/4}, \tag{17}$$

where M is the mass of Sgr A* and equals to about 4.5 million solar masses. As mentioned above or in [25–26], R_s is the radius of the three-solar-mass black hole and is equal to ~ 8.89 km; T_s is the temperature of the three-solar-mass black hole and is usually chosen to be around one trillion Kelvins, i.e. $T_s \sim 10^{12}$ K. Then we have the temperature of inside Sgr A* is $T \sim 2.3 \times 10^7$ K.

For the massive black hole Sgr A*, $d_L \sim 2.46 \times 10^{20}$ m or 26,000 light-years. The radius of radiation area r_0 can be considered to be about the radius of the object accreted by Sgr A* times a factor, $r_0 = br$. The factor b is equal to the unity if the full area of radiation faces towards to the observer or the Earth. Otherwise, we have $b < 1$ or $r_0 < r$. In addition, since the object is usually broken by the tidal force during the falling, the factor b should be smaller. An X-ray flare occurred at the opposite side of Sgr A* cannot be directly observed by an observer on the Earth. In this case, the factor b is zero. The 400 brighter than normal emission X-ray flare caught by Chandra on September 14, 2013 flares its X-rays in the upright direction according to the image [9,19]. Considering that an asteroid, whose density is usually given by about 2000 kg/m^3, has mass of 10^{17} kg, we can find its radius $r \sim 23$ km and choose r_0 equal or less than 23 km. Figure 5 plots the average spectral flux of an X-ray flare from Sgr A* as a function of the X-ray photon energy. In this plot, we have chosen $r_0 = 200, 2000, 20000$ m, respectively, and $T_s = 10^{12}$ K. It is seen that the spectral flux of X-ray flares from Sgr A*, according to this new mechanism, increases with r_0. Increasing T_s also increases the spectral flux especially in high energy end. Quantitatively, the spectral flux of X-ray flares from Sgr A* obtained from this study as emissions of dynamic massive black hole can be consistent with the measurements [39].

4 Discussion and conclusion

According to this new mechanism, the duration or time scale of an X-ray flare is the time needed for the broken horizon to be recovered. It depends on the size of the object accreted and also the rate or speed of matter diffusion. In general, the bigger the events are, the longer the flares can last, which agrees

with the measurements. The rate of matter diffusion depends on the state of matter. The rate of diffusion is faster if the matter is hotter and/or less dense. The falling of the object is usually dissipated due to radiation of lower frequencies such as near infrared as measured usually prior to the X-ray flares. In addition, the observed spectral flux of X-ray flares from Sgr A* may be significantly affected by the gravitational redshift. In future, we will address these issues in more details.

We have developed a new mechanism for X-ray flares from Sgr A*, in accordance with the black hole model of the universe that Zhang [21–22] recently proposed. According to this new mechanism, we can understand X-ray flares from Sgr A* as emissions of dynamic massive black hole at the Milky Ways center. A black hole (from star-like with several solar masses through supermasive with billions of solar masses), when accreting matter, becomes dynamic and breaks its event horizon, which leads to the inside hot (or high-frequency) blackbody radiation leaking out of it and produces an X-ray flare or burst. We calculate the energies and spectra of X-rays emitted by the galactic center massive black hole when various sized objects from asteroids through comets and planets to stars fall into Sgr A*. Then, through these calculations, we explain the current measurements of X-ray flares from Sgr A* including its steady emissions, predict big events that possibly occurred in the past or will possibly occur in future at our galactic center, and compare the predicted intensive events with the measurements of strong X-ray flares from other normal and active galactic centers. This study develops a possible mechanism for the origin of the X-ray flares from galactic centers and deepens our understanding to the black hole dynamics, galactic activities, and cosmological evolutions.

Acknowledgments

This work was partially supported by the NSF HBCU-UP, REU, and AAMU Title III programs.

Submitted on December 10, 2015 / Accepted on December 12, 2015

References

1. Balick B., Brown R.L. Intense sub-arcsecond structure in the galactic center. *Astrophysics and Space Science*, 1974, v. 194, 265–270.
2. Schödel R. et al. A star in a 15.2-year orbit around the supermassive black hole at the centre of the Milky Way. *Nature*, 2002, v. 419, 694–696.
3. Gillessen S., Eisenhauer F., Trippe S., Alexander T., Genzel R., Martins F., Ott T. Monitoring stellar orbits around the massive black hole in the galactic center. *Astrophysical Journal*, 2009, v. 692, 1075–1109.
4. Degenaar N, Miller J.M., Kennea J., Gehrels N., Reynoids M.T., Wijnands R. The X-ray flaring properties of Sgr A* during six years of monitoring with Swift. *Astrophysical Journal*, 2013, v. 769, article id. 155, 7 pages.
5. Baganoff F.K. et al. Rapid X-ray flaring from the direction of the supermassive black hole at the Galactic Centre. *Nature*, 2001, v. 413, 45–48.
6. Porquet D. et al. XMM-Newton observation of the brightest X-ray flare detected so far from Sgr A*. *Astronomy and Astrophysics*, 2003, v. 407, L17–L20.
7. Neilsen J. et al. A Chandra/HETGS census of X-ray variability from Sgr A* during 2012. *Astrophysics Journal*, 2013, v. 774, article id. 42.
8. Nowak M.A. et al. Chandra/HETGS observations of the brightest flare seen from Sgr A*. *Astrophysical Journal*, 2012, v. 759, article id. 95.
9. Haggard D. et al. An update on Chandra/VLA galactic center campaigns targeting Sgr A* and G2. *American Astronomical Society 225th Meeting*, 2015, Abstract #102.09.
10. Clavel M. et al. The reflection of two past outbursts of Sagittarius A* observed by Chandra during the last decade. *Proceedings of the International Astronomical Union*, 2014, v. 303, 344–348.
11. Komossa S., Greiner J. Discovery of a giant and luminous X-ray outburst from the optically inactive galaxy pair RX J1242.6-1119. *Astronomy and Astrophysics*, 1999, v. 349, L45–L48.
12. Saxton R.D. et al. An X-ray and UV flare from the galaxy XMMSL1 J061927.1-655311. *Astronomy and Astrophysics*, 2014, v. 572, article id. A1, 9 pages.
13. Bahcall J.N., Kirhakos S., Saxe D.H., Schneider D.P. Hubble space telescope images of a sample of 20 nearby luminous quasars. *Astrophysical Journal*, 1997, v. 479, 642–658.
14. Begelman M.C., King A. R., Pringle J. E. The nature of SS433 and the ultraluminous X-ray sources. *Monthly Notices of the Royal Astronomical Society*, 2006, v. 370, 399–404.
15. Zhang T.X. Quasar formation and energy emission in black hole universe. *Progress in Physics*, 2012, v. 8, issue 3, 48–53.
16. Kusunose M., Takahara F. Synchrotron blob model of infrared and X-ray flares from Sagittarius A*. *Astrophysical Journal*, 2011, v. 726, article id. 54, 6 pages.
17. Yusef-Zadeh F. et al. An Inverse compton scattering origin of X-ray flares from Sgr A*. *Astronomical Journal*, 2012, v. 144, article id. #1, 10 pages.
18. Liu S.M., Fulvio M., and Yahe F. Stochastic electron acceleration during the near-infrared and X-ray flares in Sagittarius A*. *Astrophysical Journal*, 2006, v. 636, 798–803
19. Sagittarius A*: NASA Chandra Detects Record-Breaking Outburst from Milky Way's Black Hole (2015). Accessed from: http://chandra.harvard.edu/photo/2015/sgra
20. Li Y.P. et al. Statistics of X-Ray Flares of Sagittarius A*: Evidence for Solar-like Self-organized Criticality Phenomena. *Astrophysical Journal*, 2015, v. 810, article id. 19, 8 pages.
21. Zhang T.X. A new cosmological model: Black hole universe. *American Astronomical Society 211st Meeting*, 2007, Abstract #152.04.
22. Zhang T.X. A new cosmological model: Black hole universe. *Progress in Physics*, 2009a, v. 2, 3-11.
23. Zhang T.X. Cosmic microwave background radiation of black hole universe. *Astrophysics and Space Science*, 2010a, v. 330, 157-165
24. Zhang T.X. Key to the mystery of dark energy: Corrected relationship between luminosity distance and redshift. *Progress in Physics*, 2013, v. 5, issue 3, 1–6.
25. Zhang T.X. A new mechanism for gamma ray bursts: Emissions of dynamic black holes. *Proceeding of Gamma Ray Burst 2013 Symposium*, 2013b, SNSN-323-63.
26. Zhang T.X. Gamma ray bursts of black hole universe. *Astrophysics and Space Science*, 2015, v. 358, article id. #14, 8 pages.
27. Zhang T.X., Frederick C. Acceleration of black hole universe. *Astrophysics and Space Science*, 2014, v. 349, 567–573.
28. Zhang T.X. Anisotropic expansion of the black hole universe. *American Astronomical Society 213rd Meeting*, 2009b, Abstract #357.03.
29. Zhang T.X. Cosmic microwave background radiation of black hole universe. *American Astronomical Society 214th Meeting*, 2009c, Abstract #303.01.

30. Zhang T.X. Observation evidences of black hole universe. *American Astronomical Society 215th Meeting*, 2010b, Abstract #313.06.

31. Zhang T.X. Black hole universe model and dark energy. *American Astronomical Society 217th Meeting*, 2011, Abstract #404.05.

32. Zhang T.X. Mechanism for gamma-ray bursts and black hole universe. *American Astronomical Society 219th Meeting*, 2012b, Abstract #310.02.

33. Zhang T.X. Acceleration of black hole universe. *American Astronomical Society 220th Meeting*, 2012c, Abstract #321.07.

34. Zhang T.X. Key to the mystery of dark energy: Corrected relationship between luminosity distance and redshift. *American Astronomical Society 221st Meeting*, 2013c, Abstract #323.06.

35. Zhang T. X. Evidences for supporting the black hole universe model. *American Astronomical Society 222nd Meeting*, 2013d, Abstract #103.02.

36. Zhang T.X. Black hole universe model for explaining GRBs, X-ray flares, and quasars as emissions of dynamic star-like, massive, and supermassive black holes. *American Astronomical Society 223rd Meeting*, 2014a, Abstract #427.05.

37. Zhang T.X. The black hole universe model. *American Astronomical Society 224th Meeting*, 2014b, Abstract #304.03.

38. Yakovlev D.G., Gnedin O.Y., Kaminker A.D., Levenfish K.P., Potekhin A.Y. Neutron star cooling: Theoretical aspects and observational constraints. *Advances in Space Research*, 2004, v. 33, 523–530.

39. Barriére N.M. et al. NuSTAR detection of high-energy X-ray emission and rapid variability from Sagittarius A* flares. *Astrophysical Journal*, 2014, v. 786, article id. 46, 10 pages.

On an Apparent Resolution of the Catt Question

Stephen J. Crothers

Tasmania, Australia. E-mail: steve@plasmaresources.com

Over a number of years there have been some attempts to answer the Catt Question within the context of classical electromagnetic theory. None of the authors of these attempts agree on the answer to the Catt Question, even though they all invoke the very same theory. An attempt at answering the Catt Question appeared in the journal *Physics Education* in 2013, penned by M. Pieraccini and S. Selleri, as a mathematical rendition of their earlier non-mathematical version published in *IEEE Antennas and Propagation Magazine*, 2012. The explanation by these two Authors contains violations of classical electromagnetic theory, although they claim to have satisfactorily answered the Catt Question by means of classical electromagnetic theory. The arguments adduced by Pieraccini and Selleri are therefore invalid.

1 Introduction

In their article [1] "An apparent paradox: Catt's anomaly", the Italian authors Pieraccini and Selleri* refer to the Catt Question as "*Catt's Anomaly*". Their earlier paper is titled 'Catt's Anomaly' [2]. Although until 2001 "The Catt Question" was called "The Catt Anomaly", it was in fact *always* a question, to be answered.

The Catt Question [3] pertains to the propagation of a Transverse Electromagnitic (TEM) wave along a transmission line. Upon closure of a switch, the TEM wave (step) travels at the speed of light between the conducting wires of the transmission line, from battery to load, as depicted in Fig. 1.

An electric field **E** appears between the conductors, directed from the top wire to the bottom wire. This electric field is orthogonal to the two parallel wires and moves towards the load; thus there are positive charges on the top conductor and negative charges on the bottom conductor in the region of the transverse electric field. The Catt Question is: *Where does this new charge come from?* [3].

2 Electron current

According to classical electromagnetic theory and circuit theory, electric current in metallic wires is the flow of electrons in the wires (conductors), and a magnetic field is generated around the conducting wires according to the Right-Hand Rule. Since the TEM step travels at the speed of light towards the load, how does the current in the conducting wires keep pace with the TEM wave, if electrons cannot travel at the speed of light? The Authors [1] give the following answer,

> "*The key idea of the explanation of this apparent paradox is related to the great number of electrons in metal. Although each single electron is not able to travel at the speed of light, a great number of slow electrons are able to produce a current as fast as an electromagnetic wave travelling at the speed of light in the conductor.*"

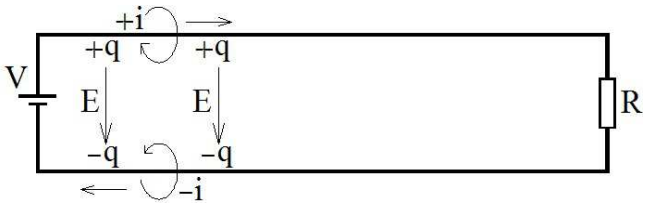

Fig. 1: An electric field points directly from the top conductor to the bottom conductor (from positive charge to negative charge). It is therefore orthogonal to the top and bottom parallel conductors. The transverse electric field travels from battery to load at the speed of light, subject to the dielectric medium between the wires.

What do they mean by "current"? They say here that electrons "*produce a current*". However, the Authors actually assume the classical electron flow along wires as the meaning of electric current in wires, and claim that this current travels along the conductors at the speed of light even though the drift speed of electrons in the wires is a snail's pace (e.g. 2mm/s in 1.0mm copper wire [1]). Strangely, the flow of electrons, although very slow, produces an electron current that is "*as fast as an electromagnetic wave travelling at the speed of light in the conductor*" [1]: after all, a current of electrons is an electron current. This impossible duality occurs, they say, because the free electron density in the conductors is very high, and they derive an equation for electron drift "velocity".

Electron drift velocity in a wire is proportional to the vector electric field $\mathbf{E}_w$ in the wire, which supposedly causes the electron drift,

$$\mathbf{v} = -\mu\mathbf{E}_w \tag{0}$$

and so the electron drift velocity and the electric field in the wire are collinear but point in opposite directions. The constant of proportionality μ is called the *mobility*.

*Massimiliano Pieraccini, Associate Professor, Department of Electronics and Telecommunications, University of Florence; Stefano Selleri, Assistant Professor, University of Florence.

The Authors begin with the following equation for electron current,

$$I = \pi a^2 v q N \,, \tag{1}$$

where a is the radius of the conductors, v "*the drift velocity of the charges (in practice electrons, and the speed is much lower than the speed of light)*" [1]*, q the elementary charge, and N the free electron density in the conductors.

Since the current, they say, travels at the speed of light, in time $\Delta t = \Delta x/c$ they obtain a passage of charge ΔQ along the top conductor, given by,

$$\Delta Q = I \Delta t = I \frac{\Delta x}{c} \,, \tag{2}$$

where Δx is the distance travelled by the TEM step in time Δt. This charge ΔQ the Authors call "*an imbalance of charge*" [1] because they say it is confined to a leading volume element of length Δx in the top conducting wire, and induces equal but opposite polarity charge on the bottom conducting wire.

Using a cylindrical Gaussian surface they next apply Gauss' Law to calculate the magnitude E of the electric field $\mathbf{E}$ due to ΔQ in the top conductor,

$$\frac{\Delta Q}{\epsilon_0} = (2\pi a \Delta x)\, E \,, \tag{3}$$

where ϵ_0 is the permitivity of free space. Substituting ΔQ from equation (2) and I from equation (1) the Authors obtain,

$$v = \frac{2c\epsilon_0 E}{qNa} \,. \tag{4}$$

From equation (4) they conclude,

> "*The notable point of this result is that the necessary speed decreases with the number of electrons per volume unit N. Therefore, a great number of slow electrons are able to generate enough unbalanced charge to follow an electromagnetic wave travelling at much higher speed.*"

Thus electrons flow slowly in the conducting wires but the electron current in the wires is nevertheless flowing along the conductors at the speed of light.

Although equation (4) follows from equations (1), (2) and (3) by purely mathematical operations, the transverse electric field $\mathbf{E}$ cannot drive electrons along the inside or outside of the wires. Equations (1), (2) and (4) imply flow of electrons along the wires, but the transverse electric field at equation (3) is orthogonal to the parallel axes of the top and bottom wires. According to classical electrodynamics, free electrons in a metallic conductor flow in the direction opposite to the direction of the electric field, according to equation (0), not orthogonal to the electric field ($\mathbf{E} \neq \mathbf{E}_w$). The Authors confound battery EMF† with the transverse electric field, and so make the transverse electric field the battery EMF to drive electrons along the wires; at equation (4).

Then they introduce the "skin effect" [1]:

> "*Up to this point, the current has been considered constant in the wire section, but in reality the current flow tends to be bound to the portion of the conductor closer to the surface.*"

The equation for current in the wires they then give as,

$$I = 2\pi a \delta v q N \,, \tag{1b}$$

where δ is the skin depth, which is frequency dependent. With the "skin effect" they still argue that electrons flowing along the wire is electric current, orthogonal to the electric field they calculated at equation (3), and continue to make that transverse electric field the driver of the electrons in the conducting wires. Using equations (1b), (2) and (3) they then obtain the electron drift speed,

$$v = \frac{2c\epsilon_0 E}{qN\delta} \tag{5}$$

although the 2 in the numerator should not in fact appear.

3 Conclusion

Pieraccini and Selleri have not answered the Catt Question. On the one hand they treat current in the conducting wires as electron current but on the other hand they invoke the transverse electric field between the conducting wires to drive this electron current at the electron drift speed. Their analysis violates the classical electromagnetic theory they use in their attempt to prove that what they call "*Catt's Anomaly*" is merely an "*apparent paradox*" [1]. The real paradox here is their claim that very slowly flowing electrons in the wires of a transmission line produce an electron current in those wires that travels at the speed of light, driven by an elecric field orthogonal to those wires.

> "*If I have promised to deliver one dozen eggs to Oxford, one hour from now, Oxford being 100 miles away, there is no point in despatching ten dozen eggs in a vehicle which travels at only ten miles/h*" [4].

Submitted on December 13, 2015 / Accepted on December 14, 2015

References

1. Pieraccini M. and Selleri S. An apparent paradox: Catt's anomaly. *Physics Education*, 2013, v. 48 (6), 718–722.
2. Pieraccini M. and Selleri S. Catt's anomaly. *IEEE Antennas Propag. Mag.*, 2012, v. 54, no. 6, 242–244.
3. Catt I. The Catt Question. http://www.ivorcatt.co.uk/x54c1.htm
4. Catt I. The death of electric current. *Wireless World*, December 1982, 79–81. Accessed from http://www.ivorcatt.co.uk/x18j100.pdf

*The Authors confound velocity with speed; the latter denoted by $|\mathbf{v}| = v$

†What EMF is, is another question.

LETTERS TO PROGRESS IN PHYSICS

The Roland De Witte Experiment, R. T. Cahill, and the One-Way Speed of Light

Joseph Catania
E-mail: jcatania1@verizon.net

In "The Roland De Witte 1991 Experiment (to the Memory of Roland De Witte)" (*Progr. Phys*, 2006, v. 2(3), 60–65), R.T. Cahill gives us a briefing on his view that interferometer measurements and one-way RF coaxial cable propagation-time measurements amount to a detection of the anisotropy in the speed of light. However, while I obtain first order propagation delays in calculations for one-way transit which would show geometric modulation by Earth's rotation, I do not agree with Cahill's simplistic equation that relates the modulation solely to the projection of the absolute velocity vector **v** on the coaxial cable, called v_P by Cahill (ibid., p. 61–62). The reader should be warned that Cahill's equation for Δt (ibid., p. 63) is crude compared with a full Special Relativistic derivation.

1 Introduction

In *The Roland De Witte 1991 Experiment (to the Memory of Roland De Witte)* [1], R. T. Cahill gives us a briefing on his view that interferometer measurements and one-way RF coaxial cable propagation-time measurements amount to a detection of the anisotropy in the speed of light. This startling conclusion is difficult to swallow in the face of rigorous light speed in vacuo measurements which are reproducible and flaunt good experimental controls. For instance, in [2] Eisele et. al. were able to limit anisotropy in c to a fractional uncertainty of 10^{-17}. It would seem apparent that, to this precision, there is no first **or** second order anisotropy in the two-way speed of light.

2 The one-way speed of light

As regards the one-way speed of light, a point of confusion in regard to spurious claims of anisotropy might be exemplified by measurements with the Global Positioning Satellite (GPS) system, which can measure the rotational speed of the Earth, v, by the way it affects the propagation time of an electromagnetic signal used in the GPS system [3]. Thus, the *apparent* velocities $c+v$ and $c-v$ would be measured instead of c. But, certainly, GPS is not to be interpreted as capable of measuring c itself. As further clarification, let us say that, through some means I could set a train moving at 20 miles per hour along a railroad track in a due Easterly direction. At some point on the track to the East of the train I have stationed a measurement instrument which reads exactly 20 mph. If I now move this measuring instrument in an Easterly direction at 5 mph I should only measure the train speed as 15 mph. If I give the measuring instrument a Westerly motion of 5 mph, I should measure for the train 25 mph. Most of us have an intuitive familiarity with this situation. In no way should there be a temptation to assign the 15 or 25 mph speed to the train velocity which is obviously 20 mph. We should not confuse actual velocity with apparent velocity. Likewise, one-way propagation times of electromagnetic signals cannot be used to calculate c, which has already been assumed constant, but they would be useful in calculating the v in $c+v$ or $c-v$, if the distance of propagation were known.

Similarly, the Michelson-Morley interferometer measurements Cahill refers to in [1] were not developed to measure the speed of light, c, but to measure relative motion, v to a postulated luminiferous ether. That Cahill admits this measurement of v was successful [4] on the one hand would seem to defy his light speed anisotropy conclusion on the other. So, I find it difficult to reconcile propagation time calculations used in interferometer measurements which assume c, a well-known constant of nature, as the speed of light in vacuo, and the explicit solution for the variable v, the motion with respect to the ether, with light-speed anisotropy in any form.

3 First order effects

Nevertheless, as pointed out, there are fringe-shifts measured in many interferometers and there is De Witte's propagation time delay (which is correlated to sidereal time). It has been established in Michelson-Morley type interferometer measurements that there is a correlation of measurements of **v** with cosmic velocity (similar to the CMB dipole velocity) accompanied with amplitude modulations with respect to rotation and revolution of the Earth. This is expected on the basis of current theory which explains fringe-shifts in interferometers as due to dielectric in the light path (no fringe-shifts are expected in vacuum interferometers) [4]. However, while I obtain first order propagation delays in calculations for one-way transit which would show geometric modulation by Earth's rotation, I do not agree with Cahill's simplistic equation that relates the modulation solely to the projection of the absolute velocity vector **v** on the coaxial cable, called v_P by Cahill [1, p. 61–62]. The reader should be warned that

Cahill's equation for Δt [1, p. 63] is crude compared with a full Special Relativistic derivation. Also the period of the modulation based on a fixed absolute motion vector in the Miller direction would not be 12 sidereal hours but 24 as can be plainly seen from the geometry. Also apparent from the geometry is that Cahill's v_P would never go negative and indeed does not attain zero. In fairness Cahill states in (ibid., p. 63) that DeWitte's data is plotted with a false zero making the periodicity appear to be 12 hours sidereal. As well, there does not seem to be sufficient support of Cahill's use of $n = 1.5$ for De Witte's coaxial cable. It's more likely that $\epsilon = 1.5$.

4 Conclusion

In conclusion I can only say that although Cahill understands De Witte's result is first order and shows correlation to the Miller direction we must be cautious in ascribing this result to unconfirmed phenomena such as light speed anisotropy especially since SR would seem to be an apt predictor of the effect.

Submitted on December 30, 2015 / Accepted on December 31, 2015

References

1. Cahill R.T. The Roland De Witte 1991 Experiment. *Progress in Physics*, 2006, v. 2, issue 3, 60–65.
2. Eisele Ch., Nevsky A.Yu., Schiller S. Laboratory test of the isotropy of light propagation at the 10^{-17} level. *Physical Review Letters*, 2009, v. 103 (9), 090401.
3. Gift S. J. G. One-way speed of light relative to a moving observer. *Applied Physics Research*, 2013, v. 5, no. 1.
4. Cahill R.T. and Kitto K. Michelson-Morley experiments revisited. *Apeiron*, 2003, v. 10(2), 104–117.

A Non-anthropic Solution to the Cosmological Constant Problem

Robin James Spivey

Biological Sciences, Bangor University, Brambell, Deiniol Road, Bangor, Gwynedd, Great Britain
E-mail: y.gofod@gmail.com

Accelerating cosmological expansion is driven by a minuscule vacuum energy density possibly seeking opportunities to decay to a true ground state. Quasar characteristics imply their central engines possess an intrinsic magnetic field compatible with the presence of an electrically charged toroidal dark hole, an eternally collapsing structure lacking an event horizon. The possibility is consistent with the inability of black holes to capture particles in a universe of finite age, Einstein's dismissal of the Schwarzschild metric as unphysical and the implausibility of the various paradoxes invoked by black hole existence. The uncloaked innards of these dark holes would expose immense vacuum accelerations at their cores, inevitably tempered by Planck scale physics. The Unruh effect predicts that intense yet highly localised heating should occur there. As thermal energy gradually amasses and dissipates, radiation would eventually start to escape into the surrounding environment. Virtual from the dark hole perspective, the emissions could not decrease the dark hole's mass: the energy source must instead be the universal vacuum, the likely repository of dark energy. In analogy with core-collapse supernovae, neutrinos should dominate the cooling flows. Red-shifting to low energies upon escape, quantum degenerate haloes should form predominantly around the largest galaxies. This mechanism is promising from the perspective of enabling the future universe to efficiently sustain aquatic life before stars become scarce, offering a biological yet decidedly non-anthropic solution to the cosmological constant problem.

1 Introduction

Despite tremendous interest in the composition, distribution and interactions of dark matter particles, the existence of only one of the candidates presently transcends speculation. This accolade belongs to the neutrino — a fermion which, by virtue of its non-zero mass [1], is capable of gravitational condensation to form quantum degenerate galactic haloes [2]. With cosmological constraints already implying hierarchical neutrino mass eigenstates, the similarity of $kT_{H_2O(aq)}$ and $|\Delta m_{13}|c^2$ is most striking. Neutrino oscillations require physics beyond the Standard Model but renormalisable extensions likely demand the existence of sterile varieties. Intriguingly, these facts could be hinting at the perpetuation of advanced aquatic lifeforms well beyond the stelliferous era [3].

Dark matter was recently overshadowed by the discovery of dark energy, a yet more pervasive and enigmatic phenomenon causing universal expansion to accelerate. Its spatial energy density is some 120 orders of magnitude smaller than quantum physics can comfortably explain [4]. Although dark energy's influence is locally imperceptible it dominates the cosmos already [5,6] and consequently represents a formidable new frontier in cosmology. Parallels can be drawn with theories of cosmic inflation, whose accelerating expansion purportedly terminated as an underlying energy field decayed into high energy particles. Whereas Mercury's orbital peculiarities provided both an impetus for Einstein's development of general relativity and a means of experimentally validating corrections to Newtonian mechanics, dark energy is far more inscrutable. Thus, insights of any kind are potentially valuable and merit careful investigation.

The goal of this work is to revisit the cosmological constant problem following the advancement of a novel model of the universe predicting the future decay of dark energy. This framework happens to incorporate the first scientific hypothesis concerning the long-standing mystery of extraterrestrial silence, yielding testable predictions for particle physics [3]. There is a very real prospect that the future universe might sustain aquatic life for $\sim 10^{25}$ years in certain locales via the annihilation of gravitationally condensed neutrinos within hexagonally close-packed iron (hcp-Fe), a material that dominates the cores of oceanic planets up to $\sim$ 15 $M_\oplus$ [7]. Active neutrinos may well have sufficient mass to maintain liquid oceans since oscillations [8] and cosmological considerations [9] imply that Σm_ν lies in the range 58–230 meV. Moreover, the hcp/fcc boundary in iron's phase diagram conveniently lends itself to planetary thermoregulation almost independently of planet size [3]. Key to the scenario is the finding that the cosmic abundance of neutrinos must first be hugely augmented, implicating the future decay of dark energy primarily to galaxy-engulfing active neutrino halos of a mass approaching 10^{21} $M_\odot$ within ~60 Gyr [10].

This particular line of cosmological investigation has not previously succeeded in venturing any suggestion as to a physical process by which dark energy might decay to neutrinos, an eventuality implied by the propensity of neutrinos to sustain aquatic life with remarkable efficiency [3]. Other avenues of enquiry have similarly failed to pinpoint specific

mechanisms for vacuum discharge capable of ending the current phase of cosmic acceleration, although a model of dark energy interacting with a neutrino-like fermion field has been considered [11]. The present approach draws heavily on developments in black hole research and observational cues from the contrasts between active and inactive galactic nuclei. A promising mechanism for dark energy discharge shall be identified here but its quantitative analysis is likely to remain challenging for some time, in large part due to the continuing lack of a theory of quantum gravity and knowledge of physics at the highest energies. The concluding discussion reflects upon the capability of this mechanism to fulfil its cosmological motivations and thereby offer a radical new approach to comprehending the minute energy density of the vacuum.

2 Theoretical motivations

2.1 The physics of biology

Life is reliant on complex biochemical interactions involving only three subatomic particles whose arrangements are stabilised by only two forces: electromagnetism and the strong interaction. The neutron is marginally more massive than the combined rest mass of a proton and an electron, allowing protons and neutrons to coexist without prohibiting the formation of neutron degenerate matter within dense stars. The strong interaction conveniently allows the assembly of heavy atomic nuclei despite intense electromagnetic repulsion between protons. Of the elements up to lead, only three lack unconditionally stable isotopes, yet most possess only one or two stable isotopes. Remarkably minor adjustments to several physical constants could radically shorten the periodic table or rule out chemistry altogether. Space might be populated mainly by neutron stars and black holes. Stars might be incapable of nuclear fusion, too short-lived to support complex evolutionary processes or so dim that planets orbiting within their habitable zones soon become tidally-locked. Supernovae might never scatter the ashes of stars into space, so that their ejecta might form elements necessary for planets and life.

Ascertaining why nature's constants might possess the values they do has been traditionally regarded as the preserve of mathematical physics — yet the approach has met with little success. One should therefore remain open to alternative possibilities. Due to the improbable compatibility of the physical laws with long-term biological evolution the 'anthropic principle' has been advanced. Although the universe existed well before life on Earth commenced, our existence imposes retrospective constraints on the physical laws and the natural constants. However, the anthropic principle does not allow one to conclude that physics would have been any different had chance chemical interactions never led to life on this planet. Furthermore, the Copernican revolution provides a historical precedent that the innate sense of human self-importance does not always provide a reliable foundation for cosmological extrapolation. Moreover, appeals to the precondition of human existence are at odds with a multitude of life-promoting characteristics in nature falling comfortably outside the gamut of the anthropic principle.

Nevertheless, the traditional perception has been that the universe is rather ill-suited to life. The Earth's living organisms only harness some 0.1% of the insolation, in turn amounting to just one billionth of the Sun's total radiation. No star liberates more than 0.008% of its rest mass energy through fusion processes. The Sun burns hydrogen to helium, yet 90% of its hydrogen will remain by the time it becomes a red giant. Planets orbiting low mass red dwarves are never habitable for very long. These considerations seem to paint a picture of a universe largely inhospitable to life. However, there is now reason to believe that impression was premature. Active neutrinos may be capable of internally heating iron-cored oceanic planets on galactic scales, sustaining aquatic life long after the stars have died with impressive efficiency [3, 10]. The fact that a technological species has evolved on this planet provides no plausible explanation for this, and neither does happenstance.

For oceans to be maintained in a liquid state by neutrino annihilation, haloes are required of a mass approaching the threshold for gravitational implosion, some 4∼7 orders of magnitude larger than the mass of a galaxy cluster. Synthesis of the available information points to dark energy decaying at a suitable juncture predominantly to active neutrinos that form dense haloes. This is expected somewhat prior to the disappearance of the last stars capable of cultivating life on orbiting planets — when the universe is approximately five times its present age. Accordingly, the continuity of life need not be endangered and there would be ample time for the evolution of technologically and ethically advanced colonising civilisations before widespread colonisation could be attempted.

The former solar neutrino anomaly was resolved when it was found that neutrinos undergo spontaneous flavour oscillations [1], demonstrating their possession of mass. The diminutive neutrino mass scale closely coincides with the energy scale associated with the temperature of liquid water. Furthermore, it is small enough to ensure that neutrinos can condense under gravity to form galaxy-enveloping structures supported by fermionic quantum degeneracy [3, 10].

The likelihood of a neutrino mutually annihilating with other neutrinos depends on the ambient neutrino concentration, but the probability of a neutrino scattering with nucleons does not. In a dense neutrino halo, annihilation events can be frequent in the presence of hcp-Fe at temperatures compatible with the presence of a 4s electron receptive to some of the annihilation energy [3, 7]. Whilst even high energy neutrinos can travel through light years of lead without scattering, low energy neutrinos are unlikely to emerge from an iron-cored planet without annihilating if the planet is immersed in a sufficiently dense neutrino halo. If, as cues from cosmology

and oscillation experiments suggest, the neutrino mass scale lies in the vicinity of ∼0.05 eV, a halo density of just one picogram per cubic kilometre can sustain liquid oceans. A neutrino mass just one order of magnitude smaller would be incapable of maintaining liquid oceans, even with assistance from a thick insulative crust of ice.

2.2 Biotic reasoning

Appreciation of the inadequacy of the weak anthropic principle as an explanation for the fine-tuning of physics inspired an investigation into whether dark matter particles might be capable of sustaining aquatic life. This led to the discovery that neutrino annihilation is capable of targeting 4s electrons in hcp-Fe, a phase transition in iron providing a natural thermoregulation mechanism as the 4s electrons transfer to the 3d subshell, assuring that the thermal flux through a subglacial ocean is essentially independent of planetary mass for relatively dense, rocky planets [3, 10].

Being polytropic, a neutrino halo expands upon depletion and, due to the resulting decline in neutrino concentration, the heating capacity eventually falls below that needed to maintain liquid oceans. A sizeable fraction of the halo energy might thereby go to waste. The energy of a neutrino halo approaching the gravitational implosion limit is inversely related to the mass of an individual neutrino. Hence, a smaller neutrino mass might support aquatic life for longer. This likely explains why the neutrino mass scale is at least one order of magnitude lower than required merely for haloes to fully surround a galaxy — reducing their ambient concentration, yet not to the degree that aquatic life cannot be maintained. Although this permits a lengthy aquatic era, the wastage this incurs as the aquatic era ends is not insignificant. This may be mitigated by another consideration, one that is potentially relevant to the current composition of dark matter.

Half the Earth's atmosphere is concentrated at altitudes below 6 km, less than 0.1% of the planet's radius. If the mass of the Earth were somehow abruptly reduced, say to the mass of the Moon, the atmospheric scale height would increase a hundred-fold. Many species, including our own, would soon die of asphyxiation. The gravitational load on the Earth's atmosphere is clearly vital to our minute-by-minute survival. By analogy, if oceanic planets are pictured as inhaling neutrinos and exhaling infrared photons, gravitationally loading an excessively large halo could locally boost the neutrino concentration over galactic scales. This could be very useful at late times when the neutrino halo would otherwise be quite rarefied within the galaxy. An inner halo of relatively low mass, roughly twice the diameter of the contained galaxy but of far greater mass than the galaxy itself, would apply an effective additional load. Ideally, this auxiliary halo would also support its own weight through fermionic repulsion but its constituent particles would be highly inert, virtually immune to all forces except gravity.

The weak interaction maximally violates parity so that right-handed particles and left-handed antiparticles are insensitive to it. Hence, particles resembling conventional neutrinos but having opposite chirality and a somewhat larger mass would be advantageous. Prior to the realisation that such particles could be biologically useful, anomalies in neutrino oscillation experiments were already alluding to the existence of sterile neutrinos at the eV-scale [12, 13]. Furthermore, gravitational lensing data for the Abell 1689 galaxy cluster strongly hinted at the presence of a cloud of degenerate 1.5 eV fermions [14, 15], inconsistent with cosmological constraints on active neutrinos but in keeping with the expectation that eV-scale sterile neutrinos would be well-suited to concentrating active neutrinos on galactic scales [3, 10].

Whilst the discovery of sterile neutrinos has not yet been formally announced and their mass remains loosely constrained, the statistical evidence for their existence already stands at 3.8σ. Active neutrinos may well have sufficient mass to maintain liquid oceans since $58 < \Sigma m_\nu < 230$ meV [16]. Moreover, the hcp/fcc boundary in iron's phase diagram beautifully lends itself to planetary thermoregulation in a manner almost independent of planet size [7, 10]. This picture testifies to the utility of biotic reasoning: a cohesive new approach to cosmology has emerged that dispenses with unsatisfactory anthropic explanations for fine-tuning and yields the first scientific resolutions of the Fermi paradox [3]. Before proceeding to apply similar logic to dark energy decay, attention shall be drawn to some other pertinent considerations.

2.3 Inferences and expectations

The potential sustainment of aquatic life by neutrinos annihilating within iron-cored oceanic planets would be sufficiently efficient as to bear the hallmarks of cosmic design, in turn implying that some coordinated strategy for life could operate at all levels throughout the universe. A swift overview of the envisaged scenario is provided here so as to facilitate expectations concerning the manner and timing of dark energy decay. The model anticipates that, following the decay of dark energy to neutrinos, oceanic planets will be populated by advanced civilisations adept at installing aquatic biospheres free of welfare-endangering perils such as carnivorous predation and avoidable disease. Photosynthesis has oxygenated the Earth's atmosphere but photochemistry would not be possible in subglacial oceans deprived of sunlight. This may not be problematic since many have speculated that complex chemosynthetic lifeforms could have evolved in Europa's dark and relatively anoxic oceans [17, 18].

Habitable planets capable of evading tidal-locking invariably orbit stars within the mass spectrum that terminate their lives as red giants, incinerating or absorbing any potentially habitable planets that may have orbited their progenitors. Given the cosmological context, this may be telling: it could

imply that lifeforms incapable of interstellar relocation are deemed too primitive to be granted survival beyond the early universe. More advanced, space-faring civilisations are likely to be skilled geneticists, especially if they have wrested control of their own biology from the clutches of haphazard evolutionary processes whether for purely ethical reasons or in an attempt to safeguard their ongoing survival [3, 19].

Galaxies frequently undergo mergers within galaxy clusters. Potentially introducing alien cultures to one another, collaboration and competition might ensue. If each galaxy spawns roughly one colonising civilisation then the ultimate outcome of a process of galactic mergers is expected to be a supercivilisation which could be confidently entrusted with colonisation [3]. The welfare of post-evolutionary lifeforms inhabiting skilfully designed aquatic biospheres could comfortably exceed that of the Earth's present lifeforms. Hence, the cosmic arrangement may seek to maximise opportunities for more advanced lifeforms subject to the need to first cultivate responsible colonists through natural selection. This impression is reinforced by the fact that formerly habitable orbiting planets would be incinerated during the red giant stage of their host stars, prohibiting the later revival even of dormant microbial organisms interred deep underground.

Statistical modelling of this scenario constrains to within a factor of two or so the rarity of advanced civilisations, not only now but also at other times [3]. This leads to three novel yet related resolutions of Fermi's paradox, all involving the future decay of dark energy to active neutrinos predominantly in galaxy clusters when the universe is ∼5 times its present age. A small fraction of life-cultivating stars will remain active until then, assuring survival for civilisations capable of interstellar relocation. It is striking that the measured energy density of empty space is compatible with this timescale, offering a hitherto elusive explanation for its tiny yet non-zero value where the $\Lambda \approx m_p^4$ guesstimate for the value of the cosmological constant has failed so spectacularly, m_p being the Planck mass. This attempt to calculate the value of the cosmological constant from quantum theory alone has yielded what is notoriously regarded as the *'worst prediction in all physics'*. Note, however, the claim that "although the magnitude of the vacuum energy remains a profound mystery, it seems clear that an understanding of how quantum-mechanical matter behaves in curved spacetime will play an important role in any eventual resolution to the puzzle" [20].

In summary, the universe may keep a tight rein on its available resources, restricting their expenditure except when it supports life — in particular post-evolutionary aquatic life. The temporary, relatively inefficient sustainment of evolutionary life during the early universe can be amortised by the vastly more efficient (∼99%) and lengthy (∼ 10^{25} year) aquatic era. Life is reliant on energy but energy conservation is a cornerstone of physics. Thus, energy cannot be the underlying currency of the universe. However, the universe could be strategically arranged so that entropy-increasing processes are restricted *unless* they either engender (via abiogenesis and evolution by natural selection) or support (via the direct internal heating of oceanic planets) advanced aquatic lifeforms.

2.4 The necessity of dark energy & its timely decay

If neutrinos are capable of efficiently sustaining aquatic life, why did the universe not provide dense neutrino haloes from the outset? Had the question instead been why did the universe not provide habitable planets from the outset, the answer would have been obvious: the primordial elements hydrogen and helium cannot form rocky planets or biomolecules. Answering the original question concerning the biological necessity for dark energy may not be so straightforward.

From a design perspective, a substantial postponement in the widespread provision of habitable environments for life could be a prudent precaution against incompetent colonisation. There may therefore be no urgency associated with the delivery of neutrinos until life-cultivating stars are becoming scarce. If dark energy must decay so that neutrino haloes capable of planetary heating can form then it can represent a temporary repository for the fuel needed by a forthcoming aquatic era. The accelerating expansion of the universe by an incongruously small cosmological constant may well be heralding the future delivery of active neutrinos.

Although some currently regard the cosmological constant as being literally responsible for cosmic acceleration, it requires an inexhaustible energy supply and its minuscule value defies theoretical explanation. Thus, independently of biotic reasoning, dynamical models of dark energy have been favoured. However, that leaves completely open the fate of the cosmic expansion. Biotic reasoning can assist here, offering clear hints concerning the future decay of dark energy, its timing, the particles it will yield and their distribution in space. A mechanism with considerable potential for satisfying all these various expectations shall now be sketched.

3 Gravitational collapse

Annual modulation in the timing of eclipses of Jupiter's moon Io allowed Ole Rømer to infer in 1676 that light travels at a finite speed. In 1783 John Michell argued for the existence of "dark stars", objects of sufficient mass that their escape velocity would exceed the speed of light. The Michelson-Morley experiment of 1887 found that light always travelled at the same speed regardless of the orientation of the apparatus relative to the Earth's passage through space. This spurred Einstein to conceive his 1905 theory of special relativity which ushered in the concept that clocks in relative motion are subjected to time dilation. When relativity was generalised a century ago to include gravitation Einstein showed that matter and energy could also affect the passage of time and indeed the entire network of temporal relationships amongst worldlines populating a spacetime manifold. Prior to this there was

no reason to suspect that nature might be capable of evading Michell's dark star expectation. We now understand that gravitational time dilation can grow without limit in general relativity: the proper time along one timelike worldline can cease to advance relative to the proper time along another. Combinations of the constants c, $\hbar$ and G cannot impose any Planck-scale restriction upon time dilation, a dimensionless quantity. It is therefore interesting to consider whether time dilation effects might be sufficient to ensure that gravitationally imploding matter is incapable of vanishing from view and becoming forever lost to the universe.

Supermassive black holes are by now widely thought to inhabit galactic nuclei, their masses occupying the range $10^6 \sim 10^{10}\ \mathrm{M}_\odot$ [21]. A Schwarzschild black hole has a surface area $A_\bullet = 4\pi R_\bullet^2 = 16\pi G^2 M^2/c^4$ which ostensibly governs its growth rate when immersed within a degenerate cloud of matter. In a galaxy hosting a million black holes of stellar mass, their combined area might be ten orders of magnitude less than that of a single supermassive black hole. Thus, if supermassive black holes did exist they would unacceptably sap neutrino haloes of biologically vital energy [3]. A mechanism for the eradication of eternal black holes is known involving the separation of virtual particle pairs via quantum tunnelling effects near the event horizon, the escape of one particle coming at the black hole's expense [22, 23]. However, the timescale for black hole evaporation via Hawking radiation is $5120\pi G^2 M_\bullet^3/\hbar c^4$ so astrophysical black holes require upwards of 10^{67} years to fully evaporate.

Rotating black holes are invariably plagued by the presence of closed timelike curves within their event horizons. The information loss paradox remains another stubborn complication [24] and locations of supposedly infinite mass density, *singularities*, hardly seem physically realistic — for example on energy conservation grounds. In addition, it has long been known that infalling particles, whether following timelike or lightlike trajectories, require infinite time to reach the event horizon of a black hole according to any arbitrarily-moving clock situated anywhere external to the event horizon. As the worldlines within a spacetime manifold must satisfy a global network of temporal interrelationships, black holes cannot grow through particle capture — rendering their dynamical formation implausible too [25–32]. No particle is better suited to the challenge of penetrating a Schwarzschild black hole event horizon than a radially ingoing photon but the metric then informs us that $|dr/dt| = c(1 - 2m/r)$ so that $dr/dt \to 0$ as $r \to 2m$ with attention confined to the regular coordinate region $r > 2m$. Evidently, the photon's motion is halted before it can reach the event horizon at $r = 2m$. It is possible to insert a mirror between the photon and the event horizon at arbitrarily late times and have it reflect back out along a radial geodesic, confirming that it never entered the black hole. Since nothing can be captured through an event horizon, the defining characteristic of a black hole, one can safely infer that gravitational collapse will always be safely arrested by the phenomenon of gravitational time dilation.

Given the enthusiasm for black hole research within modern science it may be difficult to accept that these objects are merely mathematical curiosities. Some further elaboration may thus be warranted. *Any* useful theory of gravity should be capable of predicting the trajectories of test particles in the vicinity of a gravitating point mass. If there is some maximum speed which no particle can exceed then matter straying too near the point mass will inevitably be incapable of escaping. It should therefore come as no surprise whatever that general relativity yields a stationary solution matching this expectation. But whereas Newtonian gravity would predict the existence of dark stars, general relativity departs radically since it predicts that time dilation can grow arbitrarily large even at a finite distance from the point mass responsible. Caution must hence be exercised since the fact that the Schwarzschild metric exists by no means guarantees that the solution is actually attainable through any physical process from realistic initial conditions in a universe of finite age.

Analytical solutions to Einstein's field equations can only be derived in certain idealised situations. The metrics describing the familiar eternal black holes have all been obtained by imposing the condition of stationarity: an assumption prohibiting any temporal evolution of the spacetime geometry, including of course any evolution that might be initially required to obtain the stationary configuration in question. Tracing the full *dynamics* of gravitational collapse in general relativity is hindered by the nonlinearities of the field equations. However, a pioneering work tackled this for the spherically symmetric case of a homogeneous sphere of pressureless matter [33]. If the advancement of proper time along all worldlines satisfies a very obvious constraint [31] this solution is well-behaved and time dilation asymptotically halts the collapse process just prior to event horizon formation. This constraint is compatible only with the exterior perspective on Oppenheimer-Snyder collapse — the interior perspective requiring the physically impossible advancement of proper times along *all* external worldlines. Though aware that neutron degeneracy pressure cannot always resist gravitational collapse, Oppenheimer & Snyder did at least appreciate that "it is impossible for a singularity to develop in a finite time" [33]. Hence, their collapse did not form a Schwarzschild black hole. Accordingly, gravitational collapse is expected to generate "*dark holes*", objects that may superficially resemble black holes in many circumstances but due to their lack of event horizons are free of their various pathologies. Whereas the situation considered by Oppenheimer and Snyder pertained to a particular mass distribution, a straightforward yet general proof now exists that black holes can neither form nor grow based on the inability of the Schwarzschild black hole to capture test particles of any description in a universe of finite age [31]. Furthermore, recent independent studies of dynamical collapse have also confirmed the non-formation of event horizons [27, 32].

Assertions that objects with event horizons exist cannot be verified even in principle [34] although the detection of Hawking radiation could arguably provide a counter-example. Whether or not black holes lie strictly outside the scope of science, nothing can prohibit the collection of evidence that specific black hole candidates *lack* rather than *possess* event horizons. The finite lifetimes ($10^7 \sim 10^8$ years) and the collimated jets of relativistic charged particles produced by quasars strongly suggests that their central engines have an intrinsic magnetic field — probably a dipole created by a spinning electrically charged torus [31, 35]. This interpretation calls into question the physical relevance of the *Principle of Topological Censorship*, a mathematical theorem constructed upon the assumption that trapped surfaces are present within some given spacetime [36] — a condition that no dark hole will satisfy [31] but which also belies the singularity theorems [37–39]. That is likely why, through the accrual of angular moment, dark holes are free to adopt toroidal geometry. The torus can then amass a significant net electrical charge, its rotation inducing a poloidal magnetosphere defending against charge neutralisation from the plasma of an orbiting accretion disk. Toroidal dark holes can explain the formation of relativistic jets of charged particles, the extreme energetics and the finite lifetimes of quasars [35]. Astronomers have found evidence of intrinsic magnetic fields in several black hole candidates, consistent with the absence of event horizons both in galactic black hole candidates [40–42] and in quasars [43]. When evaluating solutions of the field equations, the need to ensure that those configurations can be realistically attained without falling foul of constraints on global relationships has been generally overlooked: their formation must not involve the physically impossible advancement of time along any worldline within the spacetime manifold [31].

4 Dark energy from dark holes

The intersection of quantum mechanics and black hole physics led to the field of black hole thermodynamics. If, however, gravitational collapse is incapable of realistically producing objects endowed with event horizons, it may be more fruitful to consider the implications of quantum physics for dark holes. The complete absence of an event horizon precludes the emission of any Hawking radiation but a closely related process, the Fulling-Davies-Unruh effect [44–48], could be highly relevant to this discussion. Regarded as a fundamental and inescapable consequence of quantum field theory [49], the Unruh effect teaches us that the concept of a particle is observer dependent and that what may seem to exist in one reference frame may not exist at all in another [44]. It predicts that an accelerating detector coupled to a quantum field should perceive empty space to be seething with particles whose temperature is proportional to the acceleration of the detector [50].

According to Einstein's equivalence principle, a uniform acceleration is locally indistinguishable from a constant gravitational field. Hence, Unruh radiation is also expected if the detector/observer is stationary and, due to the presence elsewhere of a gravitating body, space is accelerating. Unruh and Hawking temperatures both share the common form $T = \hbar a/2\pi c k_B$ where T is the temperature of the perceived thermal bath of a vacuum field undergoing relative acceleration a. Although the value of the scaling factor $\hbar/2\pi c k_B$ is minute, $\sim 4\times10^{-20}\,^{\circ}\mathrm{K}/g$, it is generally accepted that the Unruh effect has already been experimentally confirmed in the observed depolarisation of electrons in storage rings [51, 52]. More sensitive measurements should be possible by exploiting Berry's phase [53].

Black body radiation from nearby galaxy clusters peaks in the X-ray spectrum, betraying the fact that gas there has been intensely heated by gravitational contraction. In the rarefied and hence transparent conditions of the intracluster medium, X-rays provide cooling. In contrast, matter exists in a dense state within stars, making their interior regions opaque to electromagnetic radiation. During core collapse supernovae, stars release large amounts of gravitational binding energy that drive runaway thermonuclear reactions. In such circumstances, cooling occurs almost exclusively through neutrino emission [54]. Even at energies above $2m_e c^2 \approx 1$ MeV at which electron/positron pairs are readily produced, neutrinos continue to dominate supernova cooling processes [55]. Some 10% of the rest mass of a collapsing star can be converted into neutrinos within a ten second interval [56]. The total luminosity during that period is $\sim 10^{46}$ W or $10^{19}L_{\odot}$, which greatly exceeds the power output of an entire galaxy. Radiated neutrinos are ultrarelativistic, a fact exploited by the Supernova Early Warning Systems to alert optical telescopes of impending supernova activity [57].

Likewise, neutrino escape will represent the main cooling mechanism for dark holes. They will copiously radiate neutrinos during their initial implosion stages but these formative outflows will soon cease as gravitational time dilation mounts, and are of no interest to this discussion. From the perspective of a stationary external observer, the internal vacuum of a dark hole whose collapse is arrested by time dilation will appear to undergo extreme acceleration — and hence, via the Unruh effect, should appear to be extremely hot. Over astronomical timescales, this intense but highly localised heating can deposit considerable thermal energy as heat percolates from the core of a dark hole to its periphery. The ordinarily prohibited proton decay process $p^+ \rightarrow n^0 + e^+ + \nu_e$ might be perceptible to dark hole onlookers whereas in the local frame it appears to be $p^+ + e^- \rightarrow n^0 + \nu_e$. The neutrino-related supernova processes $e^- + p^+ \leftrightarrow \nu_e + n^0$ and $e^+ + n^0 \leftrightarrow \bar{\nu}_e + p^0$ should also be important. Ultimately, via the Unruh effect, temperatures should become so elevated throughout the dark hole that some of the neutrinos generated by the thermal bath would satisfy the dark hole's escape requirements. A state of

pseudo-equilibrium might exist in which the neutrino cooling rate approximately balances the power in the Unruh effect.

For an observer accelerating through Minkowski space it has been speculated that the energy in Unruh radiation comes courtesy of the work that maintains the observer's acceleration [20]. Hawking radiation is thought to come at the expense of the black hole which captures negative energy virtual particles, reducing its mass. However, neither explanation satisfactorily explains the origin of the Unruh-related radiation emanating from a dark hole. According to general relativity, distortions of spacetime influence the motions of all objects because gravitation is, like the vacuum, a *global* phenomenon. Energy conservation may therefore be possible if the vacuum represents a quantum gravitational energy reservoir coupling both to gravity (consistent with accelerating cosmic expansion) and quantum mechanics (consistent with the Unruh effect). If indeed the vacuum acts as a dynamical repository for dark energy, the Unruh effect precipitated by extreme accelerations within dark holes may be uniquely capable of tapping into the cause of the accelerating cosmic expansion and eventually halting it.

Although neutrinos could dominate the cooling processes both within core collapse supernovae and dark holes, the dynamics of the latter case would be profoundly influenced by time dilation. Neutrinos escaping from dark holes would necessarily be red-shifted to low energies. This could be most advantageous to aquatic life: if the emerging neutrinos are at most mildly relativistic they could be easily retained by the gravity of the dark hole's host galaxy — thereby forming dense, inhabitable haloes.

4.1 Acceleration scales

In order to quantify the Unruh effect within dark holes there is a need to determine the acceleration of the vacuum due to gravity from the perspective of the surrounding universe. Although the Schwarzschild metric describes a black hole, by Birkhoff's theorem its exterior region can accurately represent the spacetime outside any spherically symmetric mass distribution, including a dark hole. Consider a timelike particle momentarily at rest in Schwarzschild coordinates $x^\lambda = [x^t, x^r, x^\theta, x^\phi]$. The metric reads $d\tau^2 = (1-2GM/c^2r)dt^2$ such that $dt/d\tau = 1/\sqrt{1-2GM/c^2r}$ and the particle's 4-velocity u is simply

$$u = \frac{dx^\lambda}{d\tau} = \dot{x}^\lambda = \left[\frac{1}{\sqrt{1-2GM/c^2r}},\ 0,\ 0,\ 0\right]. \qquad (1)$$

To find the particle's acceleration, $a^\lambda = \ddot{x}^\lambda$, the components of the covariant derivative of u are needed. Using the fact that dx^λ is non-zero only for dx^t and making use of the Christoffel symbols of the second kind, Γ^i_{kl} where

$$\Gamma^i_{kl} = \frac{g^{im}}{2}\left(\frac{\partial g_{mk}}{\partial x^l} + \frac{\partial g_{ml}}{\partial x^k} - \frac{\partial g_{kl}}{\partial x^m}\right) \qquad (2)$$

this simplifies to

$$du^\lambda = \left[\frac{\partial u^\lambda}{\partial x^t} + u^\sigma \Gamma^\lambda_{\sigma t}\right] dx^t. \qquad (3)$$

Since u^t is the only non-zero component of u^λ and $\Gamma^t_{tt} = 0$, it follows that $du^t = 0$. The only non-zero component of $\Gamma^r_{\sigma t}$ is $\Gamma^r_{tt} = GM(1-2GM/c^2r)/r^2$ and so

$$du^r = u^t\,\Gamma^r_{tt}\,dx^t = \left(\frac{GM(1-2GM/c^2r)}{r^2\sqrt{1-2GM/c^2r}}\right) dt. \qquad (4)$$

As Γ^θ_{tt} and Γ^ϕ_{tt} are both zero, the covariant derivative sought is $du = [0, GMr^{-2}\sqrt{1-2GM/c^2r}\,dt, 0, 0]$. The proper acceleration of the test particle can now be obtained using the fact that $dt/d\tau = 1/\sqrt{1-2GM/c^2r}$.

$$a^r = \dot{u}^r = \frac{du^r}{d\tau} = \sqrt{1-2GM/c^2r}\left(\frac{GM}{r^2}\right)\frac{dt}{d\tau} = \frac{GM}{r^2}. \qquad (5)$$

Hence, the 4-acceleration is $a = [0, GM/r^2, 0, 0]$ and for this momentarily stationary particle the magnitude of the outwardly directed acceleration is $a_s = \sqrt{a.a} = \sqrt{g_{\mu\nu}a^\mu a^\nu} = \sqrt{g_{rr}}GM/r^2$. Since $g_{rr} = (1-2GM/c^2r)^{-1}$,

$$a_s \equiv \frac{d^2r}{d\tau^2} = \frac{GM}{r^2\sqrt{1-2GM/c^2r}}. \qquad (6)$$

This acceleration corresponds to that of the vacuum at $x^r = r$, as perceived by remote observers. The Unruh temperature which this acceleration would predict, neglecting for now the influence of time dilation, would be

$$T_u = \frac{\hbar a_s}{2\pi c k_B} = \frac{\hbar GM}{2\pi c k_B r^2\sqrt{1-2GM/rc^2}}. \qquad (7)$$

Both a_s and T_u diverge as $r \to 2GM/c^2$, the radius of the event horizon. At $x^r = r$, the time dilation relative to distant objects can be readily derived from the Schwarzschild metric by setting $dr = d\phi = d\theta = 0$ to obtain $d\tau/dt = \sqrt{1-2GM/c^2R}$. Applying this correction factor to T_u, a finite temperature is obtained at the event horizon, T_{hor}. Inversely related to mass, this is the usual Hawking-Unruh temperature of a black hole:

$$T_{hor} = T_u\left(\frac{d\tau}{dt}\right) = \frac{\hbar c^3}{8\pi k_B GM}. \qquad (8)$$

Some appreciation of the variation of the matter distribution within a dynamically forming dark hole would be useful. Oppenheimer & Snyder considered the scenario of uniform density [33]. More realistically, one would expect density to decline towards the periphery of a dark hole. The mean density within the event horizon of a black hole decreases quadratically with mass, $\bar{\rho}_\bullet = 3c^6/32\pi G^3 M_\bullet^2$, and

with radius, $\bar{\rho}_\bullet = 3c^2/8\pi GR_\bullet^2$. The addition of a mass δM to a Schwarzschild black hole increases its radius by $\delta R = 2G\delta M/c^2$ and hence the density of a thin shell at radius R is $\delta M/4\pi R^2\delta R = c^2/8\pi GR^2$. This is again inversely quadratic in R, justifying the expectation that the mass density within a dark hole should generally decline with radius and be most concentrated at the core.

When the square root term in (6) is small the acceleration grows large, permitting a simplifying approximation:

$$r \approx \frac{2GM}{c^2} + \frac{c^6}{8GMa_s^2}\,. \tag{9}$$

The coordinate time of a photon falling from a modest distance outside the event horizon to this radius satisfies

$$\Delta t > \frac{2GM}{c^3}\ln\left(\frac{8GMa_s^2}{c^6}\right). \tag{10}$$

In the case of a Planck mass black hole this gives

$$\Delta t \gtrapprox 2\times 10^{-44}\ln(3\times 10^{-69}\times a_s^2)\,. \tag{11}$$

If some 10^{18} seconds (30 Gyr) are allowed to elapse after an infalling particle starts its descent, the apparent acceleration of the vacuum at the particle's final location would be approximately $10^{(10^{62})}\times$ the surface acceleration of a neutron star. It is extremely doubtful that such a huge acceleration is physically attainable. Using dimensional analysis, a quantity constructed using the constants c, G and $\hbar$ must be proportional to $c^{7/2}\hbar^{-1/2}G^{-1/2}$ in order to have the same units as acceleration. An estimate for the Planck acceleration, a_p, is therefore given by

$$a_p \sim \sqrt{\frac{c^7}{\hbar G}} \approx 10^{51} g\,. \tag{12}$$

The Planck temperature, T_p, is usually considered to be $T_p = m_pc^2/k_B = \sqrt{\hbar c^5/Gk_B^2}$. This tallies with the Unruh temperature for an acceleration of $2\pi\sqrt{c^7/\hbar G}$. However, the Hawking temperature, $\hbar c^3/8\pi k_BGM$, of a Planck mass black hole is normally assumed to be $T_p/8\pi$, yielding a Planck acceleration of $a_p \approx \frac{1}{4}\sqrt{c^7/\hbar G}$. This conforms to the Newtonian acceleration of a Planck mass from a distance matching its Schwarzschild radius. One may quibble over the best definition of a_p but it is apparent that $a_p \ll 10^{10^{62}}g$. Notice also that increments in proper time less than the Planck time, $t_p = \sqrt{\hbar G/c^5}$, are likely to be meaningless and, therefore, time dilations exceeding 10^{60} are essentially infinite within a universe less than 14 Gyr old.

If trans-Planckian accelerations are unattainable in nature then, independently of gravitational time dilation, this consideration alone would prohibit both the formation and growth of black holes. For a black hole of mass $M \gg m_p$, the ratio of the radius at which a stationary particle would experience the Planck acceleration to the radius of the event horizon would be $1 + m_p^2/M^2$. The time dilation at the Planck acceleration radius is given by

$$\frac{d\tau}{dt} = \sqrt{1-2GM/rc^2} = \sqrt{1-\frac{1}{1+m_p^2/M^2}} \approx m_p/M. \tag{13}$$

The perceived temperature of the Unruh heat bath, T_b, at radius $r > 2GM/c^2$, as reported by observers remote from the Schwarzschild black hole, requires correction for time dilation:

$$\begin{aligned} T_b = T_u\times\left(\frac{d\tau}{dt}\right) &= \left(\frac{\hbar a_s}{2\pi ck_B}\right)\left(\frac{d\tau}{dt}\right)\\ &= \frac{\hbar GM\times\frac{d\tau}{dt}}{2\pi ck_Br^2\sqrt{1-2GM/rc^2}} = \frac{\hbar GM}{2\pi ck_Br^2}\,. \end{aligned} \tag{14}$$

For a given black hole, T_b is a function of radius and declines as $1/r^2$. According to the Stefan-Boltzmann law for an ideal radiator, the radiative power, P_r, is given by the product of the area, $A = 4\pi r^2$, the Stefan-Boltzmann constant, $\sigma = \pi^2k_B^4/60\hbar^3c^2$, and the fourth power of the temperature:

$$P_r = A\sigma T_b^4 = 4\pi r^2\left(\frac{\pi^2k_B^4}{60\hbar^3c^2}\right)\left(\frac{\hbar GM}{2\pi ck_Br^2}\right)^4 = \frac{\hbar G^4M^4}{240\pi c^6r^6}\,. \tag{15}$$

Evidently, this power is highly localised since $P_r \propto 1/r^6$. Bearing in mind the Shell Theorem, it makes little difference how the mass distribution within a spherically symmetric dark hole declines with radius (whether as $\sim 1/r^2$, $1/r$, $1/\sqrt{r}$ etc) since the radiated power will be almost entirely sustained by quantum activity in the immediate vicinity of its core where the density is maximal and the gravitational acceleration of the vacuum is strongest.

By setting $r = r_s = 2GM/c^2$, the Schwarzschild radius, in (15) one recovers the power obtainable through Hawking evaporation, $P_\bullet(M_\bullet) = \hbar c^6/15360\pi G^2M_\bullet^2 \approx (M_\bullet/M_\odot)^{-2}\times 10^{-28}$ W. For a 10^{10} $M_\odot$ black hole, this comes to some 10^{-48} W, roughly 98 orders of magnitude short of what aquatic life would need. Even if a dark hole of this mass were composed of a set of concentric spherical shells, each of a thickness comparable to the Planck length ($\sim 10^{-35}$ m) and each radiating the same power, there would still be a shortfall of around 50 orders of magnitude.

The maximum power available from Hawking evaporation occurs when the black hole's mass approaches the Planck scale. The Compton wavelength of a particle of Planck mass is comparable to its Schwarzschild radius, $\hbar/cm_p \approx 2Gm_p/c^2$. Classical physics breaks down at this scale because $\hbar \neq 0$. It is customary in gravitation to work with the reduced Planck mass, $m_p \approx \sqrt{\hbar c/8\pi G}$. A black hole of this mass will be a nebulous, fuzzy object referred to here as a 'reduced Planck particle' (**rpp**). Its radiated power would be roughly $P_{\mathbf{rpp}} \approx 2\times 10^{49}$ W, though it might be somewhat larger as the Planck power is usually taken to be, $c^5/G \approx 4\times 10^{52}$ W.

It is generally thought that Planck particles are incapable of evaporating since their high Hawking temperatures preclude the black body radiation of significantly lighter particles. Hence, many imagine them to be quasi-stable remnants of black hole evaporation which is why they are now included amongst the panoply of dark matter candidates [58]. They also represent the most likely outcome of collapse processes that might otherwise result in naked singularities and the violation of cosmic censorship [59]. An **rpp** interred within a dark hole may be invulnerable to evaporation as long as the dark hole continues to exist. Thus, it is conceivable that power might be sustainably radiated at a level approaching $P_{\mathbf{rpp}}$ for spherically symmetric dark holes. Although this is very much an upper limit, it is encouraging that it yields a crude prediction that dark energy decay might terminate just as the last life-cultivating cease to be active.

4.2 Angular momentum injection

The Kerr metric represents a stationary, rotationally symmetric and asymptotically flat rotating black hole. It accommodates angular momentum through an extended singularity located within the plane $z = 0$, lying along the circle $x^2+y^2 = a^2$ in Kerr coordinates. Its radius, $a \equiv J/m$, depends on the angular momentum, J, of the black hole. At extremality, $a \to m$ and $J \to m^2$, the radius of the singularity coincides with that of the two event horizons, $r_\pm = m \pm \sqrt{m^2 - a^2}$. For $a^2 < m^2$ the singularity lies internal to both event horizons.

At high angular momentum, self-gravitating fluids bifurcate from the Maclaurin spheroids, yielding toroidal configurations [60, 61] reminiscent of the prototypical Dyson rings [62]. With analytical solutions confined to relatively simple cases, numerical techniques have now been deployed to better explore the space of axisymmetric configurations [63–67]. The assumption of homogeneity has been relaxed, differential rotation has been allowed and realistic equations of state have been modelled. Qualitatively similar results have been obtained in both Newtonian analyses and general relativity [68–70]. Ergoregions can arise even in the absence of event horizons [71], which may be of some relevance to jet formation in quasars [35].

Since topological censorship does not apply to spacetimes lacking trapped surfaces, the gravitational collapse of a rotating body can result in a toroidal mass distribution, analogous to the circular source of the Kerr geometry although visible to the surrounding universe. Whilst the angular momentum of a Kerr black hole is bounded, $a^2 \leq m^2$, there is no such restriction for a dark hole: the major radius of a self-gravitating torus can be arbitrarily larger than its minor radius. The axisymmetric Kerr geometry cannot dissipate rotational kinetic energy via gravitational waves. Moreover, since gravitational waves are incapable of superluminal propagation, any gravitational radiation due to perturbations of the singularity would necessarily remain imprisoned within the event horizon. However, deviations from axisymmetry deep within a dark hole could generate rather strong gravitational waves, and their radiation into space would sap the dark hole's energy and angular momentum.

Suppose a dense ring of radius $r = m$ is quantised by subdivision into a circular arrangement of N particles, each of roughly the reduced Planck mass, such that $N = m/m_p$. Since the ring's circumference is $C_r = 2\pi Gm/c^2$, each reduced Planck particle would then be separated from its two neighbouring particles by a distance

$$\frac{C_r}{N} = \frac{C_r m_p}{m} = \frac{2\pi G}{c^2}\sqrt{\frac{\hbar c}{8\pi G}} = \sqrt{\frac{\pi\hbar G}{2c^3}} \approx \sqrt{\frac{\pi}{2}}\,\ell_p\,. \quad (16)$$

The mean particle separation should decline as J decreases but if separations below the Planck length ℓ_p are unattainable, the idealised circular arrangement may be disrupted, resulting in localised thickening of the ring. It may help to picture the interior of the toroidal dark hole as a dense circular arrangement of knotty density existing at extreme densities approaching the Planck scale. Due to this granularity and its chaotically fluctuating nature, gravitational waves should be produced which dissipate both angular momentum and rotational energy. Ultimately, these losses should result in a topological collapse of the core.

For the purposes of this discussion, we might simply regard the core of a rapidly spinning dark hole as a circular collection of reduced Planck particles. in the case of a toroidal dark hole they would number m/m_p, a huge number. Toroidal dark holes should therefore receive enormously more internal heating via the Unruh effect than purely spheroidal dark holes of the same mass. Thus, the discharge of vacuum energy would be strongly biased towards the most massive and rapidly spinning dark holes of the cosmos — even if such objects are comparatively short-lived in astronomical terms.

Following galactic mergers, the supermassive dark holes introduced by each galaxy are generally expected to coalesce relatively swiftly since they occupy locations of least gravitational potential within their respective host galaxies. Inspiralling supermassive black hole binaries provide prime targets for gravitational wave astronomy [72]. A supergalaxy harbouring an ultramassive remnant dark hole would be the inevitable outcome of hierarchical galaxy mergers. Just as the coalescence of a pair of co-orbiting black holes is able to create a rapidly rotating black hole due to the conversion of orbital angular momentum to rotational angular momentum [73], a pair of coalescing spheroidal dark holes will often combine to produce a more massive dark hole of internally toroidal structure with a dense filamentary core. Violent galactic mergers within galaxy clusters should hence be capable of sporadically inducing episodes of intensively accelerated heating and dark energy discharge until the resulting dark holes lose their toroidal inner structure via the shedding or redistribution of internal angular momentum.

4.3 Discharge timeframes

The details of the physical cooling processes operating within core collapse supernovae are still the subject of ongoing research but it is known that even neutrinos cannot free stream away from an innermost region termed the neutrinosphere. Although the circumstances deep within dark holes will be yet more complex and involve energies well above those probed by any practical particle collider, neutrinos generated there will also be unable to free stream away into space, hindered by the exclusion principle and large interaction cross sections. In order to accurately model these situations, a working theory of quantum gravity will be needed along with an understanding of how matter behaves at near-Planckian temperatures and densities. Also, the influence of strong time dilation must be taken into account, a problem which modern numerical approaches to general relativity still grapple with. Even order of magnitude estimates to the processes involved may currently lie beyond our reach. Nevertheless, it is incumbent upon us to consider the viability of this scenario.

At energies below 1 GeV, electron neutrinos scatter onto neutrons, $\nu_e + n \rightarrow e^- + p^+$, with cross section

$$\sigma_n = \frac{(\hbar c\, G_F E_\nu)^2 (g_V^2 + 3g_A^2)}{\pi} \approx 10^{-47} \left(\frac{E_\nu}{1\ \mathrm{MeV}} \right)^2 \mathrm{m}^2. \quad (17)$$

The mean free path of a neutrino can be estimated using

$$\lambda_\nu \approx \left(\frac{\rho}{10^{18} \mathrm{kg} \cdot \mathrm{m}^{-3}} \right)^{-1} \left(\frac{E\nu}{10\ \mathrm{MeV}} \right)^{-2} \text{metres}. \quad (18)$$

Within a supernova the neutron number density may be as high as $\rho_n \sim 10^{45}\ \mathrm{m}^{-3}$. For a typical supernova neutrino energy of 30 MeV, the cross section would be $\sim 10^{-44}\ \mathrm{m}^2$ and the free path, $\lambda = 1/\rho_n \sigma_n$, may be as short as ten metres. Cooling occurs as neutrinos emerge from a thin shell surrounding a rather constipated, diffusion-limited neutrinosphere. The volumetric luminosity within that spherical escape shell could approach $10^{36}\ \mathrm{W} \cdot \mathrm{m}^{-3}$. Deep within a dark hole, degeneracy is also likely to have a profound impact on the dynamics, for instance blocking neutrino production via the Unruh effect deep within the neutrinosphere.

The limiting mass of a halo of 0.05 eV neutrinos is estimated to be $8 \times 10^{20} M_\odot$ [10], equating to an energy in excess of 10^{68} J. The power due to Unruh radiation from a reduced Planck particle was determined earlier to be $P_{\mathbf{rpp}} \approx 2 \times 10^{49}$ W so it would take some 7×10^{18} seconds to inflate a habitable neutrino halo, roughly 230 Gyr. Encouragingly, this crude and simplistic estimate has the right order of magnitude: dark energy is anticipated to decay before the universe reaches ~ 75 Gyr in age. Pauli blocking and impedance of thermal transport within the time-dilated neutrinosphere will slow the discharge, but episodic input of angular momentum generating ultra-dense rings might easily compensate by hugely accelerating the process, if only briefly.

If, by analogy with Hawking radiation from black holes, the Unruh radiation from deep within supermassive dark holes came entirely at the expense of their dark hole hosts then the lifespan of a supermassive dark hole evaporating at the rate $P_{\mathbf{rpp}}$ would be $\sim Mc^2/P_{\mathbf{rpp}}$. This is just a few hours for a $10^6\ M_\odot$ black hole and no more than a few years for a $10^{10}\ M_\odot$ black hole. The observational evidence for the ongoing existence of supermassive black hole candidates in this mass rnage within galactic nuclei confidently rules out this possibility. It instead points to the dark energy vacuum being the origin of the Unruh radiation which, according to quantum field theory, is mandatory — so much so that it needs no experimental confirmation [49].

Consider now the case of an ultramassive remnant dark hole of a galaxy cluster of a mass $\sim 10^{11}\ \mathrm{M}_\odot$ that ultimately generates a neutrino halo of mass $10^{21}\ \mathrm{M}_\odot$. The Unruh effect provides the dark hole with intense but localised heating. Over time, thermal energy accumulates and steadily diffuses throughout the dark hole. Once peripheral temperatures are sufficient to permit neutrino escape, a galactic halo can start to form. One can envisage neutrino cooling approximately balancing heating from the Unruh effect until vacuum energy is exhausted. If, instead, one assumes that neutrinos barely escape until dark energy is almost fully depleted, then the mass of the dark hole from the galactic perspective must increase by ten orders of magnitude. From the dark hole's vantage, however, its mass will not have changed since there was no Unruh effect attempting to increase the temperature of its constituent matter.

Ignoring the initial thermal energy of the dark hole, in order that it can eventually form a dense neutrino halo from the thermal energy deposited through the Unruh effect, its particles must attain Lorentz factors, γ, of $\sim 10^{10}$ where γ represents the relativistic mass ratio m/m_0. Temperature is synonymous with kinetic energy and relativistic kinetic energy is proportional to $\gamma m_0 c^2$. The Lorentz factor is closely tied to relativistic temperature according to the relationship, $T = 2(\gamma - 1) m_o c^2 / 3k_B$. If $\gamma \gg 1$, it follows that $\gamma \approx 3k_B T / 2m_o c^2$. Thus, for any given temperature, it is much easier for lighter particles such as neutrinos to attain large Lorentz factors. Consequently, irrespective of temperature, neutrinos should dominate the cooling outflows of dark holes. For a neutrino mass of 0.05 eV, a Lorentz factor of 10^{10} corresponds to a temperature of some 4×10^{12} K, comfortably below the Planck temperature $\sqrt{\hbar c^5 / G k_B^2} \approx 10^{32}$ K. Hence, there is much latitude for dark holes to discharge dark energy by only temporarily adopting internally toroidal matter distributions. Dark holes will of course undergo heating via gravitational contraction during their initial formation, and this energy will contribute to their effective mass, pushing the required Lorentz factors and temperatures for neutrinos to escape somewhat higher. Nevertheless, there seems to be ample scope to accommodate this particular consideration.

5 Discussion

Via the Unruh effect, dark holes may well be capable of tapping into the energy of the vacuum and, in due course, fostering the total discharge of dark energy. A minority of class K stars will continuously host habitable planets until the universe is five times its present age. As the neutrino haloes produced would be capable of efficiently sustaining aquatic life, the model offers a new and biological resolution of the cosmological constant problem. This provides a long-sought alternative to the relatively loose bounds imposed by anthropic arguments [4] which, if neutrinos are well-suited to the task of planetary heating, would be as untenable as the Ptolemaic system. However, this new approach can only explain the value of the present vacuum energy density, not how it might have been tuned to 120 decimal places. That question is akin to asking how the constants of physics in general might have been manipulated to be propitious towards life — something perhaps for string theorists to mull over.

The supermassive dark holes of galaxy clusters are likely to play a prominent role in the decay of dark energy due to their frequent adoption of an internally toroidal structure following violent coalescence events pursuant to galaxy mergers. With circular 'heating elements' operating at temperatures potentially approaching the Planck scale, $T_p \approx 10^{32}K$, these rapidly spinning dark holes would accumulate thermal energy far more rapidly than their counterparts in field galaxies. Such extreme conditions will help combat the intense gravitational time dilation and Pauli-blocking deep within dark holes, in time facilitating a radiative flux from their cores. Eventually, peripheral temperatures should rise sufficiently to provide opportunities for neutrinos to escape completely, albeit after redshifting to low energies as they do so. Dense haloes should thereby form around the remnant galaxies of galaxy clusters — in accordance with the expectation that neutrinos might sustain aquatic life into the distant future [3].

During the initial phase of dark hole heating the neutrino luminosity is likely to remain negligible for billions of years. This is no cause for concern: a lengthy delay would be biologically advantageous, usefully prohibiting the widespread colonisation of the universe until ethically mature civilisations are on hand to undertake such daunting responsibilities. Very loosely, the situation might be likened to the conversion of liquid water within a lake into steam by the vigorous agitation of a single water molecule over a prolonged period. It is not inconceivable that galactic nuclei within galaxy clusters may be currently generating gentle outflows of neutrinos. It may even be that changes in the dynamics of galaxies orbiting within clusters may be perceptible over time.

A promising line of enquiry has been outlined concerning dark energy without any radical departure from the scientific method. Whilst alternative proposals capable of predicting the timing, outcome and mechanism of dark energy decay have not been forthcoming, this scenario dovetails remarkably well with a recently advanced cosmological framework, reinforcing its potential to unravel the composition of dark matter, anticipate the fate of the accelerating cosmic expansion and decipher the mystery of extraterrestrial silence. This same framework offers much scope for understanding why the constants of nature assume the values they do without recourse either to mathematical arguments or anthropic reasoning [74]. Physical fine-tuning influences all aspects of the universe — from the simplest microscopic scales to the most complex macroscopic scales, including multicellular lifeforms, symbiotic ecosystems and the imponderable workings of the human mind. If the fine-tuning of physics cannot be apprehended through mathematical physics alone then alternatives can and should be explored, even if that entails a holistic synthesis of all scientific knowledge. Support has emerged here for the contention within superstring theory that there may exist a vast underlying landscape of physical configurations. A biological resolution of the cosmological coincidence problem, the naïvely surprising similarity between Ω_Λ and Ω_M is also apparent.

Although this dark energy decay scenario must be regarded as tentative for now, it exhibits many compelling features and the remaining uncertainties mainly pertain to timescales. Vacuum discharge by dark holes fulfils the original cosmological expectation that dark energy might decay predominantly to neutrinos of sufficiently low energy as to permit their retention by host galaxies. Whereas black holes have an unhealthy appetite for neutrinos, dark holes are incapable of recapturing the neutrinos they discharge on account of the Unruh effect: those particles belong to a different reality. Dark energy decay should strongly track supermassive, rapidly-rotating dark holes and hence galaxy clusters where collisions between galaxies and supermassive dark holes are commonplace events. There is no reason at present to suppose that significant errors exist in the basic timing constraints based on the measured vacuum energy density, calculated neutrino halo implosion threshold and the necessity of habitable neutrino haloes before the last life-cultivating stars expire. However, much work remains before decay timescales can be confidently calculated. In the meantime, as long as neutrinos continue to bear the hallmarks of cosmic design the efficient sustainment of aquatic life remains a very real possibility and the timely decay of dark energy just before life-cultivating stars die out is a likely outcome of cosmic evolution.

In the field of black hole thermodynamics, the entropy of a black hole is given by $S = kA/4\ell_P^2$. With the entropy of a Sun-like star being $\sim 10^{35}\ \mathrm{J \cdot K^{-1}}$ [75] a single black hole of some 10^4 solar masses would possess as much entropy as all the stars of all the galaxies within the visible universe. Therefore, the existence of even one supermassive black hole in nature would be catastrophic for a universe attempting to judiciously manage entropy-increasing processes for the benefit of advanced lifeforms [3]. Dark holes may not chime

with the scientific orthodoxy of recent decades but they accord with the original 'frozen star' interpretation of stellar collapse and dispense with the theoretical shortcomings of black holes. It is not that the stationary black hole metrics are mathematically invalid, merely that they are unobtainable via physically admissible processes: global constraints on the evolution of spacetime manifolds have been generally overlooked [31]. Although one expects any useful theory of gravity to possess within it a solution resembling that obtained by Karl Schwarzschild in 1916, i.e. a bizarre object describing a point mass surrounded by a region from which light cannot escape, the architect of general relativity did not rush to dismiss it. His considered opinion was offered after decades of careful reflection. That his views on black holes now carry so little weight within academic science is disturbing.

On a far more positive note, it comes as some surprise that particles which are able to propagate unperturbed through light years of lead may be of any potential benefit to life. That sterile neutrinos, their yet more inert counterparts, might provide further assistance also defies intuition. However, if this universe is exquisitely configured to host life, all natural phenomena would ideally have something positive to contribute. It would nevertheless be astonishing if the objects lurking at the heart of each galaxy, which many currently believe to be destructive black holes, can serve as portals to a crucial biological energy resource capable of efficiently sustaining aquatic life far into the distant future. We learnt from special relativity that mass and energy are interrelated, a breakthrough necessary to explain how stars could remain active for billions of years, sufficient time for the Sun to support the evolution of complex organisms. Ultimately, the take-home message from general relativity, if only apparent at present to extraterrestrial civilisations, may be that gravity is benign and free of pathologies precisely because time dilation provides a robust mechanism preventing the formation and growth of trapped surfaces — essential for the discharge of dark energy so that aquatic lifeforms might thrive long after the expiry of the stars, harnessing the full promise of $E = mc^2$. If Einstein were still with us he might regard the current fascination with black holes as a pathological science, further affirmation of his 1920 remark to Marcel Grossmann that the world is a "strange madhouse" [76]. It very much seems there is now a contagious misunderstanding of his theoretical legacy. It may be preventing humanity from collectively converging towards a comprehension of the universe capable of providing much-needed guidance for future policy-making.

Submitted on January 3, 2016/Accepted on January 5, 2016

References

1. Fukuda Y., Hayakawa T., Ichihara E., Inoue K., Ishihara K., Ishino H. et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 1998, v. 81(8), 1562.
2. Cowsik R., McClelland J. Gravity of neutrinos of nonzero mass in astrophysics. *ApJ*, 1973, v. 180, 7.
3. Spivey R. J. A cosmological hypothesis potentially resolving the mystery of extraterrestrial silence with falsifiable implications for neutrinos. *Physics Essays*, 2015, v. 28(2), 254–264.
4. Weinberg S. Anthropic bound on the cosmological constant. *Phys. Rev. Lett.*, 1987, v. 59(22), 2607.
5. Riess A. et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *ApJ*, 1998, v. 116, 1009.
6. Perlmutter S. et al. Measurements of Ω and Λ from 42 high-redshift supernovae. *ApJ*, 1999, v. 517, 565.
7. Stixrude L. Structure of Iron to 1 Gbar and 40,000 K. *Phys. Rev. Lett.*, 2012, v. 108, 055505.
8. Olive K. A., & Particle Data Group. Review of particle physics. *Chinese Physics C*, 2014, v. 38(9), 090001.
9. Planck Collaboration. Planck 2015 results. XIII. Cosmological parameters. 2015, arXiv:1502.01589.
10. Spivey R. J. Planetary heating by neutrinos and possible long-term habitats for aquatic life if dark energy decays favourably. *J. Mod. Phys.*, 2013, v. 4(12A), 20–47.
11. De la Macorra A. Interacting dark energy: Decay into fermions. *Astroparticle Physics*, 2007, v. 28(2), 196–204.
12. Kusenko A. Sterile neutrinos: the dark side of the light fermions. *Phys. Reports*, 2009, v. 481(1), 1–28.
13. Kopp J., Maltoni M., Schwetz T. Are there sterile neutrinos at the eV scale? *Phys. Rev. Lett.*, 2011, v. 107(9), 091801.
14. Nieuwenhuizen T. M. Do non-relativistic neutrinos constitute the dark matter? *Europhysics Lett.*, 2009, v. 86(5), 59001.
15. Nieuwenhuizen T. M., Morandi A. Are observations of the galaxy cluster A1689 consistent with a neutrino dark matter scenario? *MNRAS*, 2013, v. 434(3), 2679-2683.
16. Bousso R., Katz D. M., Zukowski C. Anthropic origin of the neutrino mass from cooling failure. *Phys. Rev. D*, 2015, v. 92, 025037.
17. Jackson B., Barnes R., Greenberg R. Tidal heating of terrestrial extrasolar planets and implications for their habitability. *MNRAS*, 2008, v. 391(1), 237–245.
18. Chyba C. F. Energy for microbial life on Europa. *Nature*, 2000, v. 403(6768), 381–382.
19. Savulescu J. Procreative beneficence: why we should select the best children. *Bioethics*, 2001, v. 15(5–6), 413.
20. Carroll S. M. Spacetime and Geometry. An Introduction to General Relativity (Vol. 1). *Addison Wesley*, 2004.
21. Celotti A., Miller J. C., Sciama D. W. Astrophysical evidence for the existence of black holes. *Class. Quant. Grav.*, 1999, v. 16(12A), A3.
22. Hawking S. W. Black hole explosions. *Nature*, 1974, v. 248(5443), 30–31.
23. Hawking S. W. Particle creation by black holes. *Comm. Math. Phys.*, 1975, v. 43(3), 199–220.
24. Mathur S. D. The information paradox: a pedagogical introduction. *Class. Quant. Grav.*, 2009, v. 26(22), 224001.
25. Einstein A. On a stationary system with spherical symmetry consisting of many gravitating masses. *Annals of Mathematics*, 1939, v. 40, 4.
26. Kiselev V. V., Logunov A. A., Mestvirishvili M. A. Black holes: theoretical prediction or fantasy? *Physics of Particles and Nuclei*, 2006, v. 37(3), 317-320.
27. Vachaspati T., Stojkovic D., Krauss L. M. Observation of incipient black holes and the information loss problem. *Phys. Rev. D*, 2007, v. 76(2), 024005.
28. Kiselev V. V. E., Logunov A. A., Mestvirishvili M. A. The physical inconsistency of the Schwarzschild and Kerr solutions. *Theor. and Math. Phys.*, 2010, v. 164(1), 972-975.

29. Weller D. Five fallacies used to link black holes to Einstein's relativistic space-time. *Prog. Phys.*, 2011, v. 7, issue 1, 93.
30. Chafin C. E. Globally Causal Solutions for Gravitational Collapse. 2014, arXiv:1402.1524.
31. Spivey R. J. Dispelling black hole pathologies through theory and observation. *Prog. Phys.*, 2015, v. 11, 321–329.
32. Piñol M. A model of dust-like spherically symmetric gravitational collapse without event horizon formation. *Prog. Phys.*, 2015, v. 4, 331.
33. Oppenheimer J. R., Snyder H. On continued gravitational contraction. *Phys. Rev.*, 1939, v. 56(5), 455.
34. Abramowicz M. A., Kluzniak W., Lasota J. P. No observational proof of the black-hole event horizon. *Astron. Astrophys.*, 2002, v. 396, L31.
35. Spivey R. J. Quasars: a supermassive rotating toroidal black hole interpretation. *MNRAS*, 2000, v. 316(4), 856–874.
36. Friedman J. L., Schleich K., Witt D. M. Topological censorship. *Phys. Rev. Lett.*, 1993, v. 71(10), 1486.
37. Penrose R. Gravitational collapse and space-time singularities. *Phys. Rev. Lett.*, 1965, v. 14(3), 57.
38. Hawking S. W. The occurrence of singularities in cosmology. III. Causality and singularities. *Proc. Roy. Soc. A*, 1967, v. 300(1461), 187.
39. Hawking S. W. and Penrose R. The singularities of gravitational collapse and cosmology. *Proc. Roy. Soc. A*, 1970, v. 314(1519), 529.
40. Robertson S. L., Leiter D. J. Evidence for intrinsic magnetic moments in black hole candidates. *ApJ*, 2002, v. 565(1), 447.
41. Robertson S. L., Leiter D. J. On intrinsic magnetic moments in black hole candidates. *ApJ Lett.*, 2003, v. 596(2), L203.
42. Robertson S. L., Leiter D. J. On the origin of the universal radio–X-ray luminosity correlation in black hole candidates. *MNRAS*, 2004, v. 350(4), 1391-1396.
43. Schild R. E., Leiter D. J., Robertson S. L. Observations supporting the existence of an intrinsic magnetic moment inside the central compact object within the quasar Q0957+ 561. *ApJ*, 2006, v. 132(1), 420.
44. Fulling S. A. Nonuniqueness of canonical field quantization in Riemannian space-time. *Phys. Rev. D*, 1973, v. 7(10), 2850.
45. Davies P. C. Scalar production in Schwarzschild and Rindler metrics. *J. Physics A*, 1975, v. 8(4), 609.
46. Unruh W. G. Notes on black-hole evaporation. *Phys. Rev. D*, 1976, v. 14(4), 870.
47. Unruh W. G. Origin of the particles in black-hole evaporation. *Phys. Rev. D* 1977, v. 15, 365.
48. Unruh W. G., Schützhold R. Universality of the Hawking effect. *Phys. Rev. D*, 2005, v. 71(2), 024028.
49. Matsas G. E., Vanzella D. A. The Fulling–Davies–Unruh effect is mandatory: the Proton's testimony. *Int. J. Mod. Phys. D*, 2002, v. 11(10), 1573-1577.
50. Crispino L. C., Higuchi A., Matsas G. E. The Unruh effect and its applications. *Rev. Mod. Phys.*, 2008, v. 80(3), 787.
51. Bell J. S., Leinaas J. M. The Unruh effect and quantum fluctuations of electrons in storage rings. *Nucl. Phys. B*, 1987, v. 284, 488.
52. Akhmedov E. T., Singleton D. On the physical meaning of the Unruh effect. *JETP Lett.*, 2008, v. 86(9), 615–619.
53. Martin-Martinez E., Fuentes I., Mann R. B. Using Berry's phase to detect the Unruh effect at lower accelerations. *Phys. Rev. Lett.*, 2011, v. 107(13), 131301.
54. Bahcall J. N., Piran T., Press, W. H., Spergel D. N. *Nature*, 1987, v. 327, 25.
55. Hartmann D. H., Woosley S. E. The cosmic supernova neutrino background. *Astroparticle Physics*, 1997, v. 7(1), 137–146.
56. Woosley S., Janka T. The physics of core-collapse supernovae. *Nature Physics*, 2005, v. 1(3), 147–154.
57. Antonioli P., Fienberg R. T., Fleurot F. et al. SNEWS: The supernova early warning system. *New Journal of Physics*, 2004, v. 6(1), 114.
58. MacGibbon J. H. Can Planck-mass relics of evaporating black holes close the Universe? *Nature*, 1987, v. 329, 308–309.
59. Christodoulou D. Violation of cosmic censorship in the gravitational collapse of a dust cloud. *Comm. Math. Phys.*, 1984, v. 93(2), 171–195.
60. Bardeen J. M. A reexamination of the post-newtonian maclaurin spheroids. *ApJ*, 1971, v. 167, 425.
61. Eriguchi Y., Sugimoto D. Another equilibrium sequence of self-gravitating and rotating incompressible fluid. *Prog. Theor. Phys.*, 1981, v. 65, 1870
62. Dyson F. W. The potential of an anchor ring. *Phil. Trans. Roy. Soc.*, 1893, v. 53, 372-375.
63. Hachisu I. A versatile method for obtaining structures of rapidly rotating stars. *ApJ Suppl.*, 1986, v. 61, 479–507.
64. Ansorg M., Kleinwächter A., Meinel R. Relativistic Dyson rings and their black hole limit. *ApJ Lett.*, 2003, v. 582(2), L87.
65. Petroff D., Horatschek S. Uniformly rotating homogeneous and polytropic rings in Newtonian gravity. *MNRAS*, 2008, v. 389(1), 156–172.
66. Horatschek S., Petroff D. A Roche model for uniformly rotating rings. *MNRAS*, 2009, v. 392(3), 1211–1216.
67. Horatschek S., Petroff D. Uniformly rotating homogeneous rings in post-Newtonian gravity. *MNRAS*, 2010, v. 408(3), 1749–1757.
68. Ansorg M., Fischer T., Kleinwächter A., Meinel R., Petroff D., Schöbel K. Equilibrium configurations of homogeneous fluids in general relativity. *MNRAS*, 2004, v. 355(3), 682–688.
69. Fischer T., Horatschek S., Ansorg M. Uniformly rotating rings in general relativity. *MNRAS*, 2005, v. 364(3), 943–947.
70. Meinel R., Ansorg M., Kleinwächter A., Neugebauer G., Petroff D. Relativistic Figures of Equilibrium. *Cambridge Univ. Press*, 2008.
71. Butterworth E. M., Ipser J. R. On the structure and stability of rapidly rotating fluid bodies in general relativity. *ApJ*, 1976, v. 204, 200–223.
72. Enoki M., Inoue K. T., Nagashima M., Sugiyama N. Gravitational waves from supermassive black hole coalescence in a hierarchical galaxy formation model. *ApJ*, 2004, v. 615(1), 19.
73. Buonanno A., Kidder L. E., Lehner L. Estimating the final spin of a binary black hole coalescence. *Phys. Rev. D*, 2008, v. 77(2), 026004.
74. Livio M., Rees M. J. Anthropic Reasoning. *Science*, 2005, v. 309(5737), 1022.
75. Bekenstein J. D. Black holes and entropy. *Phys. Rev. D*, 1973, v. 7(8), 2333.
76. van Dongen J. On Einstein's opponents, and other crackpots. 2011, arXiv:1111.2181.

Further Insight Relative to Cavity Radiation III: Gedanken Experiments, Irreversibility, and Kirchhoff's Law

Pierre-Marie Robitaille

Department of Radiology, The Ohio State University, 395 W. 12th Ave, Columbus, Ohio 43210, USA
E-mail: robitaille.1@osu.edu

Recently, gedanken experiments have been proposed in order to examine the validity of Kirchhoff's Law of Thermal Emission (*P.-M. Robitaille, Further Insight Relative to Cavity Radiation: A Thought Experiment Refuting Kirchhoff's Law, Prog. Phys., 2014, v. 10, no. 1, 38–40; P.-M. Robitaille, Further Insight Relative to Cavity Radiation II: Gedanken Experiments and Kirchhoff's Law, Prog. Phys., 2014, v. 10, no. 2, 116–120*). In the second of these works, real materials (i.e. graphite and silver) were utilized in order to construct two separate cavities at the same temperature which are then placed in thermal contact with one another. It was hypothesized that the graphite cavity initially contained blackbody radiation and that the silver cavity was devoid of radiation. In the case of the silver cavity, all of the energy of the system was assigned to the phonons in its walls. When the cavities were brought together and a small hole introduced between the cavities, it was hypothesized that thermal contact between the cavity walls would enable the transformation of phonon energy into photon energy, eventually resulting in filling the silver cavity with black radiation. Energy contained within the wall of the silver cavity was believed to be *reversibly* trapped. However, in allowing energy to flow reversibly out of the walls of the silver cavity in this context, it has been assumed that the silver conduction bands could be neglected and that only phonon energy need be considered. However, the reflectivity attributed to the silver cavity should be considered uniquely as a result of energy associated with the formation of its conduction bands. Such formation must be considered *irreversible*. It will be demonstrated that under these conditions Kirchhoff's law, once again, does not hold. The lack of thermal radiation within the silver cavity does not lead to a violation of the second law of thermodynamics.

> *If a space be entirely surrounded by bodies of the same temperature, so that no rays can penetrate through them, every pencil in the interior of the space must be so constituted, in regard to its quality and intensity, as if it had proceeded from a perfectly black body of the same temperature, and must therefore be independent of the form and nature of the bodies, being determined by temperature alone. . . In the interior therefore of an opake red-hot body of any temperature, the illumination is always the same, whatever be the constitution of the body in other respects.*
>
> Gustav Robert Kirchhoff, 1860 [1]

1 Introduction

Gedanken experiments have played a major role in building support for Kirchhoff's Law of Thermal Emission [1, 2]. If this is the case, it is because Kirchhoff proposed his law without any experimental verification [1,2]. This remains a significant departure from the other laws of thermal emission [3–6] which have been confirmed through the construction of laboratory blackbodies. In addition, The Law of Equivalence, first formulated by Balfour Stewart [7], has also been confirmed experimentally. However, Kirchhoff's law, namely the belief that the radiation contained within an arbitrary cavity will always be black, or normal, independent of the nature of the cavity wall, has never been demonstrated experimentally [8–12]. Furthermore, Kirchhoff's law knows no proper theoretical proof [13]. Even Max Planck's theoretical proof of Kirchhoff's law can be shown to be invalid [14]. As such, the real justification for Kirchhoff's law falls on thought experiments, all of which can be shown to contain logical omissions and errors.

A powerful sentiment remains in the physics community that should Kirchhoff's law be invalid, then a violation of the second law of thermodynamics would exist and perpetual motion machines of the second kind could be constructed. The arguments typically involve the consideration of two cavities isolated from the outside world by exterior adiabatic walls. The inner walls of the first cavity are then constructed from a perfect absorber (emissivity $\epsilon = 1$ and reflectivity $\rho = 0$) and should therefore contain black radiation. The inner walls of the second cavity are constructed from a perfect reflector (emissivity, $\epsilon = 0$ and reflectivity, $\rho = 1$). Both cavities are theorized to be at the same temperature. It is then argued that if the second cavity is empty of radiation, that the second law of thermodynamics would be violated as photons could

travel from the first cavity into the second cavity and do net work, even if the temperatures of the two cavities were equal. As such, the conclusion is immediately made that the second cavity cannot be devoid of radiation and indeed must contain black radiation, even if a perfect reflector has no means of generating such radiation. Obviously, a logical error exists in such arguments. The question remains to identify the error.

2 Cavity radiation revisited — reversibility

Recently, the author has proposed two gedanken experiments in order to revisit Kirchhoff's law [15, 16].

In the first of these works, two cavities are considered, wherein a perfectly reflecting cavity is placed within a perfectly absorbing cavity (see Figure 1 in [15]). The experiment demonstrates that arbitrary cavities can indeed be permanently filled with arbitrary radiation [15]. This reinforces Max Planck's statement: *"... in a vacuum bounded by totally reflecting walls any state of radiation may persist"* [6, § 51]. This gedanken experiment and Planck's statement point to a direct contradiction of Kirchhoff's law, as the radiation within all cavities is supposed to be black, independent of the nature of the walls.

In the second of these works, two cavities are once again considered (see Figure 1 in [16]). This time however, the concern is centered on the nature of the cavities themselves. Of particular significance is the realization that the perfectly reflecting cavity cannot be made solely from a theoretical adiabatic wall. That is because such a wall cannot be characterized by any temperature [16]. As such, the author moved to create the second cavity from silver, although importantly, within a footnote, he emphasized that he had neglected the conduction bands of the metal. The idea was that all of the energy of the second cavity could be placed reversibly within its walls and phonons. Thus, the interior of the second cavity would be devoid of any photons. Thermal contact could then be made with the perfectly absorbing first cavity, and the energy contained within the phonons from the second cavity could be released, such that the second cavity becomes eventually filled with black radiation through the action of the first cavity [16].

The idea of this thought experiment was to consider what would happen within the perfectly reflecting cavity, if all of the energy within this system was initially reversibly contained within the phonons of its walls. No energy was permitted to be trapped in the conduction bands.

It could be argued that this was not a proper representation of the silver cavity. As such, it is also possible to build the second perfectly reflecting cavity from a material devoid of conduction bands, but now, to enclose both its inner and outer surfaces with adiabatic walls. In this case, all of the energy of the perfectly reflecting cavity can indeed be contained within its phonons. When the second cavity is placed in thermal contact with the first cavity, by removing part of the outer adiabatic walls, the energy will flow reversibly out of its phonons. This energy would move into the walls of the first cavity, enabling a photon to be produced and then to cross through a small opening into the second cavity. Both cavities end up being filled with black radiation. No net work is done as the displacement of phonons out of the second cavity, is exactly balanced by the entry of photons into its interior space. No net temperature change is experienced by the second cavity or by the first. All that has happened is that energy initially trapped in the walls of the second cavity has been released into the radiation field. Both cavities still possess the same energy as they did initially.

In hindsight, the reversible experiment was probably not well suited to represent a perfectly reflecting cavity. In fact, it could be imagined that if one removed the inner adiabatic lining from the second cavity, that the phonons could have been used to fill the cavity directly with photons. The first cavity was not even required in this case. This serves to emphasize Max Planck's approach, in that the energy of the system could be accounted for simply through the generation of the radiation field [6]. This has now been shown to be correct when the process involved in creating the field was reversible and no other processes are involved. However, not all processes in materials are reversible and this is why Kirchhoff and Planck have stumbled. Given the state of knowledge at the time, they were unable to properly consider the effect of conduction bands.

3 Cavity radiation revisited — irreversibility

This bring us to the question of what happens when the energy of the second cavity is irreversibly trapped within the conduction bands of the silver.* Let us once again state that the exterior of the first and second cavities are surrounded by adiabatic walls. The first cavity, constructed from graphite acting as a perfect absorber [16], is assumed to be filled with black radiation. The second cavity, constructed from silver acting as a perfect reflector [16], will be assumed to be devoid of any radiation. Then, let us place the cavities in contact, but this time permitting only a small hole to link the interior of the two cavities.

It is often argued that, under these circumstances, photons can flow from the first cavity into the second cavity. However, such a proposal in itself violates the second law. The problem is evident when one considers what happens to a photon which would enter the second cavity. It is clear that at some point, such a photon would interact with the wall of the second cavity. Since a photon contains both energy and momentum, it would impart momentum and energy momentarily to

*This is a structural question, as the presence of conduction bands becomes critical to the structure of silver. It is not possible to manipulate the energy associated with the formation of these bands without destroying the very nature of the metal. Hence, the existence of the conduction bands will be considered irreversible. As for the phonons, they are now assumed, within silver acting as a perfect reflector, to contain no energy.

the wall of the silver cavity. This is strictly forbidden by the second law, because heat would be moving into the wall of the second cavity, not only within the cavity void. Alternatively, consider the entry of the second photon from cavity 1 into cavity 2. This presents a substantial problem now, since cavity 2, having already gained the first photon, has a higher energy content than cavity 1. This is because *both* the cavity wall and the radiation field are used to define the system. Movement of the second photon into cavity 2 must be strictly forbidden by the second law, because heat would be moving from a cavity with a lower temperature to a cavity with a new higher temperature.

Still, our instinct desires that photons can enter the second cavity without violating the second law. The secret to resolving this problem involves the natures of the walls themselves. Let us divide the walls of each cavity into many elements. Within the perfectly absorbing cavity, each of the elements selected possessed at one time the energy contained in the photon at the frequency of interest. However, this energy has now flowed to the interior of cavity 1, as required by Max Planck [6]. The wall elements of the first cavity can be thought of as devoid of energy, but able to absorb the energy of the photon of interest. Conversely, the elements in the silver cavity can be thought of as containing the same amount of energy as the photon of interest. That is because for the silver cavity, all of the energy is initially confined to the wall elements.

Now, the only way to permit a photon to enter the second cavity without violating the laws of thermodynamics is to simultaneously permit an element from cavity 1 to interchange with an element from cavity 2. In this way, when the photon hits the wall of the second cavity, it will actually momentarily impart its momentum and energy to a wall which has now a reduced energy by the value contained in one element of the silver cavity. The photon can enter, but the net result is that the emissivity of the second cavity has begun to rise. Simultaneously, the emissivity of the first cavity, now short one photon and with one perfectly emitting element replaced with a perfectly reflecting element, has begun to fall. Should the cavities be of equal dimensions and contain equal numbers of elements, the net result would be that the total emissivity of both cavities becomes a weighted average of the joint emissivities. Both cavities now contain gray radiation and the second law was never violated.

It is evident that when the small hole was made between the two cavities, that their walls, from a thermodynamic point of view, became one. It is in neglecting this important fact that some physicists attempt to state that the second law of thermodynamics has been violated. In fact, the law is violated only when the experiment is not fully presented. The truth is that the net emissivity of the total cavity simply becomes gray. Photons can exist anywhere within this new cavity, but their net density will not be black.

At the same time, if it is possible to drive additional heat into this system, one can built up black radiation in these two cavities, as highlighted long ago by Stewart [7] and as re-emphasized recently by the author [17, 18].

4 Summary

In the end, arbitrary cavities are not necessarily filled with black radiation. Laboratory blackbodies are specialized objects always made from relatively good emitters of radiation over the frequency range of interest, as well illustrated by the facts (see references within [8–12]). No valid theoretical proof of Kirchhoff's law has been formulated and no gedanken experiments can properly account for the existence of this law.

Dedication

This work is dedicated to my second grandchild, Daniel.

Submitted on: January 6, 2016 / Accepted on January 7, 2016
First published online on: January 7, 2016

References

1. Kirchhoff G. Über das Verhältnis zwischen dem Emissionsvermögen und dem Absorptionsvermogen. der Körper fur Wärme und Licht. *Poggendorfs Annalen der Physik und Chemie*, 1860, v. 109, 275–301. (English translation by F. Guthrie: Kirchhoff G. On the relation between the radiating and the absorbing powers of different bodies for light and heat. *Phil. Mag.*, 1860, ser. 4, v. 20, 1–21).
2. Kirchhoff G. Über den Zusammenhang zwischen Emission und Absorption von Licht und. Wärme. *Monatsberichte der Akademie der Wissenschaften zu Berlin*, sessions of Dec. 1859, 1860, 783–787.
3. Wien W. Über die Energieverteilung in Emissionspektrum eines schwarzen Körpers. *Ann. Phys.*, 1896, v. 58, 662–669.
4. Stefan J. Über die Beziehung zwischen der Warmestrahlung und der Temperature. *Sitzungsberichte der mathematischnatur wissenschaftlichen Classe der kaiserlichen Akademie der Wissenschaften*, Wien 1879, v. 79, 391–428.
5. Planck M. Über das Gesetz der Energieverteilung im Normalspektrum. *Annalen der Physik*, 1901, v. 4, 553–563 (English translation by ter Haar D.: Planck M. On the theory of the energy distribution law in the normal spectrum. The old quantum theory. Pergamon Press, 1967, 82–90; also Planck's December 14, 1900 lecture *Zur Theorie des Gesetzes der Energieverteilung in Normalspectrum*, which stems from this paper, can be found in either German, or English, in: Kangro H. Classic papers in physics: Planck's original papers in quantum physics. Taylor & Francis, London, 1972, 6–14 or 38–45).
6. Planck M. The theory of heat radiation. P. Blakiston's Son & Co., Philadelphia, PA, 1914.
7. Stewart B. An account of some experiments on radiant heat, involving an extension of Prévost's theory of exchanges. *Trans. Royal Soc. Edinburgh*, 1858, v. 22, no. 1, 1–20 (also found in Harper's Scientific Memoirs, edited by J.S. Ames: The Laws of Radiation and Absorption: Memoirs of Prévost, Stewart, Kirchhoff, and Kirchhoff and Bunsen, translated and edited by D.B. Brace, American Book Company, New York, 1901, 21–50).
8. Robitaille P.-M. On the validity of Kirchhoff's law of thermal emission. *IEEE Trans. Plasma Sci.*, 2003, v. 31, no. 6, 1263–1267.
9. Robitaille P.-M. A critical analysis of universality and Kirchhoff's law: A return to Stewart's law of thermal emission. *Progr. Phys.*, 2008, v. 3, 30–35.

10. Robitaille P.-M. Blackbody radiation and the carbon particle. *Progr. Phys.*, 2008, v. 3, 36–55.

11. Robitaille P.-M. Kirchhoff's law of thermal emission: 150 Years. *Progr. Phys.*, 2009, v. 4, 3–13.

12. Robitaille P.-M. Blackbody radiation and the loss of universality: Implications for Planck's formulation and Boltzmann's constant. *Progr. Phys.*, 2009, v. 4, 14–16.

13. Schirrmacher A. Experimenting theory: The proofs of Kirchhoff's radiation law before and after Planck. *Hist. Stud. Phys. Biol. Sci.*, 2003, v. 33, 299–335.

14. Robitaille P.-M. and Crothers S.J. "The Theory of Heat Radiation" Revisited: A Commentary on the Validity of Kirchhoff's Law of Thermal Emission and Max Planck's Claim of Universality. *Progr. Phys.*, 2015, v. 11, no. 2, 120–132.

15. Robitaille P.-M. Further Insight Relative to Cavity Radiation: A Though Experiment Refuting Kirchhoff's Law. *Prog. Phys.*, 2014, v. 10, no. 1, 38–40.

16. Robitaille P.-M. Further Insight Relative to Cavity Radiation II: Gedanken Experiments and Kirchhoff's Law. *Prog. Phys.*, 2014, v. 10, no. 2, 116–120.

17. Robitaille P.-M. On the Equation which Governs Cavity Radiation I. *Prog. Phys.*, 2014, v. 10, no. 2, 126–127.

18. Robitaille P.-M. On the Equation which Governs Cavity Radiation II. *Prog. Phys.*, 2014, v. 10, no. 3, 157–162.

LETTERS TO PROGRESS IN PHYSICS

Mansouri-Sexl Test Theory: The Question of Equivalence between Special Relativity and Ether Theories

Maciej Rybicki
Sas-Zubrzyckiego 8/27, 30-611 Krakow, Poland
E-mail: maciej.rybicki@icloud.com

The goal of this paper is drawing attention to a mistake confusing discussion upon the alternatives to special theory of relativity (STR). In the Mansouri-Sexl test theory utilized as a mathematical framework for testing the preferred frame theories, the Lorentz transformation of time has an erroneous form. This generates a false conclusion, namely that a theory based on Tangherlini transformation is empirically equivalent to STR.

Before the advent of STR, FitzGerald [1] and Lorentz [2] proposed a solution to the Michelson-Morley experiment, different from that resulting from the Einstein's theory. Their idea, extensively developed in the Lorentz's theory of electrons [3, 4] (later known as Lorentz ether theory — LET) consisted in assumption that objects moving with respect to a postulated preferred frame of reference, determined by motionless "aether", are contracted in the direction of their motion. This idea, together with the introduced by Larmor assumption that clocks moving through ether slow down by a velocity dependent factor, sufficed also to explain the modified M-M experiment, i.e. the Kennedy-Thorndike experiment. Defined in these terms, length contraction and time dilation constitute real processes of dynamic origin, connected with the impact of absolute motion on molecular forces. However, after appearing of Einstein's 1905 paper on STR [5], this idea has been ignored and abandoned by the overwhelming majority of physicists. The reason was that, in spite of its different ontology LET did not formally differ from STR, neither led to specific empirical predictions. The underlying cause binds to the space-time transformations, in fact determining the shape of theory. Namely, the Lorentz transformation (to which Voigt, Larmor, Poincare and Lorentz contributed in various degree) evolved to a symmetrical form reflecting the STR founding postulates instead of the Lorentz's assumptions. Thus, paradoxically, Lorentz transformation became the main obstacle in evolving the original Lorentz's idea to a form of consistent autonomic theory. Eventually, LET gained the status of a superfluous ontology put upon the STR formalism (so-called "Lorentzian approach to relativity"), which made the choice between LET and STR the question of simplicity ruled by the Occam's razor. Neither the (much later) space-time transformation consistent with original assumptions (Tangherlini [6]), nor the Bell's exact calculations (Bell [7]) deriving "relativistic" effects from Maxwell's equations by means of classical physics and quantum mechanics, did alter this general opinion.

The today's version of LET takes the form of test theories verifying STR by introducing free parameters instead of these resulting from definite assumptions. They are in particular the Robertson's test theory [8] and Mansouri-Sexl theory [9–11] for their basic equivalence known by the common name of Robertson-Mansouri-Sexl test theory (RMS). We shall focus on the Mansouri-Sexl (M-S) transformation presented in [9], considered to be a proper mathematical framework for experiments verifying special relativity. While the Lorentz transformation (boost) is

$$\left.\begin{aligned} t' = \gamma\left(t - \frac{vx}{c^2}\right), \quad t = \gamma\left(t' + \frac{vx'}{c^2}\right) \\ x' = \gamma(x - vt), \qquad x = \gamma(x' + vt') \end{aligned}\right\}, \tag{1}$$

where

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}$$

(while $y' = y$, $z' = z$, in all transformations here considered), Mansouri & Sexl introduced a generalization:

$$t = a\,T + \epsilon X, \qquad x = b\,(X - vT) \tag{2}$$

The coordinates, X, T are the ones measured in the postulated preferred frame Σ in which the speed of light is axiomatically isotropic. Instead, x, t are the coordinates measured in frame S being in standard configuration with Σ. The idea consists in measuring independently the factors a and b (functions of v) in experiments, and to choose one of two alternative values of ϵ: $-v/c^2$ or 0, corresponding to the alternative synchronization conventions. The first, Poincare-Einstein (P-E) "internal" synchronization, based on the axiom of isotropic one-way speed of light in any inertial frame (i.e. based on the postulate of invariant speed of light), relates to $\epsilon = -v/c^2$, a factor responsible for the relativity of simultaneity. The second, "external" synchronization, related to $\epsilon = 0$, consists in adjusting all inertial clocks to the clocks

synchronized in the preferred frame Σ according to the P-E synchronization, which entails absolute simultaneity. Beside these two, there exists a third possible convention, the (slow) clock transport, which can be classified as internal procedure. The clock-transport convention confirms P-E synchronization provided STR is correct; instead its relation with the theories involving absolute simultaneity is not unambiguous inasmuch as they basically may, or may not predict time dilation and length contraction.

Any observed deviations from the exact relativistic values of a and b in the first or second order experiments (according to Mansouri and Sexl, resulting in deviations from the isotropic two-way speed of light) would speak for the preferred frame alternatives to STR. Mansouri and Sexl state that for $a = b = 1$, $\epsilon = 0$, the Galilean transformation is obtained, which is correct. If, after employing the external synchronization, a and b equal to unity, it would mean that mechanical phenomena are ruled by Newtonian physics and subject to the Galilean principle of relativity, while the Maxwell equations (and the relevant constant speed of light) refer to the preferred frame (ether) only. This is exactly what Michelson and Morley (ineffectively) expected to detect in their experiments.

However, Mansouri and Sexl also claim that for $1/a = b = \gamma$ and $\epsilon = -v/c^2$, their transformation turns into the Lorentz transformation, which is obviously wrong. This mistake is coupled with the incorrect notation of the Lorentz transformation of time, written in their paper as:

$$t' = \frac{t}{\gamma} - \frac{vx}{c^2}\,, \tag{3}$$

whereas the correct form is

$$t' = \gamma\left(t - \frac{vx}{c^2}\right). \tag{4}$$

In fact, this mistake is not simply accidental; being trivial, it has however a dipper cause. Namely, Mansouri and Sexl intended to treat separately the questions of time dilation and simultaneity. This, however, is infeasible with respect to the Lorentz transformation in which relativity of simultaneity and relativistic effects are inseparably connected. This mistake entails false conclusion as to the question of equivalence between STR and the postulated ether theory. It also maintains a persistent myth, according to which the Michelson-Morley experiment, together with the Kennedy-Thorndike experiment provides an evidence for the invariant speed of light. What these (and other) experiments proved in fact with a high degree of probability is the isotropy of the two-way speed of light, which however is not tantamount to isotropy of the one-way speed of light. Mansouri and Sexl came to a false conclusion that the difference in one-way speed of light is a sole matter of choice of the synchronization convention. Consequently, they concluded that only violation of the two-way isotropy resulting in deviations from the relativistic values of a and b constitutes a challenge to STR.

From among various alternatives to special relativity, the preferred frame theory (PFT) here considered seems to be the only one consistent with the Lorentz's original idea (we treat PFT as a specific formulation of "ether theory"). It is based on the general assumption according to which there exists a physically substantial preferred frame of reference, of which the properties are:

1. In the preferred frame, the one-way speed of light is isotropic;
2. The bodies moving in the preferred frame shrink by the Lorentz factor in the direction of their motion; the clocks moving in the preferred frame slow down by the Lorentz factor.

The effects mentioned in the second postulate are interpreted as "real", which means that their relation to the preferred frame does not depend on the choice of reference frame in which they are described. Provided that, from these postulates one derives the following asymmetrical transformation between the preferred frame Σ (coordinates T, X) and frame S moving with respect to the preferred frame (coordinates t, x):

$$\left.\begin{aligned} t &= \frac{T}{\gamma}\,, & T &= t\gamma \\ x &= \gamma(X - vt)\,, & X &= \frac{x}{\gamma} + vt\gamma \end{aligned}\right\}. \tag{5}$$

While using the notation used in M-S transformation, this would mean: $1/a = b = \gamma$, $\epsilon = 0$. Transformation (5) determines all dynamic and kinematical properties of PFT. Formally, the above transformation and Lorentz transformation do not convert to each other. Mansouri and Sexl quote this transformation in their paper [9], rightly attributing it to Tangherlini. However, they erroneously claim Tangherlini transformation differs from Lorentz transformation only with respect to the synchronization convention employed, which is a direct consequence of a basic mistake above-mentioned. They conclude that theories determined by these transformations (i.e. STR and PFT) are empirically equivalent to each other. According to this viewpoint, the ether system can be singled out in an arbitrary manner and thus respective predictions concerning experimental results in any inertial system are identical. This false conclusion confuses discussion on the Lorentzian approach for nearly forty years.

As a matter of fact, PFT shares some empirical predictions with STR. The main similarity is that PFT predicts length contraction and time dilation by the usual Lorentz factor, provided measurements are executed in the preferred frame (in more detail Rybicki [12]). It predicts e.g. the elongation of lifetime of muons crossing the atmosphere since the Earth frame is nearly identical (compared with the muon's speed) with the postulated preferred frame. It also gives identical to STR prediction (although different interpretation) to

the twin paradox, irrespectively of the choice of the observer's "rest" reference frame. This also refers to the "realistic" version of twin paradox, namely the Hafele-Keating experiment.

PFT predicts the isotropic two-way speed of light, which makes the M-S theory ineffective in testing this alternative to STR. To show this question in details, let us return to the usually used notation with primed and non-primed coefficients, here the latter attributed to the preferred frame (thus, below, S denotes the preferred frame and S' the frame in motion). From the fact that clocks and measuring rods moving with respect to S are distorted in the definite way by the Lorentz factor it follows that, in S', the speed of light traveling along x-axis is, dependently on the (positive or negative) direction:

$$c'_1 = (c - v)\,\gamma^2, \qquad c'_2 = (c + v)\,\gamma^2, \tag{6}$$

where v denotes the velocity of the observer with respect to the preferred frame along x-axis. (In the 2D and 3D depictions, the light wave front form ellipse and ellipsoid, respectively). The averaged two-way speed of light on path l' parallel to x-axis is constant (isotropic) since the respective time is

$$t' = \frac{l'}{(c - v)\,\gamma^2} + \frac{l'}{(c + v)\,\gamma^2}\,. \tag{7}$$

After simple algebra, one gets $t' = 2l'/c$, a result identical to that predicted by STR. While the speed of light defined according to STR determines the relativity of simultaneity, the speed of light defined according to Eq. (6) forms an alternative solution, in the sense that it determines absolute simultaneity.

In general, the concept of "relative velocity" between two frames, defined in STR as identical speed (the same for the observers in S' and S), is replaced in PFT by the concept of "mutual velocities". While S' moves against S with the velocity v, the speed of S measured in S' becomes

$$v' = v\gamma^2. \tag{8}$$

This involves significant consequences, e.g. such as the following one. Assume S' and S'' are the frames in motion to each other, and that their velocities with respect to the preferred frame S are identical. Since also the Lorentz factors described in S for the frames S' and S'' are identical, the mutual velocities measured in both frames must be identical either, thus constituting the "relative velocity" in the STR sense. However, contrary to the STR predictions, neither of these frames will manifest "relativistic effects" (length contraction and time dilation) when observed (measured) from the other one, since

$$\frac{\gamma'}{\gamma} = \frac{\gamma''}{\gamma} \Longrightarrow \frac{\gamma'}{\gamma''} = 1. \tag{9}$$

This specific prediction of PFT, together with the characteristic "position" of the Earth with respect to the assumed preferred frame enables experiment settling between STR and PFT. Namely, one assumes that, if the preferred frame exists, it is likely identical with the (local) frame in which the cosmic microwave background radiation (CMBR) is isotropic. Meanwhile, from the observed Doppler effect obtained from WMAP known as "dipole anisotropy" one deduces that Solar System moves with respect to isotropic CMBR with the velocity 368 ± 2 km/sec in the direction of galactic longitude $l = 263.85^\circ$ and latitude $b = 48.25^\circ$. This translates to the Lorentz factor:

$$\gamma = (1 - 1.52 \times 10^{-6})^{-1} \approx 1 + 7.6 \times 10^{-7}. \tag{10}$$

PFT predicts that an object moving with equal velocity with respect to the isotropic CMBR, in the direction (e.g.) opposite to that of Solar System (i.e. $l = 83.85^\circ$ and $b = 228.25^\circ$) will not exhibit any relativistic effects since $\gamma'/\gamma'' = \gamma''/\gamma' = 1$. This prediction is absolute, i.e. does not depend on the choice of synchronization conventions or any other assumptions. It is quite obvious that in the lab experiments with γ reaching the value of 1,000 (thousand) and higher, the difference between 7.6×10^{-7} and zero is not identifiable. To be detected, it thus demands employing subtle methods in the specially aimed experiments. Nevertheless, it does not seem to lie beyond the scope of the today's experimental capabilities.

Conclusion

We have shown that an incorrect notation of the Lorentz transformation of time in the Mansouri-Sexl test theory entails false claims, namely:

1. Only the theories predicting anisotropic two-way speed of light differ from STR;
2. A theory maintaining absolute simultaneity is equivalent to special relativity (Mansouri and Sexl call this a "remarkable result");
3. As far as prediction of experimental results is concerned, Tangherlini transformation is completely equivalent to Lorentz transformation.

These claims confuse the discussion upon the preferred frame alternatives to special relativity. Contrary to a common belief, a theory based on the preferred frame postulate and formalized by Tangherlini transformation (i.e. PFT) is not in whole experimentally equivalent to STR. Thus settling between them two in experiments is a feasible task. The present author aims to develop this subject in the subsequent papers.

Submitted on January 6, 2016 / Accepted on January 8, 2016

References

1. FitzGerald G.F. The ether and the Earth's atmosphere. *Science*, 1889, v. 13, 390.
2. Lorentz H.A. The relative motion of the Earth and ether (1892). In: *Collected Papers*, v.IV, Hague, 1937, 219–223.

3. Lorentz H.A. Attempt of a Theory of Electrical and Optical Phenomena in Moving Bodies. Leiden, E. J. Brill, 1895.
4. Lorentz H.A. Simplified theory of electrical and optical phenomena in moving bodies. *Proceedings of the Royal Netherlands Academy of Arts and Sciences*, 1899, v. 1, 427–442.
5. Einstein A. Zur Elektrodynamik bewegter Körper (On Electrodynamics of Moving Bodies). *Annalen der Physik*, 1905, Bd. 322(10), 891–921.
6. Tangherlini F.R. *Nuovo Cimento Suppl.*, 1961, v. 20, 351.
7. Bell J.S. Speakable and unspeakable in Quantum Mechanics. In: *Collected Papers on Quantum Philosophy*, Cambridge University Press, 2004.
8. Robertson H.P. Postulate versus observation in Special Relativity. *Reviews of Modern Physics*, 1949, v. 21(3), 378–382.
9. Mansouri R. and Sexl R.U. A test theory of Special Relativity I: Simultaneity and clock synchronization. *General Relativity and Gravitation*, 1977, v. 8(7), 497–513.
10. Mansouri R. and Sexl R.U. A test theory of Special Relativity II: First order tests. *General Relativity and Gravitation*, 1977, v. 8(7), 515–524.
11. Mansouri R. and Sexl R.U. A test theory of Special Relativity III: Second order tests. *General Relativity and Gravitation*, 1977, v. 8(10), 809–814.
12. Rybicki M. Critical analysis of Special Relativity in reference to the energy transformation. *Apeiron*, 2008, v. 15, issue 3, 270–293.

The Journal on Advanced Studies in Theoretical and Experimental Physics, including Related Themes from Mathematics

PROGRESS IN PHYSICS

A quarterly issue scientific journal, registered with the Library of Congress (DC, USA). This journal is peer reviewed and included in the abstracting and indexing coverage of: Mathematical Reviews and MathSciNet (AMS, USA), DOAJ of Lund University (Sweden), Scientific Commons of the University of St. Gallen (Switzerland), Open-J-Gate (India), Referativnyi Zhurnal VINITI (Russia), etc.

Electronic version of this journal:
http://www.ptep-online.com

Postal Address

Department of Mathematics and Science,
University of New Mexico,
705 Gurley Ave., Gallup, NM 87301, USA

ISSN: 1555-5534 (print)
ISSN: 1555-5615 (online)

Standard Address Number: 297-5092
Printed in the United States of America

January 2016 **Vol. 12, Issue 2**

SPECIAL ISSUE

The Method of Physics

CONTENTS

Tselnik F. Irony of the Method: Foundations of Theoretical and Experimental Physics with Special Emphasis on the Contact Problem in Mechanics, Fields, and Particle Interactions

Preface. What is the question? 95

Part One. How and Why?

Chapter 1. Taking one step back 97
Chapter 2. Forces in terms of contacts: prediction of the link 108
Chapter 3. Fields and their propagation: prediction of the chain 113
Chapter 4. Quantum theory: repeatability of the non-repeatable 119
Chapter 5. Gravity: a forceless force 127
Chapter 6. What interactions are permitted by the Method? 129

Part Two. What For?

Chapter 7. Repeatability 136
Chapter 8. Light of expired stars 139
Chapter 9. From scratch. Uniqueness and repeatability 142

Information for Authors

Progress in Physics has been created for rapid publications on advanced studies in theoretical and experimental physics, including related themes from mathematics and astronomy. All submitted papers should be professional, in good English, containing a brief review of a problem and obtained results.

All submissions should be designed in LaTeX format using *Progress in Physics* template. This template can be downloaded from *Progress in Physics* home page http://www.ptep-online.com

Preliminary, authors may submit papers in PDF format. If the paper is accepted, authors can manage LaTeXtyping. Do not send MS Word documents, please: we do not use this software, so unable to read this file format. Incorrectly formatted papers (i.e. not LaTeXwith the template) will not be accepted for publication. Those authors who are unable to prepare their submissions in LaTeXformat can apply to a third-party payable service for LaTeX typing. Our personnel work voluntarily. Authors must assist by conforming to this policy, to make the publication process as easy and fast as possible.

Abstract and the necessary information about author(s) should be included into the papers. To submit a paper, mail the file(s) to the Editor-in-Chief.

All submitted papers should be as brief as possible. Short articles are preferable. Large papers can also be considered. Letters related to the publications in the journal or to the events among the science community can be applied to the section *Letters to Progress in Physics*.

All that has been accepted for the online issue of *Progress in Physics* is printed in the paper version of the journal. To order printed issues, contact the Editors.

Authors retain their rights to use their papers published in *Progress in Physics* as a whole or any part of it in any other publications and in any way they see fit. This copyright agreement shall remain valid even if the authors transfer copyright of their published papers to another party.

Electronic copies of all papers published in *Progress in Physics* are available for free download, copying, and re-distribution, according to the copyright agreement printed on the titlepage of each issue of the journal. This copyright agreement follows the *Budapest Open Initiative* and the *Creative Commons Attribution-Noncommercial-No Derivative Works 2.5 License* declaring that electronic copies of such books and journals should always be accessed for reading, download, and copying for any person, and free of charge.

Consideration and review process does not require any payment from the side of the submitters. Nevertheless the authors of accepted papers are requested to pay the page charges. *Progress in Physics* is a non-profit/academic journal: money collected from the authors cover the cost of printing and distribution of the annual volumes of the journal along the major academic/university libraries of the world. (Look for the current author fee in the online version of *Progress in Physics*.)

Irony of the Method

Foundations of Theoretical and Experimental Physics with Special Emphasis on the Contact Problem in Mechanics, Fields, and Particle Interactions

Felix Tselnik
Beer-Sheva, Israel. E-mail: tselnik@bgu.ac.il

The Method of Physics is not built on the basis of hypotheses about the world. It is based on the axioms of the requirements of universal reproducibility of predictions. Thus, the Method does not require confirmation in experiments: experiments are carried out in the framework of the concepts of the Method and, therefore, they are doomed to agreement with the theory (derived solely from the axioms). Critical analysis of such structures (of the Method) as time intervals, the reference systems, and distances leads to a series of rather unusual conclusions...

Contents

Preface. What is the question? 95

Part One. How and Why?

1. Taking one step back 97
2. Forces in terms of contacts: prediction of the link 108
3. Fields and their propagation: prediction of the chain 113
4. Quantum theory: repeatability of the non-repeatable 119
5. Gravity: a forceless force 127
6. What interactions are permitted by the Method? 129

Part Two. What For?

7. Repeatability 136
8. Light of expired stars 139
9. From scratch. Uniqueness and repeatability 142

Preface. What is the question?

> Mathematical science affords us a brilliant example, how far, independently of all experience, we may carry our a priori knowledge... Deceived by such a proof of the power of reason, we can perceive no limits to the extension of our knowledge. The light dove cleaving in free flight the thin air, whose resistance it feels, might imagine that her movements would be far more free and rapid in airless space.
>
> I. Kant, *Critique of Pure Reason*

The Method of reasoning is regarded an important part of our civilization. However, its very existence is paradoxical. Indeed, it is unlimited repeatability that is in its heart. But repeatability on its own doesn't as yet belong to the Method. The knowledge of a town is verily not in the competence of the Method, despite providing suitable recommendations for us to search for a house. But the town might change in time, and then such knowledge becomes useless. On the contrary, constructions and rules of the Method are claimed to be *universal*, that is, valid always and everywhere. But then, an available set of universal rules is unavoidably meager, since it is formed at the expense of disregarding everything that is uncertain, unreliable, or vulnerable by means of restricting full-fledged thinking to mere logic. The utmost form of *unambiguous* repeatability is number. A hundred of people — this is when there is no importance as to how people actually differ from each other. Even if random processes are under question, then the result is being presented in terms of their probabilities, and their — repeatable! — distributions are what is actually obtained. Repeatability is required as long as — explicitly or implicitly — one bears in mind some practical use of past experience. However, there are no completely repeatable situations in real life. Moreover, they are just unrepeatable events that are of utmost interest in it. Now, what for — and just when — are we in true need of this Method?

Since prehistoric times there have been highly valued, along with wise (sometimes) and sly (always) leaders, strong and bold warriors, skilled and nimble hunters, also those able to predict weather, to recognize a beast, while being led by hardly noticeable or completely unnoticeable for others signs, to ignite fire, to invent a tool. Frequently, these people were directed by intuition, incomprehensible even for themselves ("it seems to me" or "my bones are aching feeling bad weather"), and then their skill disappeared with them, but sometimes they managed to explain their knowledge at least to a disciple, and then it had a chance to be preserved. Thus in this way the Method has been coming into existence, and for later purposes some other people, the "philosophers", have endeavored to put all this into a system (in great many different ways), in order to make it systematically easier to understand and remember. For the large part in later-developed "natural sciences", experiment has replaced experience, and the combinations of experiments and theories have become a commonly accepted way of acquiring knowledge.

"Pure" conditions of experiments along with the prescription to use solely their combinations in applications are called upon to ensure just this universal repeatability, while getting rid of uncertainty, "turbidity" of real life, which still manages somehow to use the predictions of the Method. A confidence deserving experiment must present an unambiguous result, as well as a theory — an unambiguous deduction. The main concern and skill of the experimenter consist in this that some definite statements might be drawn from his result, whereas he mostly observes on his display (set-up) something non-repeatable, from which no definite conclusions could be drawn, and he has to update his devices and the performance of the experiment in such a way as to reach reliable repeatability. Many say that a theory is to be checked with the experiment, but then the performance of experiments is being controlled by theoretical concepts. All this is to be used further on in practice, but there is a question as to what extent the result of an experiment is ultimately conditioned by initial theoretical concepts. But what if these concepts are so restrictive that there is no need in the experiment itself: its result cannot be different being governed by the very statement of the problem, or can it?

It is commonly believed that upon perfection of experimental devices and corresponding refinement of the theory, every "reasonable" question will receive a trustworthy answer. Upon penetrating deeper still into the "structure" of matter, we shall eventually learn and understand everything about Nature. In this approach, physics and the whole science is thought of as something existing of its own being an object of unprejudiced and uninterested study.

The entire society gets accustomed to such an opinion, which has acquired the status of Kipling's "Bandar-Log criterion": "We all say so, and so it must be true". Now, the very success of technologies becomes dependent on themselves, just like advertising produces artificial needs. Such a development might turn out to be too one-sided and vulnerable with respect to future failures (apart from those inherent in the society itself), following just from the Method, while the label "reasonable" as applied to a question is often called upon to forbid curiosity that is not sanctioned by the Method.

It is desirable therefore to scrutinize the very structure of the Method, viz., its language, because the questions are always asked using a language, hence the answer is partially contained in the question itself. Since, if you are being answered in an unknown language, you will regard the answer as mere "noise". However, the language of the Method is quite different from the languages of primitive tribes, so it is to be asked on a much deeper ground as to why it is just such and to what degree its answers are determined by the requirement of repeatability alone. It turns out that this requirement is so restrictive that, at least in physics which is the example we will confine ourselves to, that we even should not expect from Nature her own answer. The answer is always completely determined by the very question, so, in principle, one could dispense with experiments at all.

The only general answer of Nature to the questions of theory is "everything might occur", whereas the Method likens to a stencil, revealing from the unlimited variety of Nature that is compatible with the structure of its own pattern as it gets finer and more sophisticated in the course of development. This has long been stated by Kant, Bergson and others. Pushkin's "monotonous beautiness" is well applicable to the theories of the Method and should be explained by the own pattern of the stencil. However, the stencil is by no means arbitrary, but, on the contrary, it meets the most important requirements of the user, while the meagerness of its constructions ("How can everything be described with so simple formulae?") results from the severe restriction due to the condition of repeatability and the difficulty of its observing, as it will become clear in the second part of this book after the explanation, in the first part, of the possibility to realize the constructions of the Method basing it solely on this condition.

One should say that physicists by no means discover the laws of Nature, which has no laws of its own, but only particular cases, while to say that this and this is not important, and then it is possible to predict what is left — this is of concern to science. Suffices it to inquire why abstract mathematical constructions, initially by no means answering the questions from physicists, later on became required, to find that both merely consider equal situations, namely, what could be unambiguously predicted. In other words, the user is being advised to "search under lamp", since nothing can be found in the dark anyway, at least if we observe repeatability. But then, whenever you succeed in rendering a practical problem acceptable for the Method, the efficacy of solutions is guaranteed, and all our high valued technology is based on it.

The image of the World, as provided by the Method, is not really a picture but rather a drawing — in projections and with dimensions. A picture creates different impressions in different people and in different times, therefore being devoid of complete universality. If not only the picture, but also its impression would always be identical, only then would it belong to the Method — though no longer to art. The products of the Method play an important though auxiliary role. So, the image in the mirror might provide slightest details, but the problem for the Method is to make a good mirror, and this is independent of the real (whole) countenance to be image-processed.

In the first part of this book, the basic geometrical structure of the Method is discussed to realize some particular ways in order to reach repeatability, which form the essence thereof is called *physics*. In order to facilitate understanding by a reader not accustomed to calculations, we present no formulae. These will be replaced by multiple figures along with qualitative explanations of the presented constructions. Infrequent cases, in which the absence of the corresponding calculation comes into conflict with the confidence in the statement, are being supplied with a short description of the gen-

eral scheme of the calculation and with auxiliary constructions.

In the second part, we discuss the requirement of repeatability itself with respect to its relation to reality. This is necessary in order to define boundaries of the applications of the Method. The main tendency here will be to define these boundaries from *inside* the Method as it is performed by means of further analyzing its basic concepts.

Part One. How and Why?

> And for mean life number was existing:
> Like domestic harnessed cattle served,
> Since the slightest shadows of meaning
> Clever number readily exposed.
>
> N. S. Gumilev, *The Word*

Chapter 1. Taking one step back

> I do not define time, space, place, and motion, as being well-known to all.
>
> I. Newton, *Mathematical Principles of Natural Philosophy*

"One of the basic concepts of mechanics is the concept of *material point*. Under this name we conceive the body, the size of which might be neglected while describing its motion. "The position of the material point in space is defined by its radius-vector $\boldsymbol{r}$, the components of which coincide with its Cartesian coordinates x, y, z." (L. D. Landau and E. M. Lifshitz, *Mechanics*.)

This or about this is the way to present the primary positions in any textbook in physics. It is implied that the reader, upon receiving a standard education and upbringing, asks no more questions in this respect. More cautious mathematician begins differently: "A number of experimental facts are a basis of classical mechanics... Our space is three-dimensional and Euclidean, and time is one-dimensional...". (V. I. Arnold, *Mathematical Methods of Classical Mechanics*.)

Newton proposed a scheme to solve some practical problems called *mechanics*, based on a generic system (devised some decades before the method of Descartes) worked out to unify algebra and geometry and on using coordinates in order to relate points to numbers. This Newtonian approach survived until now, though with an important improvement due to Einstein. Let us describe the main ideas of the scheme in general. In so doing, we intentionally scrutinize the elementary logic of the scheme, keeping in mind the development of its logical alternative in the sequel.

One has to choose a three-dimensional reference system, comprised of solid rulers or some other devices to be used for a coordinate network. The concept of *material point* is introduced as a *body* that moves along a one-dimensional smooth structure — the *path*, each point of which is specified with three numbers — coordinates — and marked with one more number — a time moment. Time flows uniformly, ensuring the absence of self-intersections in the full picture, even if the coordinates of the points of the structure are repeated.

Such a picture arose due to the observations of small or distant objects, so that their details are not important for the possible user of the scheme. This elementary description of natural events is selected for its simplicity and for the possibility of making predictions for a future position of the body. The body might change its shape, something might occur inside it, for example, a chemical or life transformation; all these are of no importance. We are interested only in this that we are, in a limited sense, able to give prediction. As S. Lem noticed in his "*The Sum of Technology*": "If everything that you want to know about the hanged is the period of his swinging on the rope, then you are a physicist".

For an actual use of the scheme, the body must be "seen", i.e. its initial coordinates must be known as well as some rules for finding its later position in the same coordinates. In Newton's time, no evidence of a top velocity would come from practice, otherwise his mechanics might have looked differently. Quite oppositely, it seemed then that for each motion a faster motion could exist. So, the body must be "seen" immediately wherever it was. Then it would be possible to follow it. Otherwise, provided the signal was retarded, the body might overtake the signal upon its acceleration to be lost from sight, and then another (similar if not marked) body might be confused with it. However, the very importance of the solution here is just in the possibility to influence the situation intentionally, which would never happen under confusion.

Further in Newton's mechanics, among all possible trajectories a particular subset is selected comprising all *uniform* and *rectilinear* ones, that is, straight lines in the four-dimensional space-time. Of these, each one is being determined by any pair of its points. One such point would be irrelevant for mechanics. If the reference frame had been chosen so poorly that any motion from an initial point uniquely determined the final point, it would then be impossible to influence this transition, and one would just unite these points. In this scheme, no explanation for this particular choice was presented. Indeed, there are many classes of lines, each one being specified by any pair of its points. However in Cartesian coordinates, a straight line corresponds to the simplest — linear — equation, having good properties with respect to linear, in particular vector-like operations.

It is assumed in Newton's scheme that these trajectories correspond to "free" motion, i.e. without external influence. Existence of such trajectories (though they never existed in practice, but rather belonged to some limiting case) is known as "the first law of Newton". "The second law of Newton" consists in representing all sufficiently smooth trajectories by

means of broken lines comprised of straight line segments, infinitely small in the limit, so that transitions between adjacent segments — acceleration — depend on an external influence — force, and on an individual parameter of the body — its mass.

The necessity of dividing the influence into external force and mass is not evident from the scheme itself. However, it is necessary for it, within its self-definition, not only to extend the variety of relevant situations, but also to predict future positions of the body according to the very statement of the problem. Indeed, in the general scheme the force must be specified at all points up to the end, and what is left to predict then? Therefore the scheme should be completed with a notion of inertia: whatever force, there exists the last segment, for which this force could be ignored so as to regard the motion as free; and this must be true for any intermediate step as well.

In practice, various particular cases are considered in mechanics, in which forces are known in advance over the whole possible trajectory, however in the full theory there must be a guaranty of the meaningful problem statement with any forces whatsoever. Forces arise from their sources, typically coming from some other bodies to weaken over a distance from them. Different forces bring about different accelerations, hence besides its mass the body to be accelerated must be provided with an additional parameter — "charge", requiring that acceleration due to this force be proportional to this charge disappearing in the limit of zero charge.

But then, what is the way to find a force in general, while knowing nothing of its source? The answer of the scheme is: this is achieved using the same second Newtonian law though inverted to determine a force by the acceleration that it provides. This looks like a vicious circle, but the point is that to determine the force along the trajectory of the body in question, the force is being measured according to acceleration of different — test — bodies, to be afterwards used in the problem. But then one must be sure that all the test bodies observe a common measurement standard. This is by no means a simple task, and a further restriction of the permitted class of forces is required. This topic will be extensively discussed in Ch. 6.

The scheme allows for extension on extended bodies, however only at expense of additional restriction on the class of permitted forces. If the body changes its shape or size, then sometimes it can be considered as consisting of smaller parts, each one moving along its own trajectory unchanged, while all these together describe the behavior of the total object. Then one more law — "the third Newtonian law" — is required to introduce, in the notions of the scheme, motions of finite length bodies (e.g., solid ones) as a whole by virtue of internal forces analogous to the external ones, thereby preventing decay of the body. Since these forces must not influence the motion of the body as a whole, they must compensate reciprocally. In terms of the Newtonian second law, this sounds as follows: "Action is equal and directed oppositely to its anti-action". Newton himself noticed that these internal forces (for solid bodies) must be very strong as compared to the external force, so that the latter only moves the body rather than deforms it.

If the source of an external influence is explicitly given in the problem, people tend to speak about interaction, and then the Newton's third law calls for the intensity of the source to be represented by the same charge. In this context the charge is coined the *constant of motion*. So, if, for example, the force comes only due to interaction of the bodies, all these together might be considered as a whole ("closed system"), that is, the law allows for only some definite class of forces for this condition to be fulfilled.

The primary concept of material point as a body of "zero" size, that is, such that its state is completely specified by its three coordinates, does not depend on a general concept of size, but it might serve as a preliminary for the latter. If a solid extended body is considered as consisting of so small pieces that each one is practically a material point, then its position is specified by means of its coordinates in the *same* reference system. Further, when approximating each acceptable path with rectilinear segments, as well as ascribing to each segment its *length* as specified by positions of its ends again in the same coordinates (since there are no other numbers in the scheme), then in a path defining limiting process this length must tend to zero independently of the orientation of the body by its own definition. This might be achieved with the definition of squared length via, e.g., the sum of their three differences squared. Now we can introduce a concept of size also for extended bodies as a maximum length of segments specified by any pair of its points, again and again in the same coordinate system that was first introduced solely for paths. We are in a position now to redefine a zero-size body as one not containing finite length segments.

Over a few centuries of its use, the Newtonian scheme became so customary that it was conceived as something belonging to Nature as her very own — her own internal "harmony" — amounting to the statement that Newton "discovered" his laws, being up to then ready though unknown, firmly hidden in Nature. A significant alternative approach (Kant, Bergson and others) denies the existence of any laws of Nature, regarding the scheme as merely a choice of objects, to which attention should be paid according to some convenient rules in order to be useful in applications. Thus, one should regard the Newtonian laws an "invention" rather than a "discovery". The user (applying the scheme) just looks around obliviously grasping only situations in which a useful action is possible. The Newtonian scheme provides the user with instructions for paying attention to particular occurrences, namely, to pick rare cases, which allow for warranted predictions. We'll examine below this question systematically, but for the time being let us look at the mentioned features of the scheme from the viewpoint of their necessity.

The final product of the scheme — three number functions: the dependence of coordinates on time. But why are these needed? How and when to use this information? A possible answer: if these functions for one body at some moment of time coincide with those for another body at the same time, the bodies might collide, that is, come into contact. So, these functions get a practical meaning only if some other bodies are present, otherwise the trajectory is found hanging, as to say, in nothing. Why then not to consider the full problem, including explicitly all the participating bodies? How reasonable is the division of the problem into separated parts, while the question is universal: whether or not the bodies collide?

Furthermore, the canonical (Newtonian) scheme is redundant in this respect that many reference systems for the solution of the same problem might be introduced on an equal footing, and some additional rules are necessary for transitions between them upon describing the same motion. The trajectories are specified with number functions, which are different as for different trajectories in one reference system, so also for one trajectory as described in different systems. Disentanglement of this ambiguity requires some special rules — "relativity principles" — to find "covariant" combinations of coordinates, such that the "form" of resulting equations will not depend on a particular reference system (in the theory of relativity these combinations inherently include also time). In this respect two types of coordinate frame transformations are introduced: "passive" transformations to change the coordinate values of a given point and "active" when the point shifts itself in a given coordinate frame (still one has to ascribe any sense whatever to the notion of motion from one point to another in empty space: what is the difference between these two?).

Moreover, reference systems themselves require a gauge to be relevant for representing actual motions of bodies. The gauge is carried out using some *standard* trajectories. For example, the rectilinearity of rods is gauged using light rays or free fall of bodies, while clocks are gauged with some stable processes. Finally, all motions are thus compared to some others, while rods and clocks are a mere intermediary for comparing motions. However, any intermediate device might either introduce something of its own to the procedure, or lose or hide something. Usually, this is harmless — still sometimes it might be important. In particular, as we will find later on, some experimental facts turn out to be unexplainable in the framework of the scheme just because a part of information is actually lost in the intermediary. For example, the "eternal" question about the dimension of the space: why only three, not seven or two (fewer still)? In the sequel we will find also some other examples.

However successful the canonical version proves to be, still a question is there concerning its possible logical alternatives. Now, what if some different schemes might exist, and those should better meet some of our needs, whereas we have just got used to this very approach upon being taught to think solely in its framework? In the second part of the 20th century, many authors strived (though generally with limited success) to get rid of clocks and rods, replacing them with the propagation of light and free fall of bodies. However, the very idea of independently existing space-time has always been considered "intrinsic to our intuition" (in the sense of Kant's judgments a priori).

It is desirable therefore to begin with something more "primary", for example, with that which makes itself evident even for the "naked eye" like the possibility of describing in a similar way such seemingly quite different events as motions of stars and flights of birds. However, a many-century-long tradition is so mighty that even the discourse on the prime position without a preliminary ripe feeling of its necessity brings about, as experience shows, a depressing effect. Therefore, we begin with the discussion of ways to reach repeatability, though a bit prior to the cited textbook (as well as others), while postponing the most primary ideas until the last chapter.

It is customarily said that bodies move along their trajectories — one-dimensional continua. It is this that we want to consider in more detail from the viewpoint of the user, all this science being ultimately destined for. In distinction to the curious researcher with his traditional "why?", the question of the user is more prosaic — "what for?". He expects recommendations for action, and it is these recommendations that only give a value to knowledge, hence the concepts of a scheme must be coordinated with its expected predictions.

The concept of motion itself depends on the statement of the problem. For instance, the orbit of a satellite might be considered as a change of its position, but sometimes (in atomic processes, say) it is more to the point to regard as motion only changes in the orbit itself. In the canonical version, it is the initial state that is highlighted as a state that later on specifies the whole trajectory along with the law of motion. Just this approach makes it indispensible to accept in advance a particular construction of space-time. Otherwise there is no reason to choose something definite for a proposed change of the state. Indeed, what is a state then the initial state has to transform at?

Quite oppositely, the final state is something known to the user already prior to addressing the Method. This is something the user wants to reach. Therefore the final state possesses its *external* description as known to the user independently of the Method, which is then committed in order to find a way (if possible) so as to reach the desirable. If the user is not able to formulate as to what is wanted, the Method cannot teach him. And only afterwards — already in terms of the Method — the problem arises concerning a relevant construction, so that both the final and initial states are now encoded accordingly to the problem statement. Thus, with respect to the relationship of initial and final states, it is just final states that are to be specified as leading ones, leaving only auxiliary roles for initial states.

So seemingly unimportant deviation from symmetrical, as it looks, relationship in a ready scheme is important for an initial formulation of the problem. Likewise, the cause-effect relationship is asymmetrical for the user. The effect is important for him in its own right, whereas the cause is important only insofar as it leads to the known-in-advance effect. Only starting with this statement does it become possible to substantiate geometrical constructions as an instrument for solving practical problems.

As for the final state, a criterion of whether or not it is reached must be formulated by the user in advance. Otherwise the problem does not exist, since upon reaching the final state it is still unknown whether or not just this was implied. Once this state is defined, we stay in need of finding a way to reach it. The required construction must somehow encode this final state now with its own *internal* for the Method description, i.e. in its internal terms serving as a tool for the solution. In this procedure, the initial state must be encoded as well, since it isn't the desired one. If these two codes are sufficient, that is, given the initial state, the final is surely achieved, then no problem at all is there, as well as if the required transition is not known as yet. It might happen, however, that in the accepted encoding scheme there is an *intermediate* state such that a way to reach this state from the initial is known and also from this state to the final. For example, it might happen that in the ripening of a plant its color in an intermediate phase defines its properties at the end. Then the farmer (biotechnician) is interested in reaching just this color.

Further development of the scheme comprises a sequence of intermediate states with transitions in between. When ascribing the index 0 to the initial state and 1 — to the final, then let us ascribe 1/2 to the intermediate state. Analogously, the intermediate state between states 0 and 1/2 receives index 1/4, while that between 1/2 and 1 — 3/4 and so on. Proceeding along this way, we obtain a structure ordered by the very statement of the problem, still not, however, acceptable as a trajectory. Indeed, the structure cannot include the last state other than the final (marked as 1). Were it that such state does exist, the transition from it to the final would be necessary, and then this state would actually render itself superfluous, and it must be identified with the final. The same is true for each intermediate state. Therefore the whole set of states must be infinite. However, in this structure sequences might exist, that have no final states further progressions to begin with. In the order-defined topology, these sequences are everywhere dense (they correspond to irrational indices). But this numbering expresses only the order in the set of states, being arbitrary in other respects. It is possible to change the indices, so that a formerly irrational index becomes rational and vice versa (this cannot be done, however, for all the indices at once). It is natural to consider every such sequence a definite state, since the sequence of its indices has an upper boundary by definition, dividing the whole construction in "before" and "after" (Dedekind). It is this very construction that will be called a trajectory.

Up to now our indices look like time moments only due to their linear order. In what follows this likeness will acquire a definite physical sense, however, this will not be in accord with the readings of some clocks. No clocks whatsoever will appear in this book; rather surprisingly, it turns out that these are not at all needed in physics (as well as rods).

In various fields of knowledge states are fixed differently. In particular, physicists suggest their own approach, which is effective, however, only within a very limited scope of real situations, while subsequently letting predictions be universally repeatable. They notice that a final state is always encoded according to the "yes-no" principle by the very problem statement. Now, it is proposed in physics to encode all other states in the same way. States encoded with this rule will be called *contacts*. The contact is either existent or not, that is, it is a point, and this definition has nothing to do with such notions as size or distance. (If a duelist missed having just one cartridge, he doesn't care how far he missed.) Contact as a state corresponds to the pictorial image of touching, and in this respect the involved entities are called bodies, but we stress that the concept of body has here but a pure informational meaning independent of an illustrating picture referring to something used in connection with contacts. This image helps the user pay attention to similar situations, in that he tries to select bodies out of the world and to reduce his problem to their touching. In general, this is a mere mental construction introduced in an ivory tower independently of any reality. But then, it allows for effective construction of transitions between relevantly encoded states. It is only afterwards, while leaving the tower armed with the scheme and using his senses, one has the opportunity to search in the surrounding world something looking like the elaborate mathematical scheme in order to make predictions in actual situations. Thus, the astronomer Galileo, who used to observe the motions of celestial objects (demanded by the practice of navigation), began to throw for some reason stones from the leaning Pisa tower, thereby founding experimental physics, in distinction from the purely observational. How funny he must have looked to the others! People used to plough, fight, bargain, whereas this eccentric man was throwing stones.

After this digression let us return to the correspondence between the scheme and the usual concepts. What is of interest in a trajectory? It is only this that, if it intersects with another trajectory, then the related bodies might come into contact. What happens in the contact is a separate aspect unrelated to the given problem. The essence of the concept of the trajectory is in this, namely that this cannot happen, provided these trajectories do not intersect, and then knowing the trajectories we are in a position to predict the occurrence of the contact. If a hare comes into contact with a wolf, it is not necessary that it will be eaten: perhaps the wolf is not hungry at the moment. But the hare-physicist, being familiar with the

basics of geometry, knows for sure that he will be safe upon avoiding any contact with a wolf. This little one would get *guaranty*. So, for him the contact possesses a meaning that is *external* to his problem, which is "to be or not to be eaten", but in order to exclusively solve a *Contact Problem* (to be abbreviated throughout the text as simply "CP", by the use of which we generally also imply the sense and instance of "CP set-up"), the user is advised to develop for this purpose some artificial *internal* mental map with respect to the Method's construction.

Therefore, the fact of the presence or absence of a contact might be taken as the starting point for a special science, namely, *physics*. Indeed, something common for an apple and stars that Newton noticed, according to the legend, belongs just to their mechanical motions — trajectories with their contacts with light entering the observer's eye, say, rather than with some changes in general, for instance, with the evolution of the star and ripening of the apple. The problem is being stated about the prediction of contacts on account of some relevant initial data.

Upon developing the Method from scratch, we should not accept all the traditional geometry as fallen from heaven. Rather, we shall first put the question about the relevance of just this structure, though perhaps just this one will spring up in some form in the course of our development. We have defined the scheme as CP. But in any CP, at least two bodies are present. Hence, it is not necessary to construct an external reference system with its coordinates fixing "positions" of a body, dropping the presence of other bodies from the outset. Now, what if we were to formulate CP directly in their relations? Perhaps then we should be able to dispense with coordinates? It turns out that this is the case, and we shall forever forget coordinates, their transformations, quantities that transform in some accord to these, relativity principles, etc.

Solutions to CP must be unambiguous and universally repeatable. These demands are so categorical and restrictive that the situations they are being fulfilled at are sufficiently rare in real life. But then, they are permanently being searched for, especially in technologies, due to efficacy of their predictions, hence looking quite widespread. The oversimplified CP statement has its consequence in the fact that for infinitely rich Nature, it is always easy to give an answer to it too, so that for all possible hardly restricted constructions of the Method, Nature will surely find applications. If the Method predicts, e.g., a particle corresponding to its scheme, it will necessarily be found in experiments, otherwise upon perceiving the particle solely in terms of the Method, we would never notice it, i.e. extract from the world as a whole. Somewhat loosely, it can be said that using the rules of the Method we "create" this particle, as well as being constructed according to the same rules by which a TV-set does exist in the world. In the development of the Method within the framework of a mental scheme, we will frequently illustrate introduced constructions by means of familiar examples. It is necessary, however, to follow the internal logic of the constructions.

Since various trajectories with their mutual relations like intersections are present in CP, we are in need of a structure for their common description. Such structure suited exclusively for CP solving we will call the *contact space* (analogous to the space-time of the canonical version). Its points are contacts as occurring in the intersections of various trajectories. The condition of universality, that is, of the possibility to formulate any CP within this structure, defines the requirements for the *geometry* of the contact space.

Trajectories of their own are already provided with their internal geometry. According to their definition as the sets of states, they are the segments of the number axis, i.e. of a simple arc. Considering their intermediate states as points, and the arcs as the trajectories of moving bodies, we should consider solely situations when the contact of two bodies A and B, with the CP being stated, takes place only if their trajectories intersect. In the same terms, we can introduce contacts of these bodies with some other bodies. In particular, the latter might be useful if they comprise a prepared-in-advance auxiliary set of *measuring* bodies specially introduced in order to predict the contact in CP. Thus, in each particular CP there might appear many (sometimes infinitely many) trajectories of bodies with or without mutual contacts.

Since the contact space as a structure has been introduced solely to solve (a combination of) CP's, the bodies that intersect the trajectory of A (for instance, at its state 1 in its own order) are considered having the contact with A at this point, i.e. the combinations of only such trajectories are being accepted in the scheme. We want to predict the contact between A and B to be denoted as (A, B), while knowing, at some pieces of their trajectories, their prior contacts with measuring bodies. In other words, we will follow A and B using their contacts with measuring bodies. To this end, we have to be sure first of all that on every piece they are the same A and B. Indeed, what for did we select the motion of a body from its one position to another among more general situations when at one place the body disappears, while in another appears an "exactly identical" one? The answer lies in that we imply a possible influence on the contacts of just this body. One could imagine bodies as marked somehow, e.g., carrying something written on them. This method might be useful sometimes, but a specific feature of the Method is the inclusion also of the impossibility of such a marking, say, if the bodies are small enough. Therefore in the Method, which is actually nothing else than a set of various combinations of CP's, we use in the following just contacts of A and B themselves and with measuring bodies (because there is nothing else in the scheme).

Let the trajectories A and B be such that (A, B) occurs. Somewhere before (A, B) we emit a bundle of measuring bodies from A, so having their common contact with A and among themselves (here and further on up to Ch. 4, it is im-

plied that contacts have no effect on the trajectories of bodies, i.e. on the existence of their other contacts, in our terms — on both A and B, as well as between the measuring bodies).

We try to choose, if possible, such measuring bodies out of the bundle that they further come into contact with B, of course, before (A, B). Each such body has a contact with B at a point, having its own index in the trajectory of B. The construction of the contact space is just an arrangement of relevant kits of measuring bodies.*

Among these, we find the body, the contact of which with B occurs before all others with respect to the order in B. Recall that no separate "clock" is required to reveal this "before": Let a contact put a "mark" on B, now any other contact will occur later if it "sees" B already marked. Not always might this be the case. For instance, how exactly to mark electrons? However, sometimes an indirect marking is still possible using auxiliary bodies. It would be reasonable to regard this first body the fastest, were it not possible that bodies might go over different paths. However, our condition "first in the order of B" means *total extremum*: The limit is being reached upon testing both paths and velocities.

If CP has a solution, hence a possibility to follow B while "sitting" on A, then such bodies must be present in a relevant measurement kit. If we were to claim the possibility of solving any CP in the framework of a single scheme, these peculiar bodies might serve as universal signals (they will be conditionally called *photons*, while their definition doesn't imply just electromagnetic implementation). Photons must exist at every point of both trajectories. Otherwise, the contact (A, B) could not be predicted, since B, say, might happen to be "faster" than bodies of the kit, thus evading the following; this kit would be irrelevant for CP. The Method is impotent in the absence of top speed. If it is not possible to distinguish two identical bodies from one that "instantly" moves from one position to another, then the user cannot control the situation by means of acting on the motion of a particular body. It is then said that this is not physics, meaning that the situation cannot be reasonably simplified in order to employ the Method with the use of the concept of body, as being defined by something allowing for CP statement. So, CP itself chooses situations, in which it is efficacious. However seldom these occur, it is recommended each time to look for the reduction of a problem to CP, because then predictions are very reliable. In practice, photons are not always necessary. For instance, in dealing with slow enough motions it is sometimes possible to use even a usual post as the top signal, using, in principle, the same theoretical scheme. The scheme of mutual contacts of bodies can further be used in a broader context. So, for instance, the steady flux of a river cannot be recognized, since its parts are being replaced by completely identical ones. In order to discern the flux and to measure its speed one has to break its uniformity placing a buoyant body there.

In the basic scheme of Newton with an infinite speed of signals, it was necessary to place clocks, synchronized in advance, in the knots of the space lattice. Just for this reason he had to ponder so laboriously on the nature of time, distinguishing the notion of "mathematical" or genuine time from the not strictly definite time, copying some, mainly astronomical, periodic process.

After this digression, let us return to CP. Let after the contact with B of a photon emitted from A another photon be instantly emitted from B back to A, then again from A to B and so on. It is convenient to say that it is one photon that oscillates between A and B up to (A, B), if this exists (Fig. 1.1). This photon realizes the following of B from A. This following is discrete, and it seems to be more reliable to emit from A more photons one after another, so that the reflected photons provide a more detailed information. However, there is a risk of confusion the returning photons. It is not obligatory that one photon emitted earlier than another will also return earlier: both their paths and velocities might differ, since our definition of the photon as a body that overtakes all others having their common contact with A is local.

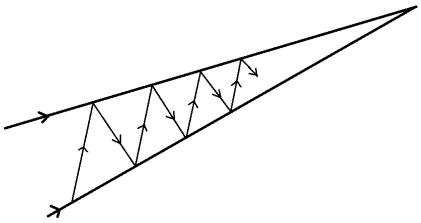

Fig. 1.1: Thin lines are the trajectories of the oscillating photon.

Wherever the counting of the oscillation numbers begins, this number as counted up to contact (A, B) is necessarily infinite. Otherwise, a last oscillation must be there, so that the next occurs after (A, B), in contradiction with the definition of the photon as the top-speed body that overtakes all others, including A and B. Such a sequence of contacts is called a *Zeno sequence* recalling his paradox about *A*chilles and the tortoise.

Let us now reverse the criterion for it to be relevant for CP-solving, considering the occurrence of (A, B) unknown (since we want to predict just this) and counting the photon oscillations. Starting from any point, it would be desirable to conclude that (A, B) will occur, provided the number of oscillations increases infinitely. However, this number will tend to infinity also if the contact does not occur. This will take infinite time, of course, but we don't have any definition of time in terms of contacts. The situation might be cured by means of introducing some multiple contacts. Suppose we

*One might keep an image of them as comprised (though not always) of bodies emitted from each point of the Euclidean space-time with all possible velocity vectors. We use the notions of velocity, acceleration, mass, charge etc. though we still have to define all these solely in terms of existence and order of contacts.

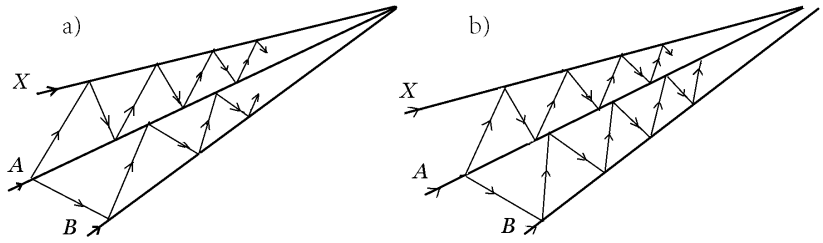

Fig. 1.2: a) (A, B, X) exists. b) (A, B, X) does not exist.

have, besides A and B, yet another body — X, that does have a contact with A, say (later on, we will include such bodies in our measurement kit). Since X is just an auxiliary body, that is, it is not one, the CP is stated about, we may specify its contacts whenever needed. And then we will change the very statement of CP, i.e. we will ask not about a contact "in general" but rather about a triple contact (A, B, X).

Now we are in a position to formulate our CP as follows. Let two photons be emitted from A at once (Fig. 1.2): one — toward B, another — toward X, and we will count the numbers of these photons' contacts only with A.

If (A, B, X) is absent, then the *ratio* of the oscillation numbers between A and B to the oscillation numbers between A and X tends to zero upon approaching the fixed (A, X), and it is this that will be the criterion of the absence of (A, B). If this ratio tends to some non-zero limit, then (A, B) does exist. Since both numbers tend to infinity, this limit depends neither on the point the counting begins from, nor on the reciprocal positions of the contacts with A of the photons reflected from B and X within neighboring oscillations. In the canonical version, this ratio can be expressed with a simple formula via local values of the velocities of A, B, and X at (A, B, X). It is important, however, that the measuring of oscillation numbers is an actual physical procedure in its own right, and it should not be regarded as something tacitly involving the "genuine" concept of velocity as a ratio of centimeters to seconds. We shall see further on that basic procedures of the Method can very naturally be expressed solely via oscillation numbers whenever they are finite and via their ratios whenever infinite.

It is just here — in the necessity of an auxiliary contact of A with a body from a measurement kit — that the concept of time, so far having appeared only in the form of order relations (basic already in the two books of A. A. Robb, written at the dawn of XX century), begins to acquire a particular meaning in measurements. We stress that the definition of a photon as top-speed body-signal implies neither its numerical value, nor even its identity in different points of the contact space, because for each pair of trajectories the photon oscillating between them are to be specified independently of all other trajectories. In this approach a numerical value of the top velocity itself is completely unessential, whereas its changes from point to point makes it possible, as will be explained in Ch. 5, to include, in the general contact scheme, also gravitation with its curved (in terms of the canonical version) trajectories of photons.

Upon corresponding photon oscillations to motions we receive an ideal realization of the Method, viz, "measuring motion with motion" devoid of any intermediary like clocks and/or rulers. By this means, we introduce a particular meaning of the very concept of motion in physics (of course, at the expense of further restriction of the field of experience). It is now not an uncertain "changing in general" but only something expressible in terms of contacts. So, considering motions of macroscopic bodies in an electromagnetic field, we ignore their internal structure, in which similar fields participate as well. But then, the so restricted approach gives us a hope that everything describable in the framework of the Method will sometime find its application in practice.

Ratios of the oscillation numbers in multiple contacts will be one of the main tools in the following. However, it might happen that in the situation in Fig. 1.2 this ratio is zero even in the triple contact due to an "unsuccessful" choice of X as a tangent (in terms of the canonical version) to the trajectory of B in the contact point. We have therefore to complete the above-given definition by an additional requirement to the measurement kit: It must include such X's ("in general position"), that the said ratio becomes non-zero. Moreover, it is possible, with an appropriate choice of X, to obtain non-zero ratios for "different orders of tangency". As will be shown in the next chapter, with an appropriate choice of the own intersection scheme in the measurement kit, it becomes possible to obtain the needed tangents in a regular way rather than just trying out various bodies from the kit.

Two arbitrary chosen trajectories might intersect many times, even infinitely many. In particular, they might be tangent at a point or even to have a common interval. The prediction of a contact using oscillation numbers counting on the trajectories implies these to contain some intervals (each one according to its own ordering indices) before the expected contact that are free of other contacts. Exactly in these very intervals the measuring photon oscillations occur. Were there so "densely" positioned contacts, the oscillations counting would begin before some (A, B), that is before (A, B, X), thus erroneously showing the absence of the expected contact.

The next task is the formulation of the properties of the measurement kit that are relevant to CP solutions, with re-

spect to its own mutual contacts. The finiteness of the top speed implies that not every pair of contacts might belong to a single trajectory. In order for CP to possess a solution, the trajectory of any single body must involve "sufficiently long" intervals around the possible contact, containing points reachable by photons emitted from other bodies in the problem (Fig. 1.3). Otherwise, some bodies would be "invisible" to others, and hence CP could not be stated.

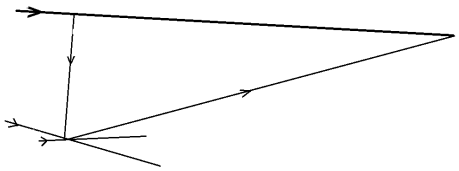

Fig. 1.3: The bodies in CP must "see" one another.

The knowledge of the full trajectory seems to provide the required prediction of the state in question to be reached. However, this trajectory, if being known up to the end state, leaves no room for an action to influence the occurrence of this state. It is then desirable to select in the set of all possible trajectories a subset of such that can be completely specified with only some of their states. Evidently, such a trajectory cannot be uniquely determined by just one of its states, since then the scheme should merely be reformulated to mark this state as the initial, so again representing a trivial no-action situation. The next possibility includes two states. If this solution is unique, i.e. all the infinity of its states can be determined in the scheme with some of its two states, and no other trajectory can include these three states (initial, final and auxiliary third) together, then the third could be chosen arbitrarily on the trajectory. Indeed, suppose that, starting from the initial state and following the trajectory up to a specified third state, we might — in this particular problem — to connect this state to the final along a set of states not belonging to the same trajectory, then this trajectory would not be unique, and the solution becomes ambiguous, bounding the user to choose among various solutions. Since this third state is sufficient to enable some choice for action, any fourth state would be superfluous. The final state being given in advance, we thus look for a broadest class of trajectories ending at this state, each one being specified with any pair of its other states. A whole possible scheme is anticipated to be defined in terms of these particular trajectories.*

The relationship of the bodies in the measurement kit defines the "geometry" in the contact space. Let us start with the simplest structure — the topology. We will define the *neighborhood* of a point of this space as a set of contacts such that any trajectory outside this set ending at this point necessarily has contacts with some other points in the neighborhood. Moreover, we require that the set of points that is common with the points in this neighborhood in each such trajectory includes some open (i.e. without its end points) interval according to its own order. Thus, nearness springs up in the neighborhood as induced by the arrangement of all possible trajectories tending to this point from outside of its neighborhoods (Fig. 1.4).

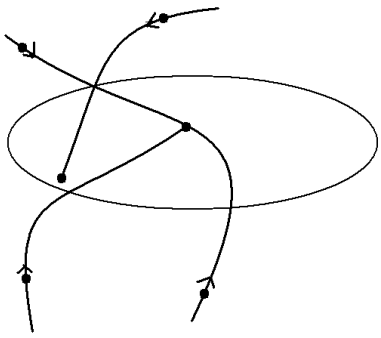

Fig. 1.4: Definition of neighborhood by means of trajectories (the boundary of the neighborhood is shown with a thin line).

This definition is in agreement with the intuitive notion of places close to the given as those not to be missed upon nearing this place from far away. The importance of this notion for practice is in this, that in order to predict the final contact it is not always necessary to know a trajectory. Sometimes, it is sufficient to know only the tendency to near the state. In the trajectory itself closeness is naturally defined by its own order as arising in the primary CP statement. Though it by no means follows from the definition that any two points of a neighborhood can be connected with a trajectory, but if, for a point in it, we take only those trajectories that pass this point, then a neighborhood of this point exists there, generated by these trajectories and completely contained in the initial neighborhood. Though a neighborhood of a point in the contact space is not necessarily a neighborhood of any of its points, as is the case, e.g., in the Euclidean space, however, it still contains a neighborhood of this point.

A particular interest for CP present so-called *spacelike hypersurfaces* to be defined as comprising points, any two of which cannot be connected with a trajectory, whereas any other point of any trajectory crossing this hypersurface at one of its points can be connected to some other of its points with a trajectory.† This condition helps to introduce some own nearness in this hypersurface as induced by the trajectories that cross it, while not belonging to it. Indeed, let us take a trajectory crossing the hypersurface at some point and an interval on the trajectory containing this point. We define the

*In the canonical version, these — initial — conditions give rise to particular "principles". Starting with the requirement of the unique trajectory to be obtained in a solution, one could invent a means to specify this trajectory with the extreme value of something like the minimal length in a metric for geodesics or, equivalently, the least action principle in dynamics.

†In this context, the trajectories themselves are also called "timelike" lines; however, we will use only the term "trajectories", thereby accentuating their primary role with respect to the space.

related neighborhood in the hypersurface as all its points that can be connected with trajectories to the points of this interval (Fig. 1.5).

The boundary of this neighborhood, formed by photons, is to be excluded from it, so that the neighborhood will be an open set, each point of which having its own neighborhood completely contained there. The boundary forming a photon set is called a light cone. In contradistinction to usual surfaces in geometry, the specification of a light cone automatically defines also its decomposition into lines — the trajectories of photons, since no other "line" here is the trajectory of a body.

Let us consider so small an interval on a trajectory containing one (then only this) point of a spacelike hypersurface, such that the neighborhood induced by this interval is completely inside this hypersurface. In accord with the order of the trajectory in this interval, there is a sequence of neighborhoods, each one including the next, thus letting us introduce continuous mappings of this interval into the hypersurface using trajectories that actually do pass the points of this interval (Fig. 1.6). Such constructed sets of points in the hypersurface we will call a *path*. Hence, strictly speaking, paths are not trajectories! They are not a subject for operations with photon oscillation numbers. In particular, they are not bound to be simple arcs, and they can have various self-intersections. However, they are lines that are continuous with respect to the structure of neighborhoods on this spacelike hypersurface.

The role of a spacelike hypersurface as an envelope of all possible configurations of paths to be relevant in CP, consists in giving them the freedom to intersect. If the paths intersect, the related bodies might either or not have a contact, but if the paths do not intersect, the contact is impossible. What is then the minimal geometry, still observing the freedom of intersections? The answer is: a three-dimensional topologically Euclidean (i.e. including, for instance, also Riemannian) space. This space allows for various combinations of one-dimensional continua — lines, since it is always possible to round one line by another, while in only two dimensions some restrictions for CP exist not due to the features of acting forces but rather on their own: A line cannot leave the region inside another closed line without intersecting it. On the other hand, four dimensions would be redundant, since for an adequate description of paths with their intersections, its three-dimensional subspace would be sufficient.*

However, this answer implies ready notions like dimension and therefore might become ambiguous in finer problems, still leaving existent effective methods of CP. It might turn out that not all paths are relevant or we will need some complex arrangements of infinite sets of paths. There are, yet, extended bodies to be considered in the Newton's scheme as if "made up" of material points, and this concept involves geometrical ideas *a priori* not to be relevant, e.g., on microscopic levels. We thus need the analysis of the commonly used concepts from the point of view readily accepted in CP.

*"Traffic interchanges" ensure the absence of collisions, while crossings require "traffic lights".

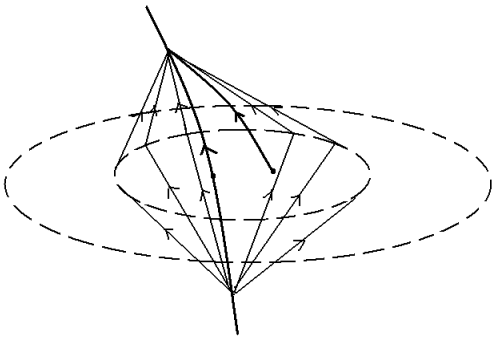

Fig. 1.5: Definition of neighborhoods on a spacelike hypersurface.

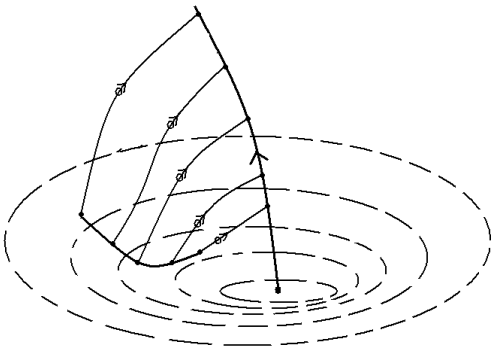

Fig. 1.6: The trajectory A is mapped (projected) into a path using a family of trajectories (thin lines) on the spacelike hypersurface, using a fixed trajectory Y. Paths are not oriented of their own. For this reason, they are shown without arrows.

Time and again, we start with the analysis of the canonical version, in which points of the space are considered as ready and specified with their coordinates. What is the way to measure coordinates? Using a ruler. The ruler is something made up of atoms, it is solid and straight, and measurements with it imply touching, i.e. contact. What is "solid and straight" will be discussed a bit later. Let us first consider the principal design of coordinate frames, i.e. what its essence and importance actually are. Indeed, what is the relevant space and how are numbers coordinated with its points?

In the related scope of mathematics, namely topology, these questions are united under the title "dimension theory". Let us briefly recall some results of this theory as applied to CP. Each point of an n-dimensional Euclidean space is being encoded with n numbers in order to distinguish one point from another, in other words, points and numbers are to be in one-to-one correspondence. However, already in the XIX century, Kantor realized that only one number is sufficient to this end. For visual simplicity, we confine the case to two dimensions ($n = 2$). Let us perform correspondence to the points of a unit square by means of the coordinates

$(0.a_1a_2\ldots, 0.b_1b_2\ldots)$, the point of a unit segment by means of the coordinate $0.a_1b_1a_2b_2\ldots$ This is a one-to-one correspondence, hence, the "quantity" of different points on one side of the square is the same (the same infinity!) as in the whole square. Why then are two coordinate sets needed? The reason is that this one-to-one correspondence is not continuous: It is not necessary that close points of the segment correspond to close points of the square. For example, two points of the segment $0.500\ldots$ and $0.499\ldots$ are infinitely close, while the "image" formed out of them — according to the said rule two points of the square $(0.500\ldots, 0.000\ldots)$ and $(0.499\ldots, 0.999\ldots)$ — will be on the opposite sides of the square. Now, it is impossible to find a correspondence between these that is both one-to-one and continuous in both directions (Brower).

But why do we require continuity? It might then happen that something else would be needed as well. Isn't it enough to find some needed numbers, upon calculating something somewhere, and so to subsequently make the prediction of the contact? And what is the meaning of "close" in CP? The answer is in this that an unlimited increase of the oscillation numbers in a contact implies the continuity (and even some smoothness, see below) of the trajectory. Discontinuity might result in the lost of the following, i.e. identity, and so a possible confusion renders CP meaningless. Frequently given examples make it possible to appreciate the danger.* Let us take a point in the square, its center, say, and define the distance between any two points as the sum of the usual distances for each of them from the center (the so-called uncountable "hedgehog"). If we encode the points of a square, as usual, by couples of coordinates, then points with adjacent coordinates though positioned on different rays aren't said to be "close" in the metrics of the "hedgehog" (Fig. 1.7).† Is this "hedgehog" two-dimensional as well?

In 1912, Poincare suggested an inductive definition of dimension allowing for the specification of a definite integer even to unusual geometrical constructions, while being equal to its dimension for a usual Euclidean space. According to his idea, "...for to partition space, one needs sets called surfaces; for to partition surfaces, one needs sets called lines; for to partition lines, one needs sets called points; we cannot step further, and a point cannot be partitioned..." (Partitioning means that the remainder is disconnected.)‡

Some other definitions of dimension have been suggested later on for various applications, and it has become popular to look for the types of spaces these definitions coincide. So, Lebesgue's notion of dimension relevant to integration problems is based on how many intersections of the topology defining ("open" or "closed") sets are there, which provide a covering of the space in their totality. So, if we cover an open (sides removed) square with small open squares, then some points of the large square fall in three small squares, and Lebesgue's dimension is defined as the integer smaller by 1 than this number, i.e. it equals 2.

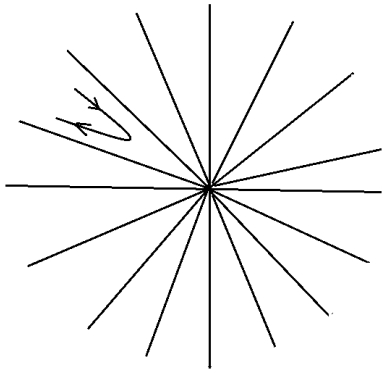

Fig. 1.7: The only way to get from one ray to another in the "hedgehog", is to pass its center.

All these concepts are based on closeness relationships, as being stated in advance in a point set, in order to give it the status of a "topological space". For a definite class of spaces, the Noebeling-Pontryagin theorem states that these might be topologically embedded in the Euclidean space of the dimension $2n + 1$, that is, so that the original closeness relationship will be intact. Intersections of neighborhoods of this Euclidean space with the embedded space form in the latter the same system of neighborhoods that it had of its own, and all neighborhoods of the original n-dimensional space will be so recovered. In particular, for the common representation of all finite and some infinite but not too complexly arranged sets of one-dimensional paths, the appropriate Euclidean space is three-dimensional. On the other hand, the "hedgehog" from Fig. 1.7, while also one-dimensional everywhere but its center, cannot be embedded into an Euclidean space, because its rays are too densely "glued" together. Although its neighborhoods can be obtained as intersections with it of open three-dimensional balls, the "hedgehog" has more neighborhoods. We have herein focused on the "hedgehog" geometry, because similar configurations will be met with in the sequel, and so we have had to outline their affordable limits.

As it turns out, this example demonstrates that topological restrictions are not very heavy for CP. Combinations comprising an infinite set of paths as well as a complexity of paths to be met with there will turn out to be even less sophisticated, so that the three-dimensional Euclidean space will always be sufficient. Most complicated situations are mainly met with in theories of propagation of various fields, in which expansions

*Close phone numbers don't necessary belong to neighbors.

†By analogy, in an environment with mountains and abysses, it might be easier to go around them.

‡This approach has been familiar to people, however, somewhat earlier. "The body, according to Apollodor in *Physics*, is something having three dimensions: length, width, and depth, this body is called volumetric. Surface is a visible limit of a body, it has a length and width, but has no depth. Line is a visible limit of surface, it has no width but only a length. Point is a limit of line, that is, the simplest sign." (Diogenes Laertius, *Lives and Opinions of Eminent Philosophers*.)

in the sets of regular functions are used. However, actual schemes based on photon oscillations require in essence still less restrictions with respect to closeness (adjacency). Various singularities, appearing mostly in the vicinity of contacts, where the numbers of oscillations increase infinitely, are potentially smoothed out automatically in the course of the measurement procedure itself, therefore *softening the needed continuity down to differentiability*. Ultimately, just the existence of a top speed is the cause of the smoothness of the Method's construction.

An important arrangement of trajectories is shown in Fig. 1.8. In the first step, for each trajectory their neighbors are found as those for which the oscillation numbers are the largest, and then the whole distribution of trajectories is so rearranged as to make the smallest of these largest numbers as small as possible. If in addition for each trajectory the ratio of the oscillation numbers between it and a pair of its neighbors equals unity, we receive a uniform distribution of trajectories over a *sphere* of constant absolute value of velocity (in terms of the canonical version). In terms of the same version, the sphere in Fig. 1.8 is pictured in the rest frame of its center. The ratios of oscillation numbers don't depend, of course, on a reference system, but the rest system is symmetric visually, hence allowing the use of a picturesque image, i.e., to introduce a fictitious central body and to count oscillations numbers between it and the bodies belonging to the sphere. Then a sphere might be defined by the unit ratio of the so-counted numbers for all pairs of its bodies. In the sequel we will use this image without additional explanations. In the same notations, it is possible to define a *ball* as a set of spheres with a common center, while having different oscillation numbers between the central body and the bodies belonging to different spheres.

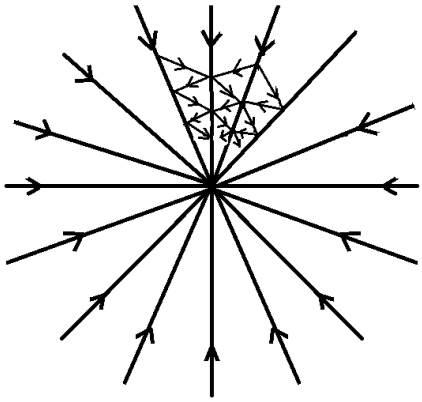

Fig. 1.8: Definition of the sphere via photon oscillations.

In two dimensions, the number of the bodies taking part in the sphere might be any. In three dimensions, however, only five strictly uniform distributions exist, since the third dimension introduces additional interrelations. These five relate the so-called "Platonic solids" or Plato bodies. Namely, four trajectories comprise the tetrahedron; six — the octahedron; eight — the cube; twelve — the icosahedron; twenty- the dodecahedron. These *stars* of trajectories have important applications in the Method to be discussed in connection with the elementary particles theory in Ch. 6. For other numbers of the bodies, a strict uniformity is impossible, but if this number is large, the deviation from uniformity is relatively small, and the distribution tends to uniformity upon increasing the number of the bodies to be accounted for in the limiting construction. In this "hedgehog" type arrangement of trajectories, their number remains only countable, and this is enough for it to be topologically embedded in the three-dimensional Euclidean space. However, in the limit of infinitely increasing the number of trajectories in the sphere, a subtle problem springs up due to the accompanying increase of the oscillation numbers.

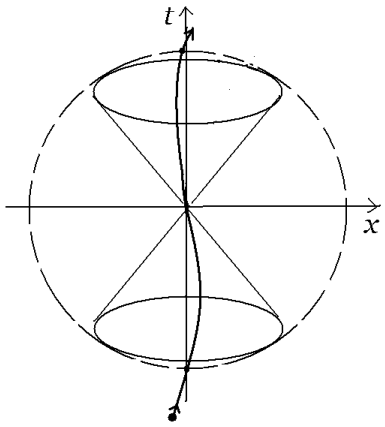

Fig. 1.9: Space-time diagram of the neighborhood; x represents the three space axes.

In terms of the canonical version, the ball consists of concentric spheres with various absolute values of their velocities. The totality of balls at all points of the spacelike hypersurface is the basic concept for further constructions of CP. An induced-by-trajectories neighborhood in the full contact space is topologically equivalent to the *space-time* of the canonical version. It is convenient to imagine this neighborhood with the diagram in Fig. 1.9 as the intersection of the interior of the light cone with Euclidean balls. The vertices of the cones fill the three-dimensional Euclidean space. This topology is uniform over the space and it is natural in CP, while it doesn't look like the familiar Euclidean topology. It is impossible to define the usual metric as a distance between any two points, the closeness of which would be specified by the smallness of its value. In particular, both mentioned definitions of dimension are equal to unity, and the spacelike hypersurfaces are zero-dimensional (discrete) for any number of coordinates, if we regard them as subspaces of the full contact space in this usual meaning that the neighborhoods on a

subspace are the intersections with it of neighborhoods of the encompassing space (Fig. 1.10). It is worthwhile to mention in this context that the notion of closeness on a spacelike hypersurface as defined above by means of trajectories is needed solely for the adequate representation of the paths, while this definition does not turn by itself the very hypersurface in the subspace of the full contact space. We recall that the only motive to introduce closeness, i.e. a topology in the hypersurface is to remove non-existing contacts as the intersections of trajectories independently of their full combinations. The three-dimensional Euclidean structure is sufficient to ensure this.

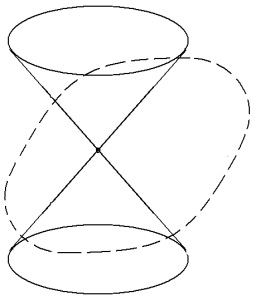

Fig. 1.10: Intersections of the light cones with the spacelike hypersurface induce only a discrete topology on it.

Further constructions of the Method will appear in this book whenever required in relevant applications. Each time it will mean an additional reduction of the scope of situations accessible within the framework of the Method. But then, each new construction provides a new possibility of effective prediction. Every time we will accurately formulate the conditions of the applicability in terms of the relevant contact schemes. The compromise between the meagerness of the initial information and the broad scope (range) of applications shall always be our main concern. We conclude finally that the actual content of the Method is just a "library" of particular cases collected according to a general approach rather than a "theory-of-everything" sometimes being dreamed about. The only requirement to a proposed construction is its realization within a relevant contact scheme. Outside this scope, one has to turn to some other science — not to physics. So certain hard restrictions on the language of the Method allow us to expect that everything compatible with its rules would necessarily be implemented in Nature's infinite self-diversification. One needs only to give a close observation to pick out in the real world, looking with a certain expectation at the surroundings, being armed with the Method in advance. Thus we have found the necessity and sufficiency of the three-dimensional Euclidean space to perform imaging of paths universally. This approach might seem as resulting from the very confined imaging of the World. Indeed, there exist, e.g., extended bodies, as it seems, besides any reference to paths. But the extension of a body reveals itself just as a restriction on possible paths. The three-dimensionality of a building is nothing but an obstacle to pass through its walls: A door is needed. A transparent glass wall for photons will not provide the impression of its extension unless you actually strike it on your very *path*.

Chapter 2. Forces in terms of contacts: prediction of the link

> The weakness of the principle of inertia lies in this, that it involves an argument in a circle: a mass moves without acceleration if it is sufficiently far from other bodies; we know that it is sufficiently far from other bodies only by the fact that it moves without acceleration.
>
> A. Einstein, *The Meaning of Relativity*

Such constructions as described in the previous chapter define only the general framework of CP, specifying the kit of tools that is sufficient for its statement, while being free of unnecessary items. However, as it was said, the prediction of the final contact (A, B) only when the whole trajectory is known makes no sense, since the result would then become known only at the (A, B) occurrence, when nothing could be changed. *Dynamic laws*, letting sometimes find a trajectory upon knowing only some its parts are of particular concern, providing actual predictions. So, even if it is found that in a given situation CP is applicable, an efficient solution to CP requires further restriction of its field.

In the idealized scheme with a material point only the exact intersection of trajectories is implied: It is not important how far they miss if they do. Then topology is enough, because any general scheme of a theory claiming universality cannot use some fixed scale of precision, hence, it is bound to confine to strict limiting sequences. It is implied that in a practical CP such a scale is conditioned by the application itself, and predictions are to be made already according to a (small) part of the trajectory. A possible approach consists in approximation of all the diversity of trajectories by means of some combinations of a special kit of *standard* trajectories, the mutual contacts in which are specified in advance, already before dealing with a particular CP, just as it is convenient to build a house using bricks or various functions using sinusoids.

In the canonical version, the role of standard trajectories is entrusted to those free of external influences. In flat spacetime free trajectories are considered uniform and rectilinear. We shall frequently refer to this image for its visual familiarity, still keeping in mind that the only issue is the scheme of the mutual contacts of bodies. The kit of standard trajectories must be capable to represent any trajectory belonging to a class of interest as a sequence of standard ones, so that in the relevant construction of the limit these sequences tend to the contact of interest in CP, if actually existing. In the canonical version this sequence is a broken line, i.e. a chain of straight

segments tangent to the given trajectory (Fig. 2.1). We anticipate to use the bodies from the measurement kit as the standard tangent bodies, and therefore we will use for these the same letters $X, Y \ldots$. Actually, of course, the problem of chains construction just shifts the prediction of the final contact to no less difficult problem of finding appropriate tangents and joining them to obtain the whole chain.

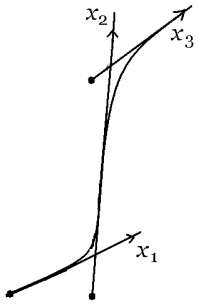

Fig. 2.1: Approximation of the trajectory with a chain of its tangents.

If a contact (A, B, X) exists, then the tangents Y to A and Z to B represent the last (inertial) links of a chain. This means that now it is (X, Y, Z) that is in question in CP, while (X, Y) is specified. The solution exists under the condition that in some neighborhood of (X, Y, Z) only this contact exists, otherwise some additional information would be needed to distinguish them. Since this information must be presented in terms of contacts as well, we would merely return to initial problem. Therefore even in general (within any particular CP) the first (1) requirement to the choice of measuring trajectories, which is at our disposal, consists in that any two of them either do not intersect or have a single contact. It follows that if any two points can be connected with a measuring trajectory, then only this one can be in the kit, since if one more existed, these two would have two mutual contacts. Besides this requirement, also the evident (2) requirement of completeness must be hold, i.e. the existence of contact at least with one measurement trajectory at arbitrary point of any trajectory. And the last (3) requirement to the measurement kit consists in this that any two contacts, which can belong to a trajectory were also belong to a measuring one (and then to only one, in view of (1)). The requirement (3) expresses the absence of a universal scale of distance a priory, in other words, a principal possibility for any interval to be the final in some PC.

The totality of uniform and rectilinear trajectories of the canonical version fulfills these requirements, of course. However, an idea might spring up as to the existence of some other properties of these in comparison to arbitrary trajectories, which idea is just one that Einstein discussed as cited in the epigraph to this chapter. In terms of contacts only these requirements are important, and they arise from the very statement of CP, unrelated to being "sufficiently far from other bodies". Simply, the solutions to CP provide predictions given initial conditions, including not only direct sources of influence on the contacts in question, but also a possible "surrounding". If this surrounding is such that it is impossible to find measuring trajectories with the required properties, CP cannot be solved. For example, one could state CP for bodies moving in an electric field in the presence of a gravitational field. Trajectories of measuring bodies as well as of photons are then no longer uniform and rectilinear, though CP still can be solvable unchanged by means of counting the oscillations numbers, provided the above requirements are still satisfied. However, outside the range of the particular CP the measuring trajectories are free to intersect many times, of course.

Only the trajectories belonging to the measurement kit are bound to intersect no more than once. Their intersections with other trajectories might be multiple, and just these are the base for dynamics. These contacts cannot be too dense as yet in order to leave the opportunity for (also infinite) photon oscillations. Consequently, the ratio of the number of these contacts to the number of photon oscillations must tend to zero in the converging sequence of the approximating chains. This corresponds to the concept of differentiability in analytical geometry.

Let us return to the prediction of (X, Y, Z). Of course, this contact might be predicted upon observing, as before, the tending to infinity of the number of oscillations with their finite ratios. However, now it is the contact of bodies from the measurement kit that is in question. Isn't possible basing on the particular properties of this kit to receive the prediction earlier? We describe first a possible procedure in canonical terms, that is, regarding the trajectories to be uniform and rectilinear and, moreover, provided with some definition of parallelism (to be discussed shortly), defining parallelism as the identity of velocity vectors rather than only of directions.

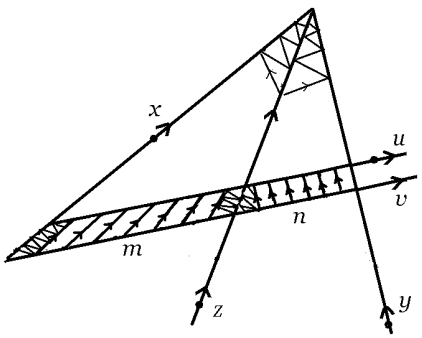

Fig. 2.2: Effective prediction of a contact.

Fixing (Y, Z), we look for a construction to predict (X, Y, Z) at a finite range, according to the order on any of the trajectories, to obtain a criterion for (X, Y, Z) to exist (Fig. 2.2). To this end, we draw an auxiliary trajectory U between X and Y and a parallel to this U', having also a contact with Y. The triangle of U, X and Y defines a plane, containing all the trajectories to be considered (If U' doesn't intersect X, the CP immediately is being solved in the negative, because

then U', X and Y don't lie in the same plane). Next, we take a point on U and draw between X, Y and U' two sets of trajectories that are parallel: one set to X, another — to Y, so that in both sets (each one starting at its side from the mentioned point) the numbers of oscillations between neighboring trajectories (the "elements") are everywhere equal. Let it be n elements between (U, Y) and (U, Z) and m elements between (U, Y) and (U, X). Upon varying m given n, we are able to achieve (X, Y, Z), letting m and n tend to infinity while keeping given m/n. So, a relevant definition of parallelism would be enough to solve CP in this case. We propose the following scheme.

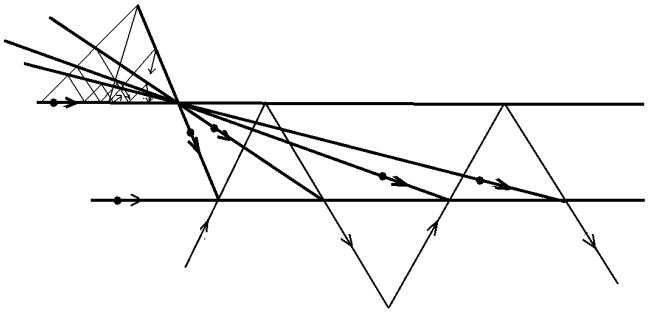

Fig. 2.3: Construction in a plane of a trajectory parallel to given, using solely the ratios of the (infinite) oscillation numbers.

In Fig. 2.3, a photon is received from infinity and emitted at a point of a first straight trajectory a parallel trajectory to be drawn to. The emitted photon is reflected somewhere to come back and to be emitted again toward infinity. In the plane, crossing all the arising light cones, a second straight trajectory is drawn, crossing all these four basic photon trajectories. We have to find the condition for the first and second trajectory to be parallel. To this end, at a point positioned before all these contacts we draw four straight auxiliary trajectories to connect this point with the contacts of the second trajectory with the basic photons. At the same point we specify the ratios of any three (infinite) oscillation numbers between the first trajectory and auxiliaries to that with the fourth auxiliary. Using these ratios, we determine from the system of three linear equations the moments of emission and returning of the two middle basic photons and also the points of contacts of the second trajectory with all four basic photons. If the second trajectory is parallel to the first, i.e. all four distances between them are the same, then these equations are homogeneous, and the equality of the determinant of the system to zero yields the dependence of its coefficients uniquely defined with the oscillation numbers ratios. Importantly, these ratios provide *the construction as a whole* rather than just to provide a condition of the trajectories to be parallel; otherwise it would be necessary to specify in advance also the contacts of emitting and receiving the photons on the first trajectory.

These schemes, solely in terms of contacts, unite for a plane space-time the concept of parallelism with the uniformity and rectilinearity of trajectories in one condition. Essentially, there is no separate definition for each of these properties. Any contact scheme regards their complex as an indivisible whole. Upon being included in the scheme of Fig. 2.2, the scheme in Fig. 2.3 directly specifies as the trajectories U and U', so also a pair of sets required for the prediction of the final contact.

Various ratios of the oscillations numbers define the set of trajectories parallel to the given, provided their combination satisfies equality to zero of the parallelism defining determinant. A simple way to select a particular trajectory out of this set is shown in Fig. 2.4. This can be done upon counting oscillation numbers between the parallel trajectories over the interval limited by the middle basic photons in Fig. 2.3, and it might be useful in constructing sequences of mutually parallel trajectories.

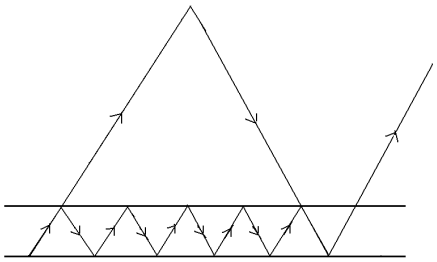

Fig. 2.4: A way to select a particular trajectory in the set of all parallel to the given.

Let us proceed in the approximation of a general trajectory by means of a chain comprised of the standard trajectories. The same measurement kit, as specified by its particular intersection scheme, might be used instead of standard trajectories in the approximations of general trajectories. According to the formulated above condition of CP applicability, for each point of a trajectory A there is a point on this trajectory, prior in the own order in A, to the first point and such that these two points are the intersections of A with one (then, only one) measuring trajectory (X_1 in Fig. 2.5), and there are no other their intersections in between. If it coincides with a measuring trajectory in this interval on A, then approximation is trivial. Otherwise, a point must be in this interval such that the trajectory between it and the final differs from a measuring one. Let us connect this point with a measuring trajectory to the final and so on. As pointed out in Ch. 1, the ratio of oscillation numbers between X_i and a measuring trajectory in general position Y to the similar ratio between X_i and A tends to zero upon nearing the final point. This will be the definition of the measuring trajectory tangent to A in this point in terms of contacts.

However, tangents in different points of a trajectory are not bound in general to have mutual contacts. So, their se-

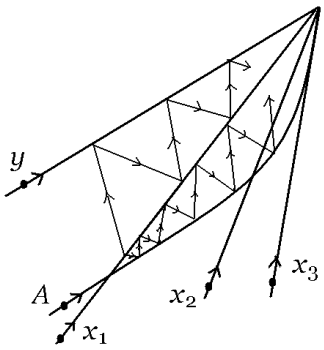

Fig. 2.5: The tangent to A is the standard trajectory that is the limit (if exists) for the sequence $X_1, X_2 \ldots$.

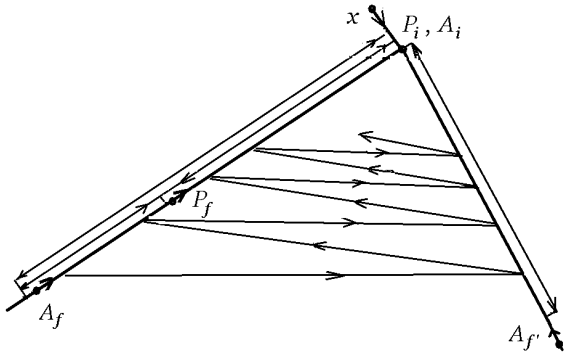

Fig. 2.6: The oscillation numbers are being counted between: 1) P_i and P_f; 2) P_i and $A_{f'}$; 3) P_f and $A_{f'}$.

quence fails to be a required chain. However, the opposite operation — the construction of a trajectory as the limit of approximating chains is possible, since in the relevant arrangement each link of each approximating chain defines the approximating tangent on its own. In other words, the sequence of approximating chains forms the sequence of the approximating tangents.

Having in hand the measurement kit, everything still needed to construct a chain is a relevant rule to define transitions between links at their contacts. The related transition is to be determined by some external influence on the motion of a body, considering the measuring bodies as not experiencing this influence. Otherwise, we cannot be sure that our conditions for the kit still holds (see, however, Ch. 5). The possibility of this separation is the next restriction on the applicability of the Method. In the canonical version, a force causes acceleration of the body, inversely proportional to its mass. But what is the way to measure force itself? As it was explained above, force is being measured according to acceleration of the bodies from a special test kit as specially constructed for this particular force. These are different from bodies in CP as stated above. The trajectories of this kit don't require a special means for encoding, because their acceleration might be measured with the same measurement kit. Indeed, we needn't construct chains for them, their use being only local to determine link transition at points of trajectories in CP, which are being determined link by link by these trajectories themselves.

A trivial solution would prescribe at each break of the broken line-chain to have a test body identical to that in CP. It is clear, however, that no prediction would then be possible, and one is left with pure observation of the motion of the body in question. Therefore, we let test bodies differ from this body, however, inasmuch as still to be expressible with a contact scheme. We denote a test trajectory as P. Then A_i and P_i will relate to the initial link and A_f and P_f to the final (Fig. 2.6).

We choose P_i to coincide with A_i to be detected by the equal zero ratio of the oscillations numbers between any of them and a trajectory in general position X to the oscillation number between them. In the common rest reference system of A_i and P_i the trajectories A_f and P_f are, generally speaking, diverging, still remaining collinear (to the first order). To determine A_f, known P_f, we should specify the ratio r oscillation numbers between P_f and A_f. However, the value of r alone doesn't determine A_f, since any A_f' belonging to the sphere with its center at the common contact $(A_i, P_i, P_f, A_{f'})$ has the same value of r. In order to find A_f, we have then to measure also the ratio of oscillations numbers between P_i and P_f to that between P_f and $A_{f'}$. The lowest value of this ratio, fixed r, specifies A_f collinear to P_f. The number r, expressing the difference between A and P, depends as on the external force, so also on the properties of A itself.*

Not sacrificing generality, it is now possible to reduce the full test kit, with any values of r whatever, to a kit, which in the canonical version corresponds to the kit comprising trajectories with various velocity vectors, because its intersection scheme is the same as that of the measurement kit. For this to be possible, we have to give to r in this kit the status of the universal standard for all CP's. The definition of this standard in terms of oscillation numbers ratios will be our main concern in Ch. 6. Moreover, practical applications of the scheme in Fig. 2.6 imply some restrictions on the smoothness of the distribution of a force in the contact space. If the force includes discontinuities on the scale of link, one should use here smaller links. In singular points, for instance, on the obstacles for the motion, CP cannot be used on its own, and then it will be a motion with ties, and these might even reduce the dimension of the motion area. A correct limiting process requires a coordination of the involved procedures solely in terms of contacts, and the necessary range of links with respect to a force should be estimated in accord with the deviation from zero of the ratio of oscillation numbers between P_i and a trajectory in general position to that between P_i and P_f. For the integration over the chain, the largest of these ratios should tend to zero. The obligatory condition to have free

*In terms of the canonical version, it would be its charge to mass ratio.

segments for the photon oscillations makes it possible to find the correct coordination of limiting sequences on many occasions, while in the canonical version this needs additional, often artificial, hypotheses.

Is it possible to set the test kit in an ordered arrangement, e.g., to provide it with coordinates? In Ch. 1 we defined a sphere comprised of trajectories with the common contact and all possible directions, though having the same absolute values of velocities. Alternatively, spheres can be defined with contact schemes (Fig. 1.8). This definition can be used to construct the whole test kit based on a finite (desirably small) number of trajectories, all others being defined using their oscillations numbers with the basics. It will then be possible to define the transitions between links requiring (again at the expense of further restriction on the class of permitted forces) these transitions to be specified only for the basis, while all other transitions are to be defined by means of counting the oscillation numbers between arbitrary trajectory and the basic trajectories.

Let us choose a non-degraded three of trajectories in the sphere, viz, such that the two ratios of the oscillations numbers between these bodies to that between them and the center don't define the third. Any such three with the common contact define a sphere, and there is the single reference system, in which they can be visualized as in Fig. 1.8. Then any other trajectory of the sphere possesses some definite ratios of the oscillation numbers between this body and each one of the basis to that between any of them and the basis center. These ratios will be the coordinates of the given trajectory. As it must be on a two-dimensional sphere, suffices it to fix just two coordinates. This definition can be extended on the whole ball, provided the basic sphere is specified. However, for each trajectory there exists its twin with the same ratios. It is easy to see this in the rest reference system of any of the basic bodies (Fig. 2.7).

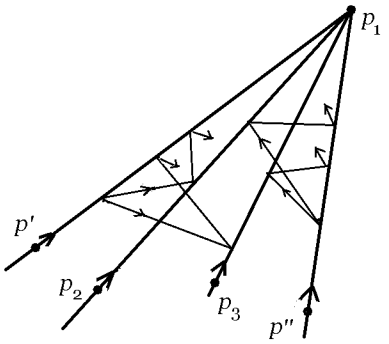

Fig. 2.7: In the reference frame, in which P_1 rests, the mirror-like positions of P' and P'' respective to the plane formed by P_2, P_3 are obvious.

For arbitrary trajectory, its "mirror" trajectory respective the plane (in general, surface) formed by two other basic trajectories will have the same ratios. Since these ratios, as any contact scheme, don't depend on a reference system, this representation is double-degenerated. The same degeneration exists, of course, in the measurement kit too.*

Let us now define the class of external forces with a naturally arising in CP uniform "conservation law" as the conservation of the oscillation numbers ratios under transition in initial-to-final links for arbitrary spheres in the test kit (Fig. 2.8). In view of the mentioned degeneration, this law should be completed with an auxiliary contact scheme to forbid spontaneous leaps to the mirror trajectory in the transitions, if it is important in a particular CP.

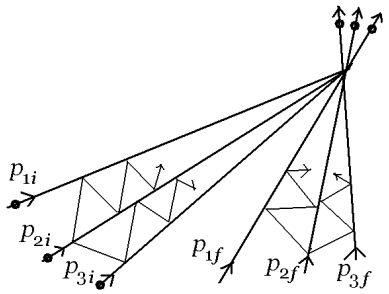

Fig. 2.8: As in Fig. 2.6, the final link is the continuation of the shown trajectories beyond their common contact.

In particular, it follows from this law that the oscillations numbers ratios between all pairs of initial and final links are equal 1. Since any trajectory can be represented via this basis, the transition for any test trajectory can be determined knowing only one of them. If the test bodies could be made identical, it remains to specify their charge and mass by some standard values (see below in Ch. 6).

Endeavoring to express via contact schemes everything in sight, we have introduced a condition on a possible force. It is interesting to look at what are the forces in the canonical version that satisfy this condition. It turns out that, for example, the Lorentz force does. In the canonical version, the three-dimensional forces are represented with the two real-valued three-component algebraic objects — the electric and magnetic field vectors. Upon considering fundamental issues, it is convenient to combine them in a complex-valued 16-component object (tensor). Algebraically, this object is represented with a four-to-four matrix, which accordingly must have only six independent components: three for electric and three for magnetic fields. This is reached with the requirement for it to be antisymmetrical: the four diagonal components equal zero, while off-diagonal components are the components of the field, and each one enters twice — with opposite signs. So happens, it is just antisymmetry that causes the conservation of the ratios of oscillation numbers under electromagnetic field influence according to the canonical version. In CP this argument should be reversed: just the only possible uniform over the whole contact space condition of

*This degeneration will further be important in the context of spin.

the oscillations numbers ratios conservation restricts the relevant forces by the requirement of their antisymmetry. As will be found in Ch. 6, antisymmetry is characteristic not only for electromagnetic, but also for the carriers of the weak and strong interactions (for bosons — integer spin particles) — gluons and heavy intermediate vector bosons, since they all are naturally being expressed with their contact schemes.

Chapter 3. Fields and their propagation: prediction of the chain

> But the properties of bodies are capable of quantitative measurement. We therefore obtain the numerical value of some property of the medium, such as the velocity with which a disturbance is propagated through it...
>
> J. C. Maxwell, *A Treatise on Electricity and Magnetism*

In order to restore the trajectory, according to only local data, suffices it to specify the force solely at the points of the trajectory itself, namely, at the next points in its progression. An alternative approach consists in the extension of the test kit, carrying information about forces, from the outset, so as to measure a force not only at the points of the trajectory but everywhere it might go. In so doing, we need not care already on the first step whatever the trajectory actually is. Being known from independent measurements also at all possible points, it would be known for the trajectory wherever it goes. If the force of its own doesn't depend on the trajectory (Such dependence might exist, e.g., due to the influence of the body in CP on the source of the force), then CP will naturally be divided in two independent steps: the determination of the distribution of the force and the construction of the trajectory under this force. Actually, no extension of the test kit is needed, provided an algorithm to determine the force is known in advance. The determination of the distribution of a "pre-force", i.e. the *field* is to be carried out by means of a universal rule independently of the charge of the body in CP itself. In this context, charge is only a factor to determine the effect of the field on the trajectory, and it does not define as yet the back influence of the body on the field, being its source. The very possibility to represent a force as a product of field and charge is being achieved in CP at expense of the introduction of the test kit in addition to the measurement kit.

Our task in this chapter will be to look for situations allowing for the prediction of field distributions in a region of the contact space according to its distribution somewhere else. In accord with the general approach, any method to make prediction might be regarded relevant as soon as it yields an unambiguous result basing solely on an initially specified distribution of contacts. Time and again: Nature scarcely refuses to answer so primitive questions.

Since everything is considered to be encoded with trajectories from the test kit as specified in some regions of the contact space, we shall have to deal with infinite sets of trajectories, and their compatibility with the constructed space geometry must be examined. The role of photons is particularly important in this respect, since the constructions they take a part in might be uniquely defined. In order to make clear the constructions themselves, we shall consider only one-component, that is, scalar fields. Later on, this variable should be defined with a contact scheme as well. However in preliminary geometrical constructions aiming to obtain the values of a field at a place via its values somewhere else, only some algebraic operations are needed, such as addition of the partial field values times real numbers. Later on these operations will be defined as contact schemes, but by now let us accept that they do exist.

First of all, we have to find the regions in the contact space, the values of the field at which uniquely determine its value at a given observation point. These regions can comprise only points that might be connected to the observation point with any trajectory whatever, in particular, with photons (Fig. 3.1). Photons form the boundary of the zone of influence for the given point, its "light cone". As mentioned, a light cone is not a usual surface but one that is defined along with its decomposition into lines — photon trajectories, containing no other trajectories.

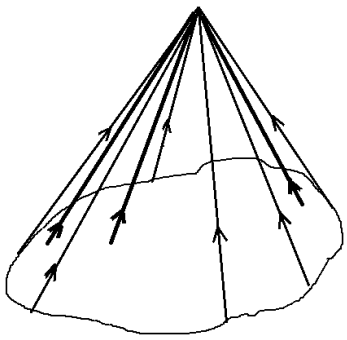

Fig. 3.1: Boundary photon trajectories form a light cone.

We begin with a partial problem, in which the value of the field at the observation point is being determined only with its arbitrary values as specified at some part of the zone of influence rather than with their differences caused by its deformations. The observation point itself cannot belong to this zone, of course, otherwise the field would be specified at this point, and nothing would be to look for there. Among all possible trajectories, coming to the observation point from its zone of influence, consider first their limiting subset, i.e. photons. Their contributions to the field value are independent of each other, since there are no trajectories that cross two photon trajectories having a contact while not joining this contact (Fig. 3.2).*

*Recall that all standard (i.e. measuring) trajectories, including photons, may have at most one contact within the kit.

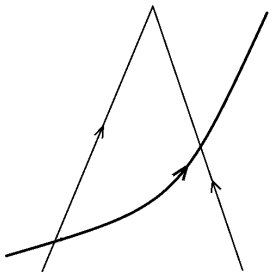

Fig. 3.2: Such trajectories never exist, otherwise the photons wouldn't be the fastest bodies, contrary to their definition.

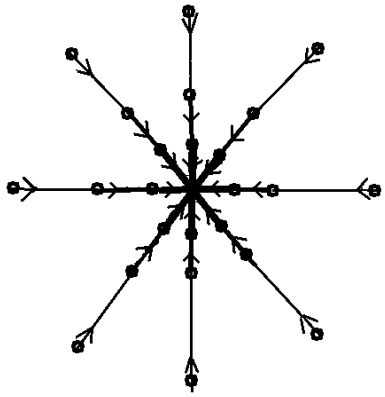

Fig. 3.3: Thicker lines show slower bodies; thin dashed lines show top-speed ones — photons.

Further, we can specify a field value only at a single point on each photon trajectory. Were there two such points, then the value at the point, which is closer over the photon ray to the observation point, would depend on the value at the second point. Indeed, if it were possible to change the field value at the observation point changing arbitrarily the value at the farther point and leaving all the others unchanged, then the field in the closer point would alter too, since taking it as an observation point of its own, while changing nothing at other points, we would get a different value at this point, although according to our condition this value is to be specified arbitrarily. For the same reason, it is forbidden to arbitrarily specify the field values on non-photon trajectories, since these have points that might be connected to other trajectories, hence again one value will depend on others. So, initial values could be specified independently of each other only on the past light cone and necessarily on every its trajectory (i.e. ray), for not to leave uncertainties, otherwise coming as the contribution from any ray that was not accounted for initially. But then, the required law of the field propagation must yield its value at the observation point given the whole (uncountable) set of independent of each other values at every ray on the light cone of the past.

An important particular case is a photon sphere formed as the limit of a sequence of massive spheres of a ball. Using the above mentioned artificial device to count oscillations with respect to the body at the ball center, we can visualize this limit as resulting from the tending to zero the ratio of oscillation numbers as counted between the center and a sphere of this ball to that for an arbitrarily chosen sphere from the ball (Fig. 3.3).

All the specified values on the rays contribute to the result with equal weights. Were the set of the rays finite, the natural solution would be to define the field value at the observation point (the center) as the mean arithmetical of its values over the limiting photon sphere. The extension of this definition on the infinite set of values implies a limiting process upon unlimited increasing the quantity of rays. In so doing, some universal measure is needed on the photon sphere. It must introduce some kind of uniformity ("democracy") in the distribution of the density of rays over the photon sphere, otherwise contributions to the value at the center would depend not only on the specified values over the cone but also on the number of rays contributing a particular value.

If it were not a photon sphere, it would be possible to introduce a uniform distribution of the initially specified values in terms of contacts like in Fig.1.8. For the photon sphere this definition cannot be applied directly. A complex limit must then be in order, including the simultaneous tending to infinity the quantity of rays, keeping their symmetry on each step, and tending the sequence of the spheres to the photon sphere (Fig. 3.4).

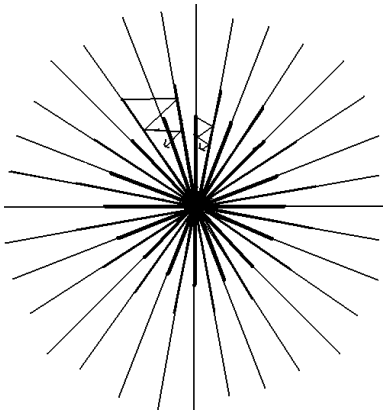

Fig. 3.4: Quantity of bodies in the spheres increases together with their symmetrization and the increase of their velocities.

An image of the construction is shown in Fig. 3.5. For the transition to the limit the moment of the contact of the photons is fictively shown as being before that of massive bodies. In the construction of the limit this outstripping, tending to zero in the limit, makes it possible to induce the uniform distribution of photons over their sphere with their imaginary contacts with the massive bodies of the ball interior.

The successive increase of the massive bodies in their spheres might be performed in accord with their "angular"

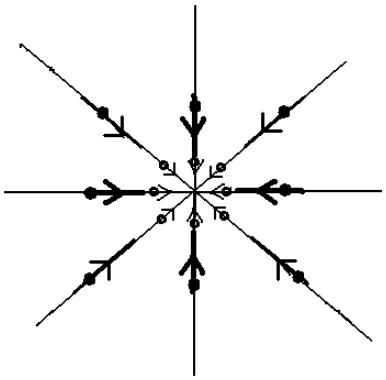

Fig. 3.5: The distribution of photons successively copies distributions on a sequence of the spheres along with the increase in their quantity and velocity.

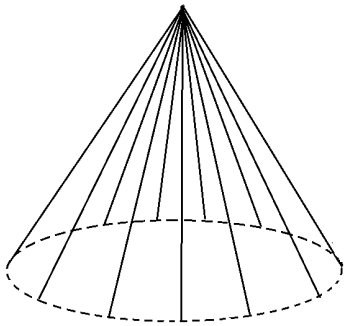

Fig. 3.6: Contributions from differences along a closed contour reciprocally compensate to zero.

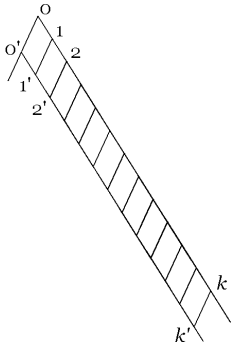

Fig. 3.7: External differences at the points on a ray as taken along the conjugated (opposite) rays.

distribution: once occupied, the angles in the closest to the center spheres are being kept in the farther ones, where the half-angle trajectories are being added in each next sphere. All this is being, of course, controlled by the photon oscillation ratios: they are being preserved in the successive (faster) spheres, while new trajectories enter to fill the angles. Since on each sphere the angles are equal, it is always possible to arrange the construction in such a way, that the oscillations, beginning in a sphere cover those in the slower ones.

This is the final construction, provided initial data are only the values of the field. However, it doesn't exhaust the abilities of contact schemes. It turns out that it is also possible to specify independently some differences of the field values, though the related procedure involves a definite coordination of algebraic operations in close points. It is impossible, as we know, to arbitrarily specify field value differences along a ray, since it would be equivalent to independently specifying these in two its points. There remain two options: either the differences between the values at the according points of the adjacent rays on the same light cone, or outside the cone (if a relevant contact scheme could be found to specify the closeness of points). We'll examine these variants separately.

In order to find the contribution to the value in the observation point from the differences of field values on different rays, we have to average these differences around the cone, i.e., first of all, to add them up. But this sum equals zero, because over rounding the cone we return to the initial point. Indeed, only one point can be taken on a ray, and then all the values will come in the sum in pairs with opposite senses (Fig. 3.6).

We are thus left with the differences between points outside the cone, i.e. the external differences (Fig. 3.7). These are to be averaged over the rays as well, but first we have to find the difference on a single ray to be then averaged. In the limit the difference will become the differential of the field, thus prior to the limit it must be non-zero and finite, otherwise nothing but zero or infinity will be obtained. In contradistinction to values of the field itself, their external differences might be specified in different points of the ray independently of each other. Even at the observation point this difference might be specified a priory, since it is the value of the field but not its differences that is the question in CP. It is therefore possible to define on each ray a depending on the differences finite value by means of adding up the external differences along the ray, so that in the limiting integral sequence with these differences tending to zero the number of the points on the ray, in which these differences are being taken, was increasing in accord.

Consider first the simplest problem, in which the difference is specified only at one point on each ray, namely, where earlier the field value has been specified itself, so it is the difference between this value and its value at the "close" point outside the cone. In order to obtain a finite quantity upon unlimited nearing these points, it is necessary to multiply this difference with a number that tends to infinity in accord with the tending of the difference to zero. On the other hand, this number must tend to zero upon nearing the observation point by the point, in which the value is specified. Indeed, then the field value here must become equal to the specified one upon any construction whatsoever, and no additional contribution able to change this value is tolerable. Therefore, the required number must reflect also the closeness of the points

along the ray and, moreover, it must be defined similarly on all rays to receive a definite result in the course of further averaging.

If we find a united contact scheme to realize also the subdivision of the ray in segments which are identical (in the same sense) as with the segments of the conjugated ray the differences are being taken over, so also with the similar segments on other rays of the cone, then just the number of the subdivision points might be this very number.

Remember in this context, that the light cone itself is nothing more than an auxiliary construction, making sense exclusively for CP solving. Dividing CP in steps, we are at risk to fall into a non-necessary abstraction, so that auxiliary at the outset concepts start "living their own life, pretending to be valuable of their own". Such are, in particular, the concepts of space, time, reference and coordinate systems, various invariance "principles", seeking their substantiation in experiment, the self-evidence of which as a basing language of the Method is declared in the canonical version. In order to be protected against non-necessary abstractions, it is useful to return time and again to the primary concepts. We recall therefore that light cone is nothing else as a tool to find field values; field is a tool to find transitions between links in chains. Hence, the initial link of this transition is always present, however non-explicitly, in all our constructions. It is its turn now to take a highly important part in the general scheme. The dependence of the field determining scheme on a particular choice of a measuring trajectory brings no questions as soon as the algorithm of the solution doesn't depend on this choice, suggesting universally definite operations, though explicitly based on a particular choice of the initial link of a transition.

So, let us take arbitrary measurement trajectory going to the observation point — the vertex of its light cone of the past. Next, take on this trajectory a point before the observation point with its own light cone (Fig. 3.8). Draw a series of trajectories parallel to this one — $00'$ — so that the last goes to the point the field initial value is specified at. Make a subdivision of the ray into k segments under the condition that all (finite) oscillation numbers as counted from one cone to another were equal each other.

One more trajectory $k_e k'$ parallel to kk' goes between past and future light cones. The number of oscillations between $k_e k'$ and kk' is taken the same as for all k trajectories of the subdivision. The position of the initial point k being kept fixed, the number of oscillations depends on both k and the relative shift of the light cones. Tend k to infinity and the relative positions of the light cones so that the number of oscillations between the segments is infinitely increasing. Upon multiplying by k at each step of the limiting process the difference of the field values between the segments, we obtain in the limit the contribution of this ray to the value of the field in the observation point. The full contribution of the external differentials will then be obtained by averaging over the rays, just as it was done before for the specified values of the field itself.

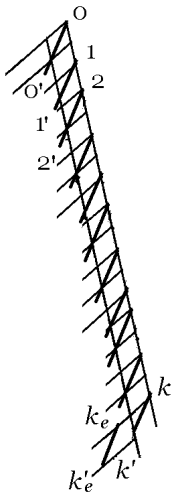

Fig. 3.8: Parallel trajectories for the construction of external differences.

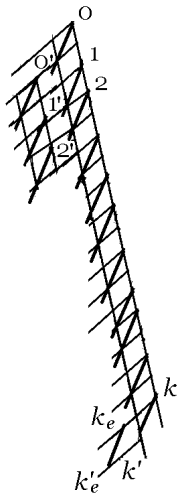

Fig. 3.9: Instead of being multiplied by k, the external differences should be added up from 0 through k.

This particular case is interesting by itself exhausting, as will be elucidated further on, all geometrically permitted contributions for "free" field propagation. However, the developed in this case device of uniform subdivision of rays is also applicable in a broader context, when the external differences are specified not only at one point on each ray, but rather on the whole light cone. In this general situation, it is also possible to get finite quantities to be averaged over the rays, adding up external differences at all the subdivision points of the ray to use its own conjugated cone at each such point (Fig. 3.9).

In the limit, these differences become infinitely small, while their number accordingly increases. The result will be finite, provided the field falls out rapidly enough from the

observation point along the rays (This condition is the next restriction on the fields acceptable in the Method.). The integral sum along a ray might also be presented in a somewhat different and more usable way. Replace the first order external differences with second order differences, that is, with differences of the first order differences taken between the adjacent points on the ray. However, if we directly add these differences up on the ray, we would get zero due to their reciprocal cancelling, as it was for the values at the points taken on different rays. In order to receive a non-zero quantity, let us first multiply the second order difference at each point by the number of the subdivision points up to this point from the observation point, and only then add up the results over the whole ray. The so obtained sum is the same as the direct sum of the first order differences, although it might be useful, e.g., provided the external differentials are being initially specified only in some isolated point on each ray: the differences should simply be set zero at all other points. It is not, however, possible to use the same device for the differences between points on the neighboring rays, since there is no initial point here similar to the observation point for a separate ray.

So exhausted is the variety of initial data as allowed for by the geometry of the contact space on light cones. All others either add up to zero, or are expressed via these. We turn now to the possibility to specify additional data *inside* the light cone. As we saw, it is forbidden to arbitrarily specify field values there, but differences could be specified in the same way as on the cone. However, in flat contact space the sum of these differences as taken over the whole interior of the cone is zero — again due their reciprocal cancellation: Whatever pair of adjacent points the difference is taken at, it is repeated with the opposite sense at another pair (Fig. 3.10).* Only over the boundary, the light cone, there is no compensation from outside (the Stocks' theorem). It seems that this compensation could be cured with the same device as was used on the cone itself, that is, to multiply each difference by the related subdivision points. As it is seen in Fig. 3.10, there is a point to start counting from: the primary light cone along the rays of its conjugated cone. Indeed, the sum will not be zero now, but then, in the limit it becomes infinite. Such a leap from zero to infinity comes from the change of dimension: inside the cone it is more by one than on it, so a finite sum of the differences on the ray is to be multiplied by the number of the subdivision points that is infinite in the limit. This is the reason for the Huygens' principle to be valid in flat (and only in flat) space, allowing to specify initial data only on the light cone.

So, the geometry of the contact space, coming into existence due to the requirement of maximum variety of the allowed trajectories, heavily restricts, in turn, the variety of allowed fields. It turns out that the constructed above solution is nothing else as the solution to the *wave* equation of the canonical version. In the usual form of this version, the initial values of the function and its differentials are specific for the second order wave equation. The construction with only the values of the function specified at one point on each ray is known as the Kirchhoff solution to the Cauchy problem for the homogeneous wave equation describing a free field, whereas the construction with the specified also differentials on each ray corresponds to the solution of the non-homogeneous wave equation, describing the field with specified sources.

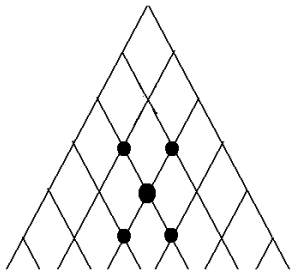

Fig. 3.10: A network inside the cone might be arranged solely with photon trajectories matching an already completed distribution of the cone in Fig. 3.9. However, the differences between the values in the network knots are being reciprocally compensated.

In view of the asymmetry of the order relation on trajectories, contact schemes automatically select only retarded solution to the wave equation, while another one is considered in CP as "parasitic". It is possible, of course, to formally construct this second solution according to the same procedure, reverting order relations on the defining trajectories. However, the very meaning of CP would be ruined, since by the same reason one could change the order for only some trajectories or even on some segments of one of them. Such constructions (the so-called "Feynman paths") are used in the quantum field theory. However, this not CP but rather a probabilistic scheme as built on its base. In this scheme, the particle is present, as it were, at different points at once, and the particles are created and annihilated, although being registered individually by means of non-annihilating (as the experimenter always hopes) classical measuring devices. It is just these that are constructed according to CP schemes.

So, in the own statement of CP half of the solution are fictive. This is not a flaw of CP, as it would be were its predictions ambiguous. Similarly, a circle is described with a square equation, the second solution of which yields a negative radius "circle", and the parasitic solution is being excluded with a separate rule. But then, our construction comprises naturally solutions when insufficient smoothness of the initial field values makes the wave equation non-existing, since no needed derivatives are there. The well known construction of solutions in terms of generalized functions realizes integration using auxiliary sets of sufficiently smooth "basic" functions to imitate differentiation. However, this construction

*What the term "flat" means, see in Ch. 5.

implies the wave equation itself as "fallen from heaven", whereas the above presented schemes immediately arise in the very CP statement.

Remember now that the field value itself must have its own contact scheme. Force is revealed in the breaks of chains (Fig. 2.6), and for the body in CP it comes via fields. Hence, the specified initial data for a field are to be expressed with inter-link transitions, though now applied to the test bodies. In order to find the field value at the observation point, where the force for CP will afterwards be determined, we have to transport all the transitions for test bodies to this point to be averaged there. In so doing, we have to realize everything we regarded insofar as known, namely, addition of the values and their products with real numbers.

We can use for this purpose the scheme of parallel trajectories (Fig. 2.3) to construct an oriented parallelogram using two such pairs (Fig. 3.11). The number of oscillations between the trajectories of one pair counted from one to another trajectory of another pair we compare to the analogous number if we exchange their roles. If, in particular, these pairs are such that the oscillation numbers are equal each other, then we'll refer the diagonals as *sums* of the trajectories.* The pairs are orthogonal, if the ratio of the oscillations numbers between the sum and the trajectories in the points of triple contacts equal 1.

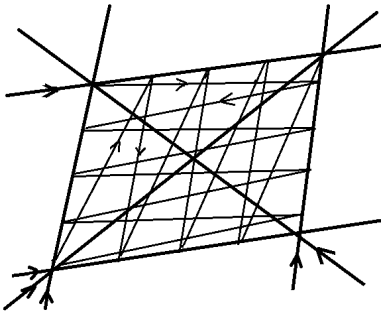

Fig. 3.11: The oriented parallelogram defines the operation of adding trajectories up.

Now we are in a position to define the sum of contributions from different rays in the observation point, so obtaining here the field value acting in CP in the limit of infinitesimal subdivisions. Initial data are no longer abstract quantities, serving earlier as a model to define the very procedure of the solution, but rather objects as defined with a contact scheme, viz., the transition between the links of a test trajectory. In order to find this in the observation point, we have to take the initial data to this point with a parallel transport. Averaging is being performed according to the scheme, once used to determine the transition in the CP trajectory on account of this transition for the test one (Fig. 2.6). In order to find the oscillations numbers defining the averaged trajectory, we have to divide the number of oscillations between the found sum and the initial link of the CP transition trajectory on the number of subdivision points. The transport of both links of the transition along the ray is being performed according to the prescription in Fig. 2.3. The external differentials are defined as the limits of differences after the parallel transport of the test bodies' links along the corresponding cones (Fig. 3.12). As shown in this figure, we have to add the third light cone to the construction in Fig. 3.8 to form the required difference. This difference is to be constructed in four steps. On the first step, the final link of the transition is transported along the conjugated ray. On the second step, the external difference is formed according to the "parallelogram rule". Then this difference is transported to the point the initial difference is to be specified at in two steps — along the conjugated and then along the basic light cones.

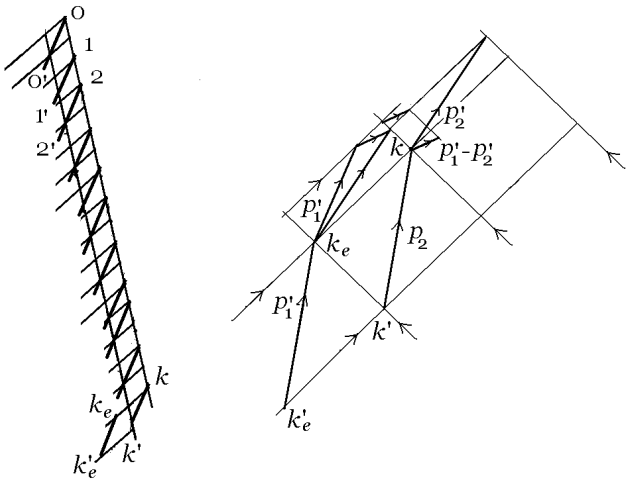

Fig. 3.12: Construction of the external difference on the ray with the parallel transport of the final link is shown in the right diagram. This difference is to be inserted as the initially specified datum in the corresponding place in the left diagram.

With all necessary procedures and quantities at hand, it is possible now to actually construct solutions for a free field. However, in a more general situation of a field with its sources, i.e. when the differentials are specified over the whole ray, the solution remains a phantom, still requiring measurements up to the final contact. A satisfactory solution is possible only if the needed differentials are known in advance, being represented with some separate contact scheme. It might happen, in particular, that the differences defining sources are the trajectories of external for CP bodies. In these cases, a self-consistent CP might include interactions, sometimes retarded, of two or more bodies. The electromagnetic Lorentz force gives the most important example of interaction admitting such problem statement. On the one hand, this force is being expressed with a uniform over the contact space scheme as preserving oscillations numbers ratios

*Otherwise, these will be weighted sums.

to provide its own description in these terms. As pointed out above, this force might therefore be represented as the antisymmetrical combination of the derivatives of potentials. On the other hand, its propagation with top velocity is expressible with the shown above schemes, hence, it is equivalent to the wave equation of the canonic version. In this version just the antisymmetry of the Lorentz force causes its spacetime derivatives (corresponding to the differentials of the field in the contact schemes) to be not arbitrary, but satisfying an additional continuity equation. However, the same equation expresses the condition of the field generating charged bodies to be non-disappearing upon comprising, e.g., some flux. Such fluxes might therefore play the role of sources in the wave equation, since in the scheme with second differences these are presented with antisymmetric combinations of the derivatives of the flow of bodies-sources. In the canonical version, it is just antisymmetry that causes the system of the second order wave equations for the field components to split into the first order system — the Maxwell equations. These are elegantly presented as in the antisymmetric tensor equations, so also by means of the alternating differential forms.

We have now to find a contact scheme suitable to present the differences in Fig. 3.12 in the own terms of the field sources rather than via the test bodies representing known in advance external fields. In this scheme, the source must be given directly with the trajectories of the bodies the source is comprised of. Then in the limiting process upon increasing the number of rays (Fig. 3.4), these trajectories are to contact the rays that are involved on each step. Otherwise, some sources might be lost. Whereas the test bodies we could put by will in the places needed in the solution scheme itself, the sources belong to the CP statement in their own right. Indeed, on each step of the limiting process to find the integration sum, the points, in which the initial data are specified, are being fixed with the very procedures of uniform subdivisions as along the rays, so also between them. These points are distributed discretely. Hence, the trajectories of the bodies-sources that miss the subdivision points will not be accounted for. We thus need a special contact scheme to smoothen the distributions of sources, realizing the idea of an averaged trajectory in the vicinity of a subdivision point. To this end, we have to parallel transport to a subdivision point the trajectories of the "closest" to it trajectories out of the source flux, and we stay in need for a definition of closeness for this case.

In the basic scheme in Fig. 1.2, the related oscillations numbers ratio turns zero if the contact is absent (Fig. 1.2b). Let us add a third body to this scheme (Fig. 3.13) and define the ratio of the oscillations numbers for these two "missing" bodies upon measuring the oscillation numbers — one between A and B, another between A and C — up to the supporting (A, X), which fixes a point of the implied subdivision.

It is now possible to specify the sources unambiguously in the full scheme of integration, parallel transporting the trajectory of the body-source to the closest point of the subdivision. This procedure completes the contact scheme equivalent to the concept of flux density in the canonical version. It isn't necessary, however, to know the trajectories of bodies-sources at full. Frequently, it is only the total flux of these bodies (especially, if these are numerous) that is important, so that, for example, the exchange of identical bodies doesn't alter their total field, even if it doesn't result from their actual motions. In essence, it is only the flux of their active factor, i.e. of their charge that is of importance. This issue is tightly connected with the procedure of gauging the charge, that is, of establishing its universal standard at all points of the contact space in the course of a special contact scheme to be presented in Ch. 6.

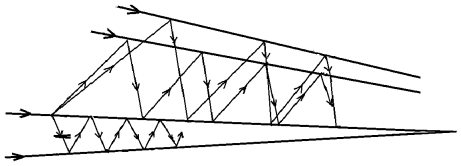

Fig. 3.13: Definition of the related closeness of the trajectories to a given point.

Thus, electrodynamics as well as mechanics might be deduced from a single condition: "Everything moving" must be an implementation of a scheme in CP. What remains doesn't belong to the Method, merely because then we would not distinguish and look for such things as space, bodies, forces, fields etc. Concepts a priori that we use to approach Nature with, must correspond to the purpose of our approach in order that its response will not be regarded as a meaningless "noise". This language is by no means arbitrary, basing instead on the main condition of the universality of predictions, on their unrestricted repeatability.

Chapter 4. Quantum theory: repeatability of the non-repeatable

> ... we have to assume that there is a limit to the fineness of our powers of observation and the smallness of the accompanying disturbance — a limit which is inherent in the nature of things and can never be surpassed by improved technique or increased skill on the part of the observer.
>
> P. A. M. Dirac, *The Principles of Quantum Mechanics*

An important condition of applicability of the above described contact schemes to clearly isolated bodies is the ignoring of the effect of measuring contacts on their trajectories as those in CP or auxiliary. For the development of the Method from scratch, this strategy looks natural, since it merely suggests a reliable way to make predictions in a limited scope of problem considered meaningful for the implied user. In other words, the inventors of the Method begin with the analysis of

practice endeavoring to elucidate what is actually desirable. However already in the course of discussing the basic constructions, the question about the limits of applicability of the Method is important. It is purposeful to analyze these limits from inside the Method, since it appears so successful, that even upon embarking on new problems the user is reluctant to reject it totally. Then the more, if these spring up within the Method itself and look as its natural continuation and refinement.

The schemes for CP solutions, as describing motions under external influence, lead on their own to the question, as to what will happen with their predictions, if the influences become weaker. Indeed, the measuring contact is an interaction as well, and it may happen that it is impossible to weaken it at will, while preserving the reliability of the registration of the very fact of contact. If CP is being stated about very light bodies or about very fine details of motions of even heavy bodies, it might happen that the effect of measuring contacts on the predictions of future contacts is no longer ignorable. But then, measuring contacts are unavoidably present in the very statement of CP, and even the concept of space and its properties were established solely as a tool for CP solutions using these contacts.

It seems that the Method fails as soon as the perturbation of motions by even most delicate measuring contacts becomes comparable with the external influence. It is just the measure of such delicacy that we have to determine, again with the contact schemes, that is, to select the cases, in which the Method could still be applied, perhaps not in its complete form.

Practice provides examples of possible extensions of the framework the Method could be applied in, while partially sacrificing the uniqueness of predictions to comply with only statistical description of actually pure mechanical situations, although in these cases it is not a measurement perturbation that is being met with but rather the complexity of the trajectories themselves, which in principle could be described as CP. These might be processes, in which along with a regular external force the body experiences multiple collisions, each one influencing the trajectories but weakly although adding up to yield an important effect. The most familiar example is the Brownian motion — the averaged effect of molecule collisions on the motion of a macroscopic body. The possibility of the description in terms of a diffusion random process is conditioned in this case by the sufficiently smooth averaged parameters of the medium allowing for taking into account the momenta only up to the variance of the random (Markov) process. Averaging methods are also useful in the description of the motion of a body in a bounded area under a force that is quite regular and simply specified with a contact scheme for the test bodies, but the collecting of deviations due to high order resonances, however small individually, leaves only probabilistic predictions to be reliable. In this case, a statistical description is relevant for sufficiently smooth initial distributions, and the particular features of the force provide the ergodicity and exponential local divergence of the trajectories.

All such cases are peculiar in this that CP in its original statement becomes no more than an auxiliary means, since the details of the force are but of minute importance, while the statistical, average features come to the first place. This means quite a new problem statement, disregarding individual predictions and implying multiple repetition of the situation with different outcomes given the precision of repeatability of the initial state.

Quantum mechanics implies the statistical approach too. However, this time the uncertainty comes from the measuring contacts themselves. In so doing, the probabilistic approach becomes intimately in touch with the basic concepts of the Method. Indeed, the very concept of trajectory has been realized in the Method with contact schemes of bodies (called particles in quantum mechanics, usually being applied to problems in micro) with some arrangements of special measuring bodies — as photons and massive bodies. It is expected that quantum mechanics might be applicable, provided the registration of measuring contacts is organized in such a way, that perturbation of the particle trajectory will be minimal, though already comparable with the external influence as found in measurements with macroscopic test bodies.

As distinct from the two above cited examples, in which a particular feature of the medium or a space distribution of the acting force immediately entered the problem conditions, now the very measuring device is to be so designed as to support the repeatability of at least probabilistic predictions with a relevant measuring procedure. This implies quite different output of experiments: we no longer predict the final contact at every case, limiting ourselves down to predicting only the probability of its occurrence in multiple measurements with "identical" initial data.

Apart from creation and annihilation of particles, we suppose the final contact of the particle to be registered unambiguously — "yes or no", since the next events are implied to occur beyond the problem limits. So, here the perturbation by measurements is no longer important according to CP statement, in which the same particle is being considered over the whole evolution. The evolution itself as well as its former realization as a trajectory is needed solely as an instrument to predict the final contact. Whenever the intermediate contacts don't influence this prediction, we can speak about a trajectory, but if the prediction of only the probability of the final contact is supposed, it is no longer obligatory to reduce the evolution down to a trajectory.

Carrying out all the discussion exclusively in terms of contacts, we have to modify the measurements in accord. So, instead of a somewhat uncertain notion of "macroscopic device", which is ultimately being reduced to the contact with a measuring body either directly or via a relevant gauge procedure, we try to extend this notion in still acceptable in mechanics way. To this end, let us analyze the structure of CP

in more details. Actually, upon constructing solutions it was suggested that measuring bodies, implicitly forming vacuum, fill the contact space so densely that the body, CP is being stated about, meets a body from the measurement kit at every point of its trajectory. This introduces no problems if the measuring contact doesn't perturb the trajectory.* We have already seen, however, that in order to construct the trajectory we don't need for it to have measuring contacts everywhere. On the contrary, these contacts should not be distributed too densely, leaving room for photon oscillations. This kind of measurements defines the differentiability of the trajectory, letting it to be approximated with a chain of separate links. For the prediction of the final contact by means of trajectories, it is then necessary for the measurement kit to keep some regular structure as defined with photon oscillations between its bodies. However, this kit must be sufficiently dense, so that still affording differentiability, it does not let the trajectories be lost, that is, we demand that the absence of the contact with one measuring body implies its occurrence with some other. But in a process similar to Brownian motion the impacts of molecules are distributed at random. Actually, just two separate random processes are here: the random position of the molecule at the impact moment and the random momentum impart in its scattering. We could reduce the randomness of our prediction removing at least one of these factors. To this end, instead of the measuring device consisting of a single measuring body, we propose a new measuring device. Namely, we will employ particle contacts with the same measuring kit, while registering now measuring contacts not with a single measuring body, but with a group of them somewhat ordered — the *order*. In this registration, *we don't determine the particular measuring body of the order this contact took place with*. A sequence of such measurements doesn't give a trajectory in the former sense, but with a relevant arrangement of the orders it is still possible to make sometimes predictions about the final contact occurrence, although now it will be only its probability. The proposed extension of the concept of trajectory consists in this that afterwards upon constructing the evolution under external force effect, it will be possible to correspond the orders to links and chains, borrowed from macroscopic contact schemes for trajectories, so as to make it possible to define the orders themselves with some contact schemes.

Considering former trajectories from this new viewpoint, we might say that there the contact with a measuring body occurs with the probability equal one. Since the measuring kit fills the whole contact space, a non-one probability of the particle contact with definite measuring bodies means its contact with some others at the same moment.

A minimal departure from the former schemes consists in this that now only the contact of the particle with an order, not with some of its bodies, has a finite probability. If according to CP statement the particle doesn't disappear, while the scheme of mutual contacts in the measurement kit is left unchanged, any intermediate state might be considered as the final, accordingly reformulating CP. Therefore beyond the limit of sensitivity it is also possible to register the fact of a contact between the particle and a particular measuring body, but now this registration will become the final for CP.

In so doing, we try to keep the construction of space intact at the expense of making the particle trajectory, as it were, "spread", allowing for the simultaneous (in terms of the canonical version) contact with more than one measuring body without their common contact. Strictly speaking we have to accordingly change the very geometry of the contact space as a minimal structure encompassing all possible trajectories. This geometry will still be valid for the averaged trajectories multiply repeated, allowing for only probabilistic CP solutions, that is, fluctuating around these average.

Avoiding for the time being complications already on the initial step of presentation, we shall suggest the external influences on the particle trajectory to be specified with the trajectories of test bodies in fields as determined with the same contact schemes as before. Then we are in a position to divide the full influence on the particle in the parts, one of which is defined by the field independent of the particle evolution, and another depending only on its measuring contacts.†

In this description of motions, we have to replace standard trajectories-links in chains, which approximated actual trajectories upon neglecting measurement interactions, with links comprised of measuring orders. The latter are specially organized contrary to chaotically distributed molecules in the Brownian motion, acting on the motion of the macroscopic particle. This process might be called "semi-random". Accordingly the statement of CP is to be altered. In a macroscopic measurement one asks: "What is the value of the variable to be measured?" In quantum theory, for each individual measurement the question is being formulated differently: "Has the variable a specified in advance value?" Interaction with orders results in a random diffusion-type process in the scattering of the particle on macroscopic measuring bodies. Unlike the Brownian motion, this will be the scattering of a light particle on heavy bodies. In this approach, no hypotheses concerning Nature are there. We just try to ask familiar questions on the verge of their applicability, and the theory is simply restricting the scope of deserving our attention cases to those, where it is still able to make predictions.‡

In a sufficiently dense flux of the measuring bodies, the scattering of the particle on them might dominate the external interaction in its influence on the final position of the particle

*Otherwise, the body could not move at all pushing through measuring bodies.

†Further on, the external field values will be corrected in accord with the probabilistic schemes.

‡If some other interesting situations would be found, in which the description is not reduced to the registration of contacts, the theory might be different.

in the contact space. It is this boundary situation that is the issue of quantum mechanics in the canonical version, in which the measurement orders copy those determining the trajectory in cases when the measurement perturbation is inessential. Let us first consider the measurement process in terms of the canonical version, insofar as they express a CP statement and correspond to their own contact schemes, i.e. we shall use such quantities as velocity, acceleration, angular momentum etc. It is desirable to keep these concepts as long as possible in the extension of the scope of applicability of the Method, while matching them to the statistical description. In the canonical version these quantities are defined as operators acting on state amplitudes. These operators copy the forms, borrowed from the macroscopic CP, and such that their mean values fulfill classical relations. It is just the keeping of the CP statement, perhaps in a probabilistic form, as directly expressing the user's concern that explains the well-known paradox: "Why quantum theory with a different, statistical, type of predictions failed to elaborate its own forms for dynamical variables?"

The measurement order of measuring bodies with parallel trajectories and the particle inside is shown in Fig. 4.1. It is a measuring device completely constructed as a CP contact scheme. It is convenient to conceive vertical rows in Fig. 4.1 and similar in the transverse direction, which move together and without distortions of the order in the direction perpendicular to these rows. This scheme is in agree with the structure in Fig. 3.11 for the case of mutually orthogonal trajectories. The positions of the bodies in the order with respect to their neighbors are defined with the oscillations numbers, which are counted in such a way that the initial and final contacts in the neighboring intervals coincide within their longest periods. The total number of such periods must be sufficiently big to avoid the collection of the error coming from the difference in the position of end contacts within one period. The motion of the particle inside the order is in itself a purely measuring procedure, depending only on its collisions with the order bodies and not depending on the external force. Since it is only a contact with the order as a whole that is being registered, this procedure should not be regarded as existing in the same "time" as the motion of the particle in external fields. To simplify the presentation we'll consider the process of registration in the rest reference system of the order as a whole.

All the measuring bodies being taken identical, the fact of registration doesn't depend on the position of the body it happened with. In particular, the probabilities of this contact are the same for all the rows. If the order is uniform, that is, all the oscillation numbers are the same, and consists of infinitely many rows, then the probability of registration doesn't depend on the position of the row the contact occurred with within the order, hence the fact of registration provides no information about the place of the particle inside it. If also the velocity of the order is exactly the same as that of the particle, then the particle once positioned at a free place inside the order, would never be registered. Reversing the argument, it would be tempting to conclude from the absence of measuring contact of the particle with the order that their velocities are exactly equal. But the same would be observed if no particles were there. Therefore some residual interaction must still be kept to register the measuring contact. In the canonical version this interaction is naturally characterized by the lowest ("boundary") momentum value of the particle to be transferred to an order body in the act of registration, i.e. its boundary velocity relative the order.

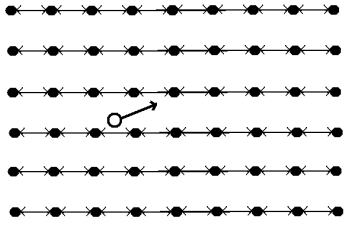

Fig. 4.1: A particle within the order of measuring bodies. Thin lines with arrows draw oscillating photons.

A single scattering contact doesn't necessarily result in the registration of the measurement contact and only multiple independent of one another collisions of the particle with the bodies of the order might at last produce the registration. In the elastic scattering of a light particle on infinitely heavy measuring body the absolute value of the momentum doesn't change. Then the existence of some fixed lowest value of the particle momentum means that the event of registration for the particle having only this absolute value might occur only in its scattering in the direction opposite to its motion. But on average the particle scatters over a small angle, and the probability of back scattering is low. Therefore, the particle having this value typically transfers but insufficient for registration momentum in a collision and only once after many collisions in the diffusion process the back scattering occurs. If only one non-disappearing particle is meant in CP, the measurement contact must occur only in inelastic scattering. Otherwise, second and more such contacts might occur, and the measurement would not be able to determine the number of involved particles. The inelastic scattering could always be interpreted as the transfer of the energy from the particle to an intermediate carrier-photon and from this photon to the body-detector. If we consider the photon as a harmonic entity, then some frequency might be ascribed to this minimum momentum, so being related to a wavelength of the photon, in turn defined by the minimum inter-row distance, in order to let photon oscillation process remain meaningful. So, we can regard the largest number of these oscillations as an equivalent in CP of the boundary particle velocity in the canonical version.

In the utmost precise measurement with the infinite number of tests the particle has only this boundary velocity. Of

course, the particle with a different velocity will come into contact with this order as well, but it is possible then to choose a different velocity for the measuring order, for which the relative velocity of the particle turns boundary. Actually, this is the same as what is met with in macroscopic measurements (Fig. 2.5), where the tangent trajectory from the measurement kit is being matched to that of the body in CP.

Measurement contacts consist in the transfer of momentum from the particle to the order. If the particle has only the boundary velocity in the reference system of the order, then — loosing this value in the measurement interaction, it becomes at rest, i.e. not discernible. This causes no difficulty for an infinite order, since the particle has already been discerned. Infinite order detects the particle with the probability 1, however this order is not relevant for dynamics requiring localization of the particle in the contact space. To this end, we could use orders with a finite quantity of rows, again with the equal inter-row oscillations numbers. But in this order the measurement contact for a particle, having the relative velocity more than the boundary, though scattering over but a small angle to transfer the boundary momentum, might be more probable than the back scattering of the particle having only its boundary value. However, we are no longer able to use the criterion for the precise matching the velocities, hence, an uncertainty in the measurement of velocity springs up in the use of orders as the only measurement devices. Actually, localization is still uncertain too, because even for a fixed quantity of rows it is still possible to vary the number of inter-row oscillations numbers left free so far. So, once the measurement contact has been registered, it is still indefinite, in which row of the order this contact occurred as well as what the relative velocities was.

The uncertainty of measurements might be reduced, so not eliminated, if we construct the measurement device-order as a non-uniform sequence of rows. This order could also be completely defined with numbers of inter-row oscillations, which start and finish at once. Only now these numbers will not be equal each other. For instance, in some part of the order these numbers might be the biggest, while so decreasing according to some law in both directions that on its ends there is only one oscillation.* Thus, we tie the measurement up to the top velocity. Now we can introduce an averaged over the order oscillation number Q as the ratio of all the oscillation numbers in the order S to the number of its rows K. Next, we can define some law of the oscillation numbers decreasing q as a function of the row number k, as counted, e.g., from the row having maximal value of $q = q_{max}$ relative to its neighbors. In both directions this numbers could be defined independently of each other. These utmost dense rows are the most important ones for the localization of the particle, while the relative oscillation numbers in different its parts are to be normalized with Q.†

It is convenient to compare this way of localization to the measurement of the position of a particle in the canonical version, upon considering the quantity of rows, for which $q > Q$ and specifying q_{max} in terms of the boundary velocity as it was done above for the uniform order. As a model, let us consider an order with a single dense part and a simple law of the density of rows decreasing in the form of the geometric progression.‡ This order could be conveniently represented with the function of the kind $q(k) = q_{max} \exp(-a|k|)$ with $q(0) = q_{max}$ possibly with different values of a for its left and right branches as along the order velocity, so also in transverse directions. Suffices it to address only one right, say, branch ($k > 0$). Here $K = a^{-1} \ln q_{max}$ (since $q(K) = 1$). This immediately defines S and Q. The value of a defines the localization of rows in the order: if a tends to zero, the order becomes uniform and accordingly K goes to infinity. Let us subdivide the order into a dense part $q > Q$ and a rarefied part — $q < Q$ with the corresponding numbers of rows to give K in the sum. It is natural to normalize the probability of the measuring contact registration to the average oscillation number Q. For a Markov random process, becoming equivalent to diffusion when the quantity of rows is so big that a discrete process might be approximated with a continuous one, the numbers of rows are replaced with their density with respect to the now continuous variable k. The probability density will be relatively high in the dense part, so that the higher probability will be here, provided the velocity of the particle (in terms of the canonical version) is close to the boundary value. However, in a regular order only the angular scattering remains, hence the probability of the contact depends only on the quantity of rows the particle is able to cross. So, if the quantity of rows in the rarefied part is much more than that in the dense, the particle with the velocity value far from the boundary might have a high probability of the contact as well.

In this case, the velocity (so also the momentum) of the particle will be inversely proportional to the relative oscillation number, so that for Q oscillations it could cross the necessary for the contact registration quantity of even rarefied rows.

In our model, the ratio of the quantities of the dense rows to that of the rarefied ones $w = \ln(f \ln q_{max}) / \ln q_{max}$, where $f = [1 - \exp(-a)]/a$. Upon varying a from zero to infinity f decreases monotonously from unity to zero. Large values of a correspond to the small total quantity of rows, and anyway it must be $f \ln q_{max} > 1$, for the quantity of the rows to be a real number. Taking q_{max} so big that even $\ln q_{max} \gg 1$, while a not too big, we may regard $f \sim 1$ for the argument of the

*Since the velocity of the particle cannot exceed the top velocity, the total quantity of rows will be finite.

†In terms of the canonical version, on the parts with comparatively small oscillations numbers the distances between the rows are longer, and a slow particle cannot pass a required quantity of rows for Q oscillations to be registered with high probability.

‡We don't address here the quantum field theory with its more complex orders.

logarithm to get $w \sim \ln(\ln q_{max}) / \ln q_{max}$, so that even the repeated logarithm representing the quantity of the dense rows, must be much more than 1. Then the value of w practically depends only on that of q_{max}, completely defined by the least detectible measurement contact. So, in CP w is the analogue of the universal constant $\hbar$ of the canonical version, since the *ratio* of the quantity of the dense rows to that of the rarefied rows in the order is analogous to the *product* of the distances and momenta in the canonical uncertainty (or indeterminacy) principle.

An important modification of the non-uniform order, completely defined with oscillations numbers ratios, is presented in Fig. 4.2. It represents the own angular momentum of the particle, i.e. its spin, and might be understood as a combination of the non-uniform order and the two-dimensional sphere. In the direction perpendicular to the plane of the Fig. 4.2, we leave inter-row oscillation numbers to be uniform, while in the radial (and the related angular intervals) prescribing the same law: $q(k) = q_{max} \exp(-ak)$ with $q(0) = q_{max}$, k being the radial number of the measuring body. The largest "radius" corresponds to one oscillation independently of the angular quantity of rows: Upon increasing this quantity this radius increases as well as the initial with $q(0) = q_{max}$.

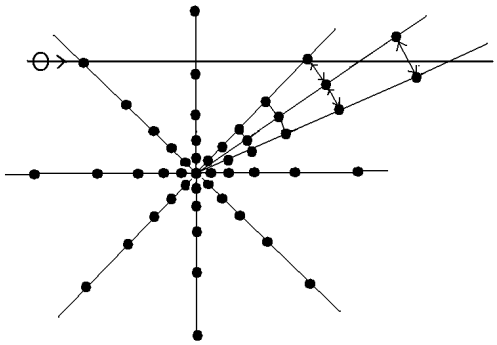

Fig. 4.2: The particle inside the order of the centered rows. Thin lines with arrows are the photon oscillations between the bodies from adjacent rows.

The analogue of the angular momentum in this contact scheme is again the ratio of the dense and rarefied rows, while now it will be interpreted as the realization with a contact scheme of the particle's spin, the existence of which has been mentioned in Ch. 2, but only as a necessary condition in the basis representations of trajectories, not supported as yet with a contact scheme. For the comparison with the canonical version we consider, as usually, the construction of spin in terms of distances and momenta. Given the initial radius r_i, the total quantity of rows K is specified by the value q_{max}, in turn, specified with the proportional to r_i distance $d(q_{max})$ between the rows for their quantity π/K. On its passing the order, the particle crosses one-by-one rows with the oscillation numbers $q(k)$. With the increase of r_i, given $d(q_{max})$ the K accordingly increases to decrease all $dq(k)$. The angular diffusion of the particle is a function of only the number of crossed rows necessary for its scattering with the average probability in a collision, corresponding to the transfer of the boundary momentum, given the absolute value of its velocity. In the vicinity of any k, the required quantity of crossed rows for the measurement contact to occur, increases with K, in turn, corresponding to a smaller velocity. Then the product of r_i and the averaged over the order velocity doesn't depend on r_i, so representing the own angular momentum of the particle, i.e. its spin. Being so defined, spin is completely specified by q_{max}, in contradistinction to the orbital momentum with its independent values of radius and velocity.

The discussed in Ch. 2 presence of twin trajectories in the sphere as represented with their decomposition in a basis, is innocuous if the diffusion accompanying measurements is neglected, since in any CP this uncertainty might be eliminated by means of a definite choice of the basic trajectories. Continuity of trajectories allows for the preservation of orientation along the chain, ascribing to the next link the orientation of the previous. However, in quantum theory indispensible measurement scattering introduces uncertainty in orientation, Spin makes it possible to remove this uncertainty, providing a "mark" on the particle.

One more example, illustrating registration schemes, is a screen with a slit (Fig. 4.3). The notion "screen" corresponds as its definition to a particular order of measuring bodies, belonging to "vacuum" in CP. Indeed, how to recognize the existence of a slit in the screen? It is "seen". This can be found upon passing bodies through it (in particular, light). In so doing, it would be incorrect to check everything on and then to turn the flux of measuring bodies off: what if the slit then gets obstructed? So, the screen itself as occupying a place in space should be considered in CP as a contact scheme, which defines limitations on the mutual contacts of moving bodies. So, the screen in Fig. 4.3 lets some trajectories pass, while blocking all others.

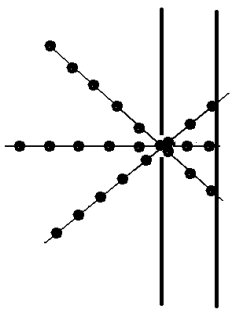

Fig. 4.3: The screen with a slit is specified by its effect on measuring trajectories.

Let us consider in this context the effect of interference in passing by the particle the screen with two slits (Fig. 4.4).

We have to consider the passing of particles through the slits on the background of the measurement order for just this

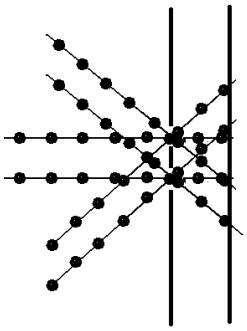

Fig. 4.4: The interference experiment with a double-slit screen.

screen and just this experiment statement: The order must have nearly the same velocity as the particles. If also this order is about uniform on the experiment scale, the scattering of the particles on its bodies will form an interference pattern in the distribution of the particles over the detector plane. Considering scattering on the measuring bodies, it is purposeful to correspond to the periodic structure of the uniform order a harmonic function with its period as defined by the numbers of the inter-row oscillations. Since the slits cut out a part of the full flux of particles, the total order to represent the screen includes also skew fluxes as shown in Fig. 4.4. Ignoring the sickness of the screen* the oscillations numbers are about the same for all the components of the total flux and equal to the velocity of the particles divided by the distance between the slits. A joint order for the two slits is characterized by the same values of these numbers for its two parts.

The trajectory of a particle as registered with its contacts with the measuring kit remains a continuous line as soon as the particle doesn't disappear, so might be detected in every neighborhood of a point it has been detected earlier. However, this trajectory is no longer a differentiable curve in the contact space. Indeed, a sequence of measuring contacts as corresponded to a trajectory cannot be locally approximated with a segment of the trajectory of a measuring body, because in diffusive collisions with measuring bodies the displacements are proportional not to time intervals but to square root of them, hence the velocity as defined with the ratio of distance to time becomes infinite as time interval tends to zero. We highlight that it would be entirely wrong to regard the "genuine" trajectory as smooth though seeming "rugged" due to an imperfect measurement procedure. The only available information about trajectories is the sequences of measuring contacts and only this. In this respect it might be possible to consider quantum mechanics as a structure *on* the basic geometry as specified with the classical scheme of mutual contacts. Then even in a free motion, i.e. in the absence of external influences, the particle moves over non-smooth diffusion trajectory due to its collisions with measuring trajectories, since were these absent, no information concerning this trajectory could exist. However, the measurement procedure is peculiar. Just this semi-random measuring process with the scattering on regular orders enables such phenomenon as interference, not to be met with in the Brownian motion or alike.

Multiply repeated measurements using orders identical in their distribution of the oscillations numbers create the statistical representation of the motion of a particle. It is then possible to describe the measuring interaction of the particle with an order in terms of a function defining the probability distribution of its trajectory as being measured with an order. Namely, each order as a measuring device, corresponding to some classical observable, appears as selecting from this function one of its own eigenvalues depending on whether or not this order had registered its contact with the particle. In terms of the canonical version, e.g., the fact of registration of the particle's contact with a non-uniform order, having a definite velocity vector and a value of a, defines the related variables within the precision limited by the uncertainty relation. Mean values as determined in the course of numerous measurements will coincide with the motion of the order in view of the symmetry of the scattering in the collisions. The particle being suggested non-disappearing and the kit of the orders full, the probability for it to be registered with at least one of them that is to be equal to the sum of the probabilities for all orders under independent measurements must be one.

Starting the description, in terms of contacts, of quantum processes under external influences, we will limit ourselves to the classically specified external fields as being measured with macroscopic test bodies ignoring their scattering. In so doing, the inter-link transitions in chains are defined as having the same ratios of their oscillations numbers as the related orders: Orders as the measurement devices move in the same way as measuring bodies were in the absence of measuring perturbations. On average, the particle in CP follows the chain transitions of the orders in accord with the trajectories of the test bodies.

However, because of uncertainty of the momentum of the particle inside the order its particular momentum in a probability distribution in the initial link in a transition depends on the place in the order the measurement contact took place at. So, in the part of an order with relatively small oscillations numbers the particle with an insufficient velocity will not be registered over the average oscillations number for this order, being unable to cross the required for the measurement contact quantity of rows. Therefore, a peculiar quantum dependence arises in the momentum transition upon averaging over the quantum ensemble, which doesn't depend on the external influence but rather being determined in the each transition by the probability distribution of the particle velocities respective the order.

In terms of the canonical version, the average velocity vector might be determined by means of decomposition of

*Screen is not a collimator!

the distribution into pieces, the distribution being then averaged over these pieces. If a piece is sufficiently small, the registered velocity is nearly constant over it, and in the total averaging each piece is to be accounted for with the weight equal to its part of the whole probability (Fig. 4.5). Just this is the velocity, the related order must move with.

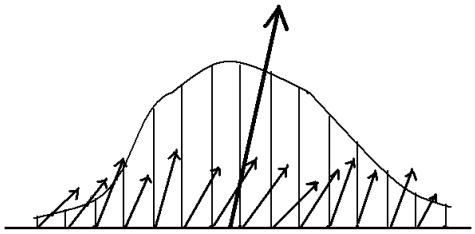

Fig. 4.5: The operation of velocity averaging.

In order to find the transition in a chain comprised of links-orders, adequate to the local change of the probability distribution for the particle (Fig. 4.6), it is necessary first to add up velocity vectors with their probability weights distributed over x_1 at the moment t_1 in the intermediate moment t_2 at the point x_2, so obtaining the average velocity on the initial link.

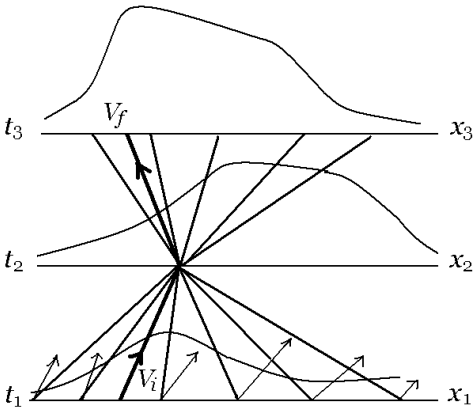

Fig. 4.6: Particle positions probability distribution functions in three close time moments. The x axis represents all three coordinates.

On the initial link, the particle is scattered by the measurement order having the velocity V_i. On the final link, the average velocity differs from V_i under the external acceleration on the interval between t_1 and t_2, and the contribution from each point on the axis x_2 on the part between t_2 and t_3 in the distribution along x_3 depends on the scattering by the order with the velocity V_f.

So the diffusive scattering by the orders on the initial and final links of the transition is important even on the infinitesimal intervals. Because of this the average velocities and variances are no longer some prescribed functions as is the case of Brownian-type motions, but now depending on the distribution itself. Moreover, even apart from external forces, upon virtually dividing the trajectory into small parts, we would obtain its effective perturbation as if due to an external force that alters its momentum over a given time interval.*

A developing quantum process could be conveniently described with the formalism of the "Madelung fluid", in which the substitution of the wave function in its exponential form in the Schrödinger equation provides the representation of this process with a classical Hamilton-Jacobi equation, though including in addition to the external potential the so-called "quantum potential" that depends on the wave function itself. In our construction of the order, this potential is directly connected with the uncertainty principle, because the averaged over the order addition to its velocity for a specified time interval, depending on the probability distribution, is an equivalent to altering the external force to be defined as a cause of changing the momentum for this time interval. A complex-valued and possibly multi-component ("wave") function of a point in the contact space is sufficient for probabilistic predictions in CP, provided the sum of their amplitude absolute values squared is interpreted as the probability for the particle to be found at a point.

The constructions of orders solely by means of the photon oscillations counting would correspond in canonical version to the relativistic quantum mechanics, in which the Schrödinger equation is to be replaced with the Dirac equation. The dependence of the full potential on the amplitude of the wave function in the Madelung formalism is still relevant, but now the diffusive addition to the probability of the inter-link transition is proportional not just to the second derivative of this amplitude absolute value as it is in the Schrödinger equation, but to a more complicated combination of this amplitude with the external field including also their derivatives. In our representation with orders this contribution results not only from the distribution of velocities over the order but also from the distribution of the external force there. Therefore, the scheme becomes nonlocal. Depending on a particular CP statement, this nonlocal behavior might appear also in a macroscopic measurement as soon as it might be represented with a relevant order.

The solutions to quantum equations define possible distributions of the particle states. In particular, if two or more identical particles are confined by an external field (as the field of the nucleus in an atom) in a phase space region of the order of quantum uncertainty, then they can contact with the same order at once, and therefore this order cannot be regarded as belonging to a definite distribution of a single particle. The measurement scattering thus masks their state inside the region. Once registering a contact with the order, it would be impossible even to say how many particles are there, and a separate "Pauli principle" is needed to for-

*Such is, e.g., the spreading of a wave packet in its free motion.

bid such a situation.* Even placing the particles in separate though partially interlacing states, it is not possible to exclude their exchange. This transposition could be considered as an additional "exchange" interaction, since it might select stationary states similarly to an external potential out of those defined by the external potential alone.

Chapter 5. Gravity: a forceless force

> Proposition 7, part III.
> That there is a power of gravity pertaining to all bodies, proportional to the several quantities of matter which they contain.
>
> I. Newton, *Mathematical Principles of Natural Philosophy*

The necessary component for CP solution — an external force must be presented with trajectories of the test kit bodies to be expressed, in turn, in terms of contacts with trajectories from the measurement kit. In this context, a question arises as to the existence of the utmost general scheme of free contacts for a single link, not using a separate kit of test bodies, though dispensing with chains with their inter-link transitions.

If no test kit is there, then no coupling constant like charge is needed any longer to define trajectories using forces. Consequently, this force must be directly expressed with photon oscillation numbers. Keeping the idea of inertia to predict contacts with CP solutions, we have to keep also the related concept of mass, and then the only possibility to eliminate coupling constants consists in this that to make the charge of a body to be equal to its mass. Then their ratio becomes a universal constant, and the motion will not depend on the individual properties of bodies, so generalizing the concept of motion that is free of external influences.

In the constructions of the last two chapters we systematically used the concept of the trajectories' parallelism. In flat space-time of the canonical version this concept is realized with rectilinear trajectories having equal velocity vectors. Sets of straight lines possess also a remarkable feature that any pair of them either don't intersect or have only one point in common. However, not all features of straight lines are so exhausted. Suffices it to draw straight lines on a sheet, which will then be arbitrarily deformed, but neither cut nor (its parts) glued together. Straight lines become curved but the scheme of their intersections remains the same: If initially the lines either intersected only once or not intersected, their images feature the same. However, a particular way to construct parallel trajectories presented in Ch. 2 doesn't guarantee all the properties of their sequences. Let a sequence consisting of parallel standard trajectories is such that for each of its trajectories the next is closer to it than the previous. In terms of contact schemes like that shown in Fig. 5.1, this means that the oscillations number in a scheme between this trajectory and the next is bigger than that between it and the previous.

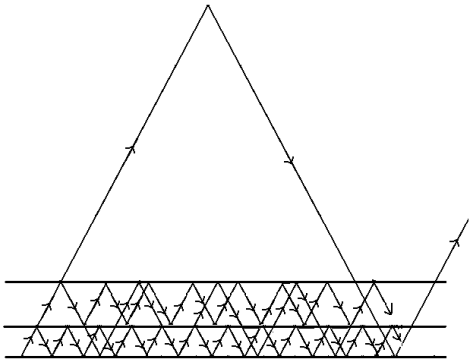

Fig. 5.1: The converging sequence of parallel trajectories.

The limiting for this sequence trajectory is determined upon tending these numbers to infinity. In flat space-time simple transitivity exists: If we specify the numbers of oscillations between the first and the second and the same between the second and the third trajectories as shown in Fig. 5.1, then a linear combination of these determines the number of oscillations between the first and the third trajectories. In a curved space-time this relation will be non-linear. So, it is possible to characterize this deviation from linearity as violating the transitivity of parallelism, which might be different in different parts of the contact space.

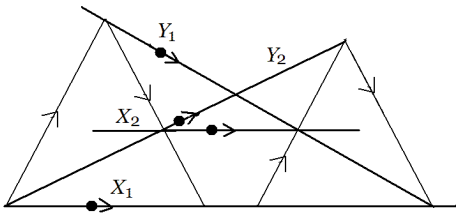

Fig. 5.2: In this scheme the parallelism of X_1 and X_2 depends on the existence of (Y_1, Y_2).

A different from shown in Fig. 2.3 construction of a trajectory that is parallel to a given was proposed by Marzke and Wheeler (1964). Their contact scheme is shown in Fig. 5.2. The parallelism of X_1 and X_2 follows from the similarity of the triangles formed by the contact of Y_1 and Y_2. In a curved space-time these two could, generally speaking, fail to come into contact, so the proof fails. This scheme cannot be generalized on a curved space-time, because it no longer provides a definite trajectory, continuously tending to a parallel to the given trajectory upon the decrease of curvature. Therefore, the authors were bound to construct a chain as being comprised of piece-wise straight links. But in an acceptable contact scheme, a trajectory that is not influenced by an external force must be a single link devoid of inter-link transitions.

*It is still possible to distinguish two particles with their spins orientations.

The scheme in Fig. 2.3 is satisfactory in this respect.

But what is bad, if the trajectories in the sequence shown in Fig. 5.1 were not parallel? Since the convergence of the trajectories is specified with the infinite increase of the numbers of the oscillations between them, the limiting trajectory would not be uniquely determined, provided these numbers could tend to infinity due to a different cause. In particular, a false sequence might occur if some its adjacent members come into contact somewhere, so that the oscillations number becomes infinite already here. Just to eliminate this possibility the parallelism is needed.

So far the concept of test bodies as a source of information was basic to determine the force that accelerates a body in CP, that is, the transitions between the tangents to its trajectory. Of course, the violation of transitivity of parallelism is able to simulate an external force on its own. It seems that it is always possible to distinguish these situations in the same way as it was done in Ch. 2, that is, to choose for the test bodies neutral or heavy ones, the motion of which is not affected by an external force. In so doing, it is implied that oscillating photons don't feel this force. But if the main ingredient of contact schemes, i.e. the trajectories of top velocity bodies (photons) depend on curvature as well, then a curved contact space itself as a contact scheme for free trajectories might be applied unchanged. It is possible logically, hence must exist in Nature, because CP is that primitive.

This peculiar interaction requiring no test bodies is known as gravitation. Its source is the presence of bodies, which influence in the canonical version as the motions of bodies themselves, so also the propagation of fields. In particular, the propagation of light possesses an interesting property (the so-called Huygens' tail): The contact space curvature lets the inside of the light cone contribute to the solution.* In Ch. 3, the inside of the light cone was deprived of the initial data specification, because this contributes to the solution with the relevant contact scheme nothing but either zero or infinity. It might easily be found, however, that a finite contribution, impossible in the degenerate plane case, can exist, provided that instead of multiplying initial values by discrete natural numbers (although tending in the limit to the dense subset of the compact) to use a continuous function of points $V(x, y)$. Indeed, the differences of this function values are also approaching zero, when the points near each other upon increasing their quantity in the compact area inside the cone. In general, this possibility cannot be ignored, if it is possible to find a contact scheme to realize the required function. The initial value to construct this function might be (see Fig. 5.1) the ratio of the oscillation numbers between X_i and X_{i+2} to that between X_{i+1} and X_i for the symmetric positions of X_i and X_{i+2} respective X_{i+1} if (X_i, X_{i+1}, X_{i+2}) exists. In particular, these trajectories might be parallel, i.e. with their triple contact at infinity. In the limit, all these numbers go to infinity, while their ratio might remain finite presenting the local contact space curvature. It is 1/2 for plane contact space, and the deviation from 1/2 will be taken as the measure of the space curvature $K(x)$.† The construction goes in steps. On the first step, an auxiliary function $U(x, y)$ for a pair of points of a ray depending on the space curvature is defined (Fig. 5.3). To this end, on the trajectory X_x a point x' with its light cone is taken close to x. As above (in Ch. 3), k trajectories are so taken between x and y that each of them is parallel to its preceding. The distribution of these trajectories along the ray is specified with some fixed number of oscillations n between the neighboring trajectories "from cone to cone". Then the own cone of the point y' positioned on the intersection of X_y with the cone of x' is taken. On both cones at x and y' we construct uniformly distributed sets, consisting of sufficiently big (not necessary equal) quantities of photon trajectories. We chose out of these a bunch of trajectories close to that from x to y on the cone of x and the reciprocal bunch from the light cone of the future of y just covering the first one. The ratio of the quantities of the trajectories in each bunch to their total quantity in the uniform distribution over the cone is analogous to the ratio of the area occupied by the bunch to the total area of the two-dimensional sphere of photons. In flat contact space, their solid angles, hence the relative quantities would differ by k^2 factor. Therefore the basic function U should be the ratio of the portions of the bunches in both spheres times k^2. Then the deviation of U from 1/2 will determine $K(x)$ via the transitivity violation in the parallelism of measuring trajectories as expressed with the ratios of photon oscillations numbers.

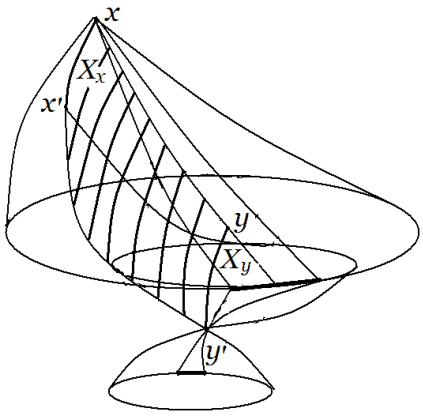

Fig. 5.3: The auxiliary function on a photon trajectory.

So far, all the construction involved only light cones, since just the limiting status of photons provided the uniqueness to the constructions. In the structure of the solution to the equations of fields' propagation, this part corresponds to the sin-

*We proceed addressing the cone as "light", since already in the next chapter just usual electromagnetic photons will play the decisive role in the contact schemes for the weak and strong interactions, though in principle one could also use the fronts of gravity field itself for the oscillations.

†We drop here rather tedious details of constructing $V(x, y)$ basing on $K(x)$, limiting ourselves to a general description of the derivation.

gular component, corresponding to initial values specified on the light cone itself. For the given definition of $U(x, y)$, suffice it to multiply by U the specified initial data at all the points y to find the solution at x. But U could also be used to construct the function $V(x, y)$, which is needed to account for the contribution from the inside of the light cone. It turns out that $V(x, y)$ might be constructed with a contact scheme by means of iterations the schemes already developed for the singular component of the full solution. Using in turn the cones of the past and future, it is possible to reach all the points inside the cone just as was done above (Fig. 3.10). Using the introduced there operations of differentiation with coherent subdivisions as of the photon trajectories themselves (null geodesics), so also the constructions of their uniform distributions on light cones, $V(x, y)$ is being determined as the limit of the convergent sequence of the functions determined by the iteration on going to consequently denser sets of points inside the cone. Two first points and the related light cones are shown in Fig. 5.4. The function $V(x, y)$ is being constructed as an "inverted tree" model. In order to determine its value at x, which depends on the values of the same function inside the light cone of x, we need to find first its values at all points of this cone (conveniently, in the intersections of an initial uniform lattice). Let this lattice already formed on the light cone of the past of x and that of the future for y. To list the operations on the first of these cones, we choose a point of the lattice y_1 and take a similar pair of cones between y_1 and y, and form their own lattice on these. For the next iteration step, we take y_2 and so on. In order to find their own $V(y_1, y)$ on each pair of cones, we need its values in their insides. So, the functions V on the smaller cones are the initial values for the larger ones, in which also the singular components of the each step's light cone take a part. All these infinitely getting smaller and smaller, cone-pairs and lattices provide $V(x, y)$ at all points, where the initial data are specified.

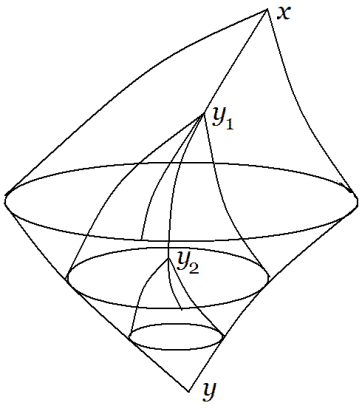

Fig. 5.4: First two initial points with their light cones pairs.

In the canonical version, the presented procedure corresponds to the solution of the wave equation in a curved space-time for the values of $V(x, y)$ at x with the initial data at the light cone of the future of y (the so-called problem with characteristic initial values). These are the values of $V(x, y)$ itself, where x's as being taken now at this cone, are to be determined with the integration of $U(x, y)$ and its derivatives. We stress that in all contact schemes the trajectories of photons are their actual trajectories, and it is only in the canonical version they look curved and non-uniform. However, in contact schemes such pictures are redundant, and they might even be deceitful, corrupting the real problem with extra decorations.

The presence of the "tail" in the propagation of the top velocity signal, violating the Huygens' principle, is characteristic for a curved space-time. This tail owes its existence to the contribution from inside the light cone. The cause of this well-known fact is evident in Fig. 5.3: The propagating wave is being "scattered on the curvature" of the space partially and multiply, so that the scattered fraction is retarded with respect to its front upon going the longer ways. In the canonical version this tail constitutes the non-singular part of the Green's function.

All this discussion belongs to the classical, that is, not quantum contact space. This space has been constructed as a tool for CP predictions, as a minimal structure enveloping all possible trajectories of bodies. However, if the space curvature is so strong that scattering on the curvature becomes comparable with the scattering on the measurement orders, then $K(x)$ becomes probabilistic itself. So, also the presentation of contacts via trajectories becomes uncertain, and the question arises as up to what degree of curvature it is still possible to regard the contact space as the same geometrical structure that only fluctuates about some average. In particular, whether or not quantum effects might violate even the topology of the space, enabling stochastic transitions across the singular light cone?

Chapter 6. What interactions are permitted by the Method?

> To earth, then, let us assign the cubical form; ... the pyramid [tetrahedron] is the solid which is the original element and seed of fire; and let us assign the element which was next in the order of generation [octahedron] to air, and the third [icosahedron] to water.
>
> Plato, *Timaeus*

Everything expressible in terms of contacts is within the Method and must exist in Nature. This is the principle the classification of interactions compatible with the required for CP geometry must follow from. It is thus natural to inquire their accepted variety. In the framework of the Method, our purpose is not in reducing this variety to a single interaction, but rather to reveal the universal construction of all interactions in terms of contacts. Actually, for the representation of forces with contact schemes it is only the possibility to

uniquely correspond at any point the trajectory of the body in CP to that of a test body is important. Being already capable to determine a field structure, we still stay in need to define everywhere the universal standard of the charge for any interaction. In other words, this standard, once specified at a point, must then be unambiguously transported to each point that is reachable with a trajectory starting at the initial point. We have to devise a relevant contact scheme.

It is tempting to use for the transport the above defined uniform distribution of trajectories in a sphere. In the constructions of integration, we used almost uniform distributions of a big (infinite in the limit) quantity of trajectories. It was not required therefore to keep the uniformity precisely, since small deviations from this might be ignored in the limiting distributions of oscillations numbers ratios.

Quite a different situation is met with if only a few trajectories are involved, while the required uniformity is precise. In the three-dimensional flat space, as mentioned in Ch. 1, there are only five such exactly uniform distributions of trajectories over a sphere. Recall that these correspond to the vertices of the regular polyhedrons known as the Plato bodies, namely, tetrahedron (4 vertices), octahedron (6), cube (8), icosahedron (12), dodecahedron (20). The cube might be presented as the two interlaced tetrahedrons, and the dodecahedron as consisting of a cube and a group of six "dipoles". Although the latter is not a regular polyhedron, it might be of interest as a complementation of the cube up to dodecahedron, the richest with respect to regular substructures sphere. Neither octahedron, nor icosahedron possesses regular substructures.

Let us correspond to each polyhedron its sphere-star comprised of the trajectories of test bodies passing the star vertices. It follows from the symmetry of a polyhedron that these bodies have equal unity ratios of the maximum (that is, between the neighboring pairs) oscillation numbers. If the charges and masses of all these bodies are the same, still due to the symmetry these ratios will remain unity in the presence of electromagnetic, say, interaction of the bodies.

Let the star comprising bodies go exactly through its center to depart from the star afterwards. Though in the classical theory with the singularity at the center this is impossible, it is possible for quantum wave packets. For the oscillations numbers counting, only the motion of the wave packet center is important, while its spreading (in addition, being smoothened for relativistic velocities) is usually of no importance, since the packet center moves classically. However, quantum effects are important for radiation upon accelerating or decelerating the bodies.

It is possible to transport the standard of charge in sequences to other points of the contact space (Fig. 6.1), so constructing a lattice to specify the charge value in CP. The exact copying of the prime symmetry in the descendent star would ensure the correct charge transport — its equality to the prime value in all the descendent star generations.

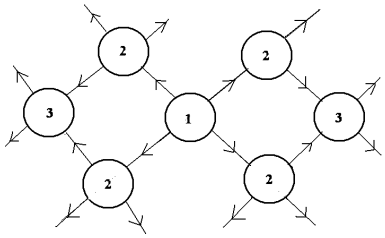

Fig. 6.1: The sequences of the stars in a regular lattice.

Inside each star, the identity of the charges (and masses) of its bodies is guarantied by the observation of the symmetry in their motions toward the center as measured with the ratios of their oscillations numbers. The copying of the charge in generations is realized with the use as a seed for the next star some bodies that leave the decaying previous stars. The correctness of the copying might be checked on, provided that in every star several bodies coming here along different paths take a part. Then the observation of its own symmetry means that the charge has been correctly transported along the sequence. The lattices possessing this property will be called *regular*, and their constructions will be our main concern throughout this chapter. In so doing, suffice it to take just three bodies out of the preceding stars as a seed, since these can form a basis.*

For visual convenience, we again shall carry the analysis out in terms of the canonical version, though actually all the constructions might be presented solely with contact schemes. The potential of the interaction of the star comprising bodies must satisfy the general for all relevant forces condition to preserve the ratios of the oscillations numbers, while to have a sufficiently long range to ensure the detection of symmetry breaking with photon oscillations counting for any size of the star — the cell of the lattice.

Among the Plato solids the cube alone possesses this property that in the motion of comprising it bodies under their interaction the trajectories keep straight and have the contact at the center not only if their charges and masses are equal but also if these are equal only in each of its two tetrahedrons separately (Fig. 6.3). If in addition these tetrahedrons differ only in the sense of their charges, then the bodies are being equally accelerated by their interaction, and the symmetry as detected with the oscillations counting is observed.

The neutrality of the star as a whole results in the common contact at its center also in the classical picture. All eight bodies are being accelerated toward the center along straight lines, and the symmetry remains intact under any radial dependence of the interaction potential. However, this dependence is not arbitrary: The potential must decrease with radius. Otherwise, even observed symmetry in stars would not allow for the regular lattice with these stars, since the star

*Remember the existence of the degeneration, however!

will not decay into separate trajectories after their contact at the center to take a part in descendent stars. The most important example of the relevant potential, enabling the detection of symmetry break in the cube by means of counting the oscillations numbers for any star size, is the electromagnetic interaction as the only available for being directly measured with detectors in experiments. In this implementation, the charges of bodies are the usual electric charges, as presented in Fig. 6.2, and the oscillating "photons" are the usual photons. (Recall that in their definition in Ch. 1, it was only important for them to move with top speed.) Note that due to the cube symmetry magnetic field is zero on the trajectories, and it is only electric field that effects on the motion. We call attention to the fact that contrary to the gauging charge in an external field, where only the charge-to-mass ratio is being measured, the star symmetry detection requires the identity both the masses and the absolute values of charges.

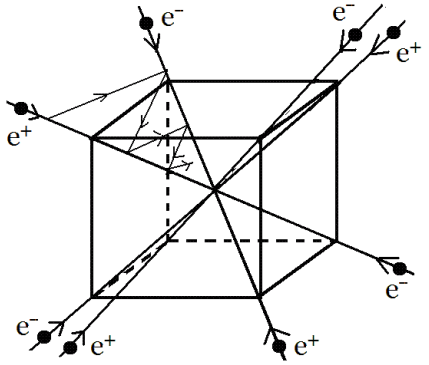

Fig. 6.2: The cube symmetry of trajectories consisting of the two oppositely charged tetrahedrons.

This contact scheme allows for the ideal gauge of charge upon gauging "motion-to-motion" without intermediate rods and/or clocks, which are prone either to add something of its own or to hide something. In order to detect asymmetry, suffice it to detect the difference in the oscillations numbers by just one oscillation. However, the infinite number of oscillations brings a problem about as to how to be sure that the symmetry is observed. Indeed, whatever number had been counted, it might happen that upon going the process on we would still detect asymmetry. In macroscopic measurements no serious problems will be met, since the desired precision is being determined in a practical CP by its application. However, if we address the issue of the minimum possible size of the star as a cell in the regular lattice, there is no a priory precision now, since everything is to be determined by the limiting process itself.

In this limiting case the particles must be the lightest, to maximize their acceleration in the interaction, so increasing the gauge sensitivity for the smallest charge. In this gauge procedure the mass and charge (its absolute value) must be discrete. Indeed, if the charge, say, were to vary continuously, then however big number of oscillations had been counted showing the symmetry as observed, a small deviation of the charge value is still possible, such that asymmetry could be detected, provided the counting was continued. Therefore, some particles with the smallest charge value must exist. These are electrons and positrons presented in Fig. 6.2.

We proceed in terms of the canonical version to keep correspondence with the familiar values. In order to find the smallest radius of the star, we consider asymmetry detection in the star, comprised of particles with the same mass, while the charges are the same within each of the tetrahedrons but the charges of one of them twice as large then those of the other. In this case, the interaction leaves the trajectories rectilinear. The double charged particles will be accelerated less than the single charged ones because of their larger repulsion. The smallest star size corresponds to the case when, given initial radius and velocity, the single-valued tetrahedron nears the center down to the distance corresponding to two oscillations, while the double-charged tetrahedron reaches only one oscillation distance. In so doing, we start counting oscillations from the initial radius which is to be minimized.

The process proceeds as follows. On the large radii the acceleration is small. In spite of the increase of the force at smallest radii, the difference in the velocities of the tetrahedrons is small here because of the already large relativistic factor. The main contribution to the inter-tetrahedron distance is being thus collected in the vicinity of the initial radius. Ignoring radiation reaction under acceleration, the estimation yields the initial radius of the order some tenths of the classical electron radius (3×10^{-13} cm), and initial relativistic factor $\gamma \sim 3$. The smallest radius the oscillation counting stops at is of the order of 10^{-16} cm.

The account of the radiation reaction, relatively low for the longitudinal acceleration, yields an additional equalizing of the tetrahedrons' velocities, so increasing the initial radius of the star. It is well-known that this radiation consists of photons with the energies up to the full energy of the particle and the directions of emission are within the angle $\sim\gamma^{-1}$, so it is only on average over many tests the particle can reach the center. In the limit of radius much larger than the classical electron radius, radiation consists mainly of multiplicity of "soft" (low energy) photons, emitted independently of each other. These might be described with a single Feynman diagram with many lines, infinite in the limit of low energy, where it amounts to the classical formula of radiation power. The relatively low probability of the emission of high energy photons makes their average radiation reaction much lower than that of soft photons. Therefore, to estimate the upper limit of radiation reaction we can account for only classical radiation reaction even on the small distances, where, strictly speaking, the classical field theory fails.

Were radiation absent, we could construct the whole regular lattice out of symmetric stars-cells, using for the descendent cells the particles slowed down to the initial velocity by

the output potential barrier of the star. Three such particles are enough for the basis to determine the other five in the cube, if the twin degeneration were absent (Fig. 2.7). The orientation of spin — along or opposite the particle velocity — makes it possible to select the needed orientation in the cube. However, even low in each individual cell radiation, not precluding symmetry detection inside it but steadily decreasing the kinetic energy of the particles, might ruin the regular lattice upon collecting the error in the initial velocity for the descendent cells, since even on the atomic range the quantity of minimal cells exceeds 10^5.

In order to overcome this difficulty, we have to improve the gauge procedure. As it was found, it ends up at the radius of the order of 10^{-16} cm. Over smaller radii we are free to introduce any new interaction, not precluding the gauge. In particular, this interaction might alter the charge of the particle. If this charge becomes zero, then the electromagnetic radiation would disappear just in the vicinity of the center, where it is the strongest, and its small remnant in the cell would not prevent the gauge even over the whole lattice range. Returning the charge to its initial state for the next cell, we anticipate the construction of the lattice to become possible though not practically but at least in principle.

The question of practical realization isn't that important as soon as it is only a limiting situation like the minimum size of the cell that is under consideration. We deal here just with the language reflecting CP statement. On a macroscopic and even on atomic scales the motion-to-motion gauge is quite practical, since radiation is ignorable there. In macroscopic measurements there is no need in the charge altering since the required quantity of cells is not that big, and it is thus possible to use the bodies passing the star center as a seed for next stars. In atomic scale gauge the charge might be altered in the processes of charge exchange and stripping (ionization).

However, in the limiting situation a separate particle with zero charge (*neutrino*) is needed to connect the stars. Since in the descendent star we detect the charged particles' symmetry anew, we need a process of recovering the charge. To this end, "blind" stars are introduced, consisting of neutrino and antineutrino, in which their annihilation creates electron/positron pairs to participate in the next charged star (Fig. 6.3). In the blind star, the neutral particles cannot be gauged with usual photons, but this is not necessary because the detection of the observed symmetry in the next charged star is sufficient to claim that in the intermediate blind star it was observed as well.

Minimum three trajectories are to be received from the ancestor stars to construct a basis of a descendent star. In so doing, we have yet to choose an appropriate trajectory out of the twins, otherwise no cubic star would arise at all. For this choice the spin contact scheme developed in Ch. 4 is in order. Indeed, any trajectory in the cube together with the three its neighbors define the full cube star. Even two of these three would be enough, provided the order of rounding them

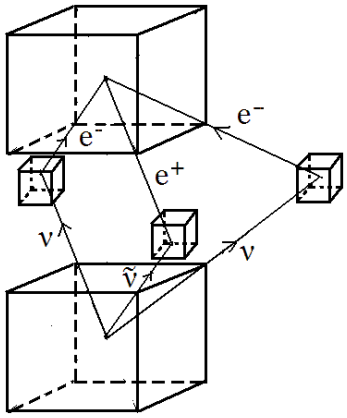

Fig. 6.3: The fragment of the lattice of cubic stars.

in turn is specified in addition to distinguish the third trajectory from its mirror in the basis decomposition. This order is just equivalent to selecting one of two spin projections on the direction of the particle motion toward the star center. However, for the definite arrangement of the descendant star out of its predecessor we have to translate the correct spin orientation through all the steps of this transition. For this to be possible the particles of the predecessor must already have a definite ("left" as it used to be said) orientation of its spin projection, i.e. to be left-polarized, and also their neutrinos must have the same spin to be kept unchanged in all transitions. Besides, the dynamical correspondence of the stars is required as well: The final velocity of the particle in the gauge process in the predecessor must turn in the initial velocity at the entrance of the descendant charged star.

To realize all these constructions, a new *weak* interaction turns the electron/positron in the neutrino/antineutrino and vice versa, so introducing a *doublet* of the weak interaction, since its role in the gauge needs just two charge states. This interaction must be short-ranged — of the order 10^{-16} cm for not to spoil the gauge in the star by altering its charge value. On the other hand, it acts within the same cubic symmetry, sharing this zone with the electromagnetic interaction, so giving rise to the united "electroweak" interaction. It is therefore natural to use for its formulation the same for all interactions structure of the wave equation, as substantiated above with the condition of preserving oscillations numbers ratios, although now with an additional "mass" member (Yukawa potential), providing a short range to the weak interaction. However, this potential, though being exponentionally small at a large distance as compared to electromagnetic interaction, still penetrates into the gauge region. For not to deteriorate the charge gauge, we must bound the upper limit of the coupling constant (its effective "charge") of the weak interaction as compared to the electric charge. On the one hand, the switching of charge and the dynamical effect of the weak interaction cannot be allowed to deform the gauge so heav-

ily as to produce an effect at the gauge minimal radius of the order of one oscillation. On the other hand, intensity of the weak interaction must be sufficiently high for the charge to be switched on a short distance. To meet all these conditions, the weak charge must be of the same order as the electron charge.*

Down to the radius in the weak interaction zone, at which the switch of the charge occurs, the charged particle (the electron, say) moves under both weak and electric forces. After the switch event only the weak force is active on the newborn neutrino, while the electric force becomes active again on the newborn electron after the opposite switch, now in the blind star. For the electrical Coulomb potential, the total dynamic effect of acceleration/deceleration is determined with the difference of the reciprocal values of the radii of direct and opposite switches. It seems that the required dynamics for matching the final velocity in the charged star to the initial velocity in the next charged star could be entrusted to the electrical potential alone. However, apart from the unnaturally specified difference of the average values of switching in both charged and blind stars, we would also have to neglect the much more probable process of the annihilation of the electron/positron pair into two photons. The ratio of the probabilities for annihilation and charge switch scales as γ^{-4}, since the cross-section for annihilation scales as γ^{-2}, whereas that for charge switching scales as γ^2 according to the general features of the wave-like equations. The involvement of the weak interaction not only in the charge switching but also in dynamics helps to solve both problems — to appropriately match the velocities of the electron and to suppress annihilation. For this to be possible, the weak interaction itself must be able to accelerate the electron in the charged star and accordingly decelerate the newborn electron in the blind star to obtain the necessary for the descendent star initial velocity.

However, given the initial and final points for the series of these switches and matches, the net dynamical effect of the weak Coulomb potential would be zero, and to actually involve the weak interaction, it is needed to turn it on and off somewhere, just like it was for the electrical potential. Turning a force on and off is equivalent to the appearance of an intermediate particle connecting these events. This particle appears when the weak potential turns off and disappears (decays) when it turns then on. For the small radius of the weak interaction, this particle will be heavy its mass being proportional to the argument of the exponent in the Yukawa potential. For the typical energies in the stars — γ less than 10^2 — this particle might be only a virtual one, and the correct translation of neutrino polarization requires it to have an integer spin. The minimal spin of this boson must be 1, since spin 0 cannot transport polarization. We are still free to choose the radius of the charge switch. This might be chosen under the condition for the electron to reach a maximum velocity at the switch still compatible with the required velocities matching to suppress the effect of its annihilation adverse for the gauge.

Different behavior of the left- and right-handed particles in the weak interaction is called its parity violation. Their dynamics remains similar in the weak interaction, provided the reverse of polarization is combined with that of the charge sense. Indeed, the opposite positions of electrons and positrons on the diagonals of the cube star correspond to their opposite polarizations, since it is the same interaction that turns the electron into its neutrino and the positron into its antineutrino.

For the short range of the weak interaction, the charge switch cross section is small and great many left-handed electrons fail to turn into neutrinos. They pass the center of the charged star, and loosing a part of their energy in radiation, however low, are not able to pass the exit potential barrier. They return to the star center changing polarization, so not being able to become neutrinos. Being reflected once again they now become able to turn into neutrinos. However, for this to be possible the opposite positrons must move quite similarly, and the probability of this event, equal to the product of their probabilities, is low: Typically their radiation losses differ, and they will not reach the weak zone at once. Even if this happens, their neutrinos will be belated with respect to those experienced normal transitions.

We have also to consider the destiny of right-polarized electrons (if these are present in the star). These produce no neutrinos, but being reflected by the exit barrier, come back to the center as left-polarized and together with analogous positrons might admixture false neutrinos to normal anti-neutrinos that have been created by positrons. We can essentially suppress this adverse process providing the weak interaction with an additional dynamical property to slow the right-polarized electrons down, so favoring their annihilation even with normal right-polarized primary positrons.

We are now in the position to complete the analysis estimating the weak interaction coupling constant in terms of the electron charge. Currently this value is specified with the "weak interaction angle" θ_w. The above considered transition conditions yield $\sin\theta_w \sim 0.5$, in agreement with the measured value.

The symmetrical cube star provides the ideal gauge of the electric charge standard by means of transporting its value along the regular lattice. However, in an individual star-cell it might be possible to "simulate" the symmetry, substituting some of the particles with "false" ones in such a way that photon oscillations counting would not "notice" the substitution. We could try to change, for instance, the mass of two particles positioned on the same diagonal of the cube. If these particles are sufficiently heavy, while their charges are the same as those of other particles, then the remaining electrons and

*Of course, all these conditions are to be understood in the probabilistic meaning of quantum mechanics.

positrons will influence their motion but weakly, and the false particles might reach the weak interaction zone at once with the electrons. The back influence of the false particles on the electrons might be small far from the center where the main part of the difference in the photon oscillations numbers between the tetrahedrons is being collected, were the symmetry broken.

For so heavy particles their acceleration, so also radiation, might be small, and contrary to the electron they might move from one star to the next not needing to cancel their charges. Perhaps, some such arrangements are able to support the symmetry test in the individual cell with the photon oscillations counting at some particular initial data. And provided this is possible, is it also possible to develop the whole regular lattice comprised of such cells or, at least, a considerable fragment of it?

To address this issue, we have to solve a rather complicated full system of the nonlinear equations of motion accounting also for radiation. We will limit ourselves to estimations with the following assumptions. First, we'll search for the false particles' trajectories disregarding their radiation and only afterwards calculating their radiated energy with the classical field theory formula that is correct for the small classical radius of a heavy particle. Second, the motion of the false particles being our main concern, we'll describe the effect of electrons on them with some averaged "background", checking afterwards the validity of this assumption varying this background within some reasonable limits. Third, we carry out the numerical solution under the same initial data for electrons as it was mentioned above for their original star. Using the final value of the particle's velocity at the center as an initial for the next cell, we'll follow its destiny in the lattice.

The full cube symmetry allows for two sub-symmetries. In the first sub-symmetry a diagonal is replaced with the false particle/antiparticle, while in the second sub-symmetry two diagonals are such. There are no other sub-symmetries in the cube, since the next replacement just brings us to the first, though with the star radius that exceeds the limiting value. In the first sub-symmetry, the trajectories of the false particles are still straight lines, whereas the remaining three electron/positron pairs move along curved trajectories, however, being identical — each one with respect to its own plane, these planes intersecting over the trajectory of the false particle. It turned out in the numerical calculation that at some value of the ratio of the false particles' mass to that of the electron, the final values of its radius and velocity in the cell, though differing from that of electrons, return to their initial values at the next cell, and this behavior repeats itself infinitely within the precision of the calculation (10^{-6}). For other mass values the disparity increases monotonously, and the construction of the regular lattice is impossible. In this equilibrium cycle, the mass of the false particle is close to that of the τ-meson. The radiation is therefore low, and it doesn't shift this value within the calculation precision.

In the second sub-symmetry, two pairs of false particles and two pairs of electrons move in their own planes: one for false particles, another for electrons. The planes intersect at right angles, and in each of them the particles move over differently curved trajectories, though in each plane they are symmetrical with respect to the center. The numerical calculation also demonstrates the existence of infinitely repeated cycles, though now consisting not of two but of four successive cells. The value of the related mass ratio in the equilibrium cycle was found roughly equal to the μ-meson mass. However, this time the calculation is by far not that reliable as for the first sub-symmetry, since for the curved trajectory and lighter μ-meson its radiation is no longer negligible. Therefore, now our result yields only the rough upper boundary of the meson mass' ratio to that of the electron. However, in view of strong differences in the ratios of the mesons' masses to that of the electron, these results look sufficiently reliable to explain the existence of the lepton families. Both mesons must possess their own neutrinos to avoid false intersections in long series of cells.

The six-dipole system of "roofs" completes the cube up to the richest Plato solid, i.e. having the maximum quantity of trajectories — the dodecahedron with its twenty vertices. These six dipoles are positioned under right angles to each other. Though not being a regular polyhedron, this system, if considered as a separate star, still possesses its own, independent of the cube equilibrium state — the common contact at the center. As distinct from a regular star, keeping its equilibrium upon the motion of the particles toward the center for any dependence of the interaction potential on distance, the six-dipole system keeps the equilibrium only if this potential increases as the distance squared. This property is not something specific to this system only. It holds for any sphere with equally charged particles if the sum of all their position vectors is zero. However in this case, it is important that after the removal of a regular structure — the cube — from an also regular structure — the dodecahedron — something capable of supporting equilibrium still remains. Of course, this potential must be attractive; otherwise the particles could not reach the center.

Such radial increasing potential has properties peculiar to the *strong* interaction, namely, confinement and asymptotic freedom. The first prevents the star comprising particles to leave the cell. These particles are called *quarks*. The second property suggests the trajectory of a body scattered in the vicinity of the star center to move freely experiencing no influence from the quarks, that is, to behave much like measuring trajectories.

Just as for the weak interaction, we have first of all to take care for this new force not to destroy the cube symmetry. The influence of the strong interaction on the cube particles is removed simply with the condition for leptons not to feel this force. However, the own arrangement of the six-dipole

system doesn't define as yet the full dodecahedron symmetry. First, we have to specify in a way the angular position of this system relative to the cube by means of photon oscillation counting. Second, the reciprocal angular positions of the quarks themselves isn't specified completely, because the equilibrium of their arrangement under the strong interaction would not be destroyed in deformations leaving the opposite quarks on their common diagonal upon its rotations.

The correct positions of the quarks as respective to the basic cube, so also among themselves, by means of oscillations numbers counting with the usual electromagnetic photons implies the quarks to be electrically charged. No distribution of this charge is able to completely remove their adverse influence on cube symmetry, and it is also impossible to remove the back influence of the leptons on the quarks. However, under some conditions these perturbations might be small, while the oscillations numbers counting might be carried outside the individual star, using the regularity of the full lattice (see below — Fig. 6.4 and the related discussion).

Let us try to arrange a regular lattice comprised of stars-dodecahedrons. We had already the lattice of cubes (Fig. 6.3). Now we are to complete it with the inter-cell transitions connecting the six-dipole systems. But this is forbidden with the quarks confinement by the radial increasing strong attractive potential. Therefore, this potential must weaken at a larger radius. Only then the united twelve quark system would be able to decay in six dipoles, since the distance between quarks in a dipole is less than that between each of them and any other quark in the system. With an appropriate choice of intensity and radial function of the attraction law, it is possible to reach the minimum energy quantum bound state of the dipole, while the bound states of its quarks with other quarks were impossible. The necessary suppressing of the strong force on large distances is interpreted as reciprocal compensation of the strong charge outside the dipole. Hence, the dipole must consist of quark and anti-quark just as it takes place on the diagonals of the twelve quarks system inside the dodecahedron. The bound state of quark and its anti-quark are called π-mesons. They are just these sub-systems of the full dipole system that are suggested to tie together the sequences of the stars in the total regular lattice (Fig. 6.4).

To reduce the influence of charged quarks on the own cube symmetry measurements, their merging in mesons is to be completed in the vicinity of the center where the disparity between the tetrahedrons cannot be collected as yet for one photon oscillation. The meson must be electrically neutral (π^0) in their disparity collection dominating zone. On the other hand, the strong interaction must not destroy the weak interaction. For this to be the case, the characteristic range of the strong interaction should be of the order of 10^{-15} cm.

Then the π^0-mesons are to be transformed in the auxiliary octahedrons to create the charged $\pi^{\pm}$-mesons for the symmetry to be checked on with the oscillating photons ratios counting at their intersection point. These ratios must be equal 1.

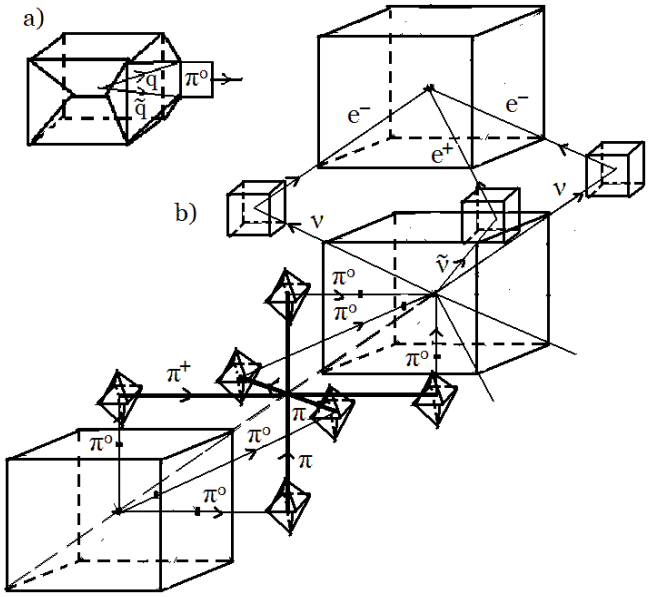

Fig. 6.4: a) Creation of the π-meson out of quark and anti-quark. b) Inter-cell transitions with π-mesons in the full regular lattice.

Besides this condition violation, the symmetry could also be destroyed if the triple contact decays into two simple, so that some of these ratios turn infinite. The total lengths of the three-leg meson trajectories could always be so chosen that in the next dodecahedron its center be reached by all the particles at once.

Again, it turns out that this task cannot be achieved exactly. Even in the neutral as a whole octahedron, there is no equilibrium distribution of its six charges. An approximate equilibrium, as specified by the condition of their missing the center at a distance less than that for the utmost short-range weak interaction, is possible only with the increase of the meson mass. In particular, it is just the weak interaction that is responsible for the charge exchange in quarks upon the transformation of the neutral mesons in the charged ones and vice versa. The estimation of the deviation of the π-meson from the center within the weak interaction zone yields for it mass the value of the order of 200 electron's mass.

In Fig. 6.4, the meson trajectories needed for the gauge of the strong interaction have not two legs as it was for the weak interaction but three legs, so allowing for three strong charge switches to constitute its triplet. In analogy with the human color vision these charge states are named as red, green and blue. The color exchange is realized in the triple contact of the charged π-mesons. In these terms, the slowing down of the increase of the strong potential at large distances might be interpreted as the reciprocal compensation of the colors, for example, that of the red quark and its anti-red partner bounded in the π-meson.

With the values of the strong interaction range and the π-meson's mass, it is possible to estimate the masses of the

quarks belonging to the lightest family, considering the ground state of the full six-dipole system near the star center and their bound states, mesons, at larger distances. In the intermediate range, the interaction potential must change in such a way that the correct contact scheme be observed.

Importantly, it is just the irregularity of the six-dipole system that makes the decay possible. It cannot happen in a regular star with its equal inter-ray distances. The possibility for the six-dipole system to decay into separate mesons determines the coupling constant value and the form of the strong potential that correspond to the first quantum level. With these data we can determine the force accelerating the quarks toward the center. This force has no transverse components, since the equilibrium is inert with respect to rotations of the oppositely positioned quark/anti-quark line around the center. On the opposite, for the electric field only the lateral component is important, while the longitudinal might be neglected as compared to the strong field. This gives the estimation of the range for "missing" the center by a quark, and its mass is determined by the condition, that this range does not exceed the weak interaction range. In the simplest model with the potential, increasing squarely to be replaced with rectangular "bag", this mass turns out to be about 10 times more than that of the electron — close to the known value.

However, the accelerated by the strong force electrically charged quarks must lose their energy by radiation under so high strong field acceleration. And then, quite similarly to what we met with for the cube, they would not be able to overcome the exit potential barrier with a noticeable probability, and the system of the dipoles would not be able to decay into separate mesons. However, the strong interaction must emit its own "photons" as well, and this radiation must be much more intensive than the electromagnetic radiation due to the higher value of the strong charge. Then, the radiation reaction would be accordingly strong, making the resulting acceleration small, so suppressing electromagnetic radiation, while the quanta of the strong interaction, gluons, must be deprived of leaving the strong interaction zone. Otherwise, this interaction could not have that short range, so violating the basic electric charge gauge. The gluons will thus return the energy to the quarks leaving the star in its decay.

The regular lattice with the dodecahedron symmetry exhausts the totality of possible interactions. In the last account, these interactions exist and are tied together solely by the procedure of the electric charge gauging, requiring compatibility with this gauge. The intrinsic for CP constructions of trajectories in terms of contacts define the interactions that we are able to recognize in Nature, to distinguish and comprehend what we should pay attention to, aiming in reliable predictions. Being designed in accord with CP, our experimental devices are capable to discern only these interactions. As for gravitation, actually it is not a force at all, but rather a general geometrical structure of force-free trajectories.

Part Two. What For?

> The answer to a question which philosophy fails to answer is this that the question should be asked in a different way.
>
> G. W. Hegel

Chapter 7. Repeatability

> Like ordinary knowledge, in dealing with things science is concerned only with the aspect of *repetition*. Though the whole be original, science will always manage to analyze it into elements or aspects which are approximately a reproduction of the past.
>
> H. Bergson, *Creative Evolution*

The demand for the universal repeatability of all constructions is fundamental in the Method. For this to be possible, the mental constructions contained in the theory must satisfy the condition of non-ambiguity. In the same way, the result of an experiment is regarded as being satisfactory only if it provides a non-ambiguous result, surviving the related checkon always and everywhere. To this end, in the very setup of the experiment, one has to ensure *pure* conditions, and it is the main concern of the experimenter to reach a result that is free from circumstances, if these are unaccounted for and/or brought about by particulars. The experiment is by no means to be confused with experience!

Structures considered in previous chapters are suitable only as a framework of the Method, as basics of the systematic approach. They do not provide predictions themselves, still requiring knowledge of external influences up to the final contact. They are formal, hence useless for immediate applications. This is a property of all mental — mathematical — schemes. For example, expansions in infinite functional series yield nothing, since to specify all the coefficients is the same as to specify the initial function. Only limited precision (for appropriate convergence of the series) makes sense, letting one to take into account only a finite number of the expansion members. For the same reason, true physics consists of particular cases, such as the field of a point charge, oscillations, collisions etc. In all such cases external interaction is given everywhere in advance. The Method is useful only as a general line of thought.

Unlike simple repeatability that is in the heart of any experience, some sufficiently artificial conditions on experiments related to basic constituents of the Method are to be fulfilled to ensure universal and non-ambiguous repeatability. Therefore, the very question, which is intended to be answered by the experiment, should be sufficiently primitive for the result of the study to be somehow used in real life. Even then, the restrictions on the setup coming from the non-ambiguity requirement are so heavy, that usually only thoroughly arranged set-ups, made up of diverse and not exactly repeatable elements of Nature, are capable to withstand them.

However meager, the set of satisfactory constructions is in the base of all our technology, just because of the possibility of unlimited repeating and combining simple and standard operations, each one being diligently verified in the related experiment.

It is of no wonder therefore (though having frequently been under discussion) that almost *any* mathematical device, though arising initially as pure mental exercises, happens to find further on some applications to the theories of the Method. This is because they are being developed under *the same* conditions of non-ambiguity. Just as the experiment depends on a priory conceptions (Einstein: "One cannot measure the velocity of light otherwise than having in advance a ready concept of velocity."), so also the theory rejects any ambiguity. Just as physics studies what "is", i.e. what do we wish and are capable to discern, so also mathematics studies what "could be", i.e. what we are capable of recognizing. It is the most important deduction from the first part of this book, that the demand for absolute repeatability results in this that the language of the Method is so meager, squeezing words down to terms, so that everything capable of being expressed with it, would necessarily be found in the stock of reality. It follows, that also the results of experiments should, in general, be conditioned by the very language of the Method, so being predicted in advance.

As Kant noticed ("*Prolegomena to Any Future Methaphysics*"): "Even the main proposition expounded throughout this section — that universal laws of nature can be distinctly known a priori — leads naturally to the proposition: that the highest legislation of nature must lie in ourselves, i.e., in our understanding, and we must not seek the universal laws of nature in nature by means of experience, but conversely must seek nature, as to its universal conformity to law, in the conditions of the possibility of experience, which lie in our sensibility and in our understanding. For how were it otherwise possible to know a priory these laws, as they are not rules of analytical cognition, but truly synthetic extensions of it?

Such a necessary agreement of the principles of possible experience with the laws of the possibility of nature, can only proceed from one of two reasons: either these laws are drawn from nature by means of experience, or conversely nature is derived from the laws of the possibility of experience in general, and is quite the same as the mere universal conformity to law of the latter. The former is self-contradictory, for the universal laws of nature can and must be known a priori (that is, independent of all experience), and be the foundation of all empirical use of the understanding; the latter alternative therefore alone remains."

However paradoxical this might seem, all the fundamental structures of the Method could be dreamt up, and it is not necessary to perform experiments for their checking on. Indeed, whenever we fail to reach non-ambiguous repeatability, we merely say with disapproval that this is not science, because it is not suitable for the expected applications. But one might clearly find in this an *upbringing* in the spirit of rejecting as inessential everything not governed by the Method, and then of believing that everything deserving attention will sometime be "explained" by science. The widespread opinion on the *objectivity* of the Method expresses its independence of a *particular* point of view. The latter, however, plays a decisive role in applications.

Let us imagine an uneven, rough surface. Suppose it is the relief of a place. A giant with his soles much larger than the highest mountains is interested in this relief only in respect to its friction, for not to slip; a dwarf is concerned with the nearest mountain as to how to climb up; a pilot should look at the highest peaks; while somebody suffering travel sickness pays attention to the periodicity of the road profile. Which of them sees the surface "genuinely"? It might be said that there is a point-wise description of the surface as a function of some variables, and everybody could draw from this function whatever he is interested in. But how should this function be found in practice? Every measurement has some finite precision. Now, will the giant measure every mountain, or will the dwarf look for the harmonics of the structure? Everyone deals with the same surface. Everyone reveals some its features from his point of view. These features would just not be found in either a different profile, so also with a different method of analysis. It is senseless therefore to claim that the surface "objectively" possesses some particular properties. It only *admits* them. In any observation, referring to peaks and depressions, a priory intention is present to look for them, to select them out of infinite variety of features and shadows. On the other hand, the surface is not something amorphous, such that any analysis, like a stencil, would find in it everything it is tuned for. Our surface is a *unique* thing, and as such, it is not bound to obey some general rules.

In the same sense, the World as a whole is unique and it is such as it is. It might be convenient, in theoretical constructions, to imagine different worlds, e.g., a class of them, containing our World as its element. However, there are infinitely many such schemes, and they might be of value only in as much as one can draw some conclusion about our unique World with his a priory accepted point of view. By the same argument, it is impossible to "objectively" regard the World as either or not changing. Taking once again the mentioned above surface for a model of the World, assume, for instance, that its profile is changing with time. How could we not only measure, but even to detect this changing? We cannot apply an external ruler, because there is nothing external to the World. A comparison of different parts of it requires time. Over this time both the World and the ruler might change.

How to distinguish between change and measurement? This depends on the measurement precision as *assumed in accord with a priory position*. The acceptable precision is defined by the implied application of the solution, but the very fact that in every problem some finite precision must be accepted is a principal property of the Method. The main *axiom*

of CP requires the meaning of the final contact to be recognized by the user *prior* to addressing the Method to receive its recommendations. The user has first to define as to what actually means for him the occurrence or non-occurrence of this contact. Recall that the Method is based just on axioms, and not on hypotheses. It contains *mental* schemes to provide non-ambiguous recommendations, if the user succeeds in according the basic concepts of the Method with his actual circumstances, just like a tool is being usually chosen for the purpose. The Method itself makes no assumptions on the "construction of Nature", and every time the user should be asked whether or not some axiom meets his purpose, and only then he might be provided with a solution to recommend him a way of action. It is in this respect only, that the putative cause-and-effect problem makes sense in the Method. The effect is something of interest of its own, while its cause might be of importance only insofar as it is capable of realizing just this effect. The effect is being defined by the user as something external to the problem, while its causes are being examined inside the problem and only in this respect. An effect might have different causes, as well as the cause might bring about different effects, but the cause-effect relationship of its own belongs to the Method.

In CP, as we have defined it in the very beginning, it is just a contact that is in question, which is a point in the contact space, rather than an "event in general" in the canonical version with its space-time given in advance. On the contrary, the contact space with its points and features, as we have found, is to be defined by the requirements of CP. One has to recognize what are *bodies*, participating in contacts. The main axiom consists just in that the user does know this, i.e. that he understands the results of the contact and develops a *definition* of body as something to take part in it. This is by no means possible in every situation: A cloud moves along its trajectory and it is expected to obscure the sun, but the cloud diffuses or thaws out. In many cases, even an experiment is needed just to learn whether or not some axiom is of interest in the given situation. The wide realm of CP applications stems, as already mentioned, from the warranty of its predictions, as soon as the problem might be reduced to CP.

However in an absolute sense, neither bodies nor their contacts — nothing like these do actually exist. Each time one has to isolate something from the participation of everything everywhere and from the interaction of everything with everything. All bodies emit various fields penetrating all others; the body might change in a way, still being considered the same in the problem; also the motion itself might have different meanings. The World is a whole, and it is the only real "thing", that possesses its absolute and perfect reality. The picking up of a *particular* body out of the World is only being possible by means of ignoring its infinite "inessential" connections and influences. Upon denoting various mountains with the same word "mountain" one implies (potentially) a definite action in respect to something so denoted in spite of various dissimilarities between particular mountains. This doesn't mean that the mountain doesn't exist objectively, independently of its perception by the subject. This does mean that without his intention to use in a way this term, he might simply not notice this mountain, it might be "of no importance" for him. Three hairs on my head — it is too few, whereas three hairs in my soup — it is too many. What is important and what is not, even if there is a many-order difference of a value, each time depends on a particular problem. How small should be the mountain for not to appear as a mountain? The wholeness of the World consists exactly in the absence of a universal measure for this importance.

The concept of contact is the only one "entering" device that is offered in CP to the user as a possible tool to reach his purpose; it is an *operational* concept, an equivalent of the *bit of information* on whether or not the contact exists. In this sense, the contact is always point-wise, even though the bodies taking part in it are extended. The contact is not something discovered in Nature, but only a recommendation to approach the problem: What are the means the attention should first be paid to? Like all other constructions in CP, contacts are recommended for trying to single them out in order to obtain reliable predictions with the general methods of CP.

If the scheme is devoid of the top speed, then time is to be defined with the use of an external independent device — the clock. In the Einstein's relativity theory this necessity is mitigated just due to the existence of an upper limit for velocity (not necessarily being the same in different points). However, clocks are not eliminated completely, but they are only being correlated with the top speed signal. The contact schemes used above are "relativistic" from the outset, according to the very logic of the operations, so there was no need in either rods or clocks (even as an affine parameter). In the canonical version, the related to these schemes measuring procedures give rise to the existence of a *universal constant*, i.e. a top speed (light). In the same way, the minimum disturbance of motion with measurements implies another universal constant, that of Planck. This similarity in the structures of the theory of relativity and the quantum theory has been constantly appreciated by Bohr. In general, any universal constant springs up from the related measurement procedure either on its extremes or on its discontinuities. So in CP, the discreteness of the electron charge and mass results from the difference in just one photon oscillation required for the difference in these oscillations numbers to detect the cube star symmetry.

The putative successes of the technology based on the Method brought about the perception of its almighty, its capability to "explain" everything, to answer all "reasonable" questions. Complex systems, as constructed by means of simple operations of the Method, demonstrate its efficacy. Along with this, however, a question springs up about utmost capabilities of the Method, since the pressure of successes along a

definite line of thought has a tendency, as a rule, to suppress the development of alternative solutions, and all the more, the asking of new unusual questions.

As an example, we consider artificial construction of a living creature. This process seems to be available at some level of technology. At least, no certain prohibitions seem to exist there, since the coherence of the quantum state of a whole system decreases with its complexity. Therefore, the individual process might be multiply repeated with arbitrary precision using the operations of the Method, so obtaining identical copies. We emphasize, that they are not just organisms with some generally described "desirable" properties that are in question, but completely identical copies uniquely reacting on everything. Is it possible to regard them as living indeed? How, for instance, will they communicate among themselves? Whatever difference will be found, this would mean the interference of something beyond the Method. And what if a further difference in the external circumstances destroys their identity, should the Method account for these as well? Then the whole World should be under control.

The process of creating a complex system inside the Method, i.e. by means of simple combining its operations, is quite arbitrary, and it is only by chance that it can produce a result to be of interest for the user. As a rule, some preliminary description of the desirable in *external* to the Method terms is needed, so that only afterward one can correspond (if possible) the definitions and structures of the Method to this particular problem. To this end, the thought scheme itself as a scheme of the transition from one state to another is important in its own right quite independently of what actually is being considered to be a state, provided it is meaningful for the given application to pick up a situation suitable for the concept of the final contact to be introduced, which might correspond to the externally defined aim. As well as a collision, it is also a transition from one state of motion to another or something else that might be in question. The thought scheme of its own might be the same, while its content, that is, the identification of required elements in the World, each time presents a practical problem. CP schemes are recommended for applications in virtue of their definite predictions.

In the description of the Method in this book, we restrict ourselves to only one of existing ways to realize repeatability, considering some basic contact schemes. These belong to that part o the Method that is called *physics*. It consists, as we have seen, in applications of CP in various situations of interest for practice. In order to get an idea about the place of physics in the general framework of the Method, it is useful to delineate, however superficially, some other means to reach a complete or partial repeatability, just for not to make impression that physics exhausts all the content of the Method. It is not always necessary to avoid contacts with a cobra; one might develop insensitivity to its poison or tame the snake.

The tendency to represent a practical situation as a combination of simpler elements, of which it "is comprised", is not a feature of only physics with its decomposition into separate contacts. Repeatability is reached, for instance, in written word, in which the decomposition elements are signs, and not at all the atoms these are being comprised of. A poem as written with different typing or spoken aloud is the same poem, though its physical realization is quite different.

In contrast to CP, in which they seek to distinguish the objects of the study according to minimum possible information, in other applications of the Method just the opposite is useful, viz., the detection of a great many fine details. For the examples, one might think of the methods in zoology, archeology or art.

Repeatability is required to make predictions basing on past experience. The sharply specified method using universally applicable elementary structures of CP implies a *strictly identical* comprehension of all its operations by all users. However, in practice, information referring to a particular person is of no less importance. Such are messages in tribe languages with their hints and reticence, not to be understood by foreigners. Such are also the items of art, as differently perceived by different people and intentionally referring to individual responses. These evident for all examples we have presented here exclusively for "to show CP its place" in the Method.

Chapter 8. Light of expired stars

> **Wagner**. Excuse me! But it is a great delight
> To enter in the spirit of the ages and to see
> How once a sage before us thought and then how we
> Have brought things on at last to such a splendid height.
>
> I. W. Goethe, *Faust*

Specifically orientated extractions (as selected according to the author's preference) from some works this chapter is devoted to, have the only purpose to present the examples of the line of thought from the past that has initiated the analysis developed in this book. The author hopes that internal logic of the above discussion is convincing of its own. However, any approach that is different from the canon creates an impression of unexpectedness, the absence of predecessors. It often occurs, however, that their particular ideas, even though of little appreciation in the following generations, create nevertheless a general intellectual air, which influences the very way of studies. To reveal this support from the past, though perhaps only indirectly, it is useful not just to appreciate the contribution of the predecessors, but also to connect the Method, narrowly specialized, as we have seen, with the ideas and trends beyond its limits and in other realms of knowledge.

The most popular question that was a demarcation between philosophers' trends ever since the ancients, concerning the materialistic or idealistic perception of the World as to "what is the first, and what is the second", refers actually not to the World of its own, but rather to *things*, comprising it in a

"self-evident way". This self-evidence has been considerably shattered with the introduction of field theories in XIX century and especially of the quantum theory in XX. Unlimitedly spreading, even in bound states, wave functions form a united system upon interlacing. However many orders of magnitude these functions fall off, the condition of independence of the objects of each other completely depends on a particular problem, and no general criterion does exist. Therefore, the very subdivision of the whole World into separate, sharply isolated from each other entities, i.e. "things", that has been perceived over centuries as an unquestionable fact, became questioned. Though the roots of these problems might be followed already since ancient authors, we deem that particularly clearly they have been stated by Kant in his *Critique of Pure Reason*. We shall begin with this work, widely using (by necessity, lengthy) citations and each time comparing these with CP.

In contrast to the tradition, dating back to Plato and common to Plotinus, Spinoza, Leibniz, Hegel and many others, of "external" (though some own for each philosopher) glance at the "self", all the Kant's analysis is being carried out "on behalf of the self". In this respect, he follows the philosophic credo of Socrates, and, in turn, finds his successors, explicitly or implicitly, in the works of existentialists.

People sharing the external approach are mostly interested with general "systems of universe", as being given from scratch. They consider it (most) important to *define*, in the framework of a suggested general scheme, what is self and his destiny, what is life, what is its "meaning" and so on. On the contrary, those supporting the existentialism hold a paradigm "existence precedes essence". Contrary to Descartes, they feel no need to prove the very fact of the existence of self, even with his "doubtlessness of doubt". The existence of self is being taken as an initially given fact, requiring no further examination, and it is just from this point of view that all the other world should be perceived (including all other persons as well as "external" perceptions of the self: "my hand"), if only for this reason, that for the already existing self there is nobody to perceive this world instead of him, and, most of all, there is no need to. And the essence, i.e. "properties" of the self, as revealed by the self or somebody else and possibly changing, though not destroying its self-identification, is secondary. This approach belongs to the founded by Kant tradition, as confirmed, in particular, by a true, no doubt, existentialist Heidegger in his *Kant and the problem of metaphysics*.

Kant begins with the observation: "But, though all our knowledge begins with experience, it by no means follows that all arises out of experience. For, on the contrary, it is quite possible that our empirical knowledge is a compound of that which we receive through impressions, and that which the faculty of cognition supplies from itself..."

But it is just this duality of knowledge that is in the origin of the Method, as it has been presented above. CP is nothing else as a basing on pure logical procedures a priory scheme, the advantage of which is in the guarantee of its prediction, so long as the correspondence between the elements of the scheme and the real situation has been carried out reasonably. This is quite opposite to the approach in "phenomenology" (Husserl and his successors), in which it is proposed to choose first some actual situation, for example, some object, and further to remove, in thought, its "inessential" features, striving to extract its "ideal meaning". Such a procedure is impossible in the Method, because there is no way even to single out a particular object from the whole Nature without some a priory accepted operations that express intentions of the self.

Kant first of all subdivides knowledge on "empirical", that is, supplied by senses, and "a priory", not depending on experience. The latter might, however, turn out to result from experience (perhaps, unintentionally or partially), or from much earlier or a more general meaning. He calls some knowledge "pure", if it depends on *no experience at all*. As an example of a priory though not a pure proposition, Kant presents: "any change has its cause", though noticing that the concept "change" might be drawn from experience only.

According to the Kant's classification, a "synthetic" judgment connects things or phenomena of different kinds, as distinct from "analytical" judgments, which follow immediately from the given definitions. Thus (according to Kant's opinion) the judgment "all bodies possess extension" is analytical, whereas "all bodies possess weight" is synthetic, because in the first case the negation of the predicate ("possess extension") leads to a logically impossible judgment, while it is not so in the second. Kant considers extension as implied in the very idea of body, whereas weight is its property not contained in its definition and known only from experience. In CP, the first judgment is synthetic as well. For bodies, the sizes of which are small as compared to the distance between them, their own extension is of no importance; hence, it is not contained in the idea of "body", considered point-wise. Though in CP approach extension is not primal, but rather coming as a scheme brought into existence via point-wise — by definition — contacts, the Kant's classification of judgments as either analytical or synthetic is fundamental for the Method in general and for CP in particular.

The main question of Kant is how possible "synthetic a priory judgments", arising in mind still before sense data are presented to it. His problem roots in the impossibility of matching arbitrary "fantasies" of conscience to practical circumstances. In his XVIII century, Kant still held the traditional conception of the clear separation of "things". This conception was not shattered as yet by the ideas, due to Newton, on the gravity as connecting bodies. Newton, as well as his successors before Maxwell, never considered field an independent substance, but rather a mere property of things like color or smell.

Independently of being perceived by the subject, existing things by only some of their features acting on senses to

form impressions, are being framed in the mind as intuitions, while something is always left unperceived, hidden. And it is not known how this "something" will come into play under comparing representations a priory with reality. Convinced in this that things are "objectively" separated, Kant pays nevertheless attention to the fact that the notions of reason cannot base on abstractions from something belonging to things themselves, since the senses do not perceive everything (so, the eye doesn't see ultraviolet component of light), and there is a censorship by the conscience of their data for its own need. Kant therefore distinguishes a "thing-in-itself" from a "thing-for-us", actually operating solely with the latter.

In this respect, there arises the problem of a priory possibility of the cause-effect connection, as had been stated by Hume and incited Kant to think on all these problems (according to his own confession). For the Method, as presented here, things-in-itself simply *do not exist*, because the very process of singling a thing out of the World (the only really existing thing-in-itself) depends on him who perceives, and he always do this *purposefully* ("intuits"). Then, the cause-effect relationship means nothing else than the declaration of the initial position of the self. Being interested in the event-effect, he *searches for* its events-causes, filtering away as deserving no attention everything that doesn't bring about the event of his interest. This is his perception of the World in this actual situation. The incompleteness of perception is, in effect, the same thing as the uncertainty of singling out, whereas an attempt of absolutely perfect perception implies the account of all connections, however small, that is, the involvement of the whole World in any phenomenon to be considered. "The bodies we perceive are, so to speak, cut out of the stuff of nature by our *perception*, and the scissors follow, in some way, the marking of lines along which *action* might be taken." (Bergson, *Creative Evolution*). The ten space-time conservation laws, as expressing the condition of repeatability "always and everywhere", are valid for *closed* systems (or, in virtue of the Noether's theorem, for those with some symmetry of the external field). But there are no closed systems in the World besides the same World. Each time, the cutting out an approximately closed system is performed according to a particular problem statement.

In order to prove the possibility of synthetic judgments a priory, as for any proof of existence, at least one actual example is needed. In *Critique of Pure Reason*, Kant considers the concepts of space and time as such an example:

"(a) Space does not represent any property of objects as things in themselves, nor does it represent them in their relations to each other; in other words, space does not represent to us any determination of objects such as attaches to the objects themselves, and would remain, even though all subjective conditions of the intuition were abstracted. For neither absolute nor relative determinations of objects can be intuited prior to the existence of the things to which they belong, and therefore not a priori.

(b) Space is nothing else than the form of all phenomena of the external sense, that is, the subjective condition of the sensibility, under which alone external intuition is possible. Now, because the receptivity or capacity of the subject to be affected by objects necessarily antecedes all intuitions of these objects, it is easily understood how the form of all phenomena can be given in the mind previous to all actual perceptions, therefore a priori, and how it, as a pure intuition, in which all objects must be determined, can contain principles of the relations of these objects prior to all experience. It is therefore from the human point of view only that we can speak of space, extended objects, etc."

("Form" means here a factor to organize intuitions. "That which in the phenomenon corresponds to the sensation, I term its matter; but that which effects that the content of the phenomenon can be arranged under certain relations, I call its form.")

And further: "But propositions of this kind cannot be empirical judgments, nor conclusions from them. Now, how can an external intuition anterior to objects themselves, and in which our conception of objects can be determined a priori, exist in the human mind? Obviously not otherwise than insofar as it has its seat in the subject only, as the formal capacity of the subject's being affected by objects, and thereby of obtaining immediate representation, that is, intuition; consequently, only as the form of the external sense in general."*

However, in these theses Kant means not a general and uncertain "philosophical" idea of space, but the quite particular to be used in physics: "For geometrical principles are always apodeictic, that is, united with the consciousness of their necessity, as: "Space has only three dimensions." Considering straight lines in this space, Kant stresses the possibility of their unlimited continuation, and, as an example of synthetic judgment, points out that the connecting two points straight segment is also the shortest, — the property, which is not directly contained in some probably implied by him definition of the straight line. Of all this Kant never asked why it is just so.†

It is only in the framework of CP the structure of physical space is developed out of its *future-owned* sub-structures and is *substantiated* by the very statement of the problem, but for

*Denying the reproaches in idealism, Kant expounds his position in *Prolegomena to any Future Methaphysics*: "My doctrine of the ideality of space and of time, therefore, far from reducing the whole sensible world to mere illusion, is the only means of securing the application of one of the most important cognitions (that which mathematics propounds a priori) to actual objects, and of preventing its being regarded as mere illusion. For without this observation it would be quite impossible to make out whether the intuitions of space and time, which we borrow from no experience, and which yet lie in our representation a priori, are not mere phantasms of our brain, to which objects do not correspond, at least not adequately, and consequently, whether we have been able to show its unquestionable validity with regard to all the objects of the sensible world just because they are mere appearances."

†One should recall, but then, that at Kant's time nobody operated with various extensions and generalizations of the concept of space, such as Riemann, symplectical, multi-dimensional and other spaces.

this to be possible, the question "Why?" is to be replaced by "What for?" in the sense of the already cited Hegel's aphorism.

Going over to the concept of time, we point out that unlike his reasoning on space, Kant implies not the physical time but rather a mere sequence of events. As a matter of fact, although time is the main concept of the Method by directly expressing repeatability, i.e. the use of past experience, but it doesn't refer to CP alone, like that of space, hence, its definition with only contact schemes might be doubtful. Everybody knows that sometimes a year is too short, while sometimes a minute is too long.

Since CP is only one particular way to reach repeatability, not exhausting a general problem of the transition to a concrete state, whatever it means in the case and however it is extracted as such from the World, no a priory measure can be there, which fits any transition, any *change* whatsoever. However, if we mean the problem for the Method, i.e. using the past to draw recommendations concerning the future, one needs to define the conditions of the very possibility to *pre*dict. In this respect, the past is something that in no way could be influenced upon, and therefore nothing in the past might be a consequence of the present or the future. Of course, in a particular situation some aim in the future might be non-reachable as well, but this is a separate problem. Also in CP, from the statement that there is no final contact, it is only the warranty for the effect of interest for the user, as it is meant in his problem, not to occur that follows. Therefore, only "affine", devoid of measure time sequence has a general meaning of warranty, while a particular way to realize this measure by means of photon oscillations is only suitable when the problem might be reduced to CP.

Only then the constructions of mechanics begin to function. According to Newton: "And if the meaning of words is to be determined by their use, then by names time, space, place, and motion, their sensible measures are properly to be understood; and the expression will be unusual, and purely mathematical, if the measured quantities themselves are meant." (I. Newton, *Mathematical principles of Natural Philosophy*.)

Also in the general structure of the Method, beyond CP, no universal measure corresponding to "change in general" exists there; for any perception whatsoever, time "stays still", if *nothing* changes. In the *Critique*, the concept of time is given as a condition of the internal perception by the self as of him, so also of the world: "Time, no doubt, is something real, that is, it is the real form of our internal intuition. It therefore has subjective reality, in reference to our internal experience, that is, I have really the representation of time and of my determinations therein. Time, therefore, is not to be regarded as an object, but as the mode of representation of myself as an object. But if I could intuit myself, or be intuited by another being, without this condition of sensibility, then those very determinations which we now represent to ourselves as changes, would present to us a knowledge, in which the representation of time, and consequently of change, would not appear. The empirical reality of time, therefore, remains, as the condition of all our experience. But absolute reality, according to what has been said above, cannot be trusted to it. Time is nothing but the form of our internal intuition. If we take away from it the special condition of our sensibility, the conception of time also vanishes; and it inheres not in the objects themselves, but solely in the subject (or mind) which intuits them."

In the Method, this general idea of time sequence meets the recommendation expected by the user for his actions, provided his aim has been recognized as a distinctly fixed final state. We repeat that all the work for the recognition of this aim he must carry out *before* addressing the Method, and this is by far not always easy. Even the experience, if reduced to words or some other conserved or reproducible cognition, mostly possesses no universal content similarly understood by all. As a rule, a message brings about different response in others or even in the self, if it comes later. Therefore, only a minor part of our experience belongs to the Method. An important role of the Method in technologies, their results being so highly valued, provides it with common appreciation and, by the way, also with trust and the feeling of "Truth". Yet, it is only the accumulated culture as a whole, not oversimplifying life that dramatically, conveys a more *precise* description of life than the Method in general, and all the more than CP do, upon encompassing also "uncertain details".

Chapter 9. From scratch. Uniqueness and repeatability

> . . . yet two times two makes four — it is not a life at all, gentlemen, but is the beginning of death.
>
> F. M. Dostoevsky, *Notes from the Underground*

> If we assume that human life might be governed by reason, then the very possibility of life would be annihilated.
>
> L. N. Tolstoy, *War and Peace*

Yet, why do we so much insist on repeatability, while no absolutely repeatable situations are there in real life? Moreover, life, which is the most valuable for the self, is absolutely unrepeatable, unique. Then, what is this Method important for? Each time, the use of the Method in a particular situation implies the disregard as *inessential* of infinitely variegated connections of everything with everything. Is it ever and within what limits possible to approximate the unrepeatable with repeatable?

According to the accepted in this book rules, we have no right to involve some new prejudices from outside in order to find the answer to any question, but only to proceed within the framework of the Method in the search of a solution, if exists, on its own limits of applications. To this end, let us analyze a particular example. Generally accepted opinions deem all

substances to be comprised of atoms. Leaving aside technical complications, assume it be possible to take proper sorts of atoms in proper quantities, put them in needed positions, provide them with needed velocities, and so to obtain a man, and not a "man in general", but the copy of a particular, actually existing "I". The Method, in principle, allows for this.*

We have to ensure somehow that the construction is satisfactory. Who will judge? Other people? But often they are confusing even twins. Their judgment depends on their own state, which should be "objectively" examined as well. So, steadily proceeding from one examination to another, all the World will be involved in the judgment. It is only possible to close this infinite chain of examinations by means of asking the same "I" to judge. Of course, "I" is the subject to various external and internal influences as well, but we introduce "responsibility" for his judgment. If "I" considers the copy perfect, he has to agree that nothing would change *for him*, if he were eliminated, while the copy remains alive instead. If the perception by "I" of his uniqueness doesn't allow for exact copy, the answer is already here: Absolute repeatability with the means of the Method is impossible. In the opposite case, we continue our study. Let us weaken the precision of copying. "I" with his hairs cut is not exactly the same "I". Upon continuing (at least in the course of thought experiment) to alter "I" in various respects, using plastic surgery, transplantation of organs (including those of his brain), we repeat our question at each step. If even after such a horrible procedure, "I" still insists on his identity, we would try in addition (or instead) to persuade him to recognize his transformation into something different. On the other hand, "I" as he is now and that in his childhood is being considered by "I" as the same, though, according to some foreign judgment, the copy might seem closer to the now-existing "I".

The only purpose of this offence of the common sense consists in an attempt to reach the verge of the Method, i.e. to shake the uniqueness of "I". And then the question is still here as to what is the value of the Method for so stubborn "I". If we fail, then we have to recognize the existence of something that cannot be constructed within the Method. If however, under some conditions, "I" agrees that some copy is satisfactory to replace him, then he is immortal, and various possibilities for fiction writers spring up. For instance, it becomes easy and comfortable to trip at the velocity of light: Suffice it to transmit by radio the message to make the copy. Medicine becomes superfluous, since the "I"-personality might be simply "rewritten" on some new body. Now, just like in the physics of elementary particles, there are no requirements of uninterrupted following, and "the similar" is identical to "the same".

If, however, the uniqueness, non-repeatability of "I" has been found, then the question of the value of the Method for him is still here in its entirety. So, time and again, what is it that is universally repeatable and still so important, that all everlasting over centuries efforts to develop the Method would be justified? It is only *death*, as being understood similarly by all as something absolutely final that might be the only candidate for this role. Of course, somebody might deny this position and insist on the non-finality of death in a way. It is impossible to claim an absolute judgment with respect to something not to be tested, but we may argue that for somebody *really* convinced in this, the Method is of no need and of no interest, that is, this book is not for him. Moreover, the real content of the Method refers *exclusively* to the problem of death, notwithstanding the seeming variety of applications.

Substituting the direct mentioning of death with the idea of *aim* in less important circumstances brings no important changes in this statement. Otherwise, the "scientific" communication by Mr. Pickwik, "Speculations on the Source of the Hampstead Ponds, with some Observations on the Theory of Tittlebats", would be of no less importance for science than the theory of relativity. Ultimately, the *significance* of the problem always relates solely to death, and the Method is able to say nothing else about the particular life, infinitely variegated and non-repeatable. According to Plato (*Phaedo*), Socrates emphasized: "For I deem that the true disciple of philosophy is likely to be misunderstood by other men; they do not perceive that he is ever pursuing death and dying..." The same idea, though in a broader context, has been expressed by Pasternak (*Doctor Zhivago*): "...art always, without interruptions, is occupied with just two things. It persistently thinks of death and persistently creates life by way of this." This is particularly true in the limited framework of the Method.

Death became a matter of main concern for everything alive because of this simple cause, that everything that had not been striving to survive, and even those, that had not been striving sufficiently hard, died out long ago in the process of evolution. Only those survived, that had been striving *very* hard. Extremities and all other organs have been formed just to protect the creature from death, but while developing its paw in the course of evolution, the creature had to protect the paw itself. The senses serve to protect vulnerable life. However, what a Method would elaborate an "I", which is somehow separated from the world, e.g., being blind, deaf or else? In essence, it must be the same Method, because, even lacking some means for his action, he still pursues the same purpose to protect his life, however difficult would it be for him to apply the constructions of the Method to his own practice. However, even the most primitive organisms possess some means for their orientation. Imperfect, as compared to dogs, hearing and smell of the human, though restricting his reactions, never obscure for him this main purpose of every living thing. While not suitable to describe the life completely, the Method might well be used also in situations not relating to

*In the sequel, "I" should be considered as a generic name for the subject, the problem of the Method is being stated on behalf of. As it was already mentioned, some results of the method are available for any living creature; also the "user" might be conceived as a community, provided its individuals take care solely of its destiny.

the life-and-death problem immediately, also those far from it and often even quite trifling. So, for a professional artillerist, it might be more convenient to fire sparrows with a gun than with a catapult. However, the line of thought, as developed for the life-and-death problem solving, forms a standard approach to any problem in sight, however unintentional. Just these problems constitute the content of the Method. A part of the Method belonging to the realm of physics is of interest for the user inasmuch as it is implied that the occurrence or non-occurrence of the final contact is, in a way, being connected with death, perhaps, only supposedly or probabilistically.

According to our position, we should first of all examine the firmness of the main ideas of the Method on the limits of its application. For the living creature, the principal vulnerability of the Method as applied to the life-and-death problem is in this that eventually death is unavoidable. As a matter of fact, to decide on whether or not to apply the Method, one has first to recognize the actual problem. Only *after* this step all the sophisticated machinery could come into action, including the arrangement of the system of standard bodies, sufficient to detect all the relevant measuring contacts; defining with them the space-time relations; constructing one more set of test bodies, reacting to the external forces that might effect the final contact occurrence and so on. Particularly and importantly, relevant order relations are to be established in accord with the demanded by the supposed user concrete recommendations for his actions *already before the formulation of the problem*. But as soon as death, in general, is not avoidable, and there is no a priory conception of time as yet, the task cannot be solved. What should "I" act for, if he dies anyway? Intuitively, the absurd artificiality of the so stated problem is doubtless. However, for the Method to be convincing a "proof of reason" (Pushkin) is necessary. It is clear but then, that it is one thing — death now, and quite, quite another — sometime afterwards, however small the time gap be according to any measure for it, whatsoever. Even in his last second, the hare still hopes. However, what if no time relations are necessary at all, but they are rather a mere habit, and one might get used to think somehow else? In order to overcome this difficulty, an *external* or "*mathematical*" time had been introduced, as opposed both to the measured and to the personally perceived time. This is clearly expressed in the Newton's *Mathematical Principles of Natural Philosophy*: "And if the meaning of words is to be determined by their use, then by names time, space, place, and motion, their sensible measures are properly to be understood; and the expression will be unusual, and purely mathematical, if the measured quantities themselves are meant." In the same sense, the personal I-time is replaced, in the Einstein's *The Meaning of Relativity*, with the *Einstein's simultaneity* — a means to synchronize clock readings at different points with an appropriate exchange of light signals.

However, the measurable quantitatively time in canonical version by no means depreciates the much more general and important individual I-time, which might progress quicker or slower according to personal perceptions. This time belongs to such concepts as past, now, and future without obligatory connection to seconds or years. It is used in life and communications much broader, and it is by no means less essential than time measured with clocks. Moreover, this perception is peculiar and necessary not only for humans but to any living creature.

For both Newton and Einstein clocks remain the basis of the whole system, though checked, in a way, by motion. But as we found, the existence of top speed makes it possible to dispense with clocks for any particular problem in physics. Counting photon oscillations serves well for all these problems. And beyond physics, in different fields, there is no measure of time at all. Phases of various processes are sometimes tied to astronomical or some other conditions, but never to straight mathematical time. Nature looks then in science merely as a set of particular cases united rather by a common approach than by a single structure equipped with its own universal laws, the "World". Then the very concept of universal inexorably flowing time diffuses, being devoid of its traditional definiteness. It turns out that time does not flow on its own, but, conversely, it depends on everything. And then some very unusual questions might spring up, which touch our deepest ideas, even those of life and death. So, the principal tool in our previous analysis — infinite sequences of photon oscillations, analogous to the Zeno sequences, — brings about a doubt in the habitual understanding of the life duration.*

In respect to the Method, the fundamental difference between the "not its own" problem of the unavoidable death in general and the death possibly to occur with the final contact in CP, is quite analogous to the mathematical notion of *compactness* as the necessary presence of an exact limit. Hamlet-like meditations on the unavoidable death are always in the focus of philosophy and art. In its finality, it is just as unperceivable as the usual falling asleep, according as to philosophical considerations ("Thus, death exists neither for the living nor for the dead, since for the former it doesn't exist in itself, while the latter doesn't exist for it himself." — Epicurus, *The Letter to Menoeceus*), so also to the literature insights ("He was looking for his past habitual fear of death and found none. Where is it? What death? There was no fear at all, because no death was there. There was light instead of death... For him all this occurred in an instant, and the meaning of this instant didn't change any further. However for the witnesses, his agony elapsed two hours more." (Tolstoy, *The Death of Ivan Ilyich*). It is just the threat of death perceived as *an incentive to action*, that gives the Method its meaning for "I" at any moment of his realized life. On his way to the scaffold, the condemned was thinking: "It is not now as yet, one

*According to an aphorism, life is measured not by the number of your breathes but rather by the number of your breathes been taken away.

more turn ahead, and a long street after it...” Not being felt by “I” himself, his death is perceived by other people as the disappearing of the particular person. The accumulated common experience yields the doubtless statement: “Everybody will die”. But the uniqueness of “I” is incompatible with this “everybody”. The fact that all people have died until now, doesn't prove that “I” will die too, because *he is not the same as all others*, and he feels his uniqueness personally. Should “I” deem himself identical to all others in this most important respect, he would be immortal just for this reason, being replaceable by somebody identical.

> “Not eternal for times,
> I'm deathless for myself:
> Perhaps, just to imagination
> Their threat has anything to say,
> I own the moment, and it may
> Enjoy me in the same relation.”
> (Baratynsky, *Finland*)

We conclude that, apart from philosophizing, it is only compact life-and-death problem that is actually of interest for “I”. It might be said, that the Method deals only with death, but only within the limits of life. However in the Method itself, there are at least hints of its limiting situations. In terms of CP, a Zeno sequence formed by photon oscillations should be completed with the single point determined by the sequence of its own, though *not belonging* to it. In this respect, a segment of the straight line is not equivalent to this line, whereas without its end points it is. And this statement remains true upon any its one-to-one order preserving mappings into some other straight line. Therefore, the actual meaning of the wise Zeno's paradox is in this, that initially, prior to introducing auxiliary, adapted for a particular CP external measure, it is just the sequence of discretely perceived events that is a primary, because the final contact is such. I-time “stays still” if nothing *interesting* occurs. In CP schemes, the sequence of contacts may well be infinite, and *according to this measure* Achilles will not overtake the tortoise *indeed*, and this fact in no way prevents practical applications of the Method.

As we have seen in the first part, in applications of the Method something that might occur if the final contact does occur, is always supposed to be known to the user in advance. Only when this became clear, might he ask the Method for recommendations. Otherwise, he would necessarily be asked time and again: “What do you want indeed?” However, what is good in the near future might not be that good later on. It is commonly known that the technologies, based on the Method and providing the impressive technical progress, hide many nuisances of their own. Aside from various dangers, stemming from quarrels of people and their communities, we point out only these that root in good intentions, stimulating the development of the Method.

The mentioned above uncertainty in the self-identification of “I”, renders some allowed in the Method efforts to prolong the life of “I” a quite practical problem. The methods of replacing worn tissues and producing fresh organs make it possible to shift the unavoidable death still further and further in the future, while “I” goes on considering himself the same “I”, endlessly keeping up his health and capabilities. In this case, if the rate of birth of new people drops to zero, then according to the perception of the living ones nothing would change from what it already was the case before. Everything reduces to a mere change in the time scale, as if the Earth was to orbit the Sun quicker, and then birthdays were celebrated more frequently. This is because the unavoidable death brings about no natural time scale. Indeed, what is infinite life? Would a thousand years be enough? How about a million? In general, to ask “naive and silly” questions is the best way to elucidate the actual meaning of conceptions. An ephemeron that is doing everything in just one day, does he lives long? Or if lived a thousand years, while sleeping on average nine hundreds of them, is this longer than hundred years straight? (Remember “*Rip van Winkle*” by Irving); or if returned from a journey in the fast space vehicle? Other people would consider him long-living indeed. However, they would notice nothing, if tripped together. And how is it “indeed”? The very variety of these questions means that they are not to be answered within the Method with its narrow universal unambiguousness.* The Method begins with the explicit presenting of the final state as the goal for all further auxiliary operations. But final state “is not seen” from the given state of a particular life. Once the birth of individuals is permitted, a life time scale is naturally defined by aging, bringing about either the expiring of any progress (and this is easier to be discerned from stagnation than from degradation) or mutual misunderstanding and non-acceptance between generations. The wretched existence of the “eternal”, bothered with themselves elders of Luggnegg, as described by Swift in his “Gulliver's travels”, is not in their impotence, but rather in their old-fashion mind. Then death itself becomes desirable for “I”, hence, the Method is no longer of his interest. And then what are new people needed for? So all the community is steadily loosing incentive to exist, as it was already guessed (two centuries ago) by Baratynsky in his poem “The last death”. Being perfected further still, the Method is steadily expanding its boundaries and claiming a more detailed approximation of life in all its variety by means of universally repeatable constructions, persuading, by the way, the users that only things that deserve heed are those that the Method is capable of providing. However, “...the man, always and everywhere, whoever he was, strove to act just as he wanted, and not at all as the reason and profit ordered him.” (Dostoevsky, *Notes from the Underground*). And if the man acts “just as he likes”, accounting for no circumstances, he never addresses the Method for its advices. Indeed, if it were

*Some people regard their life as somehow continued in their descendants, works, communities, and so on.

that protection from death be of only concern, then the organisms would be simply mechanisms to solve a particular problem; no other difference among them would be important, and nothing would prevent the substitution of one by another, so reaching "genuine" immortality. We conclude that even in its own realm, i.e. the life-and-death problem, the Method loses its importance just on the climax of its successes, where the very transition from life to death becomes indistinct.

This is its irony.

Submitted on February 26, 2015 / Accepted on October 22, 2015

The Journal on Advanced Studies in Theoretical and Experimental Physics, including Related Themes from Mathematics

PROGRESS IN PHYSICS

A quarterly issue scientific journal, registered with the Library of Congress (DC, USA). This journal is peer reviewed and included in the abstracting and indexing coverage of: Mathematical Reviews and MathSciNet (AMS, USA), DOAJ of Lund University (Sweden), Scientific Commons of the University of St. Gallen (Switzerland), Open-J-Gate (India), Referativnyi Zhurnal VINITI (Russia), etc.

Electronic version of this journal:
http://www.ptep-online.com

Postal Address

Department of Mathematics and Science,
University of New Mexico,
705 Gurley Ave., Gallup, NM 87301, USA

ISSN: 1555-5534 (print)
ISSN: 1555-5615 (online)

Standard Address Number: 297-5092
Printed in the United States of America

April–July 2016 **Vol. 12, Issue 3**

CONTENTS

Yurkin A.V. On the Descriptive Geometric Interpretation of Pauli Principle, Elements of the Mendeleev Table of Chemical Elements, and the Newtonian Laminar Current of a Liquid 149

Eid S.A., Diab S.M. X(5) Symmetry to ^{152}Sm 170

Johnson R.J. A Re-examination of Kirchhoff's Law of Thermal Radiation in Relation to Recent Criticisms (*Letters to Progress in Physics*) 175

Robitaille P.-M. A Re-examination of Kirchhoff's Law of Thermal Radiation in Relation to Recent Criticisms: Reply (*Letters to Progress in Physics*) 184

Rybicki M. Errata to "Mansouri-Sexl Test Theory: The Question of Equivalence between Special Relativity and Ether Theories" (*Letters to Progress in Physics*) 204

Esmail S.H. and Taha M.M. Application of the Differential Transform Method to the Advection-Diffusion Equation in Three-Dimensions 205

Millette P.A. On the Applicability of Bell's Inequality 211

Spivey R.J. Coincident Down-chirps in GW150914 Betray the Absence of Event Horizons 216

Marshall T.W. Repulsive Gravity in the Oppenheimer-Snyder Collapsar 219

Daywitt W.C. The Dirac-Electron Vacuum Wave 222

Muralidhar K. Mass of a Charged Particle with Complex Structure in Zeropoint Field . . 224

Brodet E. The Relationship Between the Possibility of a Hidden Variable in Time and the Uncertainty Principle 231

Daywitt W.C. A Modern Interpretation of the Dirac-Electron Continuity Equation 234

Marshall T.W. Optics of the Event Horizon Telescope 236

Matwi M. The Dual Behavior of Quantum Fields and the Big Bang 241

Consiglio J. On Quantization and the Resonance Paths 259

Belyakov A.V. On the Nature of Ball Lightning 276

Feinstein C.A. Dialogue Concerning the Two Chief World Views (*Letters to Progress in Physics*) 280

Spivey R.J. Criteria for Aerial Locomotion in Exoplanetary Atmospheres: Revisiting the Habitable Zone for Flying Lifeforms 284

Muralidhar K. The Structure of the Photon in Complex Vector Space 291

Obagboye L.F., Howusu S.X.K., Chifu E.N. and Omaghali E.J.N. Gravitational Waves from a Sinusoidially Varying Spherical Distribution of Mass 297

Daywitt W.C. Gravitational Shielding as Viewed in the Planck Vacuum Theory 301

Information for Authors

Progress in Physics has been created for rapid publications on advanced studies in theoretical and experimental physics, including related themes from mathematics and astronomy. All submitted papers should be professional, in good English, containing a brief review of a problem and obtained results.

All submissions should be designed in LaTeX format using *Progress in Physics* template. This template can be downloaded from *Progress in Physics* home page http://www.ptep-online.com

Preliminary, authors may submit papers in PDF format. If the paper is accepted, authors can manage LaTeXtyping. Do not send MS Word documents, please: we do not use this software, so unable to read this file format. Incorrectly formatted papers (i.e. not LaTeXwith the template) will not be accepted for publication. Those authors who are unable to prepare their submissions in LaTeXformat can apply to a third-party payable service for LaTeX typing. Our personnel work voluntarily. Authors must assist by conforming to this policy, to make the publication process as easy and fast as possible.

Abstract and the necessary information about author(s) should be included into the papers. To submit a paper, mail the file(s) to the Editor-in-Chief.

All submitted papers should be as brief as possible. Short articles are preferable. Large papers can also be considered. Letters related to the publications in the journal or to the events among the science community can be applied to the section *Letters to Progress in Physics*.

All that has been accepted for the online issue of *Progress in Physics* is printed in the paper version of the journal. To order printed issues, contact the Editors.

Authors retain their rights to use their papers published in *Progress in Physics* as a whole or any part of it in any other publications and in any way they see fit. This copyright agreement shall remain valid even if the authors transfer copyright of their published papers to another party.

Electronic copies of all papers published in *Progress in Physics* are available for free download, copying, and re-distribution, according to the copyright agreement printed on the titlepage of each issue of the journal. This copyright agreement follows the *Budapest Open Initiative* and the *Creative Commons Attribution-Noncommercial-No Derivative Works 2.5 License* declaring that electronic copies of such books and journals should always be accessed for reading, download, and copying for any person, and free of charge.

Consideration and review process does not require any payment from the side of the submitters. Nevertheless the authors of accepted papers are requested to pay the page charges. *Progress in Physics* is a non-profit/academic journal: money collected from the authors cover the cost of printing and distribution of the annual volumes of the journal along the major academic/university libraries of the world. (Look for the current author fee in the online version of *Progress in Physics*.)

On the Descriptive Geometric Interpretation of Pauli Principle, Elements of the Mendeleev Table of Chemical Elements, and the Newtonian Laminar Current of a Liquid

Alexander V. Yurkin
Puschino, Russia. E-mail: alvl1yurkin@rambler.ru

This work presents a two-dimensional and three-dimensional geometrical research of a ray system. We consider trajectories of motion of the particles having a half-integer spin. Interpretation of Pauli Principle showing distribution of electrons on power levels of the atom is given herein. The number of the electron shells in our model of the atom doesn't exceed 8. We give a geometric interpretation of the main, azimuthally, magnetic and spin numbers in the form of angles and distances. We show forth that the hyperbolic dependence of energy on the main quantum number n of the hydrogen atom ($E_n \sim -1/n^2$) known from experimental spectral studies, Bohr's theory and Quantum Mechanics can also be obtained from our geometrical formulation of Pauli Principle. Also, in the framework of research of the suggested ray model, the step structure of the layers at a laminar current of a liquid is deduced.

Contents

Introduction 149
1. Half-integer system of eight groups of rays 150
1.1. Two-dimensional projection of Gaussian (paraxial) rays 150
1.2. Three-dimensional projection of Gaussian (paraxial) rays 150
1.3. Periodic and acyclic trajectories 150
1.4. Periodic trajectories and step layers in the laminar current of a liquid 152
2. Gaussian (paraxial) rays and Pauli Principle 159
2.1. Angles, distances and quantum system 159
2.2. Periodic tables and geometrical constructions 160
2.2.1. Creation of the first shell of a quantum system 160
2.2.2. Creation of the second shell of a quantum system 160
2.2.3. Creation of the third and fourth shells of a quantum system 160
2.2.4. Pauli Principle and the geometric system of the hydrogen atom 162
Conclusions 167
Acknowledgements 167
References 167
Appendix 168

Introduction

Descriptive geometric models are used for the evident description of various phenomena, including quantum phenomena [1].

In works [2, 3], we already introduced a geometric model on the plane consisting of systems of paraxial rays describing distribution of light in lasers, turbulent and laminar flows of a liquid on pipes, and also finding an electron in the infinite deep potential. In work [3] we noted that the aforementioned model can be used for a descriptive interpretation of moving particles with the integer or half-integer spin.

In the works [3, 4] it was devoted to study the integer ray system (see [3]) by such means that possible to describe moving particles having the integer spin. However, even in the works [2, 5] we actually investigated a systems of ray trajectories which can be characterized as a half-integer ray system [3] by means of which it is possible to describe moving particles having a half-integer spin.

We aim, in the present work, to study a half-integer ray system, two-dimensional and three-dimensional geometric models of motion of the particles having a half-integer spin.

A geometric interpretation of Pauli Principle showing distribution of electrons on energy levels of the atom (such as those described in the physics textbooks [6, 7]) is suggested herein.

The geometric interpretation of the main, azimuthally, magnetic and spin numbers is given in the present work in the form of small angles and distances.

Also, we show a possibility of the existence of the final number of electron shells in the elements of the Mendeleev Periodic System of Chemical Elements. The shells and subshells of the atoms are interpreted as a system of the wave trajectories consisting of direct inclined pieces.

Geometric interpretations of the hydrogen atom and its power levels respectively are separately given in the work as well.

So forth, on the basis of the research of the half-integer ray model, we introduce the step structure of layers in a laminar current of a liquid (such a liquid is described in most textbooks, see [8]).*

*The laminary liquid current was first described long time ago by Newton. The Netwon theory was rechecked many times (see [8]).

Numerical calculations, presented in the present work, as well as those published in [3], were represented by means of three-dimensional tables created in Excel.

For the convenience of readers, reference drawings taken physics textbooks are given in Appendix, while the research part of our publication contains only originally calculated drawings and tables.

1 Half-integer system of eight groups of rays

1.1 Two-dimensional projection of Gaussian (paraxial) rays

In the work [3], we briefly described a paraxial binary (sharing in two) flat system of trajectories. This system consists of groups of rays, in which the rays are inclined under p angles, small to an axis, multiple to the angle γ:

$$p = \left(i + \frac{1}{2}\right)\gamma, \qquad i = 0, \pm1, \pm2, \ldots \tag{1}$$

We called this system of rays: "$(i + 1/2)\gamma$-system" or *half-integer ray system* [3]. We will describe this system in more detail in Fig. 1.

This binary system of rays consists of eight groups of the rays and their links. The rays and links of each of these groups aren't imposed on the rays of other groups, but can cross them.

Branching points of the rays will be spaced from a symmetry axis on small distances of q, multiple to $\frac{1}{2}k$ length:

$$q = \frac{jk}{2}, \qquad j = 0, \pm1, \pm2, \ldots \tag{2}$$

Further, we more precisely will refer to "$(i + 1/2)\gamma$-system" as "$[\, p = (i + 1/2)\gamma,\ q = jk/2\,]$-system".

In this work, as well as in the previous works [2–5] we assume that the rays extend along the branching links; therefore the number of the rays $\mathbb{N}$ can be summarized. We also assume that $\mathbb{K}$ is a number of the links generally $\mathbb{N} \geqslant \mathbb{K}$.

In Fig. 1 (a-d, f-i) eight groups of rays of the aforementioned $[\, p = (i + 1/2)\gamma,\ q = jk/2\,]$-system are shown: $\boldsymbol{K''}$, $\boldsymbol{L''}$, $\boldsymbol{M''}$, $\boldsymbol{N''}$, $\boldsymbol{O''}$, $\boldsymbol{P''}$, $\boldsymbol{Q''}$, $\boldsymbol{R''}$.

This system is placed on a rectangular *coordinate grid*. The size of a cell of a grid has height of $\frac{1}{2}k$ and length of L, $L \gg \frac{1}{2}k$, $L \gg \frac{1}{2}jk$.

Groups in Fig. 1 (a-d) and in Fig. 1 (f-i) are shifted from each other down on the $\frac{1}{2}k$ distance. Groups in Fig. 1 (f-i) are shifted concerning groups in Fig. 1 (a-d) on distance of L.

In Fig. 1 (e) and Fig. 1 (j) the image of groups of the rays $\boldsymbol{K''}$, $\boldsymbol{L''}$, $\boldsymbol{M''}$, $\boldsymbol{N''}$ and $\boldsymbol{O''}$, $\boldsymbol{P''}$, $\boldsymbol{Q''}$, $\boldsymbol{R''}$ respectively, are combined altogether. In Fig. 1 (k) all eight groups of rays are combined together.

1.2 Three-dimensional projection of Gaussian (paraxial) rays

In the work [3] we considered the three-dimensional image of a binary paraxial system of rays in the form of a nonlinear arithmetic parallelepiped: *In a nonlinear arithmetic parallelepiped all numbers are located in the rectangular planes of identical sizes, and these planes are located layer-by-layer one under another since parallelepiped top*.

In this case, the nonlinear arithmetic parallelepiped [3] has a $\mathfrak{N}L$ height, a length of $D = \frac{1}{2}km' + 1$ and a width $\Gamma = \gamma m + 1$, where L, k are distances, while γ is a small angle in the two-dimensional binary ray system (Fig. 1) and at the same time a small distance in a three-dimensional nonlinear parallelepiped [3], and $\mathfrak{N}$, m, m' are natural numbers or zero.

After a large number of passes (iterations) of $\mathfrak{N} \to \infty$ and $\mathfrak{N}L \gg L$, we write down the rule of consecutive filling with numbers of a nonlinear arithmetic parallelepiped as well as in [3]:

$$A = B + C, \tag{3}$$

where

$$A = \begin{pmatrix} \mathfrak{N} \\ p \\ q \end{pmatrix}, \quad B = \begin{pmatrix} \mathfrak{N} - 1 \\ p - 1 \\ q + p - 1 \end{pmatrix}, \quad C = \begin{pmatrix} \mathfrak{N} - 1 \\ p + 1 \\ q + p + 1 \end{pmatrix}.$$

For creation of various types [3] of nonlinear arithmetic parallelepipeds it is necessary to set various additional boundaries and initial conditions.

1.3 Periodic and acyclic trajectories

The system $[\, p = (i + 1/2)\gamma,\ q = jk/2\,]$ of rays generally consists of periodic and acyclic trajectories. In Fig. 2, one of eight groups of the rays of this system are shown for the case of $D = 4k$, $\Gamma = 7\gamma$.

We will set the *first boundary conditions* [3] for number A in formula (3) for nonzero $\mathfrak{N}$-layers:

$$A = 0 \tag{4}$$

for $q = |\, q_{max}|$, where $q_{max} = \frac{1}{2}D$.

Further we will set the *first boundary conditions* for numbers B and C in formula (3) for nonzero $\mathfrak{N}$-layers:

$$B = 0, \qquad C = 0 \tag{5}$$

for $|\, q + p - 1\,| > q_{max}$ and $|\, q + p + 1\,| > q_{max}$ rectively.

We now set the initial conditions [3] for the numbers B and C in formula (3) for the sequence of numbers q of a zero layer ($\mathfrak{N} = 0$):

$$\begin{pmatrix} 0 \\ p \\ q \end{pmatrix} = 1 \tag{6}$$

for $|q| \leqslant q_{max}$ and

$$\begin{pmatrix} 0 \\ p \\ q \end{pmatrix} = 0 \tag{7}$$

for other q.

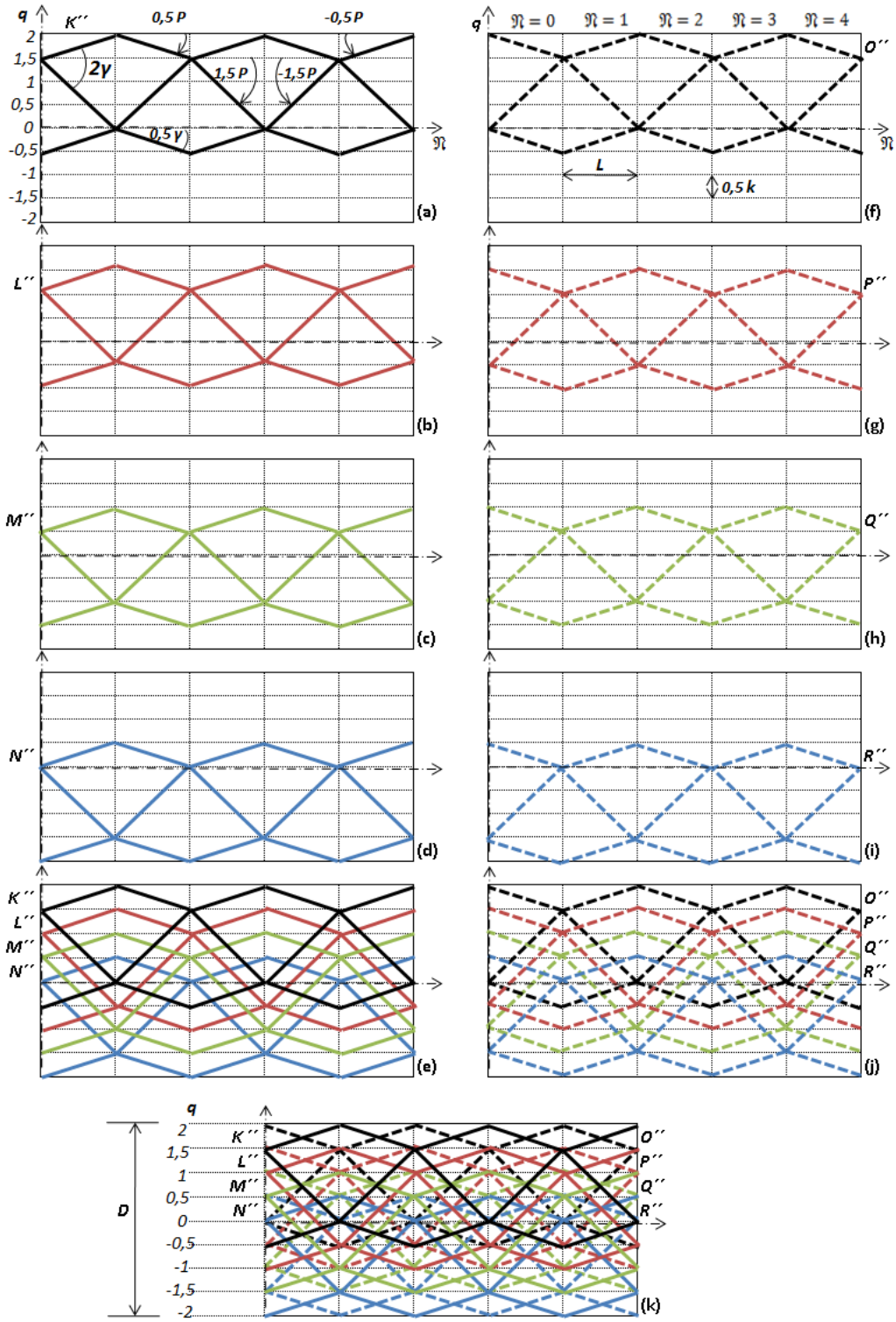

Fig. 1: Periodic trajectories. Eight groups of rays: $\boldsymbol{K''}$, $\boldsymbol{L''}$, $\boldsymbol{M''}$, $\boldsymbol{N''}$, $\boldsymbol{O''}$, $\boldsymbol{P''}$, $\boldsymbol{Q''}$, $\boldsymbol{R''}$ of the $[\, p = (i + 1/2)\,\gamma,\ q = jk/2\,]$ system. $k/2$, and L are the minimum distances on a vertical and a horizontal respectively. $\mathfrak{N}$ is the number of pass of rays (the number of iteration). Dash-dotted lines with arrows showed axes of a coordinate grid in which trajectories are placed.

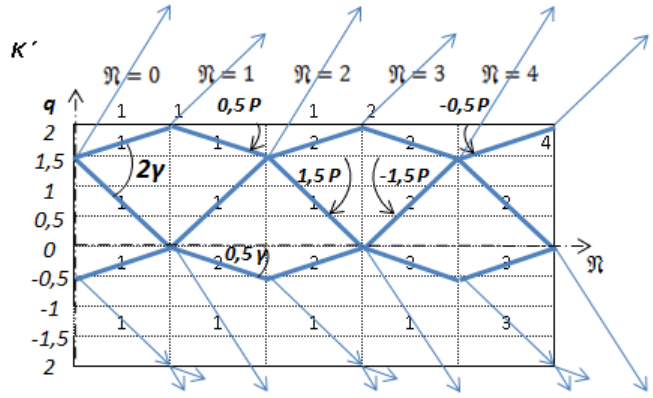

Fig. 2: Periodic and acyclic trajectories. One of eight groups of rays (K' group) of the [$p = (i + 1/2)\,\gamma,\ q = jk/2$] system is shown here. Figures about the shown links illustrate the number of the rays of $\mathbb{N}$ and the summation process of number of the rays extending along the number of the links $\mathbb{K}$ for the first five passes ($\mathfrak{R} = 0 - 4$).

In Fig. 3, calculation formulas (given in MS Excel) of a nonlinear arithmetic parallelepiped (Figs. 1, 2) for $D = 4k$, $\Gamma = 7\gamma$ case for zero, the first and second passes of the ray system, i.e. $\mathfrak{R} = 0, 1, 2$. The calculation was made according to the rule (3) of the consecutive filling with the numbers of an arithmetic rectangle taking into account the boundary (4, 5) and the initial (6, 7) conditions. Three rectangles in Fig. 3 are the layers of a nonlinear arithmetic parallelepiped.

Results of numerical calculation for the first five passes of rays, i.e. $\mathfrak{R} = 0 - 5$ are given in Fig. 4. Five rectangles are layers of a nonlinear arithmetic parallelepiped.

Results of numerical calculations for 32 pass of rays, i.e. for $\mathfrak{R} = 32$ (a) are given in Fig. 5. The envelopes of distribution of number of rays of $\mathcal{K}(q)$ on the section (b) and $\mathcal{K}'(p)$ at the angle (c) are provided.

1.4 Periodic trajectories and step layers in the laminar current of a liquid

In a specific case, the $[p = (i + 1/2)\gamma, q = jk/2]$ system of rays consists only of *periodic trajectories*. Fig. 6 shows one of the eight groups of rays of this system for the case, where $D = 4k$, $\Gamma = 3\gamma$.

In this case, we need to further set special initial and threshold conditions to create the appropriate nonlinear arithmetic parallelepiped [3].

Let's consider here a simple and illustrative (as compared to the description given in [3]) way (an Excel algorithm) of setting special initial and threshold conditions for the parallelepiped that describes the system consisting only of periodic trajectories.

Let's set the *second* threshold conditions for A, B, and C in formula (3) for nonzero $\mathfrak{R}$-layers:

$$A = 0, \tag{8}$$

if $B = 0$ and $C = 0$, and

$$B = 0 \quad \text{and} \quad C = 0, \tag{9}$$

if $A = 0$.

Let's set additional initial conditions for B and C for a zero layer ($\mathfrak{R} = 0$):

$$B = 0 \quad \text{and} \quad C = 0, \tag{10}$$

if $A = 0$.

The offered way (the algorithm) can be easily implemented in numerical calculations in Excel.

At first, we completely fill with units a numerical rectangle of the zero layer ($\mathfrak{R} = 0$) according to formula (6) and formula (7).

Then we fill with numbers a numerical rectangle of the first layer ($\mathfrak{R} = 1$) according to formulas (3 to 5). Some zeroes appear in the first layer.

Then we delete numbers (units) from the cells of the zero-layer rectangle which don't influence cells of the first-layer rectangle.

Then we delete numbers from the cells of the first-layer rectangle which don't depend on the cells of the zero-layer rectangle. We have some new zeroes in the first layer again.

Then again we delete numbers (units) from the cells of the zero-layer rectangle which don't influence the cells of the first-layer rectangle.

And so we repeat this process several times. As a result, we still have cells filled with meaningful numbers which influence other cells, and the cells which depend on other cells. The remained cells describe the $[p = (i + 1/2)\gamma, q = jk/2]$ system consisting only of periodic trajectories.

$[p = (i + 1/2)\gamma, q = jk/2]$ is the system of rays consisting only of periodic trajectories as shown in Fig. 7. The results of calculation of a nonlinear arithmetic parallelepiped (Figs. 1 and 6) are for $D = 4k$, $\Gamma = 3\gamma$ for the zero, first, and second passes of the ray system, i.e. ($\mathfrak{R} = 0, 1, 2$). The calculation was made according to the rule (3) of consecutive filling with numbers of an arithmetic rectangle taking into account the *first* and the *second* threshold (4 and 5; 8 and 9) and initial (6, 7, and 10) conditions, including the algorithm (8 to 10). The three rectangles shown in Fig. 7 are the layers of a nonlinear arithmetic parallelepiped.

Fig. 8 shows the images of layers of a nonlinear arithmetic parallelepiped and a numerical example of calculation of the $[p = (i + 1/2)\gamma,\ q = jk/2]$ system of periodic trajectories (Fig. 7) for $D = 4k$, $\Gamma = 3\gamma$ for zero and the subsequent four passes of rays, i.e. for $\mathfrak{R} = 0 - 4$.

Fig. 9 shows numerical calculations and graphics made in Excel. Numerical calculation for the 32nd pass of rays, i.e. for $\mathfrak{R} = 32$, is given in (a). It also shows the envelopes of distribution of the number of rays of $\mathcal{N}(q)$ on the section (b) and $\mathcal{N}'(p)$ at the angle (c) for $D = 4k$, $\Gamma = 3\gamma$.

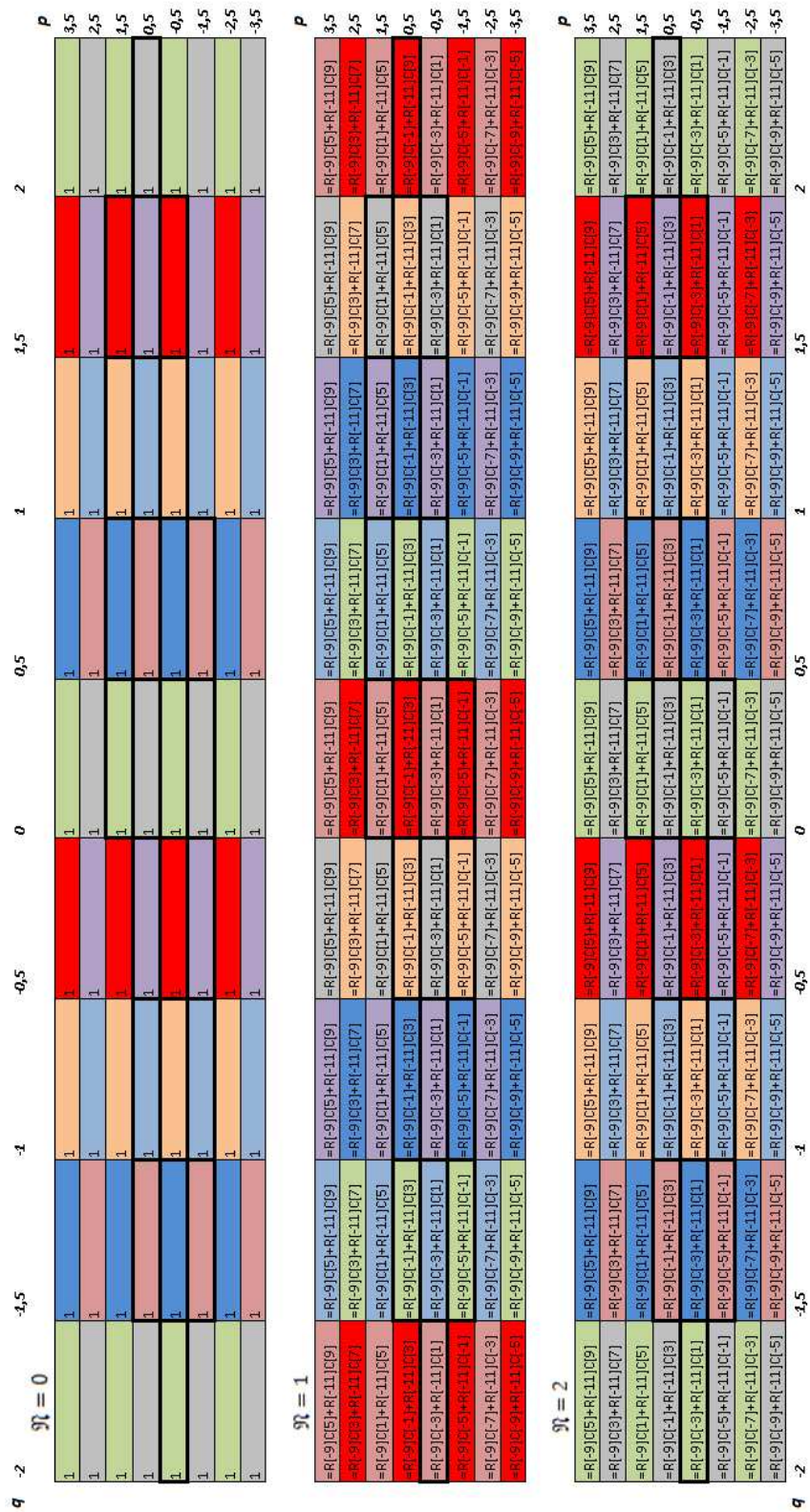

Fig. 3: Calculation of the filling with the numbers of a nonlinear arithmetic parallelepiped for $D = 4k$ case in Excel. The $[\, p = (i + 1/2)\,\gamma,\ q = jk/2\,]$ system of 8 groups of rays of periodic and acyclic trajectories; the first three pass through the rays. Each of the eight groups is marked by an own color.

Fig. 4: Results of numerical calculation in Excel for the first five passes of rays (iterations). Arrows showed dependent cells. Each of eight groups $\boldsymbol{K'}$, $\boldsymbol{L'}$, $\boldsymbol{M'}$, $\boldsymbol{N'}$, $\boldsymbol{O'}$, $\boldsymbol{P'}$, $\boldsymbol{Q'}$, $\boldsymbol{R'}$ of the $[\, p = (i + 1/2)\,\gamma,\ q = jk/2\,]$ system of periodic and acyclic trajectories. Each of the eight groups is marked by an own color.

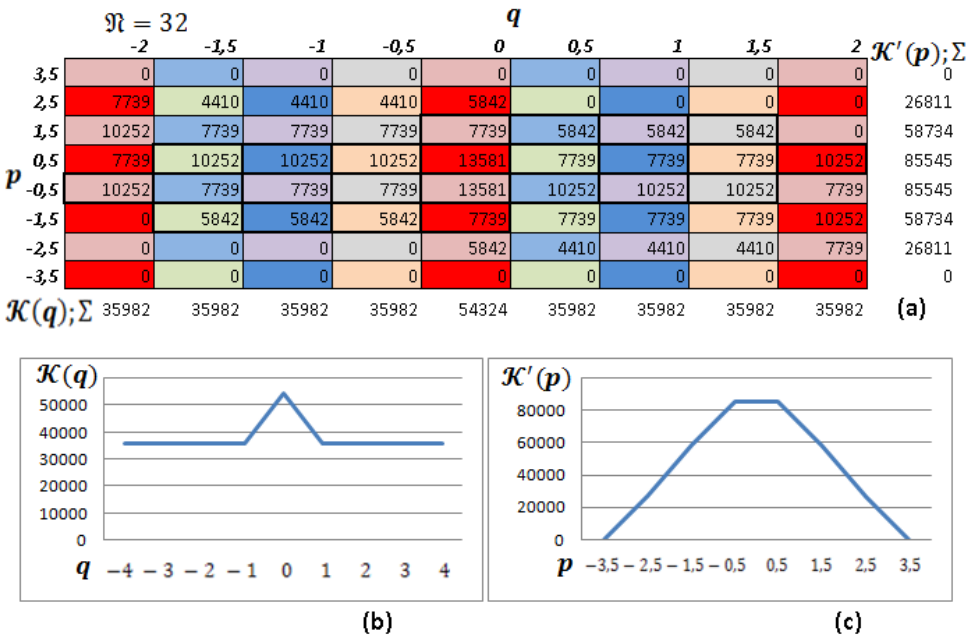

Fig. 5: Results of the numerical calculations for 32 pass of rays, i.e. for $\mathfrak{N} = 32$ for periodic and acyclic trajectories of the considered $[\,p = (i + 1/2)\,\gamma,\ q = jk/2\,]$ system (a); each of the eight groups of the system is marked by an own color; a thick framework in the central part is noted the system of periodic trajectories. The envelopes of distribution of the number of the rays of $\mathcal{K}(q)$ on the section (b) and $\mathcal{K}'(p)$ at the angle (c) are given. We note that for this case, as show our calculations, the form of envelope (b, c) practically doesn't change approximately after the 15th pass.

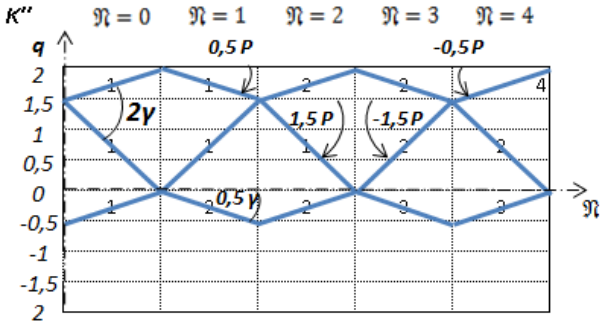

Fig. 6: Periodic trajectories. It shows one of eight groups of rays (***K″*** group) of the $[p = (i + 1/2)\gamma,\ q = jk/2]$ system.

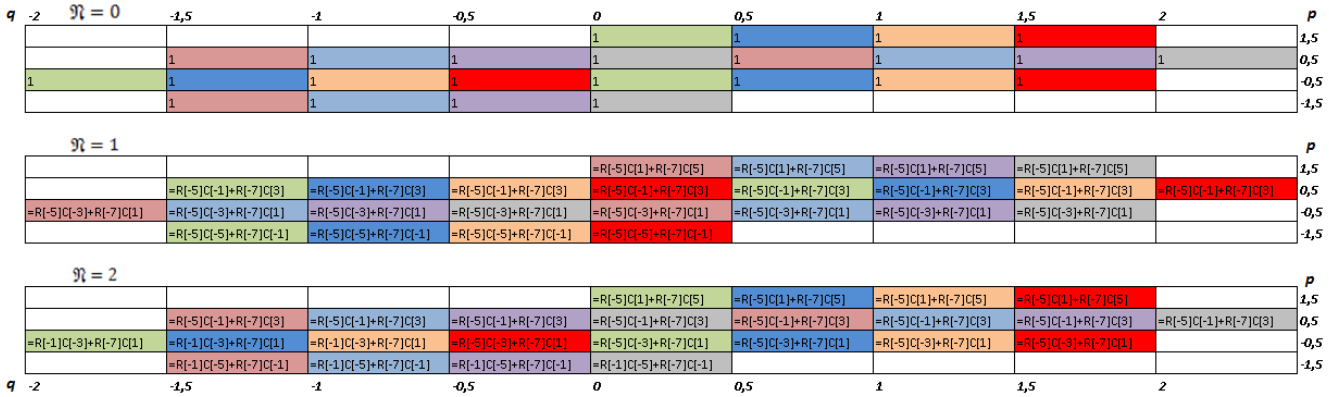

Fig. 7: Calculation of filling with numbers of a nonlinear arithmetic parallelepiped for $D = 4k$ in Excel, and the $[p = (i + 1/2)\gamma,\ q = jk/2]$ system of eight groups of rays of periodic trajectories — the first three passes of rays. Each of the eight groups is marked in a separate color.

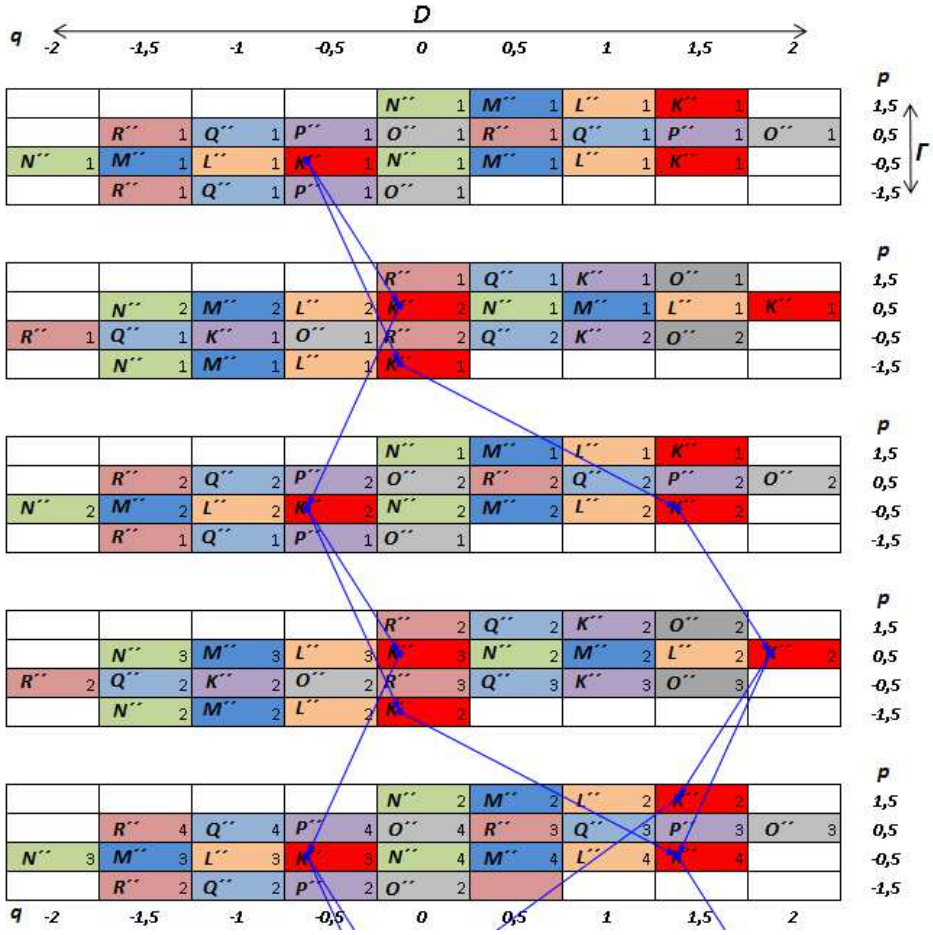

Fig. 8: Results of numerical calculation in Excel for the first five passes of rays (iterations). The arrows point to dependent cells. Each of the eight groups — $\boldsymbol{K''}$, $\boldsymbol{L''}$, $\boldsymbol{M''}$, $\boldsymbol{N''}$, $\boldsymbol{O''}$, $\boldsymbol{P''}$, $\boldsymbol{Q''}$, and $\boldsymbol{R''}$ — of the $[p=(i+1/2)\gamma,\ q=jk/2]$ system of periodic trajectories is marked in a separate color.

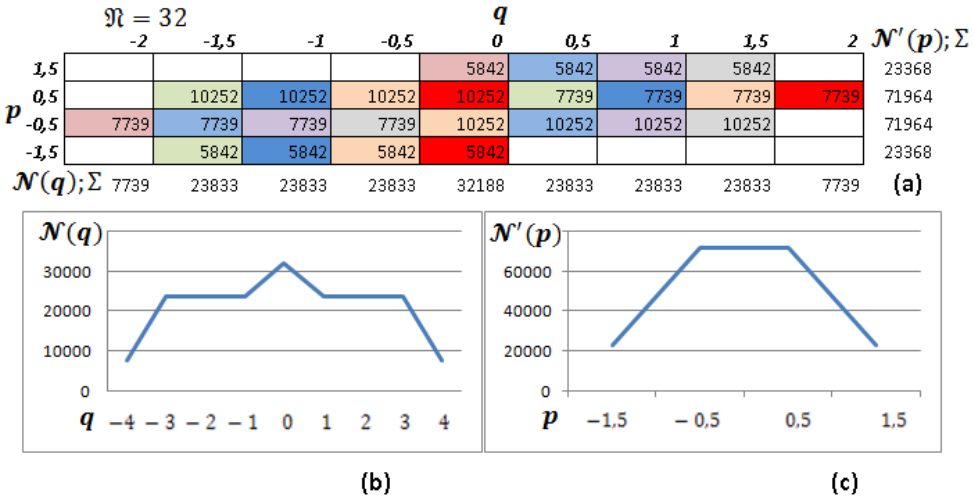

Fig. 9: Results of numerical calculations for the 32nd pass of rays, i.e. for $\Re = 32$, for periodic trajectories of the $[p=(i+1/2)\gamma,\ q=jk/2]$ system (a). Each of the eight groups of the system is marked in separate color. It also shows the envelopes of distribution of number $\mathbb{N}$ of rays of $\mathcal{N}(q)$ on the section (b) and $\mathcal{N}'(p)$ at the angle (c). Note, according to our calculations, in this case, there is virtually no change in the form of the envelope, (b) and (c), approximately after the 15th pass.

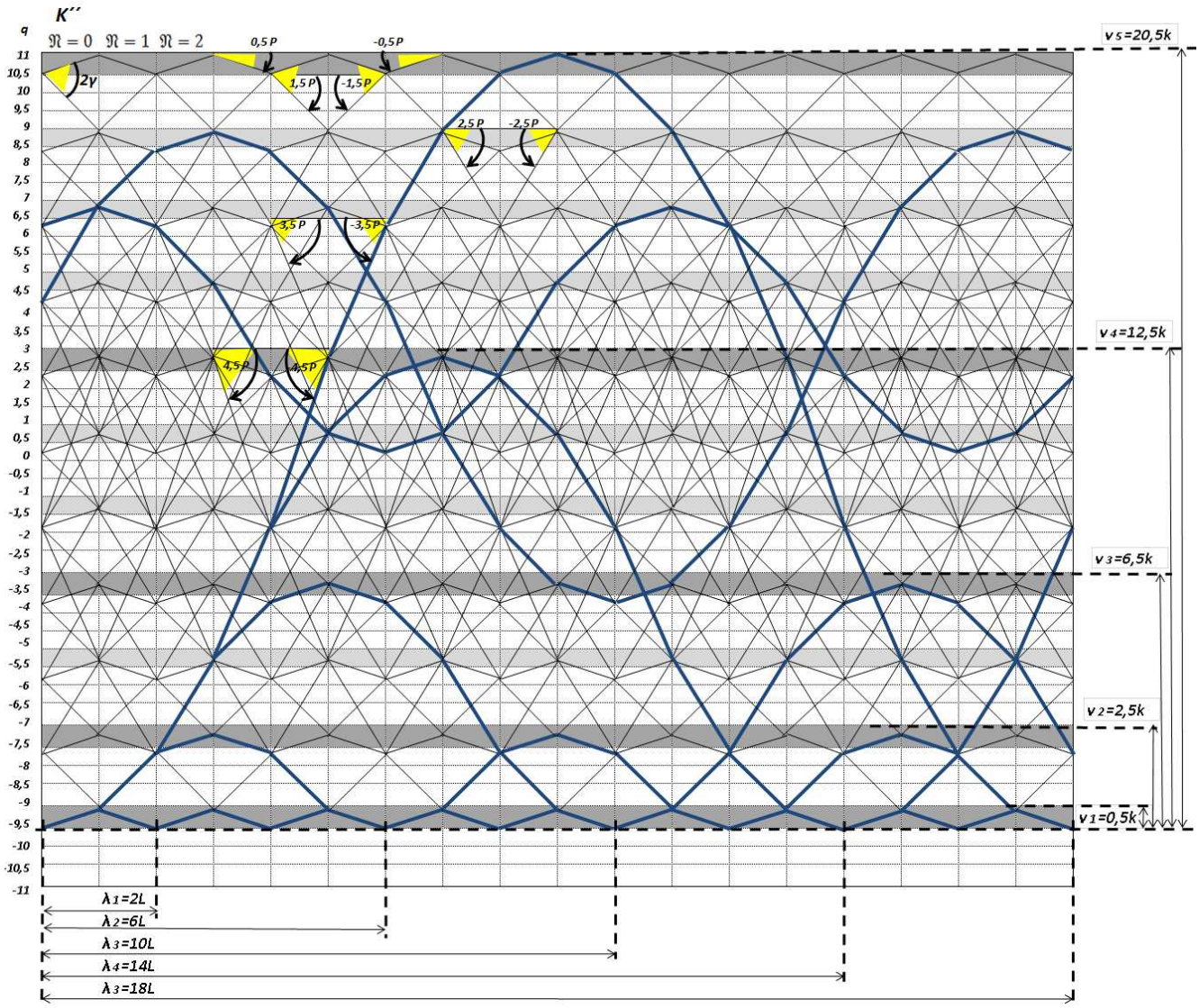

Fig. 10: Periodic (wavy) trajectories. It shows one of the eight groups of rays (***K''*** group) of the $[p=(i+1/2)\gamma,\ q=jk/2]$ system. The group contains 35 links within one pass of $\mathfrak{R}$. Crests and troughs of the "waves" are located (attached) between the horizontals marked in dark color. These horizontals have thickness of $0,5k$ and are located at identical distance of $2k$ from each other.

Fig. 10, similar to Figs. 1 and 6, shows one of the eight groups of rays of the $[p=(i+1/2)\gamma,\ q=jk/2]$ system of rays of periodic trajectories for $D=22k$, $\Gamma=9\gamma$.

In Fig. 10, some of wavy geometric trajectories for the considered $[p=(i+1/2)\gamma, q=jk/2]$ system of rays are shown as heavy lines. Wavy trajectories consist of the links inclined at small angles of $p=(i+1/2)\gamma$. The D size of the binary ray system can accommodate one "wave" or packages of "waves" of different length.

Let's denote the length of wavy trajectories by λ_n. With increasing D this λ_n is growing discretely:

$$\lambda_n = 2\,(2n-1)\,L, \tag{11}$$

where $n=1,2,\ldots$

Let's denote the height of this "wave" by ν_n. The ν_n height is proportional to the squared λ_n length:

$$\nu_n = \frac{\left(n^2-n+\frac{1}{2}\right)}{k} \sim \lambda_n^2 . \tag{12}$$

Wavy trajectories in Fig. 10 can settle down in any part of the *coordinate grid* between horizontals within D.

Fig. 11, similar to Fig. 9, shows numerical calculations and graphs made in Excel. Numerical calculation for the 128th pass of rays, i.e. for $\mathfrak{R}=128$, is given in (a). There are also envelopes of distribution of number $\mathbb{N}$ of rays of $\mathcal{N}(q)$ on the section (b) and $\mathcal{N}'(p)$ at the angle (c) given for the

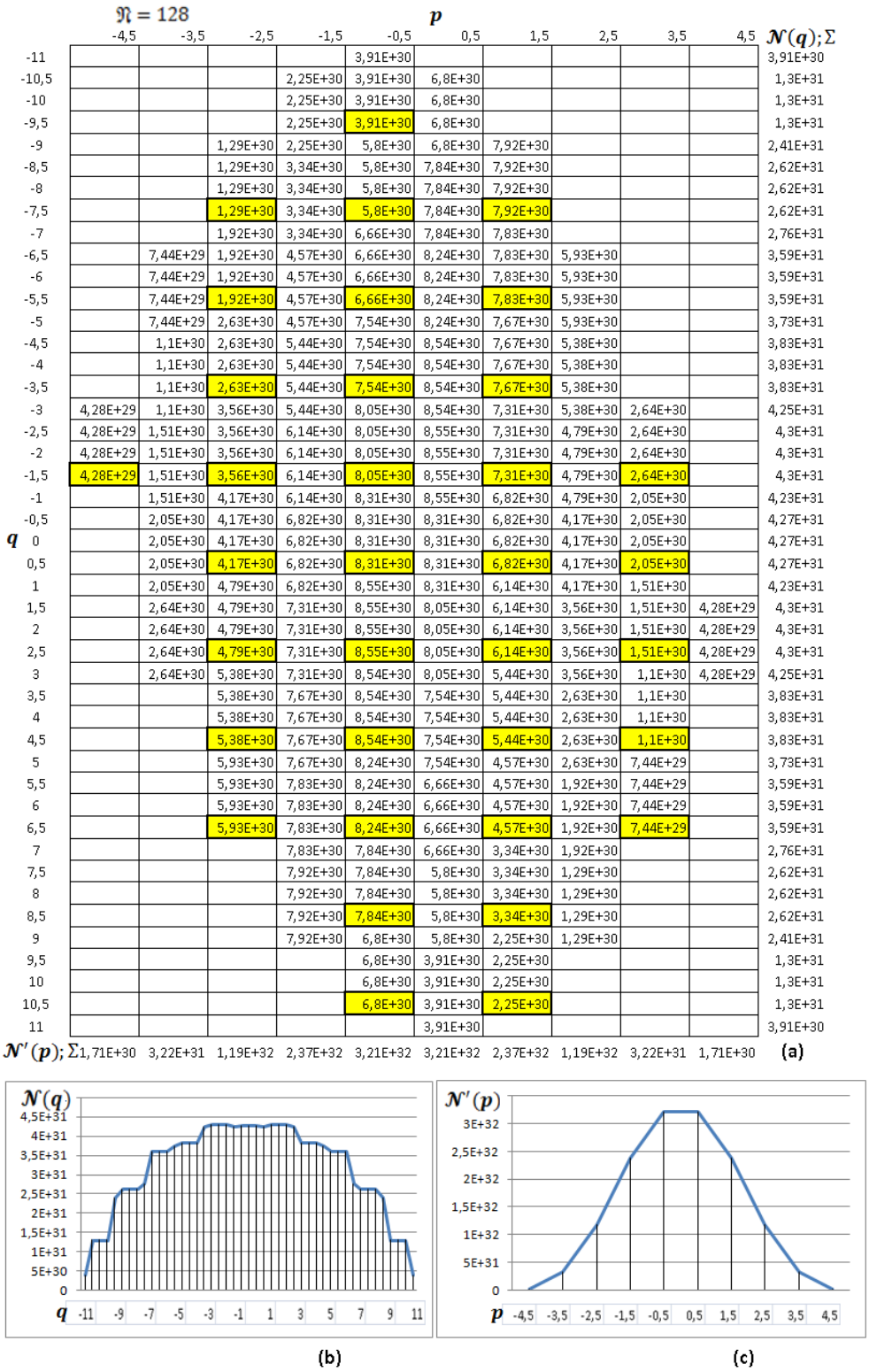

Fig. 11: Results of numerical calculations for the 128th pass of rays, i.e., $\Re = 128$, for eight periodic trajectories of the $[p = (i + 1/2)\gamma,\ q = jk/2]$ system. Thirty five cells (corresponding to 35 rays within one pass in Fig. 10) of one of the eight groups of the system (***K″*** group) are highlighted with the darker color and heavy external borders of the cells (a). The figure also shows the envelopes of distribution of number $\mathbb{N}$ of rays of $\boldsymbol{\mathcal{N}(q)}$ on the section (b) and $\boldsymbol{\mathcal{N}'(p)}$ at the angle (c). Note, according to our calculations, in this case, there is virtually no change in the form of the envelopes, (b) and (c), approximately after the 70th pass.

$[p = (i + 1/2)\gamma,\ q = jk/2]$ system for all the eight groups for $D = 22k$, $\Gamma = 9\gamma$.

The darker cells with heavy external borders shown in Fig. 11 correspond to one of the eight groups of the system (***K″*** group) shown in Fig. 10.

Forms of the envelopes of distribution of the number of rays for the *half-integer system* are similar to the forms of the envelopes for the *integer system* described in [3]. The form of the envelope $\mathcal{N}(q)$ in Fig. 11(b) on the section after a large number of passes (Fig. 11b) is close to a parabola of the fourth degree, and the form of the envelope $\mathcal{N}'(p)$ at the angle (Fig. 11c) is close to Gaussian distribution.

In [3] we noted that the form of the envelope $\mathcal{N}(q)$ on the section for periodic trajectories corresponds to the form of the envelopes of speed distribution at zero pass ($\mathfrak{N} = 0$) and volume distribution after a large number of passes ($\mathfrak{N} \to \infty$) of liquid in pipe section at laminar flow.

In Fig. 11(b) we can see that the envelope $\mathcal{N}(q)$ has a stepped structure compared to the more smooth form of the envelope $\mathcal{N}'(p)$ (Fig. 11c). Similar results were received from numerical calculations for the integer system in [3], but the half-integer model gives the more accurate image of the "steps" compared to the integer model.

It can be assumed that such a stepped structure of the envelope $\mathcal{N}(q)$ explains the existence of layers of final thickness in liquid at laminar flow [8]. The speed and volume of liquid do not change within each of these layers of a certain final thickness.

2 Gaussian (paraxial) rays and Pauli Principle

2.1 Angles, distances and quantum system

Pauli Principle [6, 7] is correct for electrons and other particles with half-integer spin in a quantum system.

The condition of each electron in an atom is characterized by four quantum numbers [6, 7]:

$$\left.\begin{array}{rll} \text{Principal} & n & (n = 1, 2, 3, \dots) \\ \text{Azimuthal} & l & (l = 0, 1, 2, \dots, n-1) \\ \text{Magnetic} & m_l & (m_l = -l, \dots, -1, 0, +1, \dots, +l) \\ \text{Spin} & m_s & \left(m_s = +\tfrac{1}{2},\ -\tfrac{1}{2}\right) \end{array}\right\}. \quad (13)$$

Fig. 23 of the Appendix illustrates an example from [6] of spatial quantization.

In monographs [6] and [7] the spin is also denoted by one letter "s":

$$m_s = s = \pm\frac{1}{2}. \quad (13a)$$

According to Pauli Principle in a quantum system, for example in an atom, there can't be two electrons possessing identical quantum numbers: n, l, m_l, m_s. That is, two electrons cannot be in the same state simultaneously. No more than $2n^2$ electrons can be in a state with n value in an atom [6, 7]. If

$$\left.\begin{array}{ll} n = 1 & \text{there can be 2 electrons} \\ n = 2 & \text{there can be 8 electrons} \\ n = 3 & \text{there can be 18 electrons} \\ \dots\dots\dots\dots\dots\dots\dots\dots & \end{array}\right\}. \quad (14)$$

Electrons having identical value of the quantum number n form a *shell*. Shells consist of *subshells*, differing in value of the quantum number l.

Shells are denoted by characters according to value of n [6] and [7]:

$$\left.\begin{array}{rl} \text{Value of } n & 1\ 2\ 3\ 4\ 5\ 6\ 7\dots \\ \text{Designation of the shell} & K\ L\ M\ N\ O\ P\ Q\dots \end{array}\right\}. \quad (15)$$

The electron which is in condition of $l = 0$ is called an s electron, $l = 1$ — p electron, $l = 2$ — d electron, $l = 3$ — f electron, followed by g, h, etc. alphabetically. The value of the principal quantum number n is specified before the symbol of the azimuthal quantum number l [7].

The division of possible conditions of an electron in an atom into shells and subshells [7] is presented in the form of a *periodic table of conditions* of an electron (see Fig. 24 of the Appendix).

The process of building electron shells [7] (according to Pauli Principle) of the first 36 elements of the Mendeleev Periodic System is presented in the form of a *periodic table of elements* (see Fig. 25 of the Appendix).

Now let's give an algorithm of creation of another specific case of the binary $[p = (i + 1/2)\gamma,\ q = jk/2]$ system of rays of periodic trajectories (considered in Section 1.4). We will begin with the minimum quantity of rays consistently passing to the more complicated configurations of the system. Thus, we will compare the properties of our system to the data provided in periodic tables in Figs. 24 and 25 of the Appendix.

We accept that, for our paraxial beams, all the angles of γn, are small and multiple to the small angle of γ, and the small distance of k is as follows:

$$k \approx \gamma L. \quad (16)$$

For perfect correspondence between our geometric constructions and expressions (13 and 13a), including the data provided in periodic tables in Figs. 24 and 25 of the Appendix, we will enter the following assumptions:

$$\text{Principal number} \quad n \sim \gamma i \sim kj, \quad (n = 1, 2, \dots;\ i = 1, 2, \dots;\ j = 1, 2, \dots), \quad (17)$$

$$\text{Azimuthal number} \quad l = n - 1 \sim \gamma(n-1), \quad (18)$$

$$\text{Magnetic number} \quad m_l = \pm l \sim \pm k(n-1), \quad (19)$$

$$\text{Spin number} \quad s = \pm\frac{1}{2} \sim \pm\frac{\gamma}{2} \text{ and } m_s = \pm\frac{1}{2} \sim \pm\frac{k}{2}. \quad (20)$$

Electron shell	n	l	m_l	m_s	Sub shell
K	1	0	0	↑↓	$K(1s)$

(a)

Element	K	L		M		
	$1s$	$2s$	$2p$	$3s$	$3p$	$3d$
1 H	1	-	-	-	-	-
2 He	2	-	-	-	-	-

(b)

Fig. 12: K shell and the first parts of the periodic tables of (a) conditions of an electron, and (b) elements.

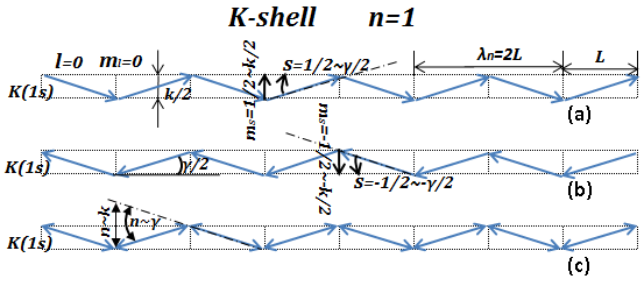

Fig. 13: One of the eight groups of rays of the $\{p=(i+1/2)\gamma\ ,\ q=jk/2\}$ subsystem of periodic trajectories, and K shell. (a) and (b) correspond to an atom of hydrogen, (c) — to an atom of helium. $n=1\sim\gamma\sim k$, $l=0$, $m_l=0$, $s=\pm1/2\sim\pm\gamma/2$, $m_s=\pm1/2\sim\pm k/2$. Dash-dotted lines show axes from which sizes of angles and distances are counted.

Electron shell	n	l	m_l	m_s	Sub shell
L	2	0	0	↑↓	$L_1(2s)$
		1	−1	↑↓	$L_2(2p)$
			0	↑↓	
			+1	↑↓	

(a)

Element	K	L		M		
	$1s$	$2s$	$2p$	$3s$	$3p$	$3d$
3 Li	2	1	-	-	-	-
4 Be	2	2	-	-	-	-
5 B	2	2	1	-	-	-
6 C	2	2	2	-	-	-
7 N	2	2	3	-	-	-
8 O	2	2	4	-	-	-
9 F	2	2	5	-	-	-
10 Ne	2	2	6	-	-	-

(b)

Fig. 14: L shell and the second parts of the periodic tables of (a) conditions of an electron, and (b) elements.

2.2 Periodic tables and geometrical constructions

2.2.1 Creation of the first shell of a quantum system

Fig. 12(a) shows the first (top) part of the periodic table of conditions of an electron (Fig. 24 of the Appendix) describing the first shell of K. Fig. 12(b) shows the top part of the periodic table of elements (Fig. 25 of the Appendix) describing the first two elements:

Fig. 13 shows one of the eight similar to (Figs. 1 and 6) groups of rays of the $[p=(i+1/2)\gamma,\ q=jk/2]$ *system* of periodic trajectories.

These trajectories correspond to the first electron shell of K shown in Fig. 12(a).

To be specific, let's call this *system* of periodic trajectories an $\{p=(i+1/2)\gamma,\ q=jk/2\}$ *subsystem of periodic trajectories* of the $[p=(i+1/2)\gamma,\ q=jk/2]$ *system* of periodic trajectories.

Fig. 13 (a and b) shows two trajectories with opposite (↑ symbol and ↓ symbol) orientation of a spin (one 1s electron). These trajectories correspond to an atom of hydrogen with random orientation of the spin (Fig. 12b).

Fig. 13c shows a trajectory with anti-parallel (↑↓ symbol) spin orientation (two 1s electrons). This trajectory corresponds to an atom of helium (Fig. 12b).

The atom of helium is closing filling of the K shell.

2.2.2 Creation of the second shell of a quantum system

Fig. 14 (a) shows the second part of the periodic table of conditions of an electron (Fig. 24 of the Appendix) describing the second cover of L. Fig. 14 (b) shows the second part of the periodic table of elements (Fig. 25 of the Appendix) describing the elements number three to ten.

Fig. 15 shows one of the eight similar to (Figs. 1 and 6) groups of rays of the $\{p=(i+1/2)\gamma,\ q=jk/2\}$ subsystem of periodic trajectories. These trajectories correspond to the second electron shell of L in Fig. 14 (a and b).

Fig. 15 (a) shows the L shell in the form of a periodic trajectory consisting of the subshell L_1 (one 1s electron) for Li. The form of this shell is the same as in Fig. 13(a) or in Fig. 13(b). The form of K shell for Li is the same as in Fig. 13(c).

Fig. 15 (b, c, d, e, f, g, and h) shows the L shells ($L_1(2s)$ and $L_2(2p)$ subshells) for Be, B, C, N, O, F, and Ne respectively (Fig. 14a and b):

The K shell for these elements is the same as that for Li (see Fig. 13c).

The K shell and L shell of the elements (Figs. 12 to 15) can settle down in our geometric model similar to arrangement of $\boldsymbol{K''}$ group of rays and $\boldsymbol{L''}$ group of rays respectively (Figs. 1, 7 and 8).

The atom of Ne is closing filling the L shell.

2.2.3 Creation of the third and fourth shells of a quantum system

Geometric schemes of Pauli Principle and elements of periodic table are further constructed in compliance with the above algorithm. Therefore, we will confine ourselves to giving specific examples.

The second part of the periodic table of conditions of an electron (Fig. 24 of the Appendix) describing the third M

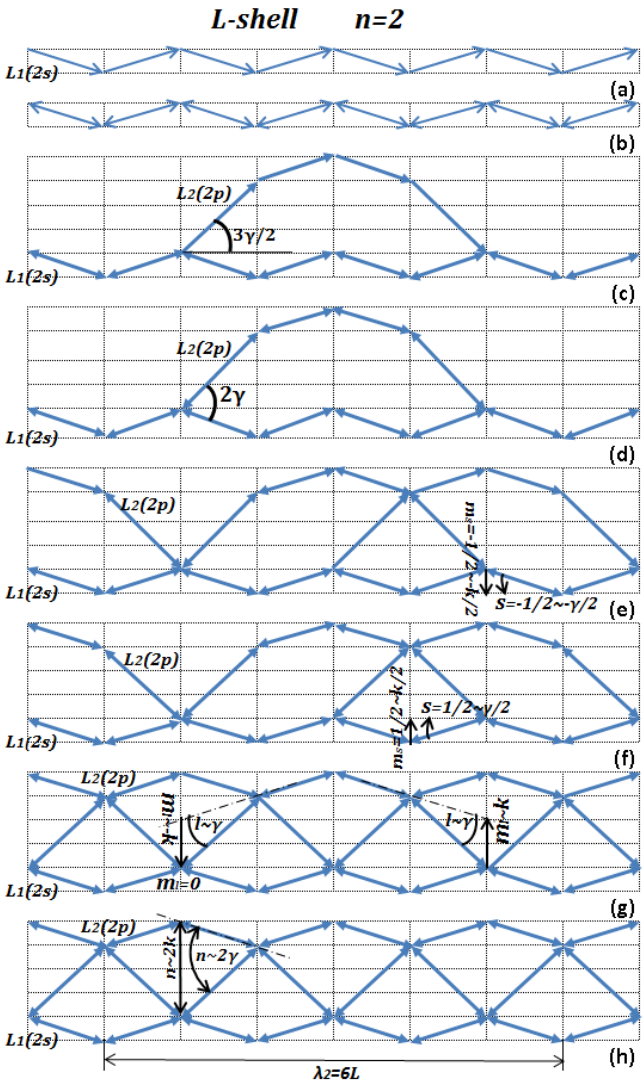

Fig. 15: One of the eight groups of rays of the $\{p=(i+1/2)\gamma,\ q=jk/2\}$ subsystem of periodic trajectories and consecutive process of filling L shell according to Fig. 14 (a and b). Designations of quantum numbers are similar to these in Fig. 13. Dash-dotted lines show axes from which sizes of angles and distances are counted.

shell and the fourth N shell is given in Fig. 16(a). Eight elements of the third part of the periodic table of elements (see Fig. 25 of the Appendix) are given in Fig. 14(b) on a selective basis (see Fig. 16).

Shells of K and L (Fig. 16a) for all the elements shown in Fig. 16(b) are the same as in Fig. 13(c) and Fig. 15(h).

The shell of M ($3s$ subshell) for Na is the same as in Fig. 13(a) or Fig. 13(b).

The subshells $3s$ and $3p$ of the shell of M for Ar and K (Fig. 16b) has the same forms as shown in Fig. 15(h).

The shell of N ($4s$ subshell) for K and Cr (Fig. 16b) is the same as in Fig. 13(a), Fig. 13(b), and Fig. 15(a).

The shell of N ($4s$ subshell) for Sc and Ni is the same as

Electron shell	n	l	m_l	m_s	Sub shell
M	3	0	0	↑↓	$M_1(3s)$
		1	−1	↑↓	$M_2(3p)$
			0	↑↓	
			+1	↑↓	
		2	−2	↑↓	$M_3(3d)$
			−1	↑↓	
			0	↑↓	
			+1	↑↓	
			+2	↑↓	
N	4	0	0	↑↓	$N_1(4s)$
		1	−1	↑↓	$N_2(4p)$
			0	↑↓	
			+1	↑↓	
		2	−2	↑↓	$N_3(4d)$
			−1	↑↓	
			0	↑↓	
			+1	↑↓	
			+2	↑↓	
		3	−3	↑↓	$N_4(4f)$
			−2	↑↓	
			−1	↑↓	
			0	↑↓	
			+1	↑↓	
			+2	↑↓	
			+3	↑↓	

(a)

Element	K	L		M			N	
	$1s$	$2s$	$2p$	$3s$	$3p$	$3d$	$4s$	$4p$
11 Na	2	2	6	1	-	-	-	-
18 Ar	2	2	6	2	6	-	-	-
19 K	2	2	6	2	6	-	1	-
21 Sc	2	2	6	2	6	1	2	-
24 Cr	2	2	6	2	6	5	1	-
28 Ni	2	2	6	2	6	8	2	-
31 Ga	2	2	6	2	6	10	2	1
36 Kr	2	2	6	2	6	10	2	6

(b)

Fig. 16: Shells of M and N, and the third parts of the periodic tables of (a) conditions of an electron, and (b) eight elements.

in Fig. 13(c) and Fig. 15(b).

The shell of N (subshells $4s$ and $4p$) for Ga is the same as in Fig. 15(c).

The shell of N (subshells $4s$ and $4p$) for Kr is the same as in Fig. 15(h).

Fig. 17 (a, b, c, and d) shows the shell of M (subshells 3s, 3p, and 3d) for (Sc), (Cr), (Ni), (Ga and Kr) respectively (Fig. 16a and b):

Three or four shells of K, L, M, and N (Figs. 13 to 17) for the elements can settle down in our geometric model similar to the arrangement of groups of rays of $\boldsymbol{K''}$, $\boldsymbol{L''}$, $\boldsymbol{M''}$, and $\boldsymbol{N''}$ in Figs. 1, 7 and 8.

The geometric schemes of Pauli Principle and elements of the periodic table are also further created in compliance with the above algorithm.

However, our geometric model similar to (Figs. 1 and 6) of groups of rays of the $\{p=(i+1/2)\gamma,\ q=jk/2\}$ subsystem of periodic trajectories consists only of eight groups of rays. Therefore, while remaining within the offered model, it is possible to assume that the number of shells of an atom is no more than eight either. If we continue increasing the number of ray groups to more than eight, the rays will overlap, and the shells will merge.

Thus, if the number of shells does not exceed eight, the total number of elements of the periodic system (14) cannot exceed 128.

Deviations from the sequence of filling the periodic system (e.g., for the elements such as K, etc.) (Fig. 16b) hypothetically reduce (or increase) the total number of elements of the periodic system.

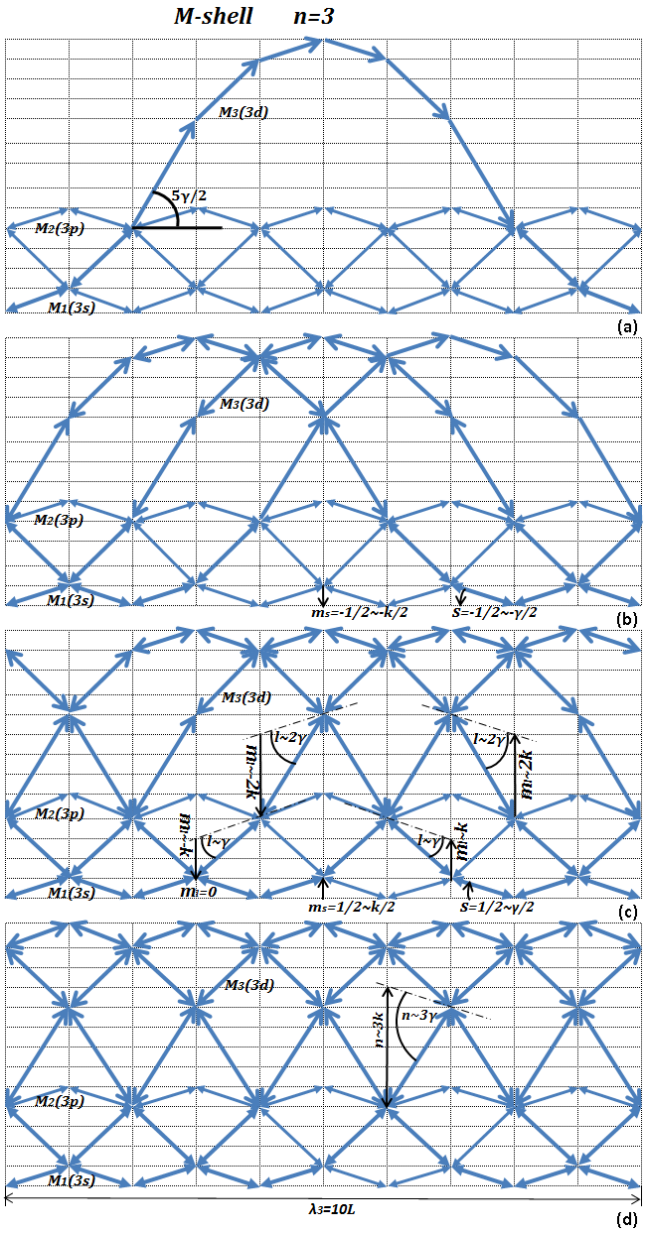

Fig. 17: One of the eight groups of rays of $\{p = (i + 1/2)\gamma,\ q = jk/2\}$ subsystem of periodic trajectories and consecutive process of filling M and N shells for eight elements according to Fig. 16 (a and b). Designations of quantum numbers are similar to these in Figs. 13 and 15. Dash-dotted lines show axes from which sizes of angles and distances are counted.

2.2.4 Pauli Principle and the geometric system of the hydrogen atom

Monographs on quantum mechanics [6] and [7] consider the simplest quantum mechanical system of an atom of hydrogen (Figs. 26 to 28 of the Appendix). Let's make a review of this example too.

Fig. 18 shows the fifth shell of O:

Electron shell	n	ℓ	m_ℓ	m_s	Sub shell
O	5	0	0	⇅	$O_1(5s)$
		1	−1	⇅	$O_2(5p)$
			0	⇅	
			+1	⇅	
		2	−2	⇅	$O_3(5d)$
			−1	⇅	
			0	⇅	
			+1	⇅	
			+2	⇅	
		3	−3	⇅	$O_4(5f)$
			−2	⇅	
			−1	⇅	
			0	⇅	
			+1	⇅	
			+2	⇅	
			+3	⇅	
		5	−4	⇅	$O_5(6g)$
			−3	⇅	
			−2	⇅	
			−1	⇅	
			0	⇅	
			−1	⇅	
			−2	⇅	
			−3	⇅	
			−4	⇅	

Fig. 18: Shell of O of the periodic table of conditions of an electron.

Fig. 19 shows one of the eight groups similar to (Figs. 1 and 6) of rays of the $\{p = (i + 1/2)\gamma,\ q = jk/2\}$ subsystem of periodic trajectories. These trajectories correspond to the fifth electron shell of O shown in Fig. 18. In this example, the O shell is filled completely and contains five subshells.

The $\{p = (i + 1/2)\gamma,\ q = jk/2\}$ subsystem for an atom of hydrogen can be constructed geometrically in accordance with Pauli Principle and similar to the construction method described in previous Sections.

The $\{p = (i + 1/2)\gamma,\ q = jk/2\}$ subsystem shown in Fig. 19 is in many respects similar to the $[p = (i + 1/2)\gamma,\ q = jk/2]$ system in Fig. 10, but contains the smaller quantity of rays and the smaller quantity of the wavy trajectories consisting of these rays.

The wavy trajectories shown in Fig. 19 settle down in the lower part of the coordinate grid and are "attached" to the lower horizontal unlike the wavy trajectories in Fig. 10, which can settle down in any part of the coordinate grid within D size.

In principle, the creation of the O shell in Fig. 19 does not differ from creation of other shells shown in Figs. 13, 15, 17.

Upon comparison of angles multiple p in Fig. 10 and multiple n in Fig. 19, it can be seen that the relationship between these angles is as follows:

$$n \sim |p| + \frac{1}{2}\,. \tag{21}$$

In Fig. 19 we illustrated the allowed quantum transitions [6] and [7]:

$$\Delta n = \pm 1 \quad \text{and} \quad \Delta l = \pm 1 \tag{22}$$

in the form of angles, but not distances as in Figs. 26 to 28 of the Appendix. However, considering ratios (16 to 20) for small angles, sizes (22) can be illustrated (in principle) in the form of distances as well, since:

$$\Delta n = \pm 1 \sim \pm\gamma \sim \pm k \quad \text{and} \quad \Delta l = \pm 1 \sim \pm\gamma \sim \pm k\,. \tag{23}$$

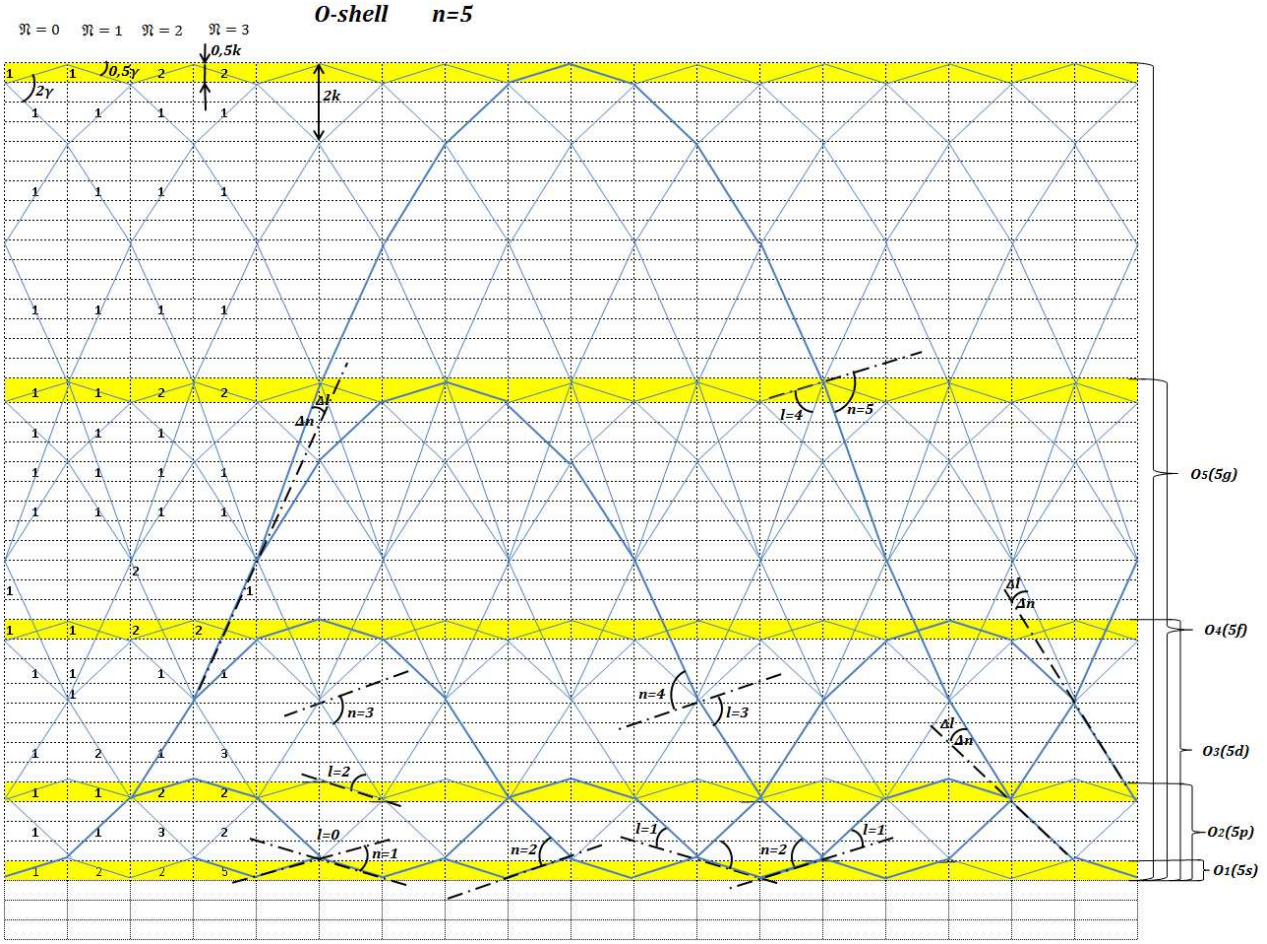

Fig. 19: One of the eight groups of rays of the considered $\{p = (i + 1/2)\gamma,\ q = jk/2\}$ subsystem of periodic (wavy) trajectories. The shell of O is completely filled according to Fig. 18. This group contains 15 links ($\mathbb{K} = 15$) within one pass of $\mathfrak{R}$. Troughs of the "waves" are located ("attached to") only on the lower horizontal, while crests of the "waves" are located ("attached to") on several higher horizontals. The horizontals are marked in separate color. These horizontals have thickness of $0,5k$ and are located at the increasing distance of $2kn$ from each other from the bottom upwards. The designations of the quantum numbers are similar to these in Figs. 13 to 17. Dash-dotted lines show axes from which sizes of angles and distances are counted. Quantum transitions $\Delta n = \pm 1$ and $\Delta l = \pm 1$ are shown in the form of angles between dash-dotted lines. On the right, location of five subshells is indicated by curly braces. Figures put next to links (for $\mathfrak{R} = 0 - 3$) show the number of rays of $\mathbb{N}$ and the process of summation of rays spreading along the links.

Fig. 20 gives images of layers of a nonlinear arithmetic parallelepiped and a numerical example of calculation of O shell of the $\{p = (i + 1/2)\gamma,\ q = jk/2\}$ subsystem of periodic trajectories (Fig. 19) for zero and the subsequent three passes of rays, i.e. for $\mathfrak{R} = 0 - 4$. Four rectangles shown in Fig. 20 are the layers of the nonlinear arithmetic parallelepiped.

n and l values are given on the right in Fig. 20 with taking into account the ratios (18 and 21).

The calculation was made according to the rule (3) of consecutive filling of an arithmetic rectangle with numbers taking into account the *first* and the *second* threshold (4 and 5; 8 and 9) and initial (6, 7, and 10) conditions, including the algorithm (8 to 10).

Each of the four layers of the arithmetic parallelepiped shown in Fig. 20 is similar to the layer represented in Fig. 11, but there are differences as well. Therefore, it is necessary to set the *third* threshold conditions. We took these conditions for our example directly from Fig. 19. As this approach is illustrative for us, the total number of rays is not that big. All the wavy trajectories are "attached" to the lower horizontal. The *third* threshold conditions can be set in other illustrative ways, e.g., by means of special nomograms.

Thus, we made calculations in Excel according to expression (3). The appropriate formulas for the respective p and q can be taken from Fig. 3 or Fig. 7, for example.

Fig. 21 shows results of numerical calculations of layers of the nonlinear arithmetic parallelepiped for O shell (Fig. 20) in Excel are given in the form of envelopes of distribution of the ray number. In a, c, e, g, and i (the left column), you can see the envelopes of distribution of the ray number at the an-

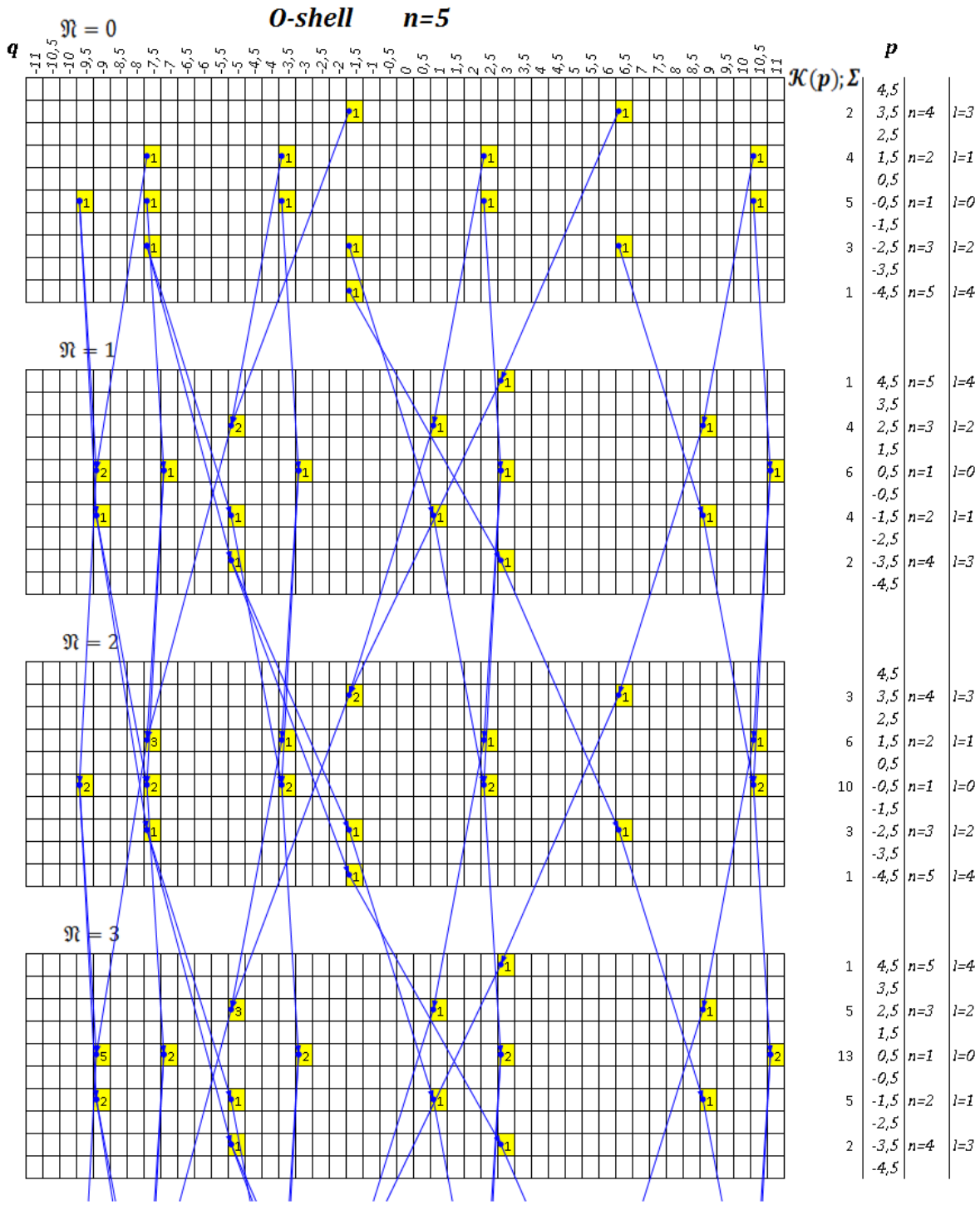

Fig. 20: One of the eight groups of rays of the $\{p=(i+1/2)\gamma,\ q=jk/2\}$ subsystem of periodic (wavy) trajectories. The O shell is completely filled according to Figs. 18 and 19. Numerical calculation in Excel was made for the first four passes of rays (iterations). Arrows show dependent cells. Fifteen highlighted cells within one pass of $\mathfrak{N}$ (one layer of a parallelepiped) correspond to fifteen links ($\mathbb{K}=15$) within one pass of $\mathfrak{N}$ shown in Fig. 19. Figures in the highlighted cells correspond to the number of rays of $\mathbb{N}$ extending along the links of $\mathbb{K}$.

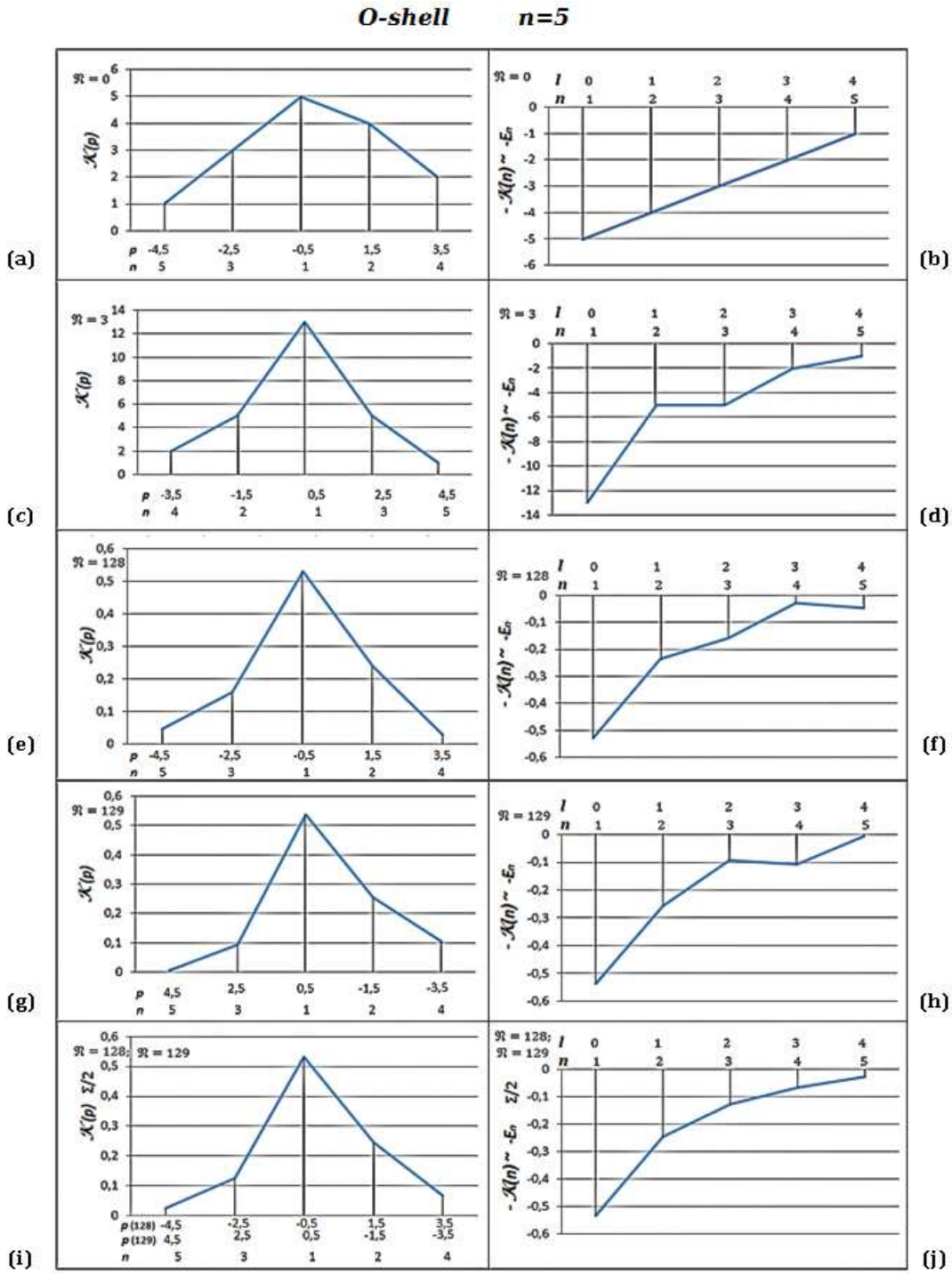

Fig. 21: This Figure shows the results of numerical calculations (Fig. 20) for $\mathfrak{N} = 0$, 32, 128, 129 (a to h) of the completely filled O shell for the $\{p = (i + 1/2)\gamma,\ q = jk/2\}$ subsystem of periodic (wavy) trajectories. The envelopes of distribution of the ray number at the angle $\mathcal{K}(p)$ are given in the left column. The envelopes of distribution of the ray number at the angle $\mathcal{K}(n)$ are given in the right column. The shared envelopes for $\mathfrak{N} = 128, 129$ are given in (i and j). Graphs (e to j) are shown in a normalized form. Note, according to our calculations, in this case, there is virtually no change in the form of the envelope approximately after the 70th pass.

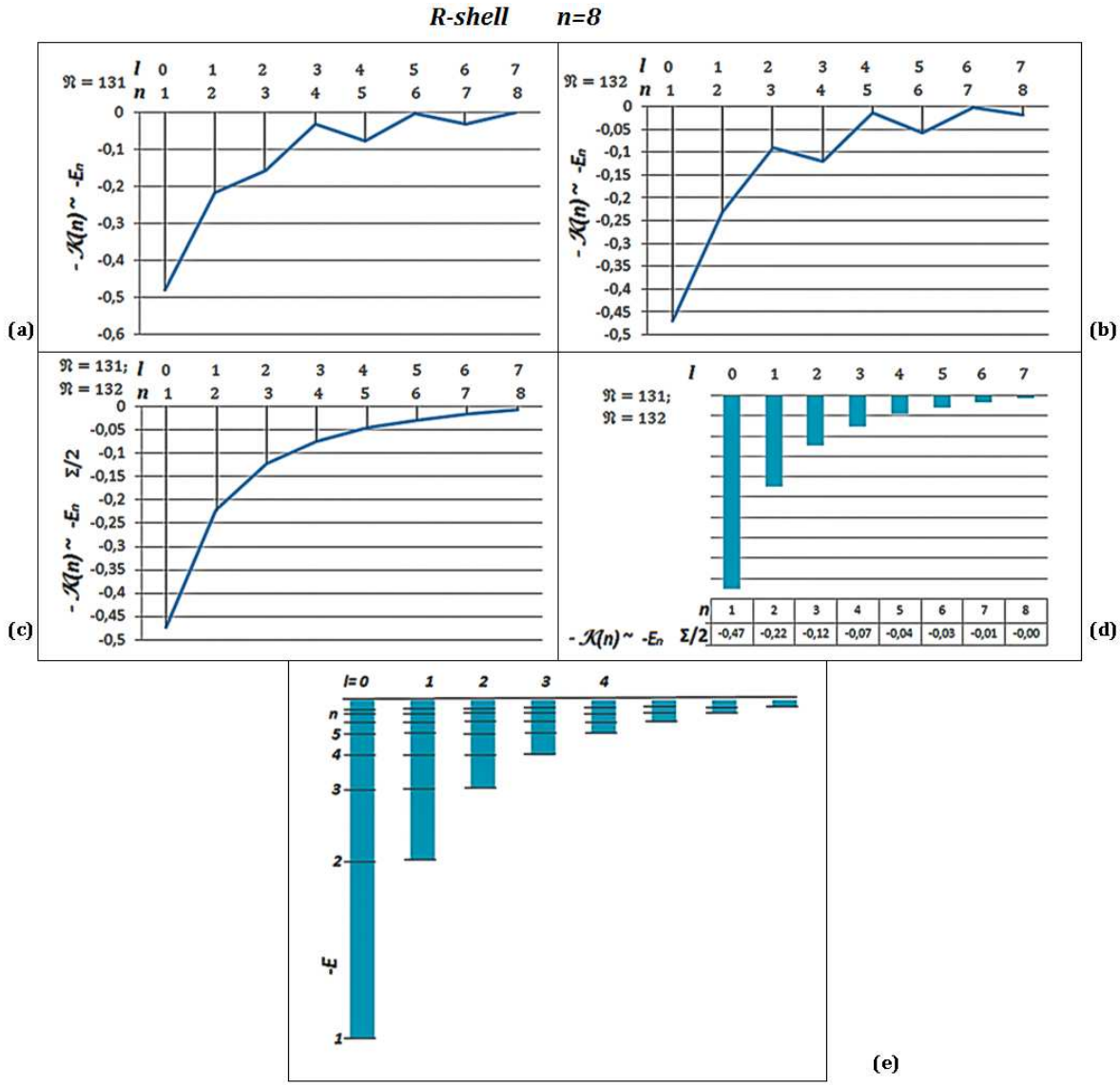

Fig. 22: This Figure shows results of numerical calculations of the completely filled 8th shell of R for the $\{p=(i+1/2)\gamma,\ q=jk/2\}$ subsystem of periodic (wavy) trajectories. In (a) and (b), you can see the envelopes of distribution of the ray number at the angle $\mathcal{K}(n)$ for passes $\mathfrak{R}=131,132$. Results of joint calculations for passes $\mathfrak{R}=131$ and $\mathfrak{R}=132$ in the form of the envelope are given in (c) and these in a form of histograms are given in (d) and (e). Envelopes (a to c) and histograms (d and e) are presented in the normalized form. Note, according to our calculations, in this case, there is virtually no change in the form of envelopes approximately after the 100th pass ($\mathfrak{R}=100$). The histogram in (e) is similar to Fig. 28 of the Appendix.

gle $\mathcal{K}(p)$. In b, d, f, h, and j (the right column), you can see the envelopes of distribution of the ray number at the angle $\mathcal{K}(n)$ (taking into account expression (21). The results of numerical calculation for zero pass of rays, i.e. for $\mathfrak{R}=0$, are given in a and b; for $\mathfrak{R}=3$ — in c and d; for $\mathfrak{R}=128$ — in e and f; and for $\mathfrak{R}=129$ — in g and h.

Shared graphs of $\mathcal{K}(p)$ and $\mathcal{K}(n)$ for $\mathfrak{R}=128$ and $\mathfrak{R}=129$ are given in i and j, namely:

$$\mathcal{K}(p)_{\mathfrak{R}=128,\ \mathfrak{R}=129}=\frac{1}{2}\left[\mathcal{K}(p)_{\mathfrak{R}=128}+\mathcal{K}(p)_{\mathfrak{R}=129}\right] \quad (24)$$

and

$$\mathcal{K}(n)_{\mathfrak{R}=128,\ \mathfrak{R}=129}=\frac{1}{2}\left[\mathcal{K}(n)_{\mathfrak{R}=128}+\mathcal{K}(n)_{\mathfrak{R}=129}\right], \quad (25)$$

where $\mathcal{K}(p)$ and $\mathcal{K}(n)$ are in the normalized form.

In this work, like in previous works [2] to [5], we assume that the number of rays of $\mathbb{N}$ extending along the number of multiplicative links of $\mathbb{K}$ is proportionate to energy. For negative energy of an electron — E extending along these rays,

we have the following:

$$|E_n| \sim \mathcal{K}(n), \tag{26}$$

$$|E_n| \sim \mathcal{K}(n)_{\mathfrak{R},\,\mathfrak{R}+1}. \tag{27}$$

If we consider Fig. 21 (f and h) and especially Fig. 21(j), we will see that the form of the envelope after a large number of passes resembles more of a hyperbole of the following type:

$$E_n \sim -1/n^2. \tag{28}$$

This ratio obtained from our geometric constructions corresponds to experimental results in spectroscopy and theoretical results of Bohr's theory and quantum mechanics [6] and [7].

Fig. 22 shows the results (similar to those that were shown in Fig. 21 f, h, and j) of numerical calculations of layers of the nonlinear arithmetic parallelepiped for the eighth R shell in the form of envelopes of distribution of the ray number. In (a) and (b), you can see the envelopes of distribution of the ray number at the angle $\mathcal{K}(n)$ for passes of rays $\mathfrak{R} = 131$, 132. Results of joint calculations for passes of $\mathfrak{R} = 131$ and $\mathfrak{R} = 132$ in the form of the envelope are given in (c), and these in a form of histograms are given in (d) and (e).

If we consider Fig. 22 (a and b) and especially Fig. 22 (c to e), we will see that the form of the envelope after a large number of passes resembles more of a hyperbole (28).

Fig. 22 (e) similar to Fig. 22 (c and d) should be compared to Fig. 28 of the Appendix.

Our numerical calculations show that with an increase in the number of passes of $\mathfrak{R}$, and an increase in the number of subshells of a shell and the main number n, the form of an envelope, Fig. 21 (j) and Fig. 22 (c), increasingly resembles a hyperbole of (28) type. If the number of subshells exceeds eight (e.g., you can construct eleven), eight of eleven subshells can be subsumed to subshells, while the rest three can be subsumed to a continuous spectrum [6] and [7]. Such creation of a continuous spectrum does not contradict Pauli Principle.

Conclusions

In our illustrative geometric researches, using just one basic summation formula of $A = B + C$ (3), Excel, and various initial and threshold conditions set, we have revealed a number of new regularities like we did in previous works [3] and [4].

It appeared that quantum systems can be geometrically interpreted by means of our model of a half-integer rays system in an illustrative way.

We have described Pauli Principle, shells and subshells of atoms of the periodic table. At the same time, the number of shells and subshells in our model does not exceed eight, and all the subshells starting with the ninth can be considered a continuous spectrum.

By means of our model, it is possible to interpret the principle, azimuthal, magnetic, and spin quantum numbers in the form of angles and distances.

By means of our model, we have given a separate geometric interpretation of an atom of hydrogen and its power levels. We have interpreted transitions of an electron from one level to another in the form of angles, but not distances as it is commonly interpreted [6] and [7]. In this work, we have also shown that the hyperbolic dependence of energy of a hydrogen atom of $E_n \sim -1/n^2$ (28) known from experimental spectral studies, Bohr's theory and quantum mechanics, can be also obtained from our geometric constructions on the basis of Pauli Principle.

Based on the research of a half-integer ray model, we have illustrated the stepped structure of layers at laminar flow of liquid [8]. The similar stepped structure was observed in research of integer ray model made by us in [3], but the half-integer model gives the more accurate image of "steps" in comparison with our integer model.

Acknowledgements

The author expresses gratitude to the late Prof. E. E. Shnoll, Prof. J. Peters (Canada), Prof. V. V. Dikusar (Russia), Prof. S. E. Shnoll (Russia), Prof. A. A. Rukhadze (Russia), Prof. V. G. Mikhalevich (Russia), Prof. A. V. Kaganov (Russia), Prof. R. Mehta (India), Dr. A. Shemetov (Russia), Dr. O. Nersesyan (Russia) and Captain V. Kabanov (Russia) for useful discussion and support. The author is also grateful to the Prof. N. J. A. Sloane (USA) for his surprising tables and to the Excel software creators for their excellent Excel.

Submitted on December 26, 2015 / Accepted on January 19, 2015

References

1. Peters J. F., Tozzi A. The Borsuk-Ulam theorem explanes quantum entangement. *Technical Reports*, November 2015, DOI: 10.13140/RG.2.1.3860.1685.
2. Yurkin A. V. Quasi-resonator a new interpretation of scattering in lasers. *Quantum Electronics*, 1994, v. 24, 359.
3. Yurkin A. V. Symmetric Triangle of Pascal and Arithmetic Parallelepiped. On Possibility of New Evident Geometrical Interpretation of Processes in Long Pipes. Lambert Academic Publishing, 2015.
4. Yurkin A. V. New Binomial and New View on Light Theory. About One New Universal Descriptive Geometric Model. Lambert Academic Publishing, 2013.
5. Yurkin A. V. Recurrence calculation of laser divergence and refractive analog of a multilobe mirror. *Quantum Electronics*, 1993, v. 23, 323.
6. Putilov K. A., Fabrikant V. A. The Course of Physics. Vol. 3. Moscow, Fizmatgiz, 1960 (in Russian).
7. Savelyev I. V. The General Course of Physics. Vol. 3. Moscow, Nauka, 1982 (in Russian).
8. Putilov K. A. The Course of Physics. Vol. 1. Moscow, Fizmatgiz, 1954 (in Russian).

Appendix: reference tables

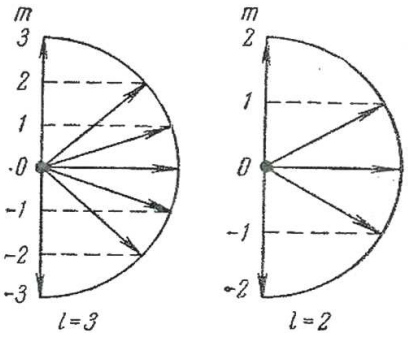

Fig. 23: Illustration of the principle of spatial quantization. Possible values of projections of the orbital momentum to the direction of a magnetic field for $l = 3$ and $l = 2$. (Fig. 231 from [6]).

Shell	n	l	m_l	m_s	Subshell	Shell	n	l	m_l	m_s	Subshell
K	1	0	0	↑↓	$K(1s)$	N	4	0	0	↑↓	$N_1(4s)$
L	2	0	0	↑↓	$L_1(2s)$			1	−1	↑↓	$N_2(4p)$
		1	−1	↑↓	$L_2(2p)$				0	↑↓	
			0	↑↓					+1	↑↓	
			+1	↑↓				2	−2	↑↓	$N_3(4d)$
M	3	0	0	↑↓	$M_1(3s)$				−1	↑↓	
		1	−1	↑↓	$M_2(3p)$				0	↑↓	
			0	↑↓					+1	↑↓	
			+1	↑↓					+2	↑↓	
		2	−2	↑↓	$M_3(3d)$			3	−3	↑↓	$N_4(4f)$
			−1	↑↓					−2	↑↓	
			0	↑↓					−1	↑↓	
			+1	↑↓					0	↑↓	
			+2	↑↓					+1	↑↓	
									+2	↑↓	
									+3	↑↓	

Fig. 24: Division of possible conditions of an electron in an atom into shells and subshells. (Table 36.1 from [7]).

Element	K	L		M			N		Basic therm
	1s	2s	2p	3s	3p	3d	4s	4p	
1 H	1	—	—	—	—	—	—	—	$^2S_{1/2}$
2 He	2	—	—	—	—	—	—	—	$^1S_0$
3 Li	2	1	—	—	—	—	—	—	$^2S_{1/2}$
4 Be	2	2	—	—	—	—	—	—	$^1S_0$
5 B	2	2	1	—	—	—	—	—	$^2P_{1/2}$
6 C	2	2	2	—	—	—	—	—	$^3P_0$
7 N	2	2	3	—	—	—	—	—	$^4S_{3/2}$
8 O	2	2	4	—	—	—	—	—	$^3P_2$
9 F	2	2	5	—	—	—	—	—	$^2P_{3/2}$
10 Ne	2	2	6	—	—	—	—	—	$^1S_0$
11 Na	2	8		1	—	—	—	—	$^2S_{1/2}$
12 Mg	2	8		2	—	—	—	—	$^1S_0$
13 Al	2	8		2	1	—	—	—	$^2P_{1/2}$
14 Si	2	8		2	2	—	—	—	$^3P_0$
15 P	2	8		2	3	—	—	—	$^4S_{3/2}$
16 S	2	8		2	4	—	—	—	$^3P_2$
17 Cl	2	8		2	5	—	—	—	$^2P_{3/2}$
18 Ar	2	8		2	6	—	—	—	$^1S_0$
19 K	2	8		8		—	1	—	$^2S_{1/2}$
20 Ca	2	8		8		—	2	—	$^1S_0$
21 Sc	2	8		8		1	2	—	$^2D_{3/2}$
22 Ti	2	8		8		2	2	—	$^3F_2$
23 V	2	8		8		3	2	—	$^4F_{3/2}$
24 Cr	2	8		8		5	1	—	$^7S_3$
25 Mn	2	8		8		5	2	—	$^6S_{5/2}$
26 Fe	2	8		8		6	2	—	$^5D_4$
27 Co	2	8		8		7	2	—	$^4F_{9/2}$
28 Ni	2	8		8		8	2	—	$^3F_4$
29 Cu	2	8		8		10	1	—	$^2S_{1/2}$
30 Zn	2	8		8		10	2	—	$^1S_0$
31 Ga	2	8		8		10	2	1	$^2P_{1/2}$
32 Ge	2	8		8		10	2	2	$^3P_0$
33 As	2	8		8		10	2	3	$^4S_{3/2}$
34 Se	2	8		8		10	2	4	$^3P_2$
35 Br	2	8		8		10	2	5	$^2P_{3/2}$
36 Kr	2	8		8		10	2	6	$^1S_0$

Fig. 25: The process of building electron shells of the first 36 elements of the periodic system. (Table 37.1 from [7]).

Appendix: reference tables (continue)

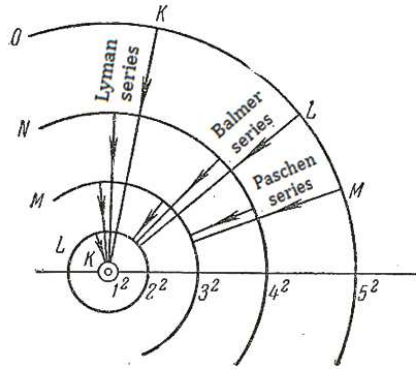

Fig. 26: Orbits of a hydrogen atom in Bohr's theory. The radial arrows located between circles show transitions of an electron from one level to another. (Fig. 228 from [6]).

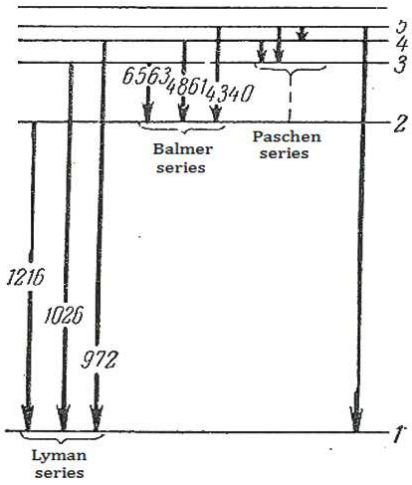

Fig. 27: Scheme of levels of energy of a hydrogen atom. The vertical arrows located between horizontal lines show transitions of an electron from one level to another. (Fig. 229 from [6]).

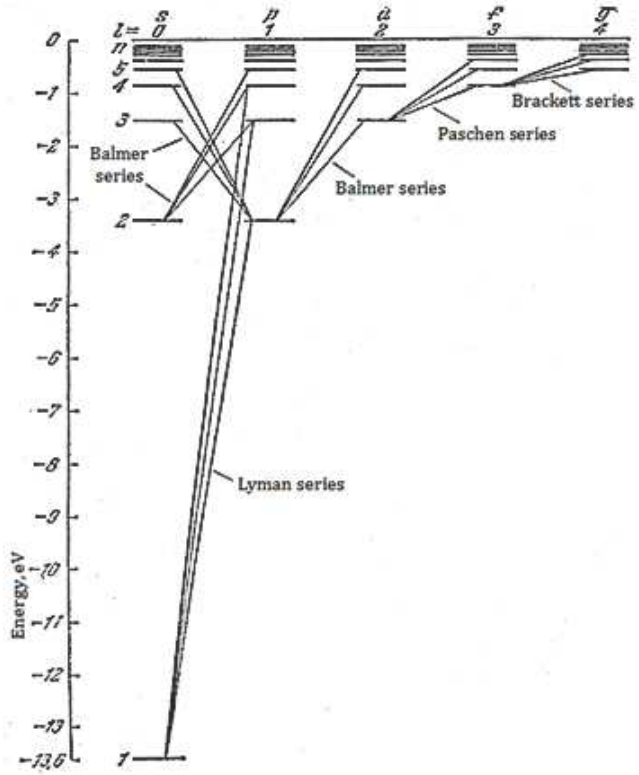

Fig. 28: Scheme of levels of energy of a hydrogen atom. The inclined lines located between horizontal lines show transitions of an electron from one level to another according to the rule of selection $\Delta l = \pm 1$. It means that only transitions upon which l changes by unit are possible. (Fig. 28.1 from [7]).

$X(5)$ Symmetry to ^{152}Sm

Salah A. Eid[1] and Sohair M. Diab[2]

[1]Faculty of Engineering, Phys. Dept., Ain Shams University, Cairo, Egypt.
[2]Faculty of Education, Phys. Dept., Ain Shams University, Cairo, Egypt.
E-mail: mppe2@yahoo.co.uk

The excited positive and negative parity states, potential energy surfaces, $V(\beta,\gamma)$, electromagnetic transition probabilities, $B(E1)$, $B(E2)$, electric monopole strength $X(E0/E2)$ and staggering effect, $\Delta I = 1$, were calculated successfully using the interacting boson approximation model IBA-1. The calculated values are compared to the available experimental data and show reasonable agreement. The energy ratios and contour plot of the potential energy surfaces show that ^{152}Sm is an $X(5)$ candidate.

1 Introduction

Phase transition is one of the very interesting topic in nuclear structure physics. The even-even samarium series of isotopes have encouraged many authors to study that area extensively experimentally and theoretically.

Experimentally, authors studied levels energy with their half-lives, transition probabilities, decay schemes, multipole mixing ratios, internal conversion coefficients, angular correlations and nuclear orientation of γ-rays[1-4].

Theoretically, different theoretical models have been applied to that chain of isotopes. One of the very interesting models is the interacting boson approximation model IBA [5-10]. Iachello [11,12] has made an important contribution by introducing the new dynamical symmetries $E(5)$ and $X(5)$.

$E(5)$ is the critical point symmetry of phase transition between $U(5)$ and $O(6)$ while $X(5)$ is between $U(5)$ and $SU(3)$ nuclei. The aim of the present work is to calculate:

1. The potential energy surfaces, $V(\beta,\gamma)$;
2. The levels energy, electromagnetic transition rates $B(E1)$ and $B(E2)$;
3. The staggering effect, and
4. The electric monopole strength $X(E0/E2)$.

2 IBA-1 model

2.1 Levels energy

The IBA-1 Hamiltonian [13-16] employed on ^{152}Sm in the present calculation is:

$$\begin{aligned} H &= EPS\cdot n_d + PAIR\cdot(P\cdot P) \\ &+\frac{1}{2}\,ELL\cdot(L\cdot L)+\frac{1}{2}\,QQ\cdot(Q\cdot Q) \\ &+5\,OCT\cdot(T_3\cdot T_3)+5HEX\cdot(T_4\cdot T_4)\,, \end{aligned} \tag{1}$$

where

$$P\cdot p=\frac{1}{2}\begin{bmatrix} \left\{(s^\dagger s^\dagger)^{(0)}_0-\sqrt{5}(d^\dagger d^\dagger)^{(0)}_0\right\}x \\ \left\{(ss)^{(0)}_0-\sqrt{5}(\tilde{d}\tilde{d})^{(0)}_0\right\}\end{bmatrix}^{(0)}_0, \tag{2}$$

$$L\cdot L=-10\sqrt{3}\left[(d^\dagger\tilde{d})^{(1)}x\,(d^\dagger\tilde{d})^{(1)}\right]^{(0)}_0\,, \tag{3}$$

$$Q\cdot Q=\sqrt{5}\begin{bmatrix} \left\{(S^\dagger\tilde{d}+d^\dagger s)^{(2)}-\dfrac{\sqrt{7}}{2}\,(d^\dagger\tilde{d})^{(2)}\right\}x \\ \left\{(s^\dagger\tilde{d}++\tilde{d}s)^{(2)}-\dfrac{\sqrt{7}}{2}\,(d^\dagger\tilde{d})^{(2)}\right\}\end{bmatrix}^{(0)}_0, \tag{4}$$

$$T_3\cdot T_3=-\sqrt{7}\left[(d^\dagger\tilde{d})^{(2)}x\,(d^\dagger\tilde{d})^{(2)}\right]^{(0)}_0\,, \tag{5}$$

$$T_4\cdot T_4=3\left[(d^\dagger\tilde{d})^{(4)}x\,(d^\dagger\tilde{d})^{(4)}\right]^{(0)}_0\,. \tag{6}$$

In the previous formulas, n_d is the number of bosons; $P\cdot P$, $L\cdot L$, $Q\cdot Q$, $T_3\cdot T_3$ and $T_4\cdot T_4$ represent pairing, angular momentum, quadrupole, octupole and hexadecupole interactions respectively between the bosons; EPS is the boson energy; and $PAIR$, ELL, QQ, OCT, HEX are the strengths of the pairing, angular momentum, quadrupole, octupole and hexadecupole interactions respectively (see Table 1).

2.2 Transition rates

The electric quadrupole transition operator employed is:

$$\begin{aligned} T^{(E2)} &= E2S\,D\cdot(s^\dagger\tilde{d}+d^\dagger s)^{(2)}+ \\ &+\frac{1}{\sqrt{5}}\,E2DD\cdot(d^\dagger\tilde{d})^{(2)}\,. \end{aligned} \tag{7}$$

$E2S\,D$ and $E2DD$ are adjustable parameters.

The reduced electric quadrupole transition rates between $I_i\to I_f$ states are given by:

$$B(E_2,I_i-I_f)=\frac{[<I_f\parallel T^{(E_2)}\parallel I_i>]^2}{2I_i+1}\,. \tag{8}$$

3 Results and discussion

In this section we review and discuss the results.

nucleus	*EPS*	*PAIR*	*ELL*	*QQ*	*OCT*	*HEX*	*E2S D*(*eb*)	*E2DD*(*eb*)
^{152}Sm	0.3840	0.000	0.0084	−0.0244	0.0000	0.0000	0.1450	−0.4289

Table 1: Parameters used in *IBA*-1 Hamiltonian (all in MeV).

3.1 The potential energy surfaces

The potential energy surfaces [17], $V(\beta, \gamma)$, as a function of the deformation parameters β and γ are calculated using:

$$\begin{aligned} E_{N_\Pi N_\nu}(\beta,\gamma) &= <N_\pi N_\nu;\beta\gamma|H_{\pi\nu}|N_\pi N_\nu;\beta\gamma> = \\ &= \zeta_d(N_\nu N_\pi)\beta^2(1+\beta^2)+\beta^2(1+\beta^2)^{-2}\times \\ &\times\left\{kN_\nu N_\pi[4-(\bar{X}_\pi\bar{X}_\nu)\beta\cos 3\gamma]\right\}+ \\ &+\left\{[\bar{X}_\pi\bar{X}_\nu\beta^2]+N_\nu(N_\nu-1)\left(\frac{1}{10}\,c_0+\frac{1}{7}\,c_2\right)\beta^2\right\}, \end{aligned} \tag{9}$$

where

$$\bar{X}_\rho=\left(\frac{2}{7}\right)^{0.5} X_\rho\,\rho=\pi \text{ or } \upsilon\,. \tag{10}$$

The calculated potential energy surfaces, $V(\beta,\gamma)$, are presented in Figures 1, 2, 3. ^{152}Sm lies between ^{150}Sm which is a vibrational like nucleus, $U(5)$, Fig. 1, while ^{154}Sm is a rotational like, $SU(3)$, nucleus, Fig. 3. So, ^{150}Sm can be an $X(5)$ candidate where levels energy, transition probability ratios as well as the potential energy surfaces are supporting that assumption (see Table 2).

3.2 Energy spectra and electric transition rates

The energy of the positive and negative parity states of ^{152}Sm isotope are calculated using computer code PHINT [19]. A comparison between the experimental spectra [18] and our calculations, using values of the model parameters given in Table 1 for the ground state, $\beta1$, $\beta2$ and γ bands are illustrated in Fig. 4. The agreement between the calculated levels energy and their corresponding experimental values are fair, but they are slightly higher especially for the higher excited states in $\beta1$, $\beta2$ and γ bands. We believe this is due to the change of the projection of the angular momentum which is due mainly to band crossing. Fig. 5 shows the position of $X(5)$ and $E(5)$ between the other types of nuclei.

Unfortunately there are no available measurements of electromagnetic transition rates $B(E1)$ for ^{152}Sm nucleus, Table 3, while some of $B(E2)$ are measured. The measured $B(E2, 2_1^+ \to 0_1^+)$ is presented, in Table 4, for comparison with the calculated values [20]. The parameters *E2S D* and *E2DD* displayed in Table 1 are used in the computer code NPBEM [19] for calculating the electromagnetic transition rates and the calculated values are normalized to $B(E2, 2_1^+ \to 0_1^+)$. No new parameters are introduced for calculating electromagnetic transition rates $B(E1)$ and $B(E2)$ of intraband and interband.

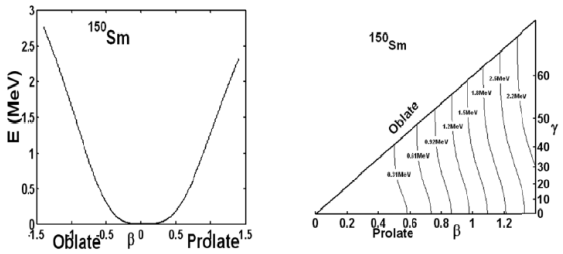

Fig. 1: Potential energy surfaces for ^{150}Sm .

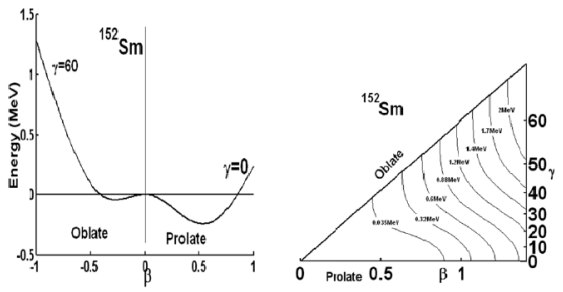

Fig. 2: Potential energy surfaces for ^{152}Sm .

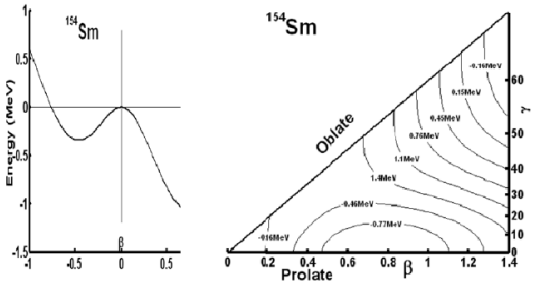

Fig. 3: Potential energy surfaces for ^{154}Sm .

nucleus	$E_{4_1^+}/E_{2_1^+}$	$E_{6_1^+}/E_{2_1^+}$	$E_{8_1^+}/E_{2_1^+}$	$E_{0_2^+}/E_{2_1^+}$	$E_{6_1^+}/E_{0_2^+}$	$E_{0_3^+}/E_{2_1^+}$	$BE2(4_1^+-2_1^+)/BE2(2_1^+-0_1^+)$
^{152}Sm	3.02	5.83	9.29	5.66	1.03	8.92	1.53
$X(5)$	3.02	5.83	9.29	5.65	1.53	6.03	1.58

Table 2: Energy and transition probability ratios.

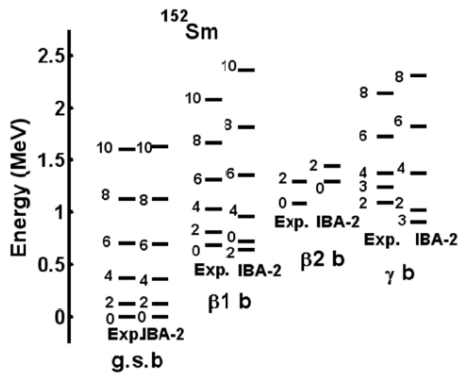

Fig. 4: Experimental[18] and calculated levels energy

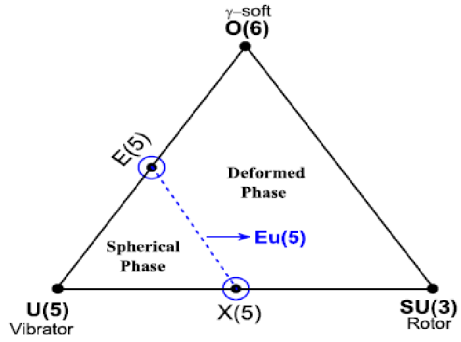

Fig. 5: Triangle showing the position of $X(5)$ and $E(5)$.

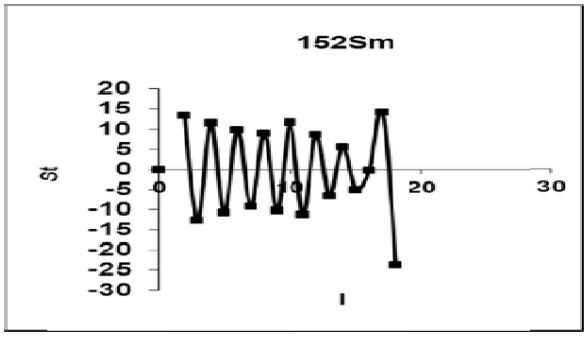

Fig. 6: Staggering effect on ^{152}Sm.

I_i^- I_f^+	$B(E1)$Exp.	$B(E1)$IBA-1
1_1 0_1	—-	0.0979
1_1 0_2	—-	0.0814
3_1 2_1	—-	0.2338
3_1 2_2	—-	0.0766
3_1 2_3	—-	0.0106
3_2 2_1	—-	0.0269
3_2 2_2	—-	0.0291
3_2 2_3	—-	0.0434
5_1 4_1	—-	0.3579
5_1 4_2	—-	0.0672
5_1 4_3	—-	0.0050
7_1 6_1	—-	0.4815
7_1 6_2	—-	0.0574
9_1 8_1	—-	0.6075
9_1 8_2	—-	0.0490
11_1 10_1	—-	0.7367
11_1 10_2	—-	0.0413

Table 3: Calculated $B(E1)$ in ^{152}Sm.

3.3 Staggering effect

The presence of $(+ve)$ and $(-ve)$ parity states has encouraged us to study the staggering effect [21-23] for ^{152}Sm isotope using staggering function equations (11, 12) with the help of the available experimental data [18].

$$St(I) = 6\Delta E(I) - 4\Delta E(I-1) - 4\Delta E(I+1) + \\ + \Delta E(I+2) + \Delta E(I-2)\,, \qquad (11)$$

with

$$\Delta E(I) = E(I+1) - E(I)\,. \qquad (12)$$

The calculated staggering patterns are illustrated in Fig. 6 and show an interaction between the $(+ve)$ and $(-ve)$ parity states for the ground state band of ^{152}Sm.

3.4 Electric monopole transitions

The electric monopole transitions, $E0$, are normally occurring between two states of the same spin and parity by transferring energy and zero unit of angular momentum. The strength of the electric monopole transition, $X_{if'f}(E0/E2)$, [24] can be calculated using equations (13, 14) and presented in Table 5.

$$X_{if'f}(E0/E2) = \frac{B(E0, I_i - I_f)}{B(E2, I_i - I_{'f})}\,, \qquad (13)$$

I_i^+ I_f^+	$B(E2)$Exp*.	$B(E2)$IBA-1
2_1 0_1	0.670(15)	0.6529
3_1 2_1	—-	0.0168
4_1 2_1	0.1.017(4)	1.0014
6_1 4_1	1.179(33)	1.1304
0_2 2_1	0.176(1)	0.3363
2_2 2_1	0.0258(26)	0.0610
2_2 4_1	0.091(11)	0.1057
4_2 2_1	0.0035(35)	0.0003
4_2 4_1	0.037(23)	0.0458
2_3 0_1	0.0163(11)	0.0141
2_3 2_1	0.0417(42)	0.0125
2_3 4_1	0.0416(32)	0.0296
4_3 2_1	0.0035(13)	0.0038
4_3 4_1	0.037(13)	0.0084
4_3 4_2	—-	0.1235
4_3 2_2	—-	0.0070
4_3 2_3	—-	0.3110
4_2 2_2	—-	0.6418
8_1 6_1	—-	1.1681
8_1 6_2	—-	0.0376
10_1 8_1	—-	1.1421

Table 4: Calculated $B(E2)$ in ^{152}Sm (* from Ref.[20])

I_i^+ I_f^+	X(E0/E2)Exp*.	X(E0/E2)IBA-1
0_2 0_1	0.7(0.1)	0.85
0_3 0_2	—-	3.68
0_3 0_1	—-	0.72
0_4 0_3	—-	4.39
0_4 0_2	—-	0.64
0_4 0_1	—-	1.27
2_2 2_1	4.5(0.5)	3.52
2_3 2_1	—-	12.23
2_3 2_2	—-	11.19
4_3 4_1	—-	1.76
4_3 4_2	—-	1.40
4_4 4_1	—-	0.44
4_4 4_2	—-	3.15
4_2 4_1	6.6(2.10)	2.02
6_2 6_1	—-	1.46
8_2 8_1	—-	1.20
10_2 10_1	—-	1.07

Table 5: $X_{if'f}$ $(E0/E2)$ ratios in ^{152}Sm (* from Ref [20]).

where I_i =I_f=0 , $I_{'f}$=2 and I_i= $I_f \neq 0$, I_f = $I_{'f}$.

$$X_{if'f}(E0/E2) = (2.54{\times}10^9)\,A^{3/4} \times \times \frac{E_\gamma^5(\mathrm{MeV})}{\Omega_{KL}}\,\alpha(E2)\,\frac{T_e(E0, I_i - I_f)}{T_e(E_2, I_i - I_{'f})}\,. \qquad (14)$$

where:
A : mass number;
I_i : spin of the initial state where E0 and E2 transitions are depopulating it;
I_f : spin of the final state of E0 transition;
$I_{'f}$: spin of the final state of E2 transition;
E_γ : gamma ray energy;
Ω_{KL} : electronic factor for K,L shells [25];
$\alpha(E2)$: conversion coefficient of the E2 transition;
$T_e(E0, I_i - I_f)$: absolute transition probability of the $E0$ transition between I_i and I_f states, and
$T_e(E_2, I_i - I_{'f})$: absolute transition probability of the $E2$ transition between I_i and $I_{'f}$ states.

3.5 Conclusions

The IBA-1 model has been applied successfully to the ^{152}Sm isotope and:

1. Levels energy are successfully reproduced;
2. Potential energy surfaces are calculated and show $X(5)$ characters to ^{152}Sm;
3. Electromagnetic transition rates $B(E1)$ and $B(E2)$ are calculated;
4. Staggering effect has been calculated and beat pattern observed which show an interaction between the $(-ve)$ and $(+ve)$ parity states, and
5. Strength of the electric monopole transitions $X_{if'f}(E0/E2)$ are calculated.

Submitted on January 10, 2016 / Accepted on January 12, 2016

References

1. Qian Y., Ren Z. and Ni D. α-decay half-lives in medium mass nuclei. *J. Phys. G*, 2010, v. 38, 1.
2. Palla G., Geramb H. V. and Pegel C. Electric and inelastic scattering of 25.6 MeV protons from even samarium isotopes. *Nucl. Phys. A*, 1983, v. 403, 134.
3. Talon P., Alamanos N., Lamehi-Rachti M., Levi C. and Papineau L. Coulomb and nuclear excitation effects on ^{16}O scattering from samarium isotopes. *Nucl. Phys. A*, 1981, v. 359, 493.
4. Palla G. and Pegei C. Inelastic scattering of hellions from even-even stable samarium isotopes at 40.9 MeV. *Nucl. Phys. A*, 1979, v. 321, 317.
5. Sabri H. Spectral statistics of rare-earth nuclei: Investigation of shell model configuration effect. *Nucl. Phys. A*, 2015, v. 941, 364.
6. Nomura K., Vretenar D., Nikie T. and Lu B. Microscopic description of octupole shape-phase transitions in light actinide and rare-earth nuclei. *Phys. Rev. C*, 2014, v. 89, 024312.
7. Nikie T., Kralj N., Tutis T., Vretenar D., and Ring P. Implementation of the finite amplitude method for the relativistic quasiparticle random-phase approximation *Phys. Rev. C*, 2013, v. 88, 044327.

8. Nomura K. Interacting boson model with energy density functionals. *J. Phys. Conference Series*, 2013, v. 445, 012015.
9. Nomura K., Shimizu N., and Otsuka T. Formulating the interacting boson model by mean-field methods. *Phys. Rev. C*, 2010, v. 81, 044307.
10. Zhang W., Li Z. P., Zhang S. Q., and Meng J. Octupole degree of freedom for the critical-point candidate nucleus Sm152 in a reflection-asymmetric relativistic mean-field approach. *Phys. Rev. C*, 2010, v. 81, 034302.
11. Iachello F. Dynamic symmetries of the critical point. *Phys. Rev. Lett.*, 2000, v. 85, 3580.
12. Iachello F. Analytic describtion of critical point nuclei in spherical-axially deformed shape phase transition. *Phys. Rev. Lett.*, 2000, v. 85, 3580.
13. Puddu G., Scholten O., Otsuka T. Collective Quadrupole States of Xe, Ba and Ce in the Interacting Boson Model. *Nucl. Phys.*, 1980, v. A348, 109.
14. Arima A. and Iachello F. Interacting boson model of collective states: The vibrational limit. *Ann. Phys.*, 1976, v. 99, 253.
15. Arima A. and Iachello F. Interacting boson model of collective states: The rotational limit. *Ann. Phys.*, 1978, v. 111, 201.
16. Arima A. and Iachello F. Interacting boson model of collective states: The $O(6)$ limit. *Ann. Phys.*, 1979, v. 123, 468.
17. Ginocchio J. N. and Kirson M. W. An intrinisic state for the interacting boson model and its relationship to the Bohr-Mottelson approximation. *Nucl. Phys. A*, 1980, v. 350, 31.
18. Martin M.J. Adopted levels, gammas for ^{152}Sm. *Nucl. Data Sheets*, 2013, v. 114, 1497.
19. Scholten O. The programm package PHINT (1980) version, internal report KVI-63. Keryfysisch Versneller Institut, Gronigen, 1979.
20. Church E. L. and Weneser J. Electron monopole transitions in atomic nuclei. *Phys. Rev.*, 1956, v. 103, 1035.
21. Minkov N., Yotov P., Drenska S. and Scheid W. Parity shift and beat staggering structure of octupole bands in a collective model for quadrupole-octupole deformed nuclei. *J. Phys. G*, 2006, v. 32, 497.
22. Bonatsos D., Daskaloyannis C., Drenska S. B., Karoussos N., Minkov N., Raychev P. P. and Roussev R. P. $\Delta I = 1$ staggering in octupole bands of light actinides Beat patterns. *Phys. Rev. C*, 2000, v. 62, 024301.
23. Minkov N., Drenska S. B., Raychev P. P., Roussev R. P. and Bonatsos D. Beat patterns for the odd-even staggering in octupole bands from quadrupole-octupole Hamiltonian. *Phys. Rev. C*, 2001, v. 63, 044305.
24. Rasmussen J. O. Theory of $E0$ transitions of spheroidal nuclei. *Nucl.Phys.*, 1960, v. 19, 85.
25. Bell A. D., Aveldo C. E., Davidson M. G. and Davidson J. P. Table of $E0$ conversion probability electronic factors. *Can. J. Phys.*, 1970, v. 48, 2542.

LETTERS TO PROGRESS IN PHYSICS

A Re-examination of Kirchhoff's Law of Thermal Radiation in Relation to Recent Criticisms

Robert J. Johnson
E-mail: bob.johnson1000@gmail.com

This paper investigates claims made by Pierre-Marie Robitaille in a series of papers from 2003 to 2015 that Kirchhoff's Law of thermal radiation does not apply to cavities made of arbitrary materials, and that Planck's theoretical derivation and apparent proof of this law in these cases is faulty. Robitaille's claims are compared to statements in the original papers by Kirchhoff and Planck. The present paper concludes that Robitaille's claims are not sustainable and that Kirchhoff's Law and Planck's proof remain valid in the situations for which they were intended to apply, including in cavities with walls of any arbitrary materials in thermal equilibrium.

1 Introduction

In a series of papers from 2003 to 2015 [1–10], Pierre-Maire Robitaille has challenged the validity of Kirchhoff's Law of thermal emission and Planck's derivation of the mathematical form of the universal function of spectral radiance absorbed and emitted by a black body. As the consequences of a failure of Kirchhoff's Law would, if proven, include the loss of universality of application of various fundamental physical constants including *'Planck's constant, Boltzmann's constant, ... "Planck length", "Planck time", "Planck mass", and "Planck temperature"'* [10, p. 121], Robitaille's claims deserve serious consideration.

In this paper, Robitaille's claims will be compared to the original works by Kirchhoff [11] and Planck [12] in order to determine whether his criticisms of these earlier works are valid. The present paper focuses initially on the arguments contained in the series of papers by Robitaille from 2003 to 2014 [1–9]; the second part will address the recent paper authored jointly by Robitaille and Crothers [10].

2 Robitaille's earlier papers [1–9]

2.1 Kirchhoff's law and Planck's proof

Kirchhoff's Law of thermal radiation dating from 1859-1860 may be stated as follows: *"For an arbitrary body radiating and emitting thermal radiation, the ratio E / A between the emissive spectral radiance, E, and the dimensionless absorptive ratio, A, is one and the same for all bodies at a given temperature. That ratio E / A is equal to the emissive spectral radiance I of a perfect black body, a universal function only of wavelength and temperature"*. This radiance, I, is often referred to simply as black radiation.

The form of the universal function was not known until Planck derived it theoretically in 1914 in what is now known as Planck's Law. Planck's derivation is seen as proof of Kirchhoff's Law. However, Robitaille points out that the above definition of Kirchhoff's Law is not complete and furthermore Robitaille maintains that the statement above should be called Stewart's Law as it was originally propounded by Stewart in 1858 [13]: *"All too frequently, the simple equivalence between apparent spectral absorbance and emission is viewed as a full statement of Kirchhoff's law, ... Kirchhoff's law must always be regarded as extending much beyond this equivalence. It states that the radiation within all true cavities made from arbitrary walls is black. The law of equivalence is Stewart's"* [5, p. 11].

According to Robitaille, in deriving his law of equivalence Stewart had considered the case of a cavity made from perfectly absorbing (i.e. black) material; he had shown that the radiation in such a cavity at thermal equilibrium must also be black, of an intensity appropriate to the equilibrium temperature.

Whilst Robitaille agrees with Stewart, he profoundly disagrees with Kirchhoff's extension of this finding to cavities made of arbitrary materials, and therefore with Planck's proof of Kirchhoff's result. Planck had based his proof on a consideration of perfectly reflecting cavities containing *"an arbitrarily small quantity of matter"* [12, § 51], arriving at the same result that Kirchhoff had obtained for perfectly absorbing cavities. Planck had thereby demonstrated that all cavities either containing some arbitrary matter, or equivalently having walls made of some arbitrary matter, must also contain black radiation when at thermal equilibrium.

2.2 Black radiation in a perfectly reflecting cavity

In following the reasoning of both sides of this disagreement, it is important to distinguish between a perfectly reflecting cavity containing a vacuum and one containing an opaque object or a partially-absorbing medium.

In the first case, Planck writes that *"Hence in a vacuum*

bounded by totally reflecting walls any state of radiation may persist" [12, § 51]; Robitaille claims that this statement is a violation of Kirchhoff's Law [10, p. 130]. However, Planck's statement should perhaps be more properly be viewed as a situation to which Kirchhoff's Law does not apply because there is no matter present which either absorbs or emits radiation.

When considering the case of a perfectly reflecting cavity containing an arbitrary object, again it is important to distinguish between two situations. The first is that the object absorbs and emits some fraction of all frequencies of radiation; this situation may be further subdivided into the special case where the object is itself a black body such as Planck's particle of carbon which is a perfect absorber and emitter at all frequencies; and the general case where the object only absorbs and emits some fraction above zero but less than unity of every frequency. The second situation is that the object only absorbs and emits over part of the spectrum i.e. there are some frequencies for which the object itself is a perfect reflector, neither absorbing nor emitting at those frequencies.

The question in both situations is, what is the nature of the radiation in the perfectly reflecting cavity at thermal equilibrium?

Starting with the special case of a black body, Robitaille, Kirchhoff and Planck all agree that the radiation is necessarily black. The disagreements start over the general case of an object imperfectly absorbing at all frequencies.

Planck maintains that *"... the radiation of a medium completely enclosed by absolutely reflecting walls is, when thermodynamic equilibrium has been established for all colors for which the medium has a finite coefficient of absorption, always the stable radiation corresponding to the temperature of the medium such as is represented by the emission of a black body"* [12, § 51], quoted in [1, p. 1263]. Note that the quoted statement covers both the situation where the object absorbs and emits over all frequencies, and the situation where some frequencies are not absorbed or emitted at all.

In contrast, Robitaille claims that *"In fact, if an object is placed within* [perfectly reflecting] *walls, an equilibrium will be established, but it will not correspond to that of a blackbody. Indeed, the radiation contained within such a device will reflect purely the emission profile of the object it contains"* [1, p. 1264].

This is Robitaille's central argument against the universality claimed by Kirchhoff and Planck i.e. that all cavities containing an object must, at equilibrium, come to contain black radiation at all frequencies absorbed and emitted by the object.

2.3 The approach to equilibrium

In effect, the argument comes down to the quantity of the radiation in the cavity at equilibrium. Both sides agree that there is some radiation at all frequencies absorbed and emitted by the object; the disagreement is over the intensity of that radiation. Does it, as Kirchhoff and Planck maintain, equal the intensity of black radiation which we can now quantify according to Planck's Law of 1914; or does the radiation density in the cavity fall short of the black body level at some or all frequencies because of the imperfect absorption and emission of the object in the cavity, as Robitaille claims?

The role played by the reflected radiation, i.e. that fraction of incident radiation which is not fully absorbed by the object, is the key. Robitaille maintains that the radiation density in the cavity cannot be increased to black body levels by what he terms "driving the reflection" because this would imply a departure from thermal equilibrium which, Robitaille argues, contravenes the initial assumption that thermal equilibrium exists.

A simplified numerical example may be helpful here in order to crystallise the arguments. Suppose an opaque object in a perfectly reflecting cavity is in thermal equilibrium at a certain temperature and has a coefficient of absorption of 0.8 (i.e. 80%) of all incident radiation at all frequencies. The remaining 20% of any incident radiation will be reflected. Suppose further that the radiation density in the cavity is already at the level at which a black body at the same temperature would be in thermal equilibrium with it, say100 units. This will represent the incident radiation on the opaque object which will then absorb 80 units and reflect 20 units. The object will also re-emit the same 80 units into the cavity. The total radiation coming off the surface of the object, consisting of the emitted and reflected components, is 100, therefore thermal equilibrium will be maintained with the radiation in the cavity. What's more, the radiation density is and remains black according to the initial assumption. This represents the situation described by Kirchhoff's Law.

Consider now the situation where the same object at the same initial temperature is introduced into the perfectly reflecting but otherwise empty cavity, i.e. there is no radiation density in the cavity initially. In this case, the object will emit 80 units appropriate to its temperature; these will be reflected off the walls and become "incident" radiation on the object. The object will now absorb 80%, or 64 units, and reflect 16 units. But it is bound by its initial temperature to continue emitting 80 units. There is therefore a shortfall between the amount absorbed and the amount emitted and the object will cool down. The energy lost by the object will be converted to additional radiation density in the cavity which will increase until equilibrium is achieved between the object and the radiation density at some new, lower, temperature. At this point, the radiation will again be black, but at the level appropriate to the lower temperature, not the initial temperature of the object.

Robitaille would object to this second example on the grounds that thermal equilibrium has not been maintained. This is correct. But Robitaille goes further and maintains that this proves that the cavity cannot contain black radiation because it is not allowable to "drive the reflection" until

a new equilibrium is reached – the object must be maintained at the original temperature throughout and therefore there is no spare energy available to "drive the reflection" up to black body densities.

In essence, Robitaille disallows the approach to thermal equilibrium between the object and the radiation density in the cavity by the mechanism outlined in the second numerical example above. As a result, Robitaille maintains that the cavity cannot contain the black radiation required by Kirchhoff's Law and therefore the law fails.

In support of his argument, Robitaille quotes Stewart [13] as follows: *"Let us suppose we have an enclosure whose walls are of any shape, or any variety of substances (all at a uniform temperature), the normal or statical condition will be, that the heat radiated and reflected together, which leaves any portion of the surface, shall be equal to the radiated heat which would have left that same portion of the surface, if it had been composed of lampblack ... Let us suppose, for instance, that the walls of this enclosure were of polished metal then only a very small quantity of heat would be radiated; but this heat would be bandied backwards and forwards between the surfaces, until the total amount of radiated and reflected heat together became equal to the radiation of lampblack"* [13, § 32] quoted in [4, p. 45].

Robitaille comments: *"These passages are quite similar to Kirchhoff's with the distinction that universality is never invoked. Stewart realizes that the lampblack surface within the enclosure is essential"* [4, p. 45]. But Stewart is quite specific – the walls may be of any variety of substance including polished metal. This implies that Robitaille's objection to what he refers to as Kirchhoff's extension of Stewart's result to cavities made of arbitrary material is unfounded; Stewart had already made the theoretical leap.

How then did Stewart conclude as he did that *"the sum of the radiated and reflected heat together became equal to the radiation of lampblack?"*

2.4 Stewart's treatment of reflection

In Stewart's original paper there is a footnote to the section quoted above which explains the calculation by which he arrived at this conclusion. Stewart considers *"two parallel plates of polished metal of the same description radiating to one another"* [13, § 32-footnote] and investigates what happens to an initial amount r of radiation emitted by each opposing plate and falling perpendicularly on the other plate, where a proportion is reflected back to the first plate. As an ever-decreasing part of the original radiation r is "bandied about" by repeated reflection between the plates, with a proportion $\alpha(< 1)^*$ of the incident radiation being reflected each time, Stewart shows that the total amount falling on one of the plates is

$$r\left(1 + \alpha + \alpha^2 + \alpha^3 + \alpha^4 \ldots\right) = \frac{r}{1-\alpha},$$

which, Stewart explains, is the same formula as results from the case where one of the plates is a black body in thermal equilibrium with the other plate.

The question then arises, can this calculation also be applied to a situation where thermal equilibrium has not yet been achieved? It turns out that it can. Note that, in modern parlance, Stewart's calculation sums the repeated reflections of the two initial pulses (one from each plate) emitted in the first interval of time δt over subsequent intervals of time. It may be supposed without loss of generality that the interval of time δt corresponds to the transmission time of radiation between the plates. Then the same sum would result from considering what proportion of a series of identical initial pulses each of emission duration δt fell on one plate in a single (later) interval of time δt. This second case represents continuous emission of radiation in thermal equilibrium. One of the plates may then be replaced with a black body at the same equilibrium temperature which emits exactly the same amount of radiation that it absorbs, or alternatively with a perfect reflector. Again, the same sum emerges from the calculation, as Stewart explained.

What's more, exactly the same result is obtained when one plate is perfectly reflecting and there is no radiation in the gap between the plates initially, i.e. there is no initial thermal equilibrium to supply the series of constant pulses prior to the arrival interval δt under consideration. In this case, all the radiation is emitted by just one of the plates; therefore double the time is required to achieve the same result that Stewart obtained but, in effect, this result shows that once a steady state has been achieved then the radiation arriving on the single partially-absorbing plate is equivalent to that coming from a black body. The only difference in this case is that during the initial period the partially-absorbing plate is absorbing less radiation than it is emitting; it is therefore cooling down and part of its initial thermal energy is being used to increase the radiation density between the plates, or, in Robitaille's terms, in "driving the reflection". However, when thermal equilibrium is established then the calculation shows that the radiation reflected back on to the emitting plate will be equivalent to black radiation at the equilibrium temperature.

What this demonstrates is that Stewart's method of calculation of the reflection being "bandied about" can also be applied to the approach to equilibrium provided that time is allowed for a sufficient number of reflections to build up the radiation density in the cavity to equilibrium levels. The total time necessary to fill the space with black radiation is likely to be short because of the extremely short transmission time δt and the limited number of reflections necessary to achieve near-perfect black body radiation in most normal situations.

*Stewart uses α to represent the proportion of reflected radiation; in Planck's usage, α represents the coefficient of absorption. To comply with Planck's usage, α should be replaced with ρ in the above equation. The derivation of the equation is unaffected.

Only in cases where the plate is nearly a perfect reflector might an appreciable time be required.

Thus "Stewart's mechanism", if we may so call it, should be interpreted as indicated in the second numerical example given above, with the walls themselves taking the part of the opaque object. Stewart's words "bandied about" can be applied to the reflected proportions of the continuing emission which build up the radiation density in the cavity until thermal equilibrium is achieved. Robitaille calls this "driving the reflection"; it may be clearer to think of the effect as "increasing the radiation energy density in the cavity" at the expense of the thermal energy of the walls. The important point, though, is that it occurs on the approach to thermal equilibrium between the walls and the radiation density in the cavity, not at the stage where equilibrium has already been achieved. However, once thermal equilibrium has been established then the radiation in the cavity will be black.

If an object in a perfectly reflecting cavity absorbs and emits some radiation at all frequencies it is clear that Stewart, Planck and Kirchhoff all held that the full black body spectrum will by achieved by the mechanism outlined numerically above and described by Stewart in the passage quoted. In contrast, Robitaille maintains throughout his series of papers [1–9] that it is necessary to include a black body in the cavity, whether by making part of the walls black or by inclusion of a black object, in order to achieve black radiation in accordance with Kirchhoff's Law.

2.5 Planck's particle of carbon

Robitaille claims that this is precisely why Planck insisted on including a carbon particle in his analysis and why Kirchhoff included one in his experiments. Robitaille dismisses Planck's assertion that the particle merely acts as a catalyst and insists that the carbon particle is responsible for producing the black radiation that Kirchhoff s Law requires. For example, as recently as 2014 Robitaille stated "[Planck's] *cavities all contained black radiation as a direct result* [of placing a carbon particle in the cavity] ... *Since he was driving reflection, all cavities contained the same radiation ...* " [9, p. 158].

However, it is important to distinguish between the *nature* of the black radiation emitted and the *quantity* of it. Planck is perfectly clear that the reason for assuming that the carbon particle is merely a catalyst is that it may be made as small as one likes and, most importantly, its thermal energy can be made so small as to not significantly change the total energy in the cavity [12, § 52]. By definition, therefore, the carbon particle cannot increase the radiation density in the cavity to the level commensurate with the black body temperature; in Robitaille's terms, the particle cannot "drive the reflection", and therefore this cannot be the reason why Planck included it.

Furthermore, if the radiation density is being increased at all frequencies by Stewart's mechanism then there is no need for the particle at all; all one needs to do is wait until thermal equilibrium has been achieved. If the object is a very poor absorber and emitter then this could take some time. In adding a carbon particle to his experiments, Kirchhoff may simply have wanted to accelerate the process.

The situation is somewhat different in the case when the object is a perfect reflector at one or more frequencies. In that case, as Planck stated, the spectrum is black for all frequencies at which the object absorbs and emits but it is indeterminate at the frequencies for which the object is a perfect reflector: *"Hence in a vacuum bounded by totally reflecting walls any state of radiation may persist. But as soon as an arbitrarily small quantity of matter is introduced into the vacuum, a stationary state of radiation is gradually established. In this the radiation of every color which is appreciably absorbed by the substance has the intensity K_ν corresponding to the temperature of the substance and determined by the universal function ..., the intensity of radiation of the other colors remaining indeterminate"* [12, § 51].

However, if the spectrum is indeterminate at any frequencies then it is not possible to properly determine a temperature which is defined in terms of the black body spectrum. See for example *"... the radiation in the new volume V' will not any longer have the character of black radiation, and hence no definite temperature ..."* [12, § 70]. It is apparently in order to avoid this situation that Planck included a particle of carbon which guaranteed that the intensity of radiation was determinate at all frequencies. Why Planck considered that this precaution was necessary is apparent from earlier sections of his work.

Planck had previously discussed the relationship between surface roughness and reflection, pointing out that whether a surface reflected or not was a function of roughness in relation to the wavelength: *"All the distinctions and definitions mentioned in the two preceding paragraphs refer to rays of one definite color only. It might very well happen that, e.g., a surface which is rough for a certain kind of rays must be regarded as smooth for a different kind of rays. It is readily seen that, in general, a surface shows decreasing degrees of roughness for increasing wave lengths. Now, since smooth non-reflecting surfaces do not exist (Sec. 10), it follows that all approximately black surfaces which may be realized in practice (lamp black, platinum black) show appreciable reflection for rays of sufficiently long wave lengths"* [12, § 11].

Thus all objects except perfect black bodies will become reflective at long enough wavelengths. It is apparently in order to avoid this situation that Planck insisted on including a particle of carbon which ensured that all frequencies were present in the equilibrium spectrum. The total radiation energy would not be affected because the particle would not have sufficient energy to do so, by definition. Thus the particle merely acted as a catalyst, as Planck insisted, to convert

the spectrum emitted by the object into a black spectrum as necessary for a proper temperature measurement to be made in accordance with the definition.

Interestingly, despite numerous repetitions in Robitaille's papers [1–8] of his claim that Planck's carbon particle was essential in order to increase the radiation density to the required black body level, Robitaille [9] hints at a change of stance, admitting that *eventually, the cavity might become filled with black radiation, provided that emission and reflection are Lambertian"* [9, p. 160] but then he negates the possibility by stating *"However, for most materials, the introduction of photons into the reflected pool will be inefficient, and the temperature of the system will simply increase. That is the primary reason that arbitrary cavities can never contain black radiation"* [9, p. 160]. In 2015, Robitaille & Crothers [10] return to this theme, stating *"Stewart recognized that, if one could "drive the radiation" in a cavity made from arbitrary materials, by permitting the slow buildup of reflected radiation, the interior could eventually contain black radiation. The argument was true in theory, but not demonstrated in practice"* [10, p. 122].

It appears that Robitaille and Crothers now accept Stewart's mechanism for building up the radiation density by *"bandying about"* the reflection, at least in principle. The authors do not give any explanation for this remarkable *volteface* from Robitaille's earlier works [1–8], but it now appears that his previous objections to Planck's particle of carbon are unfounded: the particle cannot, by definition, increase the total radiation density in the cavity, and Robitaille & Crothers apparently now accept that it is not necessary for the validity of Kirchhoff's Law that it does so.

2.6 Experimental evidence against Kirchhoff's law

Robitaille bases many of his arguments against the validity of Kirchhoff's Law on the fact that black body cavities are never constructed of arbitrary materials; on the contrary, Robitaille insists that manufacturers go to great lengths to construct cavities from special materials to ensure that the radiation is black. Equally, Robitaille points out that resonant microwave cavities cannot contain black radiation. Both these counter-examples are held to demonstrate that Kirchhoff's Law must be incorrect.

However, there appear to be alternative explanations available. In the former case, it may well be that users are concerned about the efficiency of the approach to equilibrium and therefore require black materials in order to speed up the process. It is also likely that manufacturers are concerned, as Planck himself apparently was, to ensure that there are no frequencies at which the cavity is a perfect reflector, which would preclude a proper measurement of temperature.

In the case of microwaves, the cavity is being electromagnetically forced to resonate at a particular frequency and so the radiation cannot be black. Such cases of non-thermal emission were specifically excluded by Plank in deriving his proof: *"A necessary consequence of this is that the coefficient of emission ϵ depends, apart from the frequency ν and the nature of the medium, only on the temperature T. The last statement excludes from our consideration a number of radiation phenomena, such as fluorescence, phosphorescence, electrical and chemical luminosity, ..."* [12, § 7].

Thus it is not logical to conclude that Kirchhoff's Law must necessarily fail because of these supposed counter-examples.

2.7 Challenges to Monte Carlo simulations

Robitaille states that Monte Carlo simulations apparently support Kirchhoff's Law but then he objects on the grounds that: *"Monte Carlo simulations introduce black photons into cavities. Hence, they become black. The process is identical to placing a highly emitting carbon particle, or radiometer, at the opening of a cavity. No proof is provided by computational methods that arbitrary cavities contain black radiation. It can be stated that Monte Carlo simulations obtain similar answers by modeling the repeated emission of photons directly from the cavity walls. In this case, computational analysis relies on internal reflection to arrive at a cavity filled with black radiation"* [5, p. 6].

Apparently, Robitaille's objection to the Monte Carlo simulations is that they rely on Stewart's mechanism for building up the radiation by internal reflection. As Robitaille and Crothers [10] now accept that this mechanism is valid in principle, Robitaille's previous objections to Monte Carlo simulations supporting Kirchhoff's Law should also drop away.

2.8 Super-Planckian emission

Robitaille suggests that recent research into metamaterials supports his arguments. For example, he states: *"Recent results demonstrating super-Planckian thermal emission from hyperbolic metamaterials (HMM) in the near field and emission enhancements in the far field are briefly examined. Such findings highlight that cavity radiation is absolutely dependent on the nature of the cavity and its walls. As previously stated, the constants of Planck and Boltzmann can no longer be viewed as universal"* [9, p. 157].

In relation to the near field emissions, Robitaille refers to three examples from the recent literature [14–16]. All three papers refer to experiments involving bodies with separation distances smaller than the thermal wavelength. However, experimental distances below the thermal wavelength were expressly excluded by Planck: *"Throughout the following discussion it will be assumed that the linear dimensions of all parts of space considered, as well as the radii of curvature of all surfaces under consideration, are large compared with the wave lengths of the rays considered"* [12, § 2].

Planck was concerned about the effects of diffraction at

small scales, in effect limiting his analysis to what are now known as far field effects. Near-field effects are not covered by Kirchhoff's Law and so these three papers cited by Robitaille cannot be used as examples of contraventions of the law. In fact, Guo et al point out that Kirchhoff's Law still suffices to calculate the thermal emission in the far-field and that *"the high-k waves which are thermally excited in the HMM are trapped inside and will be evanescent in vacuum (not reach the far field)"* [14, p. 2]. After comparing the behavior of HMM to other near-field phenomena of surface electromagnetic excitations and photonic crystal structures, Guo et al *"emphasize that in all the above cases including hyperbolic metamaterials, the presence of an interface is enough to guarantee that the far-field emissivity is limited to unity"* [14, p. 5], i.e. that it is Planckian.

The evidence for super-Planckian far-field emissions is not convincing either. Robitaille cites two papers by Yu et al [17, 18] and Nefedov & Melnikov [19] but he notes that Yu et al's claim of emissions in excess of the Stefan-Boltzmann Law made in their arXiv preprint were withdrawn in the published version, and that Nefedov & Melnikov's experiment was not in thermal equilibrium as required by Kirchhoff's Law.

Robitaille's conclusion that *"the universality of blackbody radiation has simply been overstated"* [9, p. 161] does not appear to be warranted on the basis of these recent experiments into metamaterials.

2.9 Robtaille's thought experiment

In [7], Robitaille postulates a thought experiment which he claims disproves Kirchhoff's Law: *"Through the use of two cavities in temperature equilibrium with one another, a thought experiment is presented ... which soundly refutes Kirchhoff's law of thermal emission"* [7, p. 38]. In this thought experiment, the outer cavity is perfectly absorbing and emitting; the second cavity, which is contained entirely within the outer cavity, has perfectly reflecting walls and one side which can be closed remotely. Starting with this inner side open, the two cavities are brought to 4 K; the inner side is then closed; the outer cavity is then heated to 300 K. Robitaille continues: *"The inner cavity walls are thus also brought to 300 K. However, unlike the outer cavity which is filled with blackbody radiation at 300 K, the inner cavity remains filled with blackbody radiation at 4 K. Thereby, Kirchhoff's law is proven to be false"* [7, p. 39].

But by making the inner cavity walls perfectly reflecting and closing the last side, Robitaille has created two entirely separate cavities; by definition, the inner cavity walls cannot emit radiation in either direction, whatever their temperature. They therefore act as boundary walls to what has become a "hollow" outer cavity. The outer cavity no longer contains the inner cavity within itself in the thermal sense; Kirchhoff's Law therefore survives this thought experiment.

3 Robitaille and Crothers 2015 paper

Robitaille & Crothers' paper [10] represents a significant departure from the previous works by Robitaille alone [1–9]. Robitaille and Crothers' *volte-face* on the viability of Stewart's mechanism for filling any cavity with black radiation has been discussed above. However, apart from a re-statement of many of Robitaille's previous objections which have also been discussed above, the thrust of the 2015 jointly-authored paper is to concentrate on criticising Planck's proof of Kirchhoff's Law, a matter only touched on briefly in previous works. Section 4 is titled *"Max Planck and Departure from Objective Reality"* and contains the authors' principal objections to Planck's proof. These will now be examined in detail.

3.1 The meaning of Planck's term "surface"

A number of Robitaille and Crothers' objections hinge on their interpretation of Planck's term "surface" which Planck himself had been careful to distinguish from Kirchhoff's earlier definition. Robitaille and Crothers quote from Planck: *"In defining a blackbody Kirchhoff also assumes that the absorption of incident rays takes place in a layer "infinitely thin". We do not include this in our definition"* [10, p. 124] quoting a footnote from [12, § 10]. In the original text, Planck later explains why he is diverging from Kirchhoff on this point: *"Heat rays are destroyed by absorption. According to the principle of the conservation of energy the energy of heat radiation is thereby changed into other forms of energy (heat, chemical energy). Thus only material particles can absorb heat rays, not elements of surfaces, although sometimes for the sake of brevity the expression absorbing surfaces is used"* [12, § 12]. It appears that Planck could not accept Kirchhoff's "infinitely thin" absorbing layer because it could not include any material particles.

In § 12, Planck is simply being consistent with his earlier discussion of emission: *"The creation of a heat ray is generally denoted by the word emission. According to the principle of the conservation of energy, emission always takes place at the expense of other forms of energy (heat, chemical or electric energy, etc.) and hence it follows that only material particles, not geometrical volumes or surfaces, can emit heat rays. It is true that for the sake of brevity we frequently speak of the surface of a body as radiating heat to the surroundings, but this form of expression does not imply that the surface actually emits heat rays. Strictly speaking, the surface of a body never emits rays, but rather it allows part of the rays coming from the interior to pass through. The other part is reflected inward and according as the fraction transmitted is larger or smaller the surface seems to emit more or less intense radiations"* [12, § 2].

In both § 10 and § 12, it is clear that Planck's use of the term "surface" refers to a geometrical surface dividing two media; the material effects of emission and absorption take

place within the adjoining media. Planck's reference to the surface radiating or absorbing heat is clearly stated as being no more than a convenient shorthand. In contrast, Robitaille and Crothers interpret Planck's term "surface" as being one composed of material particles; it appears that this misinterpretation has led them to a number of erroneous conclusions.

For example, Robitaille and Crothers ask in relation to an element $d\sigma$ of the bounding surface: *"First, what exactly was the location of $\delta\sigma$? In reality it must rest in one of the two media"* [10, p. 127]. This is contrary to Planck's own description of the bounding surface σ as a *"surface separating the two media"* [12, § 35]. Thus Robitaille and Crothers' first objection, that Planck is being inconsistent as to the location of the bounding surface, is unfounded. Similarly, Robitaille and Crothers' second objection to Planck's treatment of the bounding surface, namely *"Planck neglected the fact that real materials can possess finite and differing absorptivities"* [10, p. 127] cannot be maintained.

Robitaille and Crothers raise a third objection to the analysis of an element $d\sigma$ of the bounding surface, namely: *"Third, the simplest means of nullifying the proof leading to Planck's Eq. 42, is to use a perfect reflector as the second medium. In that case, a refractive wave could never enter the second medium and Planck's proof fails"* [10, p. 127]. However, if the surface separating the two media is itself a perfect reflector then the reflectivity on the side of the first medium is obviously equal to 1 but so is the reflectivity for any rays coming from the other side. Thus, $\rho = \rho'$ in accordance with Planck's Eq. 40 leading to his Eq. 42 (see also below) and the proof remains valid. In fact, Planck had already considered this theoretical possibility as occurring for an instant: *"Since the equilibrium is nowise disturbed, if we think of the surface separating the two media as being replaced for an instant by an area entirely impermeable to heat radiation, the laws of the last paragraphs must hold for each of the two substances separately"* [12, § 35]. Obviously the instantaneous nature of this theoretical replacement is necessary to preserve the single system being analysed; a more permanent separation would create two separate systems to which the analysis did not apply. Once again it seems that Robitaille and Crothers' objection is unsustainable.

3.2 Absorption and transmission

Following their quote from Planck's footnote departing from Kirchhoff's definition of an infinitely thin surface in which all the absorption occurred (see above), Robitaille and Crothers commented as follows: *"With his words, Planck redefined the meaning of a blackbody. The step, once again, was vital to his derivation of Kirchhoff's Law, as he relied on transmissive arguments to arrive at its proof. Yet, blackbody radiation relates to opaque objects and this is the first indication that the proofs of Kirchhoff's Law must not be centered on arguments which rely upon transmission. Planck ignored that real surface elements must possess absorption, in apparent contrast with Kirchhoff and without any experimental justification"* [10, p. 124].

However, as is obvious from the passages quoted above, Planck did recognize that absorption must be related to material particles. Once again, the apparent problem arises from the fact that Planck's surface is a geometrical one, whilst Robitaille and Crothers are obviously referring to a surface layer in which, they maintain, all absorption must take place because transmission is not permitted through a black body.

However, Planck also allows for the possibility that absorption in an opaque medium may take place at some unspecified depth below the geometrical surface, i.e. not necessarily in the particles immediately adjacent to the surface. Robitaille and Crothers quote from Planck's description in § 10 of the dependence of the absorbing power on the thickness of the black body material which ends *"The more absorbing a body is, the smaller the value of this minimum thickness, while in the case of bodies with vanishingly small absorbing power only a layer of infinite thickness may be regarded as black"*. Robitaille & Crothers object to this sentence stating that *"Now,* [Planck] *explicitly stated that bodies which are poor absorbers can still be blackbodies. Yet, we do not make blackbodies from materials which have low absorptivities, because these objects have elevated reflectivities, not because they are not infinite"* [10, p. 125] quoting [12, § 10].

But these two objections, about absorptivity and reflectivity respectively, seem to be missing the points that Planck is making: firstly, some absorption may take place by particles situated below the surface. Secondly, Planck had previously stated: *"When a smooth surface completely reflects all incident rays, as is approximately the case with many metallic surfaces, it is termed "reflecting". When a rough surface reflects all incident rays completely and uniformly in all directions, it is called "white". A rough surface having the property of completely transmitting the incident radiation is described as "black"* [12, § 10]. Note that Planck defines black materials as those with a rough surface which does not reflect; all rays falling on a black material pass through Planck's geometrical surface and are subsequently absorbed at some depth in the interior of the black body. No rays are reflected from the body even if the material is, in Planck's terms, a poor absorber. This immediately undermines Robitaille and Crothers' second objection.

Robitaille and Crothers also argue that Planck incorrectly includes transmission within the material of the black body when in fact, Robitaille and Crothers claim, absorption must all occur at the surface: *"Blackbodies are opaque objects without transmission, by definition"* [10, p. 125]. Once again, they are apparently overlooking Planck's definition of a geometrical surface and his careful consideration of where any absorption of radiation passing through that geometrical surface subsequently takes place.

3.3 Reflection

Robitaille and Crothers' § 4.2 deals further with Planck's treatment of reflection. The authors state: *"In the first section of his text, leading to his Eq. 27, ... Planck chose to formally neglect reflection, even though the total energy of the system included those rays which are both emitted/absorbed and those which would have been maintained by driving reflection. Such an approach was suboptimal"* [10, p. 125].

However in the first section of his text, Planck is expressly dealing with the situation within a medium, not with surface effects. His § 25 begins: *"We shall now, as in the previous chapter, assume that we are dealing with homogeneous isotropic media whose condition depends only on the temperature, and we shall inquire what laws the radiation phenomena in them must obey in order to be consistent with the deduction from the second principle mentioned in the preceding section ... Let us consider, first, points of the medium that are far away from the surface"* [12, § 25]. A mathematical treatment then follows, leading to Planck's Eq. 27 towards the end of § 26 which Planck follows with the words *"i.e.: in the interior of a medium in a state of thermodynamic equilibrium the specific intensity of radiation of a certain frequency is equal to the coefficient of emission divided by the coefficient of absorption of the medium for this frequency"* [12, § 26].

Note that Planck is still talking about the interior of the medium where reflection is not applicable because there is no surface; therefore Robitaille and Crothers' objection cannot be maintained.

3.4 Polarization and equality of reflection

Robitaille and Crothers then object to Planck's analysis based initially on a plane-polarised ray, stating: *"In § 5 Planck admitted that homogeneous isotropic media emit only natural or normal, i.e. unpolarized, radiation: "Since the medium was assumed to be isotropic the emitted rays are unpolarized". This statement alone, was sufficient to counter all of the arguments which Planck later utilized to arrive at Kirchhoff's Law [Eq. 42]. That is because the important sections of Planck's derivation, namely § 35–37 make use of plane-polarized light. These steps were detached from experimental reality, relative to heat radiation [Planck, § 35] ..."* [10, p. 127] quoting [12, § 35].

Yet Robitaille and Crothers themselves admit that there was method in Planck's approach, quoting Planck again: *"to prepare for his use of polarized light in later sections, Planck resolved, in § 17, the radiation into its two polarized components"* [10, p. 127], which in itself is unobjectionable. However, Robitaille and Crothers later state that *"such rays could never exist in the context of heat radiation"* [10, p. 129] and this appears to be their principal objection to this means of analysis from which Planck derives the equality of the reflectivity on either side of a geometrical surface separating two different media in his Eq. 40.

But Planck made it clear that an analysis of the special case of polarised light under consideration leads to a valid general conclusion because, as he explained at the end of § 36, the intensity of radiation K_ν, the velocity of propagation q, and the coefficient of reflection ρ at a surface dividing two different media are related by the equation

$$\frac{K_\nu}{K'_\nu}\,\frac{q^2}{q'^2} = \frac{1-\rho'}{1-\rho}\,,$$

where the accented quantities refer to the second medium. Planck continued in §37: *"In the last equation the quantity on the left side is independent of the angle of incidence and of the particular kind of polarization; hence the same must be true for the right side. Hence, whenever the value of this quantity is known for a single angle of incidence and any definite kind of polarization, this value will remain valid for all angles of incidence and all kinds of polarization. Now in the special case when the rays are polarized at right angles to the plane of incidence and strike the bounding surface at the angle of polarization, $\rho = 0$, and $\rho' = 0$. The expression on the right side of the last equation then becomes 1; hence it must always be 1 and we have the general relations:*

$$\rho = \rho' \qquad (40)$$

and

$$q^2 K_\nu = q'^2 K'_\nu \quad (41)".$$

Regarding Planck's Eq. 40, Robitaille and Crothers state bluntly that *"The result was stunning. Max Planck had determined that the reflectivities of all arbitrary media were equal"* [10, p. 129]. On the contrary, what Planck had in fact demonstrated is that the reflectivities on each side of a geometrical surface bounding two different media are equal. Clearly if a different pair of media are chosen, the value of the reflectivity of the bounding surface may be different as well. Planck had previously addressed this point in § 10: *"Since, in general, the properties of a surface depend on both of the bodies which are in contact, this condition shows that the property of blackness as applied to a body depends not only on the nature of the body but also on that of the contiguous medium. A body which is black relatively to air need not be so relatively to glass, and vice versa"* [12, § 10]. Robitaille & Crothers' interpretation that Planck had determined that the reflectivities of all media were equal is unwarranted.

4 Summary and conclusions

Stewart [13] had shown that the radiation in a cavity made from perfectly absorbing material at thermal equilibrium must be black, of an intensity appropriate to the equilibrium temperature. According to Robitaille, Kirchhoff [11] extended this finding to cavities made of arbitrary materials. In a series of papers [1–10], Robitaille has raised various objections to Kirchhoff's extension of Stewart's finding to arbitrary cavities, and to Planck's proof of Kirchhoff's Law [12].

Robitaille concludes that the Law can only be applied validly to cavities containing a black body.

The present paper has investigated Robitaille's claims in depth and compared them to the original papers by Stewart [13], Kirchhoff [11] and Planck [12]. In no instances have Robitaille's objections been found to be sustainable. Furthermore, is has been noted that one of Robitaille's key and often-repeated objections to the build-up of black radiation in an arbitrary cavity according to a mechanism first proposed by Stewart [13] has now been effectively withdrawn in the recent paper by Robitaille and Crothers [10].

Robitaille is obviously correct to point out that black body cavities are never made from reflective materials. However, this fact appears to be more a question of practicality and the need to ensure that the walls are not perfectly reflective at any wavelength so that proper measurements of temperature can be made. It does not seem to amount to a demonstration that Kirchhoff's Law necessarily fails, as Robitaille claims.

This investigation suggests that Kirchhoff's Law and Planck's proof of it remain valid in the situations for which they were intended to apply, including in cavities with walls of any arbitrary materials in thermal equilibrium, unless some other more sustainable objections can be raised in the future.

Acknowledgements

I thank Pierre-Marie Robitaille for encouraging my own research into the subject of Kirchhoff's Law and for engaging in an informative correspondence with me in 2014 regarding Stewart's mechanism for building up the radiation density. I also thank the Mainwaring Archive Foundation for their generous support of my research.

Submitted on: January 8, 2016 / Accepted on: January 11, 2016
First published online on: January 28, 2016

References

1. Robitaille P.-M. On the validity of Kirchhoff's law of thermal emission. *IEEE Trans. Plasma Sci.*, 2003, v. 31, no. 6, 1263–1267.
2. Robitaille P.-M. An analysis of universality in blackbody radiation. *Prog. Phys.*, 2006, v. 2, 22–23.
3. Robitaille P.-M. A critical analysis of universality and Kirchhoff's law: A return to Stewart's law of thermal emission. *Prog. Phys.*, 2008, v. 3, 30–35.
4. Robitaille P.-M. Blackbody radiation and the carbon particle. *Prog. Phys.*, 2008, v. 3, 36–55.
5. Robitaille P.-M. Kirchhoff's law of thermal emission: 150 Years. *Prog. Phys.*, 2009, v. 4, 3–13.
6. Robitaille P.-M. Blackbody radiation and the loss of universality: Implications for Planck's formulation and Boltzmann's constant. *Prog. Phys.*, 2009, v. 4, 14–16.
7. Robitaille P.-M. Further insight relative to cavity radiation: a though experiment refuting Kirchhoff's law. *Prog. Phys.*, 2014, v. 10, no. 1, 38–40.
8. Robitaille P.-M. On the equation which governs cavity radiation I. *Prog. Phys.*, 2014, v. 10, no. 2, 126–127.
9. Robitaille P.-M. On the equation which governs cavity radiation II. *Prog. Phys.*, 2014, v. 10, no. 3, 157–162.
10. Robitaille P.-M. and Crothers S.J. "The theory of Heat Radiation" revisited: a commentary on the validity of Kirchhoff's law of thermal emission and Max Planck's claim of universality. *Prog. Phys.*, 2015, v. 11, no. 2, 120–132.
11. Kirchhoff G. Über das Verhältnis zwischen dem Emissionsvermögen und dem Absorptionsvermogen. der Körper fur Wärme und Licht. *Poggendorfs Annalen der Physik und Chemie*, 1860, v. 109, 275–301. (English translation by F. Guthrie: Kirchhoff G. On the relation between the radiating and the absorbing powers of different bodies for light and heat. *Phil. Mag.*, 1860, ser. 4, v. 20, 1–21; DOI 10.1080/14786446008642901).
12. Planck M. Waermestrahlung (1913), Trans. M. Masius as: "The Theory of Heat Radiation". P. Blakiston' Son & Co., Philadelphia, PA, 1914.
13. Stewart B. An account of some experiments on radiant heat, involving an extension of Prévost's theory of exchanges. *Trans. Royal Soc. Edinburgh*, 1858, v. 22, no. 1, 1–20 (also found in Harper's Scientific Memoirs, edited by J. S. Ames: The Laws of Radiation and Absorption: Memoirs of Prévost, Stewart, Kirchhoff, and Kirchhoff and Bunsen, translated and edited by D.B. Brace, American Book Company, New York, 1901, 21–50).
14. Guo Y., Cortez C.L., Molesky S., and Jacob Z. Broadband super-Planckian thermal emission from hyperbolic metamaterials. *Appl. Phys. Let.*, 2012, v. 101, 131106.
15. Biehs S.A., Tschikin M., Messina R. and Ben-Abdallah P., Super-Planckian near-field thermal emission with phonon-polaritonic hyperbolic metamaterials. *Appl. Phys. Let.*, 2013, v. 102, 131106
16. Petersen S.J., Basu S. and Francoeur M. Near-field thermal emission from metamaterials. *Photonics and Nanostructures — Fund. Appl.*, 2013, v. 11, 167–181.
17. Yu Z., Sergeant N.P., Skauli T., Zhang G., Wang H., and Fan S. Enhancing far-field thermal emission with thermal extraction. *Nature Comm.*, 2013, DOI: 10.1038/ncomms2765.
18. Yu Z., Sergeant N., Skauli T., Zhang G., Wang H. and Fan S. Thermal extraction: Enhancing thermal emission of finite size macroscopic blackbody to far-field vacuum. 4 Nov 2012, arXiv:1211.0653v1 [physics.optics].
19. Nefedov I.S. and Melnikov L.A. Super-Planckian far-zone thermal emission from asymmetric hyperbolic metamaterials. 14 Feb 2014, axXiv:1402.3507v1 [physics.optics].

LETTERS TO PROGRESS IN PHYSICS

A Re-examination of Kirchhoff's Law of Thermal Radiation in Relation to Recent Criticisms: Reply

Pierre-Marie Robitaille

Department of Radiology, The Ohio State University, 395 W. 12th Ave, Columbus, Ohio 43210, USA.
robitaille.1@osu.edu

Recently, Robert J. Johnson submitted an analysis of my work, relative to Kirchhoff's Law of Thermal Emission (R.J. Johnson, A Re-examination of Kirchhoff's Law of Thermal Radiation in Relation to Recent Criticisms. *Prog. Phys.*, 2016, v. 12, no. 3, 175–183) in which he reached the conclusion that *"Robitaille's claims are not sustainable and that Kirchhoff's Law and Planck's proof remain valid in the situations for which they were intended to apply, including in cavities with walls of any arbitrary materials in thermal equilibrium"*. However, even a cursory review of Johnson's letter reveals that his conclusions are unjustified. No section constitutes a proper challenge to my writings. Nonetheless, his letter is important, as it serves to underscore the impossibility of defending Kirchhoff's work. At the onset, Kirchhoff formulated his law, based solely on thought experiments and, without any experimental evidence (G. Kirchhoff, Über das Verhältnis zwischen dem Emissionsvermögen und dem Absorptionsvermogen. der Körper fur Wärme und Licht. *Pogg. Ann. Phys. Chem.*, 1860, v. 109, 275–301). Thought experiments, not laboratory confirmation, remain the basis on which Kirchhoff's law is defended, despite the passage of 150 years. For his part, Max Planck tried to derive Kirchhoff's Law by redefining the nature of a black body and relying on the use of polarized radiation, even though he realized that heat radiation is never polarized (Planck M. *The Theory of Heat radiation*. P. Blakiston's Son & Co., Philadelphia, PA, 1914). In advancing his proof of Kirchhoff's Law, Max Planck concluded that the reflectivities of any two arbitrary materials must be equal, though he argued otherwise (see P.-M. Robitaille and S. J. Crothers, *"The Theory of Heat Radiation"* Revisited: A Commentary on the Validity of Kirchhoff's Law of Thermal Emission and Max Planck's Claim of Universality. *Prog. Phys., 2015, v. 11, no. 2, 120–132*). Planck's Eq. 40 (ρ=ρ') , as presented in his textbook, constituted a violation of known optics. Planck reached this conclusion, because he did not properly treat absorption and invoked polarized light in his derivation. Planck also made use of a carbon particle, which he characterized as a simple catalyst. This conjecture can be shown to result in a violation of the First Law of Thermodynamics, if indeed, all cavities must contain black radiation. In the end, while Johnson attempts to defend Planck's proof, his arguments fall short. Though the author has argued that Kirchhoff's law lacks both proper theoretical and experimental proof, Johnson avoids advancing any experimental evidence from the literature for his position. It remains the case that experimental data does not support Kirchhoff's claims and no valid theoretical proof exists.

If a space be entirely surrounded by bodies of the same temperature, so that no rays can penetrate through them, every pencil in the interior of the space must be so constituted, in regard to its quality and intensity, as if it had proceeded from a perfectly black body of the same temperature, and must therefore be independent of the form and nature of the bodies, being determined by temperature alone... In the interior therefore of an opake red-hot body of any temperature, the illumination is always the same, whatever be the constitution of the body in other respects.

Gustav Robert Kirchhoff, 1860 [1]

1 Introduction

Nearly two centuries have elapsed since Gustav Kirchhoff formulated his Law of Thermal Emission [1, 2]. In that time, this law has achieved unquestioned acceptance by the physics community, standing at the very foundation of thermodynamics, condensed matter physics, and astronomy. It constitutes the central pillar upon which Max Planck built his blackbody expression and his claims for universal constants [3, 4]. Eddington's theory of the stars, based on ideal gases, depends on Kirchhoff's law, in order to account for stellar spectra [5]. This remains true for stellar physics to this day [6, 7]. Kirchhoff's law constitutes a citadel for modern astronomy, defend-

ing not only the ideas that stars are gaseous plasmas devoid of lattice structure [5–7], that white dwarfs and neutron stars are highly compressed objects, and that black holes exist [8], but also the concept that a primordial atom once emitted a thermal spectrum and gave rise to the universe [9, 10]. It is precisely because Planck, Eddington, Chandrashekhar, Penzias, Wilson, Dicke, Peebles, Roll, and Wilkinson [1–10] relied on Kirchhoff's law, that they could ignore the central role of the structural lattice in helping to define the emissivity of an object.* While this could be understood in the days of Gustav Kirchhoff, it can no longer be permitted, in light of the tremendous advances made in condensed matter physics and medicine.

Hence, over the course of the past 15 years, I have turned my attention to Kirchhoff's law [13–18, 20–24, 24–26]. My interest in this law did not arise from any desire to study astronomy, but rather, as a consequence of assembling the first ultra high field magnetic resonance imaging (UHFMRI) scanner, at The Ohio State University [27–29]. It was as a direct result of questioning what it meant to say that nuclear magnetic resonance (NMR) and magnetic resonance imaging (MRI) were thermal processes. This had been highlighted long ago by Felix Bloch (Nobel Prize, physics, 1952) who was concerned with thermal processes linking the lattice and the spins [30].†

The laws of emission [1–4, 31–33], are just beginning to impact upon human medicine, as MRI scanners continue to be pushed to ever higher frequencies [27–29]. Thus, there is much more at stake here than the quest to a better understanding of the universe. Correcting Kirchhoff involves moving to a proper description of all thermal processes, not only in physics and astronomy, but in a field as seemingly remote and unrelated as radiology. I have stated that Planck's blackbody law, although valid, remains unlinked to physical reality [12, 17, 19, 23]. That is precisely because of Kirchhoff's faulty law. The physics community has not provided for thermal radiation what is evident for every other spectroscopic process, namely: 1) the setting under which emission occurs, 2) the nature of the energy levels involved, and 3) the nature of the transition species. Only 4) an equation, and 5) the emission of light, have been described [12]. Yet, in every other spectroscopic process, equations are related to physical reality. It takes a hydrogen atom, for instance, to obtain a Lyman or Balmer line. In that case, the transition species is the electron and the electronic orbitals constitute the energy levels. But, for blackbody radiation, spectra are related only to theory, unrestrained by a particular setting, such as the need to have a structural lattice.

That is how astronomers can justify the creation of blackbody spectra from any object. For instance, they have summed a large number of spectroscopic processes to account for the thermal emission from the Sun (see [34] for a complete discussion of this problem). Yet, not one of these processes can be related to the thermal emission from graphite. They have hypothesized that the Big Bang has generated the microwave monopole which surrounds the Earth [10], but have ignored the hydrogen bond from the water which makes up the oceans bathing our planet [35]. Once again, unrestricted by the need to describe thermal emission using a physical mechanism, astronomy has been left to postulate without any consideration of the central physical question in thermal emission: what causes a thermal photon to be emitted by graphite [19]?

Given all that is involved relative to the validity of Kirchhoff's Law [1, 2], Robert Johnson is to be commended, as the first duty of a scientist is to defend established science against possibly false charges. He has also been forthright in submitting a letter to this journal [36], rather than rely on anonymous attacks through social media.

At the same time, it would be an injustice to fail in one's own defense, when a proper understanding of science rests on the outcome. Therefore, I have decided to provide a point by point discussion of Johnson's letter [36]. I do so with the hope that some members of the physics community will begin to take an interest in Kirchhoff's claims and call into question many of the ideas which have been hypothesized [5–10], as a result of concepts which predate the discovery of the atom.

Before I begin analyzing the contents of Johnson's letter, it is vital to outline the setting under which Max Planck viewed a blackbody, as described in *The Theory of Heat Radiation* [4].

Throughout much of his text, Planck make use of perfectly reflecting walls to construct blackbody cavities. As I mentioned previously, this was *"an interesting approach"* [17, p. 4], precisely because such walls, in Planck's context [4], were *"adiabatic, by definition"* [17, p. 4]. They could not participate in generating, or absorbing, a single photon. Moreover, being adiabatic, they were also immune to all conductive and convective processes.

Conversely, unlike Planck, in thinking about perfectly reflecting cavities, I have invoked silver as a nearly ideal reflector of radiation in the infrared [21, 26]. Furthermore, I have insisted that cavities, constructed from such a perfect reflec-

*Nowhere is this more evident than when Eddington insisted that white dwarfs had to possess a small radius, in order to account for their lack of luminosity [5], given the well-established mass-luminosity relationship. Had Eddington considered the critical role of structure in defining emissivity, he would have seen that white dwarfs simply had a different hydrogen based lattice than the hexagonal planar arrangement shared by the Sun and the stars of the main-sequence (see [11, 12] and references therein). But deprived of the use of a lattice, when he stated that all stars could be viewed as ideal gases, Eddington had no other means of explaining the lower than expected luminosity of the white dwarf. Therefore, he was forced to reduce their radius to unreasonable values [5]. This was the first step towards hypothesizing highly dense objects, including the densities now attributed to neutron stars and black holes [8].

†Suffice it to say that the cavity experiments discussed later in this letter have relevance to both blackbody radiation and MRI. Furthermore, any valid analysis of noise power in MRI will be critically based on properly defining and modeling the processes responsible for thermal emission.

tor, possess a characteristic temperature. They are also subject to conductive and convective heat transfer in the establishment of thermal equilibrium. These are important modifications in properly addressing all thermal processes, including radiation, convection, and conduction. For while Planck properly insists that, at thermal equilibrium, there can be *"no conduction"* [4, § 25], no-one maintains that cavities cannot be subject to conductive processes in reaching thermal equilibrium. Laboratory blackbodies are usually brought to temperature using conduction. This will be important later in this letter.

As for adiabatic walls, they could never be characterized by *any* temperature, as I recently emphasized [23]. Consequently, they could never be in thermal equilibrium with anything. Planck stated *"Hence in a vacuum bounded by perfectly reflecting walls any state of radiation may persist"* [4, § 51]. That was very true. But it is also true that such cavities are devoid of any radiation, unless it had previously been injected by some outside means [15–17, 20–23, 26].

In the initial sections of his text, Planck had insisted that all of the energy could be characterized by the radiation field. In truth, the energy must have, at some time, been associated with his oscillators. Otherwise, no photons could have been produced. Thus, Planck's oscillators could be used to produce the field and set thermal equilibrium, but the energy of the system had to be considered as being irreversibly transfered to the radiation field: *"Accordingly we have frequently ... pointed out that the simple propagation of free radiation represents a reversible process. An irreversible element is introduced by the addition of emitting and absorbing substances"* [4, § 170].

This irreversibility and the need for the oscillators to have access to energy, in order to produce the photons, was vital to properly understanding this work. In addition, Planck admitted, in the very last section of his text that *"For the oscillators on which the consideration was based influence only the intensities of rays which correspond to their natural vibration, but they are not capable of changing their frequencies, so long as they exert or suffer no other action than emitting or absorbing radiant energy"* [4, § 190].

Planck insisted that he could place a minute particle of carbon within his cavities. He viewed this object as a catalyst [4, § 51–52], converting radiation within the cavity from one form to another: *"... This change could be brought about by the introduction of a carbon particle, containing a negligible amount of heat as compared with the energy of radiation. This change, of course, refers only to the spectral density of the radiation u_ν, whereas the total density of the energy u remains constant"* [4, § 71]. Planck's particle could only act on the radiation which was already in the cavity. It could not interact with the walls, introduce new energy into the cavity, or set the temperature of the system.

But to interact with the radiation, the carbon particle must have oscillators of its own, functioning over the proper frequency range. Namely, it must be a perfect absorber, characterized by a temperature and part of the thermal equilibrium problem, not a catalyst uncharacterized by any temperature. If devoid of a characteristic temperature, Planck's carbon particle would not contain the proper vibrations to even interact with the radiation in the cavity.

Neither the walls of Planck's perfectly reflecting adiabatic cavities, nor the catalytic carbon particle, could establish temperature. Planck resorted to placing all of the heat within the radiation field. None of the energy could be contained in the walls. He then altered the nature of his walls and removed the requirement that they could not interact with radiation: *"Since, according to this law, we are free to choose any system whatever, we now select from all possible emitting and absorbing systems the simplest conceivable one, namely, one consisting of a large number N of similar stationary oscillators..."* [4, § 135]. Note from this quotation, that Planck could advance no mechanism by which oscillators can actually alter the radiation distribution within the cavity. Planck's oscillators cannot convert the radiation from one form to another, as would be required in the action of Planck's carbon particle were simply catalytic. It remains the case that the radiation contained within a cavity can only be characterized by the nature of the oscillators which produced it. For all these reasons, Planck's carbon particle could never by considered as a catalyst. Indeed, if this particle is attributed with only a catalytic function, it can easily introduce a violation of the First Law of Thermodynamics, as will be seen in § 9 below.

At this point, it is time to address Johnson's submission [36]. In order to maintain the same section numbers, I begin immediately with a review of his introduction [36, § 1].

Johnson's first errors occur in his opening statement, wherein he asserts that I have *"... challenged the validity of Kirchhoff's Law of thermal emission and Planck's derivation of the mathematical form of the universal function of spectral radiance absorbed and emitted by a blackbody"*. There are actually two problems with this statement.

First, I never questioned the mathematical validity of Planck's expression, in the context of an actual blackbody. Rather, I have stated repeatedly that Planck's solution for a blackbody was correct (see e.g. [12, 16, 17, 25]). For instance, in [16, § 1] it is explicitly written that *"The accuracy of Planck's equation has been established beyond question"*. Along with Crothers, I state that *"Fortunately, in Planck's case, the validity of his equation is preserved, but only within the strict confines of the laboratory blackbody"* [25, § 4].*

Secondly, the absorbance of a blackbody does not have a functional form, contrary to Johnson assertion. When Kirch-

*It is troubling that Johnson has misrepresented my position on this matter. My concern has been exclusively centered on Kirchhoff's formulation of a law extending to objects which are not solids and which are constructed from materials lacking a good absorber [13–18, 20–24, 24–26]. I have never questioned the validity of Planck's equation in the case of proper blackbodies.

hoff formulated his law, he defined $E/A = e$ and immediately set A to 1 [1]. This enables the function "e" to have units. Johnson failed to understand, at the onset, Kirchhoff's formulation. Such errors continue throughout his letter [36].

2 Robitaille's earlier papers

2.1 Kirchhoff's law and Planck's proof

Johnson affirms in § 2, relative to Planck's law, that ... *"Planck's derivation is seen as proof of Kirchhoff's law"*. This is not correct.

There are actually several questions addressed in Planck's treatment. This was made clear in the manner in which Planck wrote his book, *The Theory of Heat Radiation* [4].

Planck was primarily concerned with two issues. First, did all arbitrary cavities contain black radiation? This is addressed in the first two chapters [4, § 1–52]. Secondly, Planck was focused on providing the functional form for the blackbody spectrum, through much of the remainder of his presentation. He did so by reviewing the laws of emission advanced by Wein [31] and Stefan [32].* He then discussed Boltzmann and entropy, and presented his oscillators and the blackbody function. In fact, the derivation of the blackbody function itself was completely independent of the derivation of Kirchhoff's law, since when setting $A = 1$, one obtains $E = e$ from Kirchhoff. The functional form of the blackbody spectrum can be obtained, without insisting that all cavities contain black radiation. Planck derived Kirchhoff's Law in the first section of his text solely because of his desire to confer, upon the blackbody expression, universal implications.

If Kirchhoff's Law would be found invalid, as it will eventually become, then Planck does not lose the functional form he supplied describing the radiation of a blackbody, as I have stated repeatedly [12,16,17,25]. However, it would imply that arbitrary cavities are not necessarily blackbodies and that the universality of the constants h and k does not hold [13–18, 20–24, 24–26].

In the next sentence, Johnson writes [36, § 2]: *"However, Robitaille points out that the above definition of Kirchhoff's Law is not complete and furthermore Robtaille maintains that the statement above should be called Stewart's Law as it was originally propounded by Stewart in 1858"*. How could Johnson make such claims?

He begins by omitting an important concept when citing my work. The complete citation is as follows: *"All too frequently, the simple equivalence between apparent spectral absorbance and emission is viewed as a full statement of Kirchhoff's law, adding further confusion to the problem. Kirchhoff's law must always be regarded as extending much beyond this equivalence. It states that the radiation within all true cavities made from arbitrary walls is black. The law of equivalence is Stewart's"* [17]. Importantly, in this citation, I had also included references wherein Kirchhoff's Law was described, solely in the context of the Law of Equivalence, and not within its full scope relative to claiming that a universal function existed. In any event, I never claimed that Kirchhoff's Law was not complete. What I did state was that people often give credit to Kirchhoff for the Law of Equivalence which properly belongs to Stewart [33]. As for Kirchhoff's Law, it is incorrect. Johnson does not seem to understand the fundamental differences between Stewart's Law [33] and Kirchhoff's [1].

The Law of Equivalence [33] simply affirms that, at thermal equilibrium, the radiation emitted by a surface will be equal to the radiation it absorbs, emissivity, ϵ, is equal to absorptivity, α. Stewart did not insist that the radiation inside all cavities was black. That is the reason Kirchhoff's Law [1] does not belong to Stewart [33]. This is an important point, as Johnson falsely asserts, throughout his letter, that Stewart recognized that cavity radiation must always be black. Rather, Stewart recognized that all cavities could become black if they could be driven (see [16] for further discussion). The problem, of course, is that cavities constructed from low emissivity materials cannot be properly driven [15–17, 22–25].

Stewart, while aware of mathematical arguments which might lead to such a conclusion, left the discussion to a footnote [33]. The reason was clear. Stewart recognized, as an experimentalist, that he was not able to prove, in the laboratory, that all arbitrary cavities were black. The experiments described in his work dealt with emission from plates and surfaces [33], not cavities [36]. That was precisely why he did not make a law for cavities, as *The Laws of Physics must be experimentally verified*. In his rebuke of Kirchhoff, Stewart had made the point plainly *"nor did I omit to obtain the best possible experimental verification of my views, or to present this to men of science as the chief feature, grounding theory upon experiment, rather than deducing the experiments from the theory"* (cited in [16]). Stewart never presented any experiments on cavities and therefore, he never made a law related to cavities, as Johnson claims I stated [36, § 2.1].

This was a central difference between the work of Stewart [32] and Kirchhoff [1, 2]. Johnson could have easily come to learn the distinction had he studied the historical review by Seigel [37], which I had cited in [16]. Siegel highlighted that ... *"Kirchhoff himself never performed any experiments which could be construed as attempts at quantitative experimental verification of his law"* [37, p. 588]. Seigel went on to state what Kirchhoff believed: *"... Kirchhoff was rightly pointing out that in this instance neither Stewart's experiments nor his own experiments sufficed to establish a quantitative law, and the burden of the priority claims would therefore have to rest on theoretical proof"* [37, p. 588]. Unfortunately, Kirchhoff was not right. Stewart's experiments were more than adequate to establish the Law of Equivalence. It was with the treatment of cavities that experimental confirmation was lacking.

*Planck never addressed the contributions of Balfour Stewart [33].

In any event, experiments take precedence over theory when it comes to formulating a new law, as our theories are not able to define nature. Furthermore, it is all too easy to accidentally omit a critical element from a theoretical discussion, as has happened when Kirchhoff and Planck unknowingly ignored the energy trapped within the walls of cavities. Such energy can remain forever unavailable to thermal emission. That is why Kirchhoff's Law is invalid. It also provides an illustration of the danger of inferring the laws of physics from theory.

In the end, Seigel also highlighted the difference between Stewart's law and Kirchhoff's claims: *"Stewart's conclusion was correspondingly restricted and did not embrace the sort of connection between the emissive and absorptive powers of different materials, through a universal function of wavelength and temperature, which Kirchhoff established"* [37, p. 84]. It is clear that Stewart's Law did not encompass the universal nature of cavity radiation which Kirchhoff sought, as Johnson attempts to inappropriately claim throughout his letter.

This section closes with Johnson quoting from § 51 of Planck's text [4] and insisting that by placing an *"arbitrary small quantity of matter"* in a perfectly reflecting cavity that *"Planck had thereby demonstrated that all cavities either containing some arbitrary matter, or equivalently having walls made of some arbitrary matter, must also contain black radiation when at thermal equilibrium"*. Yet, in § 51, Planck was placing a small particle of carbon in the cavity. The carbon particle was not an arbitrary material. It was acting as a perfect absorber. I have discussed the inappropriate introduction of a perfect absorber into cavities in detail [16] and will return to the question, once again, in § 2.2, § 2.5 and § 2.9 of this letter.

2.2 Black radiation in a perfectly reflecting cavity

As Johnson opens the third section of his letter, he objects to my conclusion that Planck's statement, *"Hence in a vacuum bounded by perfectly reflecting walls any state of radiation may persist"* [4, § 51], constituted an implicit admission against the validity of Kirchhoff's law.

In trying to defend Planck, Johnson writes: *"However, Planck's statement should perhaps be more properly viewed as a situation to which Kirchhoff's law does not apply because there is no matter present which could absorb or emit radiation."* However, Kirchhoff's law was meant to be independent of the nature of the walls, by definition. Planck associated the temperature of a cavity solely with the radiation it contained, not with any material particles.*. If Kirchhoff was correct, what difference should it make if matter was present to absorb or emit radiation? Nothing in Kirchhoff's law required this restriction and that was precisely the problem. Kirchhoff's law was devoid of all link to actual materials and nature. It was only concerned with hypothetical cavities.

In considering Kirchhoff's law, we can simply examine mathematical limits, as defined by the opaque perfectly absorbing wall (absortivity, $\alpha_\nu = 1$; reflectivity, $\rho_\nu = 0$) and the opaque perfectly reflecting wall ($\alpha_\nu = 0$; $\rho_\nu = 1$). Yet, the second condition led to an undefined expression for Kirchhoff's law, as Planck himself recognized [4, § 48]. It was not possible to claim that a law applies to all materials, when one of its limits was undefined.

Johnson goes on to cite Planck's §51 stating that the radiation within all cavities will always be black, even though Planck, in the very same section, has just introduced a particle of carbon in this cavity, which Johnson recognizes as being a *"perfect absorber and emitter at all frequencies"* [36]. But, Planck viewed the carbon particle as a catalyst [4, § 51–51]. Johnson then writes, in speaking of Planck: *"Note that the quoted statement covers both the situation where the object absorbs or emits over all frequencies, and the situation where some frequencies are not absorbed or emitted at all"* [36].

Planck reached his conclusion by inserting a particle of carbon. This ensured absorption and emission over all frequencies. Planck never demonstrated that this applied to *situations where some frequencies are not absorbed or emitted at all"*, as Johnson claims [36]. Planck placed the carbon particle within the cavity and then claimed that it acted only as a catalyst. He sidestepped the fact that this particle was acting as a perfect absorber, and thereby controlled the entire problem. I have already demonstrated this fact mathematically and the reality that arbitrary cavities, at thermal equilibrium, do not contain black radiation [15]. Importantly, Johnson's letter fails to address these simple algebreic proofs that Kirchhoff's Law cannot be valid [15].† Again, I will return to the question of the carbon particle in § 2.5 and § 2.9.

2.3 The approach to equilibrium

Johnson opens this section by pondering what was correct: Do all cavities contain black radiation, as Kirchhoff and Planck held, or do arbitrary cavities contain arbitrary radiation, as Robitaille asserted? The question was simple enough to answer, as blackbodies are always constructed from good absorbers.

In fact, had Johnson considered the history of blackbody radiation, he would have recognized that arbitrary cavities are never black. That is why those who provided Max Planck with the data used to verify his equation worked so hard to

*"Still, even Planck recognized that material objects were required to establish a temperature, *"But the temperature of a radiation cannot be determined unless it be brought into thermodynamic equilibrium with a system of molecules or oscillators, the temperature of which is known from other sources"* [4, § 144]

†Reference [15] contains a detailed analysis of some of the problems with Kirchhoff's logical arguments in advancing his proofs [1,2]. It also contains simple proofs of Stewart's Law of Equivalence [33] and clear demonstrations that arbitrary cavities, under conditions of thermal equilibrium, are not black. Johnson cannot ignore these proofs in his letter, if he wishes to honestly evaluate my work.

construct laboratory blackbodies which provided the proper functional form [38–40]. These papers, especially the review by Hoffmann [38], are important to study, because they highlight the complexity of building proper blackbodies.

As a simple example, the problem can be viewed to involve, to some extent, the behavior of graphite itself. In the visible range, some forms of graphite, which are mined, can be relatively good absorbers, but others, surprisingly, can be rather poor, as can be ascertained by examining emissivity tables [41]. However, as one becomes increasingly interested in the region towards the infrared, graphite begins to fail. This has been known since the days of Langley at the end of the 19th century [16, § 2.1]. That is why materials like the metal blacks are utilized, in this region of the electromagnetic spectrum, to assemble blackbodies [42–45].

We already have the experimental proof, but most people simply ignore these laboratory realities. For, if Kirchhoff's Law was valid, there would be no need for metal blacks in building laboratory blackbodies and German scientists would not have used rolled platinum and specialized mixtures of chromium, nickel, and cobalt oxide to blacken the interior of their cavities [38, p. 57]. Such mixtures indicate that their was nothing arbitrary in the construction of blackbodies.

This remains a specialized field and such objects are *always* sophisticated devices unavailable when Kirchhoff extended his law to all cavities.*

In this same section of his letter, Johnson goes on to consider what would happen to the radiation, within an arbitrary cavity, if the initial radiation was less than the maximal hypothesized by Kirchoff's law. The arguments he advanced are flawed at a fundamental level.

Johnson first places an opaque object within a perfectly reflecting cavity and defines that the intensity of the radiation is 100 within the cavity, the proper value for black radiation. He assumes that the object has an emissivity of 0.8 and then states that when radiation within the cavity interacts with the object, 80 units will be absorbed/re-emitted and 20 units being reflected. Johnson notes that the radiation within such a cavity will remain black at 100 units. Of course, the experiment is false, as an object with an emissivity of only 0.8 could never fill the cavity with black radiation in the first place. The radiation would have to be increased by some other means.† Deviations from this case are only permitted if thermal equilibrium has been violated, after the cavity and the object reached the temperature of interest, or if the perfectly reflecting cavity has otherwise been filled with black radiation [16, 17]. It is important to recall, that even the sampling of a cavity with a detector can act to fill it with black radiation [17, § 2]. Therefore, this situation, as described by Johnson, does not lend any support to Kirchhoff's claims. It was simply ill-conceived.

At this point, Johnson considers another scenario wherein an object with an emissivity of 0.8 can only emit 80 units initially into the cavity. These 80 units then strike the wall and reflect back towards the object, where now he claims that only 64 units are absorbed (since the emissivity is 0.8), and 16 units are reflected. Johnson notes that the object *"... was bound by its initial temperature to continue emitting 80 units"* [36, § 2.3]. He notes the shortfall in the total amount of radiation absorbed by the object, and claims that this can only be rectified by lowering the temperature of the object. The errors in logic are striking.

First, Johnson fails to recognize that it is the total radiation coming off the object at thermal equilibrium which matters. That total radiation is equal to 64 units emitted and 16 units reflected at the onset, because the cavity and the object are already at thermal equilibrium, by definition. Johnson does not get to say that the object must emit 80 units to begin his experiment and then state that only 64 units are absorbed and re-emitted. He can only sample the total radiation coming off the object. He has no means of distinguishing what was, in fact, reflected and what was emitted. He only knows that 80 units came off the object. These are then reflected off the wall and travel towards the object, where 64 units will be absorbed, then re-emitted, while 16 units will be reflected. Johnson also fails to understand that he cannot allow the temperature of the object to drop, as this is a violation of the zeroth law of thermodynamics. For my part, I would not disallow thermal equilibrium between the cavity and the object, as Johnson asserts.‡

Relative to the last experiment, it is interesting to note what Johnson has actually done. At first, he ignored reflection, stating that all 80 units leaving the object were emitted. Then, on absorption, he now considered reflection, permitting only 64 units to be absorbed and the remaining 16 to be reflected. So what has happened?

Note, for instance, that when Max Planck derived the first section of his proof of Kirchhoff's Law, he also ignored reflection (see [25, § 4.2] for a complete description of what Planck did in this instance). Robitaille and Crothers note that Planck was allowed to ignore reflection, as these terms, if retained, could be canceled out [25, § 4.2]. They also demonstrate that the full treatment retaining reflection can lead to additional insight, relative to this problem [25, § 4.2].

If Planck was allowed to ignore reflection, perhaps this can be most easily explained by examining the Law of Equivalence itself [33]. I have already highlighted that Stewart's Law can be written either as, $\epsilon_\nu=\alpha_\nu$, or as, $\epsilon_\nu+\rho_\nu=\alpha_\nu+\rho_\nu$ [15]. The use of either form will lead to the correct answer. However, what Johnson has done was to mix the two forms of Stewart's law, inventing a scenario wherein he sets $\epsilon_\nu=\alpha_\nu+\rho_\nu$,

*The author has reviewed laboratory blackbodies in [16, 17].

†I have already demonstrated mathematically, that the radiation in the cavity, in this case, will not be black but will have an intensity appropriate for the emissivity of the object it contained [15].

‡I also reject all of Johnson's other deductions relative to how I would view an experiment which I never even described in my papers.

which is clearly false.

At this point, Johnson once again tries to state that I have attributed Kirchhoff's Law to Balfour Stewart. In this, he misses the central point. Stewart's footnote does not make a law of physics. It presents a mathematical argument. Stewart recognized that, if he wanted a blackbody spectrum from a cavity, he must have recourse to lampblack. Johnson believes that Stewart was specific on this point, arguing for a *"theoretical leap"* [36]. But in so thinking, he failed to recognize what Stewart understood: cavities can only be *demonstrated* to be black experimentally if they contain a good emitter. Stewart did hypothesize extensively about bandied radiation, well after 1858 (see [16] for a complete discussion), and conjectured that cavities of low emissivity can be made to appear black. The thesis was never proven and with good reason [15–17, 22–25]. Stewart stated a theoretical idea, not a law. The point has been made clearly by Seigel, as noted in § 2.1 above [37]. That is why I wrote in my initial paper: *"Stewart realizes that the lampblack surface within the enclosure is essential"* [16]. Stewart might have had a theoretical argument, but he did not have data. It is in this aspect that he was much more prudent than Kirchhoff when he presented his work [33]. That is why I have always acknowledged this Scottish scientist. Stewart exercised wisdom in 1858 [33] and Johnson shall not deprive him of this quality.

2.4 Stewart's treatment of reflection

Johnson then goes on to describe, in detail, Stewart's footnote, as if this was central to the idea which Stewart was conveying. Stewart's paper deals with the Law of Equivalence, not with cavity radiation and universality [33]. The argument which Johnson resurrects is contained in a footnote, precisely because this constitutes its proper position in the paper. Stewart makes us aware that he understands a mathematical argument previously advanced by others (see references contained in [17]), but he does not raise them to a central part of this thesis, because these ideas were not supported by laboratory data.

In considering the bandied radiation, Johnson makes the claim that the energy required to fill the cavity can be extracted from the walls in order to drive *"Stewart's mechanism"*. In this aspect of his letter, Johnson is actually repeating ideas from my own papers on cavity radiation, wherein such processes have already been discussed in detail [21, 23, 26].* Johnson adds nothing new to this discussion. He also fails to understand, at a fundamental level, that it is by invoking the energy retained in the wall that Kirchhoff's Law can be proven to be false [21, 23, 26]. Planck specifically used an adiabatic wall which could not be characterized by any temperature to build his perfect reflector, because he wanted all of the energy to be contained in the field, not in the wall [4]. Since adiabatic walls are detached from all thermal processes (i.e. radiation, conduction, convection), they cannot be characterized by any temperature [21].

Johnson analyzes Stewart's experiments [33] with low emissivity plates in obtaining the same functional form as if the plate had been black. Yet, it is not solely a question of time elapsed, as he attempts to argue. For instance, he permits the temperature of one of his plates to drop in clear violation of the Zeroth Law of thermodynamics *"... the only difference in this case is that during the initial period the partially absorbing plates is absorbing less radiation than it is emitting; it is therefore cooling down and part of its initial energy is being used to increase the radiation density between the plates, or, in Robitaille's terms, in "driving the reflection"* [36, § 2.4]. I never permitted an object temperature to drop, in order to drive the reflection.

My papers are concerned with a law defined under thermal equilibrium, not the approach to equilibrium. I have highlighted that one cannot create photons from nothing. Scientists are not permitted to violate the First Law. What has happened in this letter is that Johnson permits the temperature to drop in order to avoid violating the First Law, as he knows that he must get photons from somewhere. The arguments are all invalid, as we are concerned with a system in thermal equilibrium, not the approach to such equilibrium.

While Johnson understands that the idea of driving a cavity is an important concept, he continues to ignore its consequences. For instance, such processes rely on access to a perfect absorber, or some temporary violation of thermal equilibrium [15–17, 22–25]. They are also prone to introduce a violation of the First Law of Thermodynamics, as energy must come from somewhere. Also, energy cannot be destroyed.

The question, relative to thought experiments, relates to the origin of the energy entering a cavity once it is already at thermal equilibrium. Provided that the cavity walls are not adiabatic, but can be represented by graphite, or silver, then there are three scenarios to consider: 1) energy enters the system from outside, 2) energy travels reversibly out of the walls of the cavity to irreversibly fill the cavity [21, 26], and 3) energy is irreversbilty trapped within the cavity walls [26]. None of these possibilities have ever been considered by Planck. They arise from the assembly of work which is currently being challenged by Johnson.

Let us assume that the energy came from outside the system. Then, once it reaches the cavity walls, it must be allowed to either 1) help fill the cavity with additional photons, or 2) dissipate additional energy into the walls of the cavity. However, the cavity walls are already at a given temperature. To permit additional energy to enter would alter this value. As such, no energy can be allowed to enter the walls, as this would violate the zeroth law. Thus, if any energy enters the cavity walls from outside the system, it must simultaneously leave and produce additional photons in the interior. It is clear that, with such a scenario, if the walls are fully reversible

*The author published [26] just a few days before Johnson submitted his letter and he was made aware of this work.

stores of energy, the cavity will become filled with radiation. The problem becomes, when does one stop? Obviously, the experimentalist can place any amount of photons in the cavity, given enough available energy from the outside and no concern for the First Law. If the radiation intensity within the cavity becomes too great, then he can simply affirm that thermal equilibrium has been violated and that the cavity must now be represented by a higher temperature.

As for the idea that the energy contained within the walls can be reversibly used to fill the cavity with even more radiation, I have already considered the concept on two occasions [21, 26]. In reality, such processes are likely to be physically impossible. Thermodynamically, the concept is allowed, but the problem is that, if the energy of the walls if fully available to build up photons in the interior, the cavity would already be black, unless specialized means are used to isolate this energy [26]. In reality, every material which is not a perfect emitter will actually possess at least some energy which is irreversibly trapped in the walls [26] relative to the ability to support emission. That is the central reason why arbitrary cavities are never black. Planck had considered that only the production of the radiation field was irreversible, as I discussed in the introduction. This may have been everyone's major stumbling block relative to cavity radiation and Kirchhoff's Law. Prior to 1906, when Planck's lectures where written [4, p. xi], neither he, nor Kirchhoff, understood that some of the energy which enters a metal will be trapped in its conduction band electrons and forever remain unavailable to emission [26].* We shall return to *"Stewart's mechanism"* in § 2.5, § 2.6, § 2.7, and § 2.9.

2.5 Planck's particle of carbon

Johnson then moves to try to defend Planck's use of a carbon particle as a simple catalyst. I have already spoken extensively on this issue: Planck's carbon particle is not a catalyst [16]. It is a perfect absorber/emitter. Planck uses carbon, not a particle of some other material, and with good reason. He needs a perfect absorber. It is not simply a question of having a particle which can absorb over all frequencies of interest. In fact, a quick study of emissivity tables would demonstrate that, if this were the case, Max Planck had many other materials available to him [41]. He wanted a perfect absorber and, when he placed it in his cavity, as I have said previously, it was as if he had coated the entire inner surface with lampblack. Otherwise, what does it mean to be "perfect"? As I stated in the introduction, the reality remains that Planck's carbon particle must have access to oscillators, otherwise it cannot even interact with the radiation. It must also be characterizable with a temperature, such its oscillators could operate over the entire range of frequencies required to make the cavity radiation black at the proper temperature. The need for this temperature directly implies that the carbon particle is a perfect absorber, not a catalyst.

Johnson claims that there is a difference between *"the nature of the black radiation and the quantity of it"*. He then argues that Planck has made the particle small such that its energy content can be neglected relative to filling the cavity with radiation. Planck's position and Johnson's defense are not well-reasoned in that they neglect that the particle and cavity must be allowed to come to thermal equilibrium. This is one of the reasons why Planck's use of an adiabatic wall to build a perfectly reflecting cavity is not appropriate. Planck also attempted to deprive the carbon particle of a specific temperature. In so doing, he was overlooking the very detail which was critical to obtaining the proper answer (see also § 2.9). Johnson states, *"... By definition, therefore, the carbon particle cannot increase the radiation density in the cavity to the level commensurate with the black body temperature; in Robitaille's terms, the particle cannot "drive the reflection", and therefore this cannot be the reason why Planck included it. Furthermore, if the radiation density is being increased at all frequencies by Stewart's mechanism then there is no need to include the carbon particle"* [36, § 2.5]. Unfortunately, for Johnson, he cannot resort to *"Stewart's mechanism"*, as he cannot practically demonstrate its validity in the context of a perfect reflector. Even Stewart, cannot generate photons from a perfectly reflecting cavity. The issue at hand is the carbon particle, not *"Stewart's mechanism"*.

As such, let us first consider the proper way of viewing the carbon particle, then return to Planck and Johnson, both in this section and in § 2.9.

The simplest means of addressing this problem is to consider that whatever light is reflected off the walls of a perfectly reflecting cavity can strike the particle. The particle must then transform the radiation and return this light back towards the cavity walls [15]. The temperature of both the cavity and the particle must be the same and the temperature of the latter must not be allowed to drop in order to respect the zeroth law. Under this condition, full equilibrium between the walls and the particle would exist and the cavity could easily be demonstrated to contain black radiation [15]. Herein was the power of equilibrium arguments.

In order to further clarify the point, let us consider what was physically occurring within the cavity when Planck introduced his small particle of carbon. Since the cavity was perfectly reflecting, we can assume that it can be best approximated by polished silver [23,26], not by an adiabatic wall [4]. The emissivity of the cavity must be 0 and it initially contains no photons. Let us surround the cavity with an adiabatic wall, in order to isolate the system.

As a result, the temperature of the cavity in this case is defined by the energy content of its walls. When the carbon particle is introduced into such a system, even if it contains no appreciable heat on its own, it also comes into thermal contact with the wall of the perfectly reflecting cavity. At this point,

*The energy can still be removed from the wall through conduction and convection.

thermal energy will become available to the carbon particle from the cavity wall. This particle can then transform the energy which would be otherwise irreversibly trapped in the walls [26] and fill the cavity with radiation. In this sense, the carbon particle acted as a transformer, converting phonon energy and/or energy associated with thermal conduction in the silver wall into photons. It was not a catalyst, as it was critical to conversion occurring. I have always modeled perfect reflectors using silver [23, 26], not using adiabatic walls. Without the carbon particle, the cavity would remain devoid of any radiation and all of its energy would remain forever trapped in its walls.

As for the case considered by Max Planck, an adiabatic wall contained no energy. Therefore, the carbon particle, devoid of significant heat, could never fill such a cavity with radiation at any temperature.

Contrary to Johnson's claim, neither Crothers, nor I, have said that the carbon particle cannot increase the radiation inside the cavity. Rather, my papers provided the only means for the carbon particle to fill the perfectly reflecting cavity. As for Johnson, he must adopt Planck's position, and remain forever unable to consider the content of the walls and the ability of the carbon particle to transform this energy into photons. He cannot be permitted to jump between my model and Planck's, as this is the entire basis of this discussion.

If Planck stated that *"Hence in a vacuum bounded by totally reflecting walls any state of radiation may persist"* [4, §51], it was because he recognized that Kirchhoff's law became undefined when A=0. But that does not mean that the cavity in this case contains forms of radiation which are blackbody, unless such radiation has been introduced by some outside mechanism.* In fact, the perfectly reflecting cavity must be considered empty, because it had no means of producing a photon and all of its energy content was trapped in its walls before the introduction of the carbon particle [26]. Johnson argues that, if the spectrum is indeterminate at any frequency, it is impossible to set a temperature. Again this is false, as the walls also contain energy [26]. Max Planck also ignored this fact, a critical error in selecting adiabatic walls.

Johnson then cites Planck's discussion [4, § 11] that all objects show significant reflection at sufficiently long wavelengths, except perfect blackbodies. He concludes that this is why Planck introduced the carbon particle [36]. But, if that was true, then Planck's introduction of the carbon particle would be acting to make all cavities perfect blackbodies, a point which supports my position.

In closing this section, Johnson makes the charge that I now accept, at least in principle, *"Stewart's mechanism"* for building up the reflection within a cavity. He alleges a remarkable *"volte-face"* on my part when I published a paper with Crothers [25].

Such a conclusion is not reasonable, as my papers have always considered Stewart's hypothesis (see e.g. [15–17, 22–24] all of which precede [25]). Furthermore, Crothers and I have restated, in no uncertain terms, that *"Stewart's mechanism"* does not work [25, § 2, 3].

Over the years, I have dealt consistently with the problem of thermal emission and have always held the position that arbitrary cavities are not black. I have examined numerous questions including 1) perfectly absorbing cavities, 2) perfectly reflecting cavities, 3) perfectly reflecting cavities containing a carbon particle, 4) perfectly reflecting cavities containing an arbitrary object, 5) perfectly reflecting cavities containing two arbitrary objects, 6) perfectly absorbing cavities containing a perfectly reflecting cavity, 7) two cavity problems (both for the reversible and the irreversible cases), 8) actual laboratory blackbodies, 9) Kirchhoff's two faulty initial proofs, 10) Planck's faulty proof of Kirchhoff's Law, 11) proper equations governing cavity radiation and 12) effects of driving the reflection term. Nowhere have I ever stated that *"Stewart's mechanism"*, as Johnson refers to bandied reflection, can ever lead to black radiation in all cavities, despite repeatedly addressing the question [15–17, 22–25]. What I have stated is that, if one tries to drive a cavity made from materials with a low emissivity, in order to build up black radiation in its interior, it is likely that the cavity will simply prefer to move to a higher temperature [23]. That is because any energy introduced into the cavity must also be available to the walls. If those walls cannot easily emit a photon, they will simply increase their temperature. Moreover, I have emphasized that the use of bandied radiation, even if possible, could only lead to filling a cavity with black radiation, in the ideal that the walls were capable of Lambertian reflection [23]. No specular reflection must have taken place and all reflection must have been diffuse. Otherwise, one risks generating standing waves, as I have previously highlighted [16] (see also § 2.6). Johnson ignores all these points when he addresses bandied radiation.

2.6 Experimental evidence against Kirchhoff's law

This is perhaps the most unusual section of Johnson's letter [36], as he tries to explain why manufacturers do not build blackbodies from arbitrary materials. Rather than concede that this constitutes direct experimental evidence against Kirchhoff's law, as I have stated, Johnson reaches for the indefensible. He argues: *"It is also likely that manufacturers are concerned, as Planck himself apparently was, to ensure that there are no frequencies at which the cavity is a perfect reflector, which would preclude a proper measurement of temperature"*.

Kirchhoff's Law demands that all cavities be black, independent of the nature of the walls. Manufacturers are not concerned with materials acting as perfect reflectors, since most solids emit continuous spectra over a wide range of frequencies. The problem is that many solids are poor emitters,

*I will return to this issue in § 2.9.

not that they are perfect reflectors.

Furthermore, the temperature of a cavity in the laboratory is determined by temperature sensors in its walls. Cavities are heated, conductively or otherwise, the temperature on the sensors in its walls are noted, and thermal equilibrium is defined by those sensors maintaining a stable temperature reading. Establishing the temperature of a laboratory cavity has nothing to do with measuring its radiation field and it would be irrelevant, if some frequencies were absent from the spectrum. This cannot affect the reading of a sensor in the wall of the cavity.* Even Planck recognized that a proper measure of temperature depends on the use of sensors or thermometers: *"But the temperature of a radiation cannot be determined unless it be brought into thermodynamic equilibrium with a system of molecules or oscillators, the temperature of which is known from other sources"* [4, § 144].

Johnson then moves to question any work in microwave cavities. He launches this new challenge precisely because these cavities are known not to contain black radiation, as I have demonstrated experimentally using UHF frequencies near the microwave region [17]. In attempting to dismiss microwave cavities, Johnson cites Planck: *"The last statement excludes from our consideration a number of radiation phenomena such as fluorescence, phosphorescence, electrical and chemical luminosity"* [4, § 7]. Johnson's use of such a quotation relative to microwave cavities demonstrates that he does not fully understand the experimental problem.

Kirchhoff's law allows for the presence of any object within the cavity. Therefore, the resonant elements used in my own work [17] are allowed, as they do not emit a single photon. They build up standing waves. Still, it remains clear that fluorescence, phosphorescence, electrical and chemical luminosity cannot be considered.†

The microwave cavity is not producing radiation by some non-thermal means, like fluorescence, phosphorescence, electrical or chemical luminosity. Rather, it is being subjected to sampling by a network analyzer which is sending microwave energy into the enclosure and noting what energy returns. If the cavity is able to reflect some of this radiation internally, then it can build up standing waves. Alternatively, if the cavity is truly black, then it should be able to absorb all the energy coming from the network analyzer with no returned energy. In any case, such return-loss measurements on cavities are routinely done throughout thermometry (see references cited in [48]).

In the infrared, cavities can be subjected to radiation from a standard blackbody, for instance, in order to verify their absorptivity by noting the returned energy.‡ In the microwave, when testing blackbodies for satellites, the source is often a network analyzer (see references cited in [48, 49]). This is a common measurement in testing the quality of blackbodies at these frequencies.

Johnson must recognize that microwave cavities are utilized on satellites such as COBE [50] and PLANCK [49]. These cavities are tested using return-loss methods, exactly as I have done in [17], when testing an MRI cavity. Many of these cavities are not black, including some which have been claimed as such and launched aboard satellites [49]. Microwave cavities often contain signs of standing waves, as radiation from the network analyzer enters the cavity. The presence of such standing waves provides solid evidence that not all cavities in the microwave contain blackbody radiation. This is an important point to recognize, as Johnson would like to build up arbitrary radiation in cavities with reflection, using *"Stewart's mechanism"*. Standing waves demonstrate that the presence of specular reflection within a cavity is always counter to the interior containing black radiation [48,49]. This highlights yet another problem with *"Stewart's mechanism"*. It is critically dependent on any reflection within a cavity being diffuse and not specular. Otherwise, the radiation will not be Lambertian, as required of a blackbody.

Contrary to Johnson's position, experiments with cavities in MRI provide strong evidence that Kirchhoff's law does not hold (see [17] and references therein). This is especially true given that Kirchhoff's Law has been generalized to treat geometries where diffraction becomes important (see [17] and references therein). Furthermore, microwave studies demonstrate that small cavities, containing only a few centimeters of Ecosorb and conductively anchored to a radiation shield, like the 4K reference loads on the Planck satellite, can never be black [48, 49]. This presents a serious problem for those interested in the LFI data produced by this satellite [49].

Once again, the fact remains that Kirchhoff's Law does not have any valid experimental support. Arbitrary cavities are not black and this reality has consequences which must not be ignored.

*However, for real blackbodies, when the temperature sensors indicate a certain temperature, one can be assured that the radiation sampled will be black.

†Surprisingly however, in Kirchhoff's initial paper [2] he actually insists that even fluorescent material could be included within the cavity and it will still be black: *"It may be observed, by the way, that the proposition demonstrated in this section does not cease to hold good even if some of the bodies are fluorescent. A fluorescent body may be defined as one whose radiating power depends on the rays incident on it for the time being. The equation E/A = e cannot generally be true for such a body; but it is true if the body enclosed in a black covering of the same temperature as itself, since the same considerations that led to the equation in question on the hypothesis that the body C was not fluorescent, avail in this case even if the body C be supposed to be fluorescent.*" [2]. These arguments are removed however, without explanation, when Kirchhoff's work is revised several years later [46]. Still, this indicates a flaw in Kirchhoff's initial derivation of his law [2], as he had thought that his derivation applied to fluorescent bodies, which was not correct. There are indeed flaws in Kirchhoff's initial derivation, as the author has independently ascertained [15]. Moreover, Schirrmacher [47] has reviewed the proofs of Kirchhoff's Law before and after Planck [48]. Even in 1912, Hilbert complained that a valid proof a Kirchhoff's law still did not exist [47], even though the Planck's lectures on the subject were given in 1906 [4, p. xi]. Such a proof is lacking, to this day.

‡Note that if Kirchoff's law was correct, there would be no need to have standard blackbodies in order to calibrate other cavities, as all cavities would be black.

2.7 Challenges to Monte Carlo simulations

Johnson then moves to briefly discuss Monte Carlo simulations in a single paragraph stating: *"Apparently, Robitaille's objection to the Monte Carlo simulations is that they rely on Stewart's mechanism for building up the radiation by internal reflection. As Robitaille and Crothers now accept that this mechanism is valid in principle, Robtaille's previous objections to Monte Carlo simulations supporting Kirchhoff's Law should also drop away"*.

Clearly, I have never accepted *"Stewart's mechanism"* for building up radiation within a cavity. First, such a mechanism, under certain circumstances, constitutes a violation of the First Law of Thermodynamics. Secondly, it is not possible to place energy into the interior of a cavity without also potentially placing energy into the walls. This is never considered by Monte Carlo simulations, and that is why they remain invalid. Such simulations agree with Kirchhoff Law, precisely because they ignore the dynamics going on in the wall and *a priori* forbid the temperature of the wall to rise in lieu of emitting a new photon.

2.8 Super-Planckian emission

Johnson's letter then examines my treatment of metamaterials [23]. He argues that Planck specifically excluded the near field by quoting: *"Throughout the following discussion it will be assumed that the linear dimensions of all parts of space considered, as well as the radii of curvature of all surfaces under consideration, are large compared with the wave lengths of the rays considered"* [4, § 2]. On the surface, this is a good point. Planck is clearly allowed to restrict his derivation. This does not mean, however, that the near field region cannot be considered today, in order to shed additional light on thermal emission.

In this regard, Kirchhoff's law has been generalized to include the limit initially excluded by both Kirchhoff and Planck [17, § 3]. The near field behavior can be considered for additional insight and the point raised by Johnson is weak at best. Science does get to move forward.

Johnson then goes on to claim that the evidence in the far field, is not convincing. He notes from Guo et al. [51] that: *"the presence of an interface is enough to guarantee that the far-field emissivity is limited to 1"* [36]. Guo's statement is noteworthy. However, Johnson neglects to cite the following from Guo's paper: *"The usual upper limit to the blackbody emission is not fundamental and arises since energy is carried to the far-field only by propagating waves emanating from the heated source. If one allows for energy transport in the near-field using evanescent waves, this limit can be overcome"* [51].

It is clear that the study of metamaterials is an area of science which is just beginning to be explored. It is also not established that far-field behavior will always adhere to the limits set forth by Planck's law. This is why I previously highlighted [23] the work by Yu et al. [52] and [53]. Yu et al. removed the claim made in the arXiv version of their paper [53] when they published their *Nature Communications* paper [52]. Here is the exact quotation from my paper on this issue: *"In that case, the spatial extent of the blackbody is enhanced by adding a transparent material above the site of thermal emission. A four-fold enhancement of the far-field emission could thus be produced."* In their *Nature Communications* article, the authors argue that this does not constitute a violation of the Stefan-Boltzmann law, because the effective "emitting surface" is now governed by the transmitter, which is essentially transparent. However, this was not the position advanced when the results were first announced and the authors wrote: *"The aim of our paper here is to show that a macroscopic blackbody in fact can emit more thermal radiation to far field vacuum than* $P = \sigma T^4$*"* [53].

In Yu's work, the emission is arising from a small blackened disk of material [52, 53]. The photons emitted from this surface greatly exceed anything predicted by Planck. At issue is the assignment of the emitting surface, from a theoretical perspective. Is it the blackened disk, which is the only possible source of photons, or the transparent shield? The key difficulty for blackbody radiation science is that blackbodies were always defined as opaque objects. Hence, it is difficult to conceive why the blackened disk should not be considered as the proper emitting surface in this problem. But assigning the emission to a transparent surface is now the only way of salvaging Kirchhoff's law. Once again, note how Kirchhoff had worded his law in the quotation at the very beginning of this reply. He was referring to opaque objects.

Then, there is the problem that, during Yu et al's experiment, the blackened disk is always heated [52, 53]. This implies that thermal equilibrium does not exist, since conduction of energy, which is heating the disk, must be considered. However, if this is to be used as an argument against these findings, then what of the problem of continuously heating ordinary cavities, in order to maintain their temperature equilibrium? As I previously stated: *"Obviously, modern experiments fall short of the requirements for thermal equilibrium, as the cavities involved are heated to the temperature of operation. But given that all laboratory blackbodies suffer the same shortcomings, the production of super-Planckian emission in the near and far fields cannot be easily dismissed. After all, in order for Planck to obtain a blackbody spectrum in every arbitrary cavity, he had to drive the reflection term, either by injecting a carbon particle or by permitting additional heat to enter the system, beyond that required at the onset of thermal equilibrium"* [23]. Johnson cannot apply his arguments to metamaterial experiments and not make them with regard to regular laboratory cavities. In light of these many considerations, he has not demonstrated that my position, relative to the universality of blackbody radiation, has been overstated.

2.9 Robtaille's thought experiment

In the next section of his paper, Johnson reviews a very simple thought experiment, which I advanced in 2014, illustrating that Kirchhoff's Law cannot be valid [20]. Briefly, the idea involved two cavities. The larger outer cavity was constructed from perfectly emitting and absorbing walls and initially placed in a helium bath. Within this cavity and *in thermal contact with its floor*, rested an inner cavity made from perfectly reflecting walls. Initially, one of the six sides of this latter cavity remained open. As such, both cavities now contained black radiation at 4K, which had been produced by the outer cavity. The open wall of the inner cavity was then closed. It thus contained blackbody radiation at 4K. Then, the helium bath was removed and the system was allowed to rise to room temperature. In that case, the inner cavity still contained radiation associated with a 4K blackbody and the outer cavity contained radiation corresponding to room temperature.

Johnson argues that: *But by making the inner cavity walls perfectly reflecting and closing the last side, Robitaille has created two entirely separate cavities; by definition, the inner cavity walls cannot emit radiation in either direction, whatever the temperature. They therefore act as boundary walls to what has become a "hollow" outer cavity. The outer cavity no longer contains the inner cavity within itself in a thermal sense; Kirchhoff's Law therefore survives this thought experiment"* [36].

There is no validity in this argument. Simply examine the quotation by Kirchhoff which opens this reply: *"... In the interior therefore of an opake red-hot body of any temperature, the illumination is always the same, whatever be the constitution of the body in other respects"*. Obviously, Kirchhoff's statement has been violated and Kirchhoff's law permits the placement of any object within the cavity interior, provided that it does not have the ability to emit photons by non-thermal means. I have not sidestepped the conditions set forth by Kirchhoff. The inner cavity, having perfectly reflecting walls, is linked to the floor of the outer cavity through thermal conduction [20]. The inner cavity is not composed of an adiabatic wall which is unable to contain or transmit heat, as Planck used. Rather, it is made of a perfect reflector, best approached by a material such as silver: *"Since the inner cavity is perfectly reflecting, it will also be highly conducting, as good reflectors tend to be good conductors"* [20, p. 38]. Therefore, conductive heat transfer was allowed [23, 26]. Silver is known to be essentially a perfect reflector in the infrared ($\rho > 0.994$ [54]), as I previously mentioned in the work under question [20]. It also possesses one of the highest electrical conductivities and has a very reasonable thermal conductivity, on the order of 400 W m^{-1} K^{-1} [55]. Johnson cannot argue that: *"The outer cavity no longer contains the inner cavity within itself in a thermal sense."* [36].

Mathematically, adiabatic walls can act as perfect reflectors, but reflectors themselves are not mathematical walls. Silver reflectors can be characterized by temperature, precisely because, though they are ideally immune to capturing radiative energy, they are able to allow energy to enter or leave either through conduction or, when applicable, convection. Johnson will not deny that thermal conduction exists. Conversely, adiabatic walls cannot be characterized by any temperature, as they are fully immune to energy transfer by radiation, conduction, and convection.

In the case of a perfect reflector, all of the energy of the system can be trapped in its walls. In the case of the perfect absorber, Planck considered that all of the energy was contained in the radiation field. Yet, Planck still needed to allow his oscillators the opportunity to have some momentary interaction with radiative energy. Otherwise, no photons could be produced or absorbed. Similarly, the perfect reflector must be allowed to have some momentary interaction with conductive energy. Johnson can no more deny the presence of thermal conduction than he can deny the presence of thermal emission and absorption. Silver, an near perfect reflector in the infrared, still has access to conductive paths of heat transfer.

Johnson tries to dismiss this thought experiment [20] and with good reason. It constitutes strong evidence that Kirchhoff's Law could never have been correct. In fact, let us revisit this setting, as it also helps to dispel Planck's ill-conceived claims relative to the carbon particle acting as a catalyst.

First, note that the radiation contained within the inner cavity depends on its history prior to the cavity being closed [20]. It will contain whatever radiation was present within the outer perfectly absorbing cavity at that time. That is, it will be defined by the temperature of the outer cavity at closure (i.e. 4 K). The radiation within the inner cavity persists as Planck claims [4, § 51], but in a state which was well-defined by history, not just any arbitrary state.

If we place a carbon particle in the perfectly reflecting cavity and if this particle does not act to transform heat from the wall into the radiative field, but can only act as a catalyst, as Planck claimed [4, § 51] (relative to the existing radiation which initially corresponded to 4 K radiation [20]), the interior of the cavity could never become black. That is because the interior of the second cavity lacks sufficient energy in its 4 K photons to adopt the proper blackbody intensity for the new higher temperature of its walls, when the both cavities have been brought to room temperature [20]. The carbon particle, can never act to shift the Wien's peak to higher frequencies because Planck denies that it can contain any significant heat on its own [4, § 51]. The cavity, in this instance, could not contain black radiation at room temperature, without violation of the First Law of Thermodynamics. That is the central problem in Planck's notion that the carbon particle was merely a catalyst. In the example provided, Planck would stand in violation of the First Law, if he persisted in insisting that the carbon particle was not transforming the energy content of the walls and if he maintained his insistence that the

cavity became filled with blackbody radiation.

Should the carbon particle be characterized with a temperature, but interaction with the walls still prevented, then it could convert the radiation within the cavity to the proper Planckian distribution for the higher temperature. But this radiation will always remain gray, as the temperature of the carbon particle cannot be allowed to fall and since it has no access to other sources of heat. Once again, Planck is restricted by the First Law. The relative distributions of frequencies might become correct, but their intensity will always be too low.

It is only when the carbon particle is allowed to transform the thermal energy contained within the wall of the perfectly reflecting cavity that we obtain the correct answer and that the interior of the second cavity can become black [15]. That is why the carbon particle was never a catalyst. Planck ignores its ability to transform thermal energy contained within the walls. He was only concerned with the radiation field and this was a crucial error.

3 Robitaille and Crothers' 2015 paper

This section begins, once again, by claiming that there was a *volte-face* relative to my position on Stewart's mechanism. As noted previously, such claims are unwarranted. I have never supported *"Stewart's mechanism"* as providing a valid means of extending Kirchhoff's claims to all cavities made from arbitrary materials. Neither has Steve Crothers.

While defending Max Planck, Johnson has failed to recognize that there can be a substantial difference between 1) what Planck claims to have done, 2) what he actually did, and 3) what nature permits. For instance, when Planck denied the absorptivity of the surface layer and inserted only reflectivity, he made claims which were *demonstrably false* in the laboratory, relative to the nature of a blackbody surface. He inappropriately applied polarized light and Brewster's Law to secure his proof, when such an approach was disallowed based on the very definition of heat radiation. Finally, he concluded that his unnumbered equation at the end of section § 36 [4],

$$\frac{\mathbf{K}_\nu}{\mathbf{K}'_\nu} \cdot \frac{q^2}{q'^2} = \frac{1 - \rho'}{1 - \rho},$$

could be satisfied by all values of ρ and ρ'. Yet, when $\rho = 1$, this expression became undefined. As such, both Crothers and I maintain that Planck's "proof" [4] of Kirchhoff's Law remains fundamentally flawed and invalid. Planck has, therefore, been deprived of any justification in claiming universality. In his initial paper [3] and in the latter portion of his text [4], Planck correctly derived an expression for the blackbody function. But Planck can never state, based on § 35-37, that interiors of all cavities contain black radiation. This remains a serious crack in the armor of modern physics and Johnson's letter has not helped to rectify the problem.

3.1 The meaning of Planck's term "surface"

Within his classic text, Planck described how he has deviated from Kirchhoff's definition of a blackbody. For Johnson, Planck's new definition was permitted, whereas, in truth, it constituted a rejection of nature itself. As we have highlighted [25, p. 124], Planck stated within a footnote *"In defining a blackbody Kirchhoff also assumes that the absorption of incident rays takes place in a layer "infinitely thin". We do not include this in our definition"* [4, § 10]. This was not footnote material, as it constitutes a critical redefinition of the blackbody. In opposition to Kirchhoff, Planck decided to write: *"The creation of a heat ray is generally denoted by the word emission. According to the principle of the conservation of energy, emission always takes place at the expense of other forms of energy (heat, chemical or electric energy, etc.) and hence it follows that only material particles, not geometrical volumes or surfaces, can emit heat rays. It is true that for the sake of brevity we frequently speak of the surface of a body as radiating heat to the surroundings, but this form of expression does not imply that the surface actually emits heat rays. Strictly speaking, the surface of a body never emits rays, but rather it allows part of the rays coming from the interior to pass through. The other part is reflected inward and according as the fraction transmitted is larger or smaller the surface seems to emit more or less intense radiations"* [4, § 2].

Was Kirchhoff actually correct? Does the absorption of incident rays take place in a layer *"infinitely thin"* [4, § 10]? Or, did Planck more closely approximate nature: *"Strictly speaking, the surface of a body never emits rays, but rather it allows part of the rays coming from the interior to pass through"* [4, § 2]. Of course, if a surface, strictly speaking, cannot emit rays, it also cannot absorb rays.

The answer to this problem has been provided in the laboratory. If one considers the hexagonal planar structure of graphite and the reality that soot (or lampblack) has always played an important role relative to the creation of blackbodies (see references within [16, 17]), then the answer is readily apparent. For soot shares, in large measure, the hexagonal planar structure of graphite, although more breaks exist in the lattice. The surface of graphite or soot, is well represented by graphene [56, 57], as this alone constitutes the outer layer of a sheet of graphite.

Mak et al [58] speak of the absorption of graphene, *"Indeed, it was the strong absorption of single-layer graphene (with its absorbance of ~2.3%, ... that permitted the initial discovery of exfoliated monolayers by visual inspection under an optical microscope"*. The authors are referring to the work of Novoselov and Geim [56] (Nobel Prize, Physics, 2010). Moreover, even a single layer of graphene has been shown to be an absolutely phenomenal emitter, when driven by current [59].

Consequently, Planck's position that, *"Strictly speaking, the surface of a body never emits rays, but rather it allows*

part of the rays coming from the interior to pass through" [4, § 2], simply cannot be upheld. Laboratory evidence is firm on this point: some of the rays will begin to be absorbed even by the first mono-layer of atoms and less than 50 hexagonal planes of atoms should result in near complete absorption.

It is a fact that, even a single layer of graphene, the only structure which can be associated with the surface of a graphite blackbody, has powerful absorbance. Max Planck cannot be permitted to neglect this layer, when discussing blackbodies. Planck knew Kirchhoff's definition and chose to ignore it, even though he recognized that he could make the entire radiation within a cavity black, by introducing even the smallest of carbon particles [4, § 51]. Planck's error was in not allowing any absorption or emission at all, not in allowing that a single layer did not have 100% absorption. In this respect, Planck's statement was imprudent at the time and laboratory experiments have now demonstrated that, indeed, it was false. Johnson cannot correct this situation. Yet, as we shall see below, this was a critical step towards Planck's faulty derivation of Kirchhoff's Law. As for Kirchhoff, given his period in history, it is clear that his definition remains valid. For a single layer of atoms is as "infinitely thin" as nature can allow and 50 layers of atoms about as "infinitely thin" as a man could conceive in Kirchhoff's days. He could have no concept of the dimensions of atoms in 1860 [1, 2].

In continuing his letter, Johnson then attempts to justify Planck's insistence that the term *"bounding surface"* referred to a geometrical surface dividing two media, and that *"the material effects of emission and absorption take place within the adjoining media"* [36]. In this respect, we return to the question of what Planck has said and what he can be permitted to say.

In § 35 of his textbook [4], Planck outlined the notation relative to primed and unprimed superscripts: *"Let the specific intensity of radiation of frequency ν polarised in an arbitrary plane be $\boldsymbol{K}_\nu$ in the first substance ..., and $\boldsymbol{K}'_\nu$ in the second, and, in general let all quantities referring to the second substance be indicated by the addition of an accent"* [4]. Planck continued in § 43, *"The most adequate method of acquiring more detailed information as to the origin and the paths of the different rays of which the radiations I_1, I_2, I_3, ... I_n consist, is to pursue the opposite course and to inquire into the future fate of that pencil, which travels exactly in the opposite direction to the pencil I and which therefore comes from the first medium in the cone $d\Omega$ and falls on the surface element $d\sigma$ of the second medium"* [4, § 43]. Here, Planck clearly assigned to the surface element $d\sigma$, properties of the second medium.

Johnson argues that Planck's bounding surface did not have to absorb any light, citing Planck's claim, *"Thus only material particles can absorb heat rays, not elements of surfaces, although sometimes for the sake of brevity the expression absorbing surfaces is used"* [4, § 12]. But what Johnson fails to understand is that, should he argue along these lines, he would be brought to accept yet another truth from Robitaille and Crothers which I now state: Only material particles can reflect light! Thus, Planck cannot be allowed an imaginary surface which reflects light, while at the same time denying that this same surface can absorb or emit light.

The truth being that when Planck placed two materials together, the bounding element, $d\sigma$, must be characterized on one side by the reflectivity and aborptivity of the first material and on the other side, by the reflectivity and absorptivity of the second material. That is because, the elements in either of the materials are not properly characterized only by reflectivity. This is precisely why Crothers and I object to Planck's use of a bounding surface which does not *fully represent* the materials which it unites.

Planck is welcome to claim that he can place a hypothetical bounding surface between two materials which considers only transmission and reflection. As for Crothers and I, we continue to object. The bounding surface which Planck envisioned was completely detached from reality. The issue is not that Planck cannot place the geometric surface between two layers. That is self-evident. The issue is that Planck cannot detach this geometric bounding layer from the material properties of those substances which he claims it characterizes. It is impossible to extract only the reflectivity of a particle, assign it to a geometric bounding surface, and at the very same time, ignore the absorptivity of this same particle. Contrary to Johnson, our statement that *"Planck neglected the fact that real materials can possess finite and differing absorptivities"* [25, p. 127] is entirely appropriate and valid. Planck's own textbook provides additional insight: *"Whenever absorption takes place, the heat ray passing through the medium under consideration is weakened by a certain fraction of its intensity for every element of path traversed"* [4, § 12]. By necessity, the element contained within the bounding section is one of the elements in the path traversed. Planck cannot ignore its absorption, because its properties can only be related to the medium to which it is linked.

Johnson then attempts to counter our statement: *"Third, the simplest means of nullifying the proof leading to Planck's Eq. 42, is to use a perfect reflector as the second medium. In that case, a refractive wave could never enter the second medium and Planck's proof fails"* [25, p. 127]. In order to counter this argument, Johnson tries to make the bounding surface perfectly reflecting, but unfortunately, he is not allowed to adopt such an approach, as Planck's proof intrinsically depends on the transmissivity of this bounding surface. Johnson cannot make it a perfect reflector, as in doing so, he optically isolates the two media. Furthermore, Johnson has failed to notice what has been mentioned above; namely, if $\rho = 1$, then $(1 - \rho) = 0$ and Planck's equation, at the bottom of Planck's § 36 (see § 3 herein) becomes undefined. Planck needs this equation to be valid in order to obtain his Eq. 41, $q^2\mathbf{K}_\nu = q'^2\mathbf{K}'_\nu$. But after he obtains Eq. 40, $\rho=\rho'$, Planck must return to the equation he lists at the end of § 36 and this ex-

pression is not always true.

It also remains the case that Planck's entire proof of Kirchhoff's Law collapses, as Crothers and I correctly highlighted in our joint paper [25, p. 127], when we replaced the second medium with a perfect reflector.

At first, it was difficult for me to even understand why Johnson would have wanted to replace the geometric bounding surface with a perfectly reflective surface. The answer rests in his use of this quote from Max Planck: *"Since the equilibrium is nowise disturbed, if we think of the surface separating the two media as being replaced for an instant by an area entirely impermeable to heat radiation, the laws of the last paragraphs must hold for each of the two substances separately"* [4, § 35]. However, in that case, Planck was referring to the treatment he had just outlined when addressing a single medium. Note that Planck writes in § 32, *"that the total state of radiation of the medium is the same on the surface as in the interior.* Then in § 33, Planck writes, *"While the radiation that starts from a surface element and is directed towards the interior of the medium is in every respect equal to that emanating from an equally large parallel element of area in the interior, it nevertheless has a different history. That is to say, since the surface of the medium was assumed to be impermeable to heat, it is produced only by reflection at the surface of radiation coming from the interior"* [4, § 35]. Planck had assumed that the surface in this case was impermeable to heat because this was the only way he could treat the isolated medium near its surface.

However, when Planck moved to two media, he no longer used a boundary impermeable to heat, but assumed that the surface of each medium was *"smooth"* [4, § 36]. In § 9, Planck had defined a smooth surface as one which can partially reflect and transmit the incoming radiation [4, § 9]. Planck required transmission for his later proof of Kirchhoff's law in § 35 and § 36. This is an essential element, which Johnson failed to consider in stating that a perfectly reflecting boundary enabled $\rho = \rho'$. In that case, as mentioned above, the equation at the bottom of Planck's § 36 would become undefined. It is for this reason that Johnson cannot support Planck's position, by making the bounding surface a perfect reflector.

Robitaille and Crothers remain correct. Planck improperly treated absorption and reflection in his derivation. Furthermore, the use of a perfect reflector for the second medium was all that was needed to shatter Planck's proof of Kirchhoff's Law, as we have previously noted [25].

3.2 Absorption and transmission

This section of Johnson's letter begins by quoting from the paper by Robitaille and Crothers: *"With his words, Planck redefined the meaning of a blackbody. The step, once again, was vital to his derivation of Kirchhoff's Law, as he relied on transmissive arguments to arrive at its proof. Yet, blackbody radiation relates to opaque objects and this is the first indication that the proofs of Kirchhoff's Law must not be centered on arguments which rely upon transmission. Planck ignored that real surface elements must possess absorption, in apparent contrast with Kirchhoff and without any experimental justification"* [25, p. 124].

Strangely, Johnson then concludes from this quotation that *"the apparent problem arises from the fact that Planck's surface is a geometrical one, whilst Robitaille and Crothers are obviously referring to a surface layer in which, they maintain, all absorption must take place because transmission is not permitted through a black body"* [36]. But we never stated that *all* of the absorption must take place from the surface layer. We stated that *"real surface elements must possess absorption"* [25, p. 124]. The surface need not have 100% absorption, as only a slight absorption is sufficient to invalidate Planck's proof. It is obvious, from our treatment of the first section of Planck's proof, that we do in fact allow transmission to take place within the medium and for elements within the blackbody to absorb, exactly like Max Planck [25, § 4.2]. We caution, however, that blackbodies are opaque objects and that Planck's proof cannot rely exclusively on transmission and reflection. Our point remains valid, as well demonstrated by the experimental realities outlined relative to graphene in § 3.1 above.

Again quoting from our paper, Johnson then attempts to argue that Planck was correct in inferring that *"... while in the case of bodies with vanishingly small absorbing power only a layer of infinite thickness may be regarded as black"* [4, § 10]. Once again, it is difficult to understand how Johnson can come to Planck's defense in this case. An opaque object which has a low absorptivity, also has a high reflectivity *by definition*. If not, it would not be opaque. As such, most photons which approach an opaque surface with low absorptivity are reflected away from the body. For Planck's argument to work, one would have to discount the surface reflection from an opaque object with a low emissivity which is counter to all laboratory experience. This highlights that Planck's new definition of a blackbody is completely outside the laws of nature. Planck cannot argue that he can neglect surface reflection, simply to salvage his derivation of Kirchhoff's Law.

Our point remains valid *"Blackbodies are opaque objects without transmission, by definition"* [25, p. 125]. Still, we have, in fact, allowed Planck to have some mathematical latitude and some level of transmission within the object, as presented in our § 4.2 [25]. But we cannot allow Planck to completely negate the presence of the reflection which is known to occur at the surface of an opaque object of low emissivity. Johnson and Planck shall not redefine nature.

3.3 Reflection

Relative to neglecting the reflection which occurs within a medium, we never stated that such an approach was invalid,

merely that it was suboptimal. In fact, in § 4.2 of our paper, we specifically outline the effects of neglecting the reflection taking place within the medium [25].

Johnson, however, is under the impression that reflection is strictly a surface phenomenon and cannot take place within the medium. At the end of this section, Johnson emphasizes the point when he states *"Note that Planck is still talking about the interior of the medium where reflection is not applicable because there is no surface; therefore Robitaille and Crothers' objection cannot be maintained"* [36, § 3.3]. Johnson is confused on this point.

Planck himself explicitly commented on scattering within media: *"The propagation of the radiation in the medium assumed to be homogeneous, isotropic, and at rest takes place in straight lines and with the same velocity in all directions, diffraction phenomena being entirely excluded. Yet, in general, each ray suffers during its propagation a certain weakening, because a certain fraction of its energy is continuously deviated from its original direction and scattered in all directions. This phenomenon of "scattering", ... takes place, generally speaking, in all media differing from absolute vacuum ... "* [4, § 8]. Later in the same section, Planck noted that, beyond diffraction, scattering also depends on reflection [4, § 8]. Hence, contrary to Johnson's claims, Planck understood that reflection is not strictly a surface phenomenon.

Crothers and I have properly considered internal reflection [25, § 4.2]. We have demonstrated that, when internal reflection is considered, powerful new insight is gained. Rather than simply obtaining Kirchhoff's formulation, $\mathbf{K}_\nu = \boldsymbol{\epsilon}_\nu/\alpha_\nu$, which is potentially undefined, we can actually extract $\boldsymbol{\epsilon}_\nu = (1-\rho_\nu)\mathbf{K}_\nu$, which is never undefined [25, § 4.2]. The insight provided by this treatment is important, contrary to what Johnson implies when insisting, without justification and in opposition to Planck's own statements, that reflection is only a surface phenomenon.

3.4 Polarization and equality of reflection

In the final section of his letter, Johnson attempts to justify Planck's use of polarized light and his assertion that the reflectivities of a pair of media at the bounding surface must be equal. He begins by quoting from our paper: *"In § 5 Planck admitted that homogeneous isotropic media emit only natural or normal, i.e. unpolarized, radiation: "Since the medium was assumed to be isotropic the emitted rays are unpolarized". This statement alone, was sufficient to counter all of the arguments which Planck later utilized to arrive at Kirchhoff's Law [Eq. 42]. That is because the important sections of Planck's derivation, namely § 35–37 make use of plane-polarized light. These steps were detached from experimental reality, relative to heat radiation [Planck, § 35] ... "*.

At this point, Johnson recalls that we have allowed Planck to resolve heat radiation into two equal orthogonal components, each plane-polarized. He objects to our statement that *"such rays could never exist in the context of heat radiation"* [25, p. 129]. Apparently, Johnson has failed to grasp that even though Planck can resolve heat radiation into two components, he is not allowed to apply only one component in his derivation. He must always consider *both* components, even if he can resolve them into two orthogonal planes.

Johnson apparently does not understand why Planck wanted to treat only one component, in part, because he seems unaware of Brewster's Law. Planck, in his derivation of Kirchhoff's Law, invoked plane-polarized radiation, such that he could set $\rho = \rho' = 0$. He could only obtain this expression, when dealing with a single plane polarized beam of light. That is because, if he sent such a beam at the proper angle and with the proper polarization towards his bounding surface, there would be no reflection, according to Brewster's Law.

However, Planck was not right in stating that there could be no reflection in the context of heat radiation. He could not obtain the plane-polarized beam of light, which he required, because the other component of the radiation, which was inappropriately ignored in his derivation, was also present. Moreover, Planck did not even test reflectivity by his argument from Brewster's law, as the latter is dependent upon the presence of a reflected ray as well as a transmitted ray. Thus, Planck could not conclude that the reflectivities of both materials were 0. The absence of a reflected ray does not imply that reflectivity is zero, as the polariscope attests. Just because Planck can resolve light into two components does not mean that he can ignore one of these components. *This is one of the most significant flaws in Planck's derivation of Kirchhoff's Law*.

Johnson then tries to defend Planck's most dramatic claim. Planck states [4, § 37]: *"Now in the special case when the rays are polarized at right angles to the plane of incidence and strike the bounding surface at the angle of polarization,* $\rho = 0$, *and* $\rho' = 0$. *The expression on the right side of the last equation then becomes 1; hence it must always be 1 and we have the general relations:*

$$\rho = \rho' \qquad (40)$$

and

$$q^2\mathbf{K}_\nu = q'^2\mathbf{K}'_\nu \quad (41)".$$

As I have just outlined, Planck cannot refer to this special case, because he does not have access to light polarized in a single plane. He must *always simultaneously treat both components*. Secondly, Planck is incorrect in asserting that the right side of the expression at the end of his § 36 [4] (also shown in § 3 of this letter), *"must always be 1"*, because it becomes undefined when $\rho = 1$. Planck was making an elementary error in mathematics. We maintain that *"The result was stunning."* [25, p. 129]. We also maintain that *"Max Planck had determined that the reflectivities of all arbitrary media were equal"* [25, p. 129].

Johnson then tries to defend Planck one last time, by insisting that what the latter *"had in fact demonstrated is that the reflectivities on each side of a geometrical surface bounding two different media are equal. Clearly if a different pair of media are chosen, the value of the reflectivity of the bounding surface may be different as well"* [36, § 3.4]. He then quoted from *The Theory of Heat Radiation*, *"Since, in general, the properties of a surface depend on both of the bodies which are in contact, this condition shows that the property of blackness as applied to a body depends not only on the nature of the body but also on that of the contiguous medium. A body which is black relatively to air need not be so relatively to glass, and vice versa"* [4, § 10].

Both Crothers and I understand what Max Planck claimed. However, we are properly concerned with what he has actually done. Planck's statement that *"the properties of a surface depend on both of the bodies which are in contact"* [4, § 10] can never be verified in the context of opaque media, precisely because his bounding surface is an abstraction. Snell's law, for instance, also relies on the interface of two media, but a bounding surface, or the changes at the surfaces, need not be introduced to obtain the proper answer. The indices of refraction of the two media alone are sufficient to treat the problem.

Planck's statements relative to the bounding surface were subject to two fundamental objections. First, they are justified by nothing; second, they constitute *"une hypothèse gratuite"* (see table presenting arguments against 19th century proofs of Kirchhoff's Law in [47, p. 16]). Planck may wish to claim that *"A body which is black relatively to air need not be so relatively to glass, and vice versa"* [4, § 10], but he had absolutely no justification for such a statement.

Rather, what Planck did possess are two isotropic media. Each of these is characterized by the absorptivity and reflectivity for each of its constitutive elements. Within his bounding surface, Planck could only introduce the reflectivity of elements contained in the media in question. When he introduced this reflectivity into his bounding surface, he had to additionally introduce some absorptivity, since this also characterized the media. Planck was not free to ignore the absorptivity. But he did so, as absorptivity in the bounding surface would prevent him from making use of Brewster's Law.

In any case, Planck could not invent a new reflectivity, which now existed only when he places the two media in contact with one another. After all, the reflectivities of the bounding surface must somehow be related to the materials under study. Furthermore, all that Planck could ever know about these materials are the reflectivities which can be measured. Neither he, nor Johnson, are allowed to hypothesize on what can never be measured in opaque media.

Planck recognized that he could not state that reflectivities of all materials are identical. As such, he postulated, *without any experimental evidence*, that his proof actually refers to something else [4, § 10]. Crothers and I dispute such claims. Planck's derivation must be taken on what the setting and the mathematics demonstrate. If we ignored Planck's mathematical errors and experimental oversights, we could much more convincingly argue that he had demonstrated that the reflectivities of all arbitrary materials were equal, using the same proof. Planck could measure nothing more than the reflectivities of each medium. Thus, he remains in violation of known optics, despite his attempts to introduce a new meaning to the reflectivity of a surface. Furthermore, Planck is forbidden from writing Eq. 40, $\rho = \rho'$, precisely because he has violated nature's rule that heat radiation is never polarized. It also remains the case that the unnumbered equation, which Planck presents at the end of his § 36 [4] (see § 3 herein), is undefined when $\rho = 1$.

4 Johnson's summary and conclusions

In opening this section of his letter, Johnson claims that, *"Stewart [33] had shown that the radiation in a cavity made from perfectly absorbing material at thermal equilibrium must be black, of an intensity appropriate to the equilibrium temperature. According to Robitaille, Kirchhoff [1] extended this finding to cavities made of arbitrary materials"*. Once again, Johnson has missed the mark.

Stewart considered plates in his experiments and Johnson is distorting what Stewart has done. It was with plates that Stewart demonstrated the Law of Equivalence (in modern notation: $\epsilon = \alpha$, or $\epsilon + \rho = \alpha + \rho$). Kirchhoff's extension to all arbitrary cavities [1, 2], went well beyond Stewart's legitimate law and has never been demonstrated to be true in the laboratory.

Johnson's claim that I have now withdrawn my objections to *"Stewart's mechanism"*, in my paper with Crothers [25], is without basis. *"Stewart's mechanism"* has numerous problems, including potential violations of the First Law of Thermodynamics, depending on the circumstances considered. It suffers from the reality that cavities made of low emissivity materials can prefer to increase the temperature of the walls, rather than emit a photon. Johnson's letter does nothing to counter this argument and that is why *"Stewart's mechanism"* cannot be realized in practice, as recognized by Crothers and myself.

Finally, Johnson admits: *"Robitaille is obviously correct to point out that black body cavities are never made from reflective materials."*. However, he then attempts to excuse the observation, in noting that, *"... this fact appears to be more a question of practicality and the need to ensure that the walls are not perfectly reflective at any wavelength so that proper measurements of temperature can be made. It does not seem to amount to a demonstration that Kirchhoff's Law necessarily fails, as Robitaille claims."*. Again, the arguments are ill-conceived. The fact that an experiment, required to establish a law of physics, still remains impractical after 150 years, well indicates that the law was never valid.

Some have argued, for instance, that when cavities are constructed from low emissivity materials, their dimensions need to be increased. This helps to augment their absorbance when sampling return losses. But Kirchhoff's law is explicit. The dimensions of the walls are irrelevant and, at a given temperature, must remain unrelated to the emissivity of the material, provided that the diffraction limit is avoided. The diffraction limit is not set by the emissivity. Furthermore, Johnson's arguments, relative to the ability to properly measure a temperature remain unfounded. Temperature sensors in the walls of cavities can easily report such information.

5 Conclusions

Throughout his letter, Johnson demonstrates that he has not carefully considered what Stewart, Kirchhoff, and Planck have written. He attributes to them positions which they never adopted. Then, he misinterprets the positions they did take. He repeatedly makes elementary errors relative to the understanding of cavity radiation. His statements on properly measuring the temperature of a cavity are but one example. He argues for *"Stewart's mechanism"*, in building up the radiation within a cavity, while not recognizing that the introduction of specular reflection within such objects can easily lead to the formation of standing waves. He also fails to understand that a cavity can simply increase the temperature of its walls and not emit a single photon.

He rejects my experimental work on MRI cavities, as unrelated to the problem of thermal emission and notes that processes, like fluorescence, have been excluded by Max Planck. Yet, the sampling of a cavity with a network analyzer does not involve such processes. In this respect, he also fails to note that Kirchhoff had mistakenly included such processes, in his initial work [1]. This was the only work of Kirchhoff which Johnson cited.

Furthermore, he fails to recognize that microwave cavities are utilized aboard modern satellites, wherein such objects are claimed to be black. Johnson also improperly and unknowingly expresses Stewart's Law as $\epsilon_\nu = \alpha_\nu + \rho_\nu$ in a thought experiment, thereby reaching conclusions which were clearly false. Then, he ignores the very existence of thermal conduction, when he attempts to invalidate my thought experiment with two cavities. He misrepresents my statements and those of Stephen Crothers, when he tries to state that we denied that the interior of a medium can have absorbance. He failed to understand the difference between resolving a heat ray into its two plane-polarized components and making use of a single plane-polarized ray, in order to infer something about heat radiation, which is never polarized. He hypothesized that replacing Planck's geometrical bounding surface with a perfect reflector could be used to validate Planck's claims, when clearly, it leads to an undefined mathematical expression and an invalid setting.

For all these reasons, Johnson cannot state that he has, in any way, nullified my objections to Kirchhoff's Law. Still, he must not be faulted for trying to defend Kirchhoff and Planck. As I stated in the introduction, it is the first obligation of a scientist to defend established science. Moreover, the study of cavity radiation is not at all simple. In this regard, Johnson's efforts are noteworthy and he is to be given credit for the time he has invested in reviewing these many papers.

Through the exchange prompted by his letter, Johnson has been indirectly responsible for bringing to the forefront many aspects of cavity radiation. Progress is often achieved, only when old ideas are first rejected, even if the process of discovery is not smooth. The process of correction, in itself, leads to scientific advancement. Hence, through such an exchange, readers can better come to understand why Kirhchoff's Law of thermal emission was never valid. Consequently, Planck's claims for universality must be rejected.

Acknowledgments

The author would like to thank the editors of *Progress in Physics*, namely, Dmitri Rabounski, Florentin Smarandache, and Larissa Borrisova, for the courage they have shown in publishing my work on Kirchhoff's Law, the Sun, and the Microwave Background. I am well aware that this has not been always easy for the Journal. I would also like to thank my colleague, Stephen J. Crothers, for valuable discussions relative to this work.*

Dedication

This work is dedicated to my youngest sister, Mireille and her husband, John.

Submitted on: January 25, 2016 / Accepted on: January 27, 2016
First published online on: January 28, 2016

References

1. Kirchhoff G. Über das Verhältnis zwischen dem Emissionsvermögen und dem Absorptionsvermogen. der Körper fur Wärme und Licht. *Poggendorfs Annalen der Physik und Chemie*, 1860, v. 109, 275–301. (English translation by F. Guthrie: Kirchhoff G. On the relation between the radiating and the absorbing powers of different bodies for light and heat. *Phil. Mag.*, 1860, ser. 4, v. 20, 1–21).
2. Kirchhoff G. Über den Zusammenhang zwischen Emission und Absorption von Licht und. Wärme. *Monatsberichte der Akademie der Wissenschaften zu Berlin*, sessions of Dec. 1859, 1860, 783–787.
3. Planck M. Über das Gesetz der Energieverteilung im Normalspektrum. *Annalen der Physik*, 1901, v. 4, 553–563 (English translation by ter Haar D.: Planck M. On the theory of the energy distribution law in the normal spectrum. The old quantum theory. Pergamon Press, 1967, 82–90; also Planck's December 14, 1900 lecture *Zur Theorie des Gesetzes der Energieverteilung in Normalspectrum*, which stems from this paper, can be found in either German, or English, in: Kangro H. Classic papers in physics: Planck's original papers in quantum physics. Taylor & Francis, London, 1972, 6–14 or 38–45).
4. Planck M. The Theory of Heat Radiation. P. Blakiston' Son & Co., Philadelphia, PA, 1914.

*Johnson and I had indeed exchanged correspondence following our meeting at a conference in 2014 [60].

5. Eddington A.S. The Internal Constitution of the Stars, Cambridge University Press, Cambridge, 1926.
6. Chandrashekhar S. An Introduction to the Study of Stellar Structure, Chicago University Press, Chicago, 1939.
7. Kippenhahn R. and Weigert A. Stellar Structure and Evolution, Springer-Verlag, New York, 1990.
8. Srinivasan G. From White Dwarfs to Black Holes: The Legacy of S. Chandrasekhar, The University of Chicago Press, Chicago, 1996.
9. Penzias A.A. and Wilson R.W. A measurement of excess antenna temperature at 4080 Mc/s. *Astrophys. J.*, 1965, v. 1, 419–421.
10. Dicke R.H., Peebles P.J.E., Roll P.G., and Wilkinson D.T. Cosmic black-body radiation. *Astrophys. J.*, 1965, v. 1, 414–419.
11. Robitaille P.M. Liquid metallic hydrogen: A building block for the liquid Sun. *Prog. Phys.*, 2011, v. 3, 69–74.
12. Robitaille P.M. Forty lines of evidence for condensed matter — The Sun on trial: Liquid metallic hydrogen as a solar building block. *Prog. Phys.*, 2013, v. 4, 90–142.
13. Robitaille P.-M. On the validity of Kirchhoff's law of thermal emission. *IEEE Trans. Plasma Sci.*, 2003, v. 31, no. 6, 1263–1267.
14. Robitaille P.-M. An analysis of universality in blackbody radiation. *Prog. Phys.*, 2006, v. 2, 21-23.
15. Robitaille P.-M. A critical analysis of universality and Kirchhoff's law: A return to Stewart's law of thermal emission. *Prog. Phys.*, 2008, v. 3, 30–35.
16. Robitaille P.-M. Blackbody radiation and the carbon particle. *Prog. Phys.*, 2008, v. 3, 36–55.
17. Robitaille P.-M. Kirchhoff's law of thermal emission: 150 Years. *Prog. Phys.*, 2009, v. 4, 3–13.
18. Robitaille P.-M. Blackbody radiation and the loss of universality: Implications for Planck's formulation and Boltzmann's constant. *Prog. Phys.*, 2009, v. 4, 14–16.
19. Robitaille P.-M. A radically different point of view on the CMB in *"Questions of Modern Cosmology"*, Mauro D'Onofrio and Carlo Burigana, eds., Springer Scientific, New York, N.Y., 2009, 93–108.
20. Robitaille P.-M. Further Insight Relative to Cavity Radiation: A Though Experiment Refuting Kirchhoff's Law. *Prog. Phys.*, 2014, v. 10, no. 1, 38–40.
21. Robitaille P.-M. Further Insight Relative to Cavity Radiation II: Gedanken Experiments and Kirchhoff's Law. *Prog. Phys.*, 2014, v. 10, no. 2, 116–120.
22. Robitaille P.-M. On the Equation which Governs Cavity Radiation I. *Prog. Phys.*, 2014, v. 10, no. 2, 126–127 (see also, Robitaille P.M. Notice of Revision: "On the Equation which Governs Cavity Radiation I, II", by Pierre-Marie Robitaille. *Prog. Phys.*, 2015, v. 11, no. 1, 88).
23. Robitaille P.-M. On the Equation which Governs Cavity Radiation II. *Prog. Phys.*, 2014, v. 10, no. 3, 157–162 (see also, Robitaille P.M. Notice of Revision: "On the Equation which Governs Cavity Radiation I, II", by Pierre-Marie Robitaille. *Prog. Phys.*, 2015, v. 11, no. 1, 88.).
24. Robitaille P.M. Blackbody radiation in optically thick gases. *Prog. Phys.*, 2014, v. 10, no. 3, 166–168.
25. Robitaille P.-M. and Crothers S.J. "The Theory of Heat Radiation" Revisited: A Commentary on the Validity of Kirchhoff's Law of Thermal Emission and Max Planck's Claim of Universality. *Prog. Phys.*, 2015, v. 11, no. 2, 120–132.
26. Robitaille P.-M. Further Insight Relative to Cavity Radiation III: Gedanken Experiments, Irreversibility, and Kirchhoff's Law. *Prog. Phys.*, 2016, v. 12, no. 1, 85–88.
27. Robitaille P.-M.L., Abduljalil A.M., Kangarlu A., Zhang X., Yu Y., Burgess R., Bair S., Noa P., Yang L., Zhu H., Palmer B., Jiang Z., Chakeres D.M., and Spigos D. Human magnetic resonance imaging at eight Tesla. *NMR Biomed.*, 1998, v. 11, 263–265.
28. Robitaille P.-M.L. Magnetic resonance imaging and spectroscopy at very high fields: a step towards 8 Tesla, in "Proceedings of Physical Phenonema at High Magnetic Fields-III", (Z. Fisk, L. Gor'kov, D. Meltzer and R.Schrieffer), eds., World Scientific, London, 1999, 421–426.
29. Robitaille P.-M.L. and Berliner L.J. (Eds.) Biological Magnetic Resonance. Volume 26: Ultra High Field Magnetic Resonance Imaging. Springer, New York, 2006.
30. Bloch F. Nuclear Induction. *Phys. Rev.*, 1946, v. 70, no. 7–8, 460–474.
31. Wien W. Über die Energieverteilung in Emissionspektrum eines schwarzen Körpers. *Ann. Phys.*, 1896, v. 58, 662–669.
32. Stefan J. Über die Beziehung zwischen der Warmestrahlung und der Temperature. *Sitzungsberichte der mathematischnatur wissenschaftlichen Classe der kaiserlichen Akademie der Wissenschaften*, Wien 1879, v. 79, 391–428.
33. Stewart B. An account of some experiments on radiant heat, involving an extension of Prévost's theory of exchanges. *Trans. Royal Soc. Edinburgh*, 1858, v. 22, no. 1, 1–20 (also found in Harper's Scientific Memoirs, edited by J.S. Ames: The Laws of Radiation and Absorption: Memoirs of Prévost, Stewart, Kirchhoff, and Kirchhoff and Bunsen, translated and edited by D.B. Brace, American Book Company, New York, 1901, 21–50).
34. Robitaille P.M. Stellar opacity: The Achilles heel of the gaseous Sun. *Prog. Phys.*, 2011, v. 3, 93–99.
35. Robitaille P.M. Water, hydrogen bonding, and the microwave background. *Prog. Phys.*, 2009, v. 2, L5–L8.
36. Johnson R.J. A Re-examination of Kirchhoff's Law of Thermal Radiation in Relation to Recent Criticisms. *Prog. Phys.*, 2016, v. 12, no. 3, 175–183.
37. Siegel D.M. Balfour Stewart and Gustav Robert Kirchhoff: Two independent approaches to Kirchhoff's Law. *Isis*, 1976, v. 67, no. 4, 565–600.
38. Hoffmann D. On the experimental context of Planck's foundation of quantum theory. In: Revisiting the Quantum Discontinuity, Max Planck Institute for the History of Science, Preprint 150, 2000, 47–68.
39. Lummer O. and Pringsheim E. Kritisches zur schwarzen Strahlung. *Annalen der Physik*, 1901, v. 6, 192–210.
40. Rubens H. and Kurlbaum F. Anwendung der Methode der Reststrahlenzur Prüfung der Strahlungsgesetzes. *Annalen der Physik*, 1901, v. 2, 649–666; Rubens H. and Kurlbaum F. On the heat radiation of longwave-length emitted by black bodies at different temperatures. *Astrophys. J.*, 1901, v. 74, 335–348.
41. Touloukian Y.S., DeWitt D.P. Thermal radiative properties of nonmetallic solids. Vol. 8 in: Thermophysical Properties of Matter, IFI/Plenum, New York, 1972.
42. Harris L. The optical properties of metal blacks and carbon blacks. MIT and The Eppley Foundation for Research, Monograph Ser., 1, New Port, R.I., Dec. 1967.
43. Nelms N., Dowson J. Goldblack coating for thermal infrared detectors. *Sensors and Actuators*, 2005, v. A120, 403–407.
44. O'Neill P., Ignatiev A, Doland C. The dependence of optical properties on the structural composition of solar absorbers: Gold black. *Solar Energy*, 1978, v. 21, 465–478.
45. Harris L., McGuiness R.T., Siegel B.M. The preparation and optical properties of gold black. *J. Opt. Soc. Am.*, 1948, v. 38, 582.
46. Kirchhoff G.R. On the relation between the emissive and the absorptive power of bodies for heat and light, In "Investigations on the Solar Spectrum and the Spectra of the Chemical Elements", 2nd Edition, Ferd. Dummler's Publishing House, Berlin, 1866; Gesammelte Abhandlugen, pp. 571-598, Leipzig, 1882; also found in Harper's Scientific Memoirs, edited by J.S. Ames: The Laws of Radiation and Absorption: Memoirs of Prévost, Stewart, Kirchhoff, and Kirchhoff and

Bunsen, translated and edited by D.B. Brace, American Book Company, New York, 1901, 73–97).

47. Schirrmacher A. Experimenting theory: The proofs of Kirchhoff's radiation law before and after Planck. *Hist. Stud. Phys. Biol. Sci.*, 2003, v. 33, 299–335.
48. Robitaille P.-M. Calibration of microwave reference blackbodies and targets for use in satellite observations: An analysis of errors in theoretical outlooks and testing procedures. *Prog. Phys.*, 2010, v. 3, 3–10.
49. Robitaille P.-M. The Planck Satellite LFI and the Microwave Background: Importance of the 4 K Reference Targets. *Prog. Phys.*, 2010, v. 3, 11–18.
50. Robitaille P.M. COBE: A radiological analysis. *Prog. Phys.*, 2009, v. 9, no. 4, 17-42.
51. Guo Y., Cortez C.L., Molesky S., and Jacob Z. Broadband super-Planckian thermal emission from hyperbolic metamaterials. [**?**], 2012, v. 101, 131106.
52. Yu Z., Sergeant N.P., Skauli T., Zhang G., Wang H., and Fan S. Enhancing far-field thermal emission with thermal extraction. *Nature Comm.*, 2013, DOI: 10.1038/ncomms2765.
53. Yu Z., Sergeant N., Skauli T., Zhang G., Wang H. and Fan S. Thermal extraction: Enhancing thermal emission of finite size macroscopic blackbody to far-field vacuum. 4 Nov 2012, arXiv:1211.0653v1 [physics.optics].
54. Diggers R.G (Ed), Encyclopedia of Optical Engineering (vol. 3). Marcel Dekker, Inc., New York, 2003, p. 2354.
55. Electrical Resistivity of Pure Metals. In, CRC Handbook of Chemistry and Physics, 2013–2014, CRC Press, Baca Raton, FL, p. 12–42.
56. Novoselov K.S., Geim A.K., Morozov S.V., Jiang D., Zhang Y., Dubonos S.V., Grigorieva I.V., Firsov A.A., Electric field effect in atomically thin carbon films. *Science*, 2004, v. 306, 666–669.
57. Geim A.K. and Novoselov K.S. The rise of graphene. *Nature Materials*, 2007, v. 6, 183–191.
58. Mak K.F., Ju L., Wang F., and Heinz T.F. Optical spectroscopy of graphene: From the far infrared to the ultraviolet. *Solid State Comm.*, 2012, v. 152, 1341–1349.
59. Kim Y.D., Kim H., Cho Y., Ryoo J.H., C.-H. Park, Kim P., Kim Y.S., Lee S., Li Y., Park S.-N., Yoo Y.S., Yoon D., Dorgan V.E., Pop E., Heinz T.F., Hone J., Chun S.-H., Cheong H., Lee S.W., Bae M.-H., and Park Y.D. Bright visible light emission from graphene. *Nature Nano.*, 2015, v. 10, 676–682.
60. Johnson Robert. "Could Electricity Explain Some'Known Unknowns' in Geology?" and "How Do Tall Trees Survive? New Models for the Flow of Sap", presented at The Electric Universe Conference, Albuquerque, N.M., March 20–24, 2014.

LETTERS TO PROGRESS IN PHYSICS

Errata to "Mansouri-Sexl Test Theory: The Question of Equivalence between Special Relativity and Ether Theories"

Maciej Rybicki
E-mail: maciej.rybicki@icloud.com

The title paper [1] contains an essential mistake committed by the present author. Namely, the Mansouri and Sexl generalized transformation of time, as well as the relevant form of the Lorentz transformation of time have been erroneously read and typed. In consequence, a certain part of the paper (indicated in the table, below) requires replacing to conform to the correct equations. The rest of the paper, except for the minor corrections indicated in the Errata, still remains valid. Let me apologize to the Readers and Editors for the inconvenience.

Page	Written	Read
89 (Abstract)	"has an erroneous form."	"is incorrectly used."
89, Eq. (2)	$t = aT + \epsilon X$	$t = aT + \epsilon x$
90, Eq. (5)	$x = \gamma(X - vt)$	$x = \gamma(X - vT)$
91 (Conclusion)	"We have shown that an incorrect notation…"	"We have shown that an incorrect use…"

Page 90, left column

Written (part to be replaced, starting from):
"Mansouri and Sexl state that for $a = b = 1$, $\epsilon = 0$ the Galilean transformation is obtained, which is correct. …"
(ending with, 33 lines down):
"…Consequently, they concluded that only violation of the two-way isotropy resulting in deviations from the relativistic values of a and b constitutes a challenge to STR."

Read (part to be introduced):
"Thus, the difference in the one-way speed of light would be a sole matter of choice of the synchronization convention.

From M-S theory it follows that for $a = b = 1$, $\epsilon = 0$, the Galilean transformation is obtained. If, after employing the external synchronization, a and b equal to unity, it would mean that mechanical phenomena are ruled by Newtonian physics and subject to the Galilean principle of relativity, while the Maxwell equations (and the relevant constant speed of light) refer to the ether frame only.

Instead, for $1/a = \gamma$ and $\epsilon = -v/c^2$, the M-S transformation of time turns into the Lorentz transformation of time:

$$t' = \frac{t}{\gamma} - \frac{vx'}{c^2}\,. \tag{E1}$$

In this form, the "rest-to-observer" coordinates appear on both sides of equation. Written in the same manner, the inverse Lorentz transformation is therefore:

$$t = \frac{t'}{\gamma} + \frac{vx}{c^2}\,. \tag{E2}$$

Consequently, the M-S transformation of time, and the inverse transformation are:

$$\begin{aligned} t &= aT + \epsilon x\,,\\ T &= at - \epsilon X\,. \end{aligned} \tag{E3}$$

Now, assuming $1/a = \gamma$ and $\epsilon = 0$, we obtain:

$$t = \frac{T}{\gamma} \Longrightarrow T = t\gamma\,, \tag{E4}$$

in contradiction with

$$T = \frac{t}{\gamma}\,. \tag{E5}$$

Mansouri and Sexl intended to treat independently the questions of time dilation and simultaneity. This, however, is infeasible with respect to the Lorentz transformation in which relativity of simultaneity and relativistic effects are inseparably connected. In the Lorentz transformation, one cannot obtain time dilation without taking into account the relativity of simultaneity. Likewise, the self-consistence of reciprocal equations in the Lorentz transformation involves the mutual dependence between $\gamma = 1/\sqrt{1 - v^2/c^2}$ and v/c^2. The incorrect use of Lorentz transformation (in particular, not including the inverse transformation) led to a false conclusion as to the question of equivalence between STR and the postulated ether theory."

Submitted on January 30, 2016 / Accepted on February 10, 2016

References

1. Rybicki M. Mansouri-Sexl Test Theory: The Question of Equivalence between Special Relativity and Ether Theories. *Progress in Physics*, 2016, v. 12 (1), 89–92.

Application of the Differential Transform Method to the Advection-Diffusion Equation in Three-Dimensions

Samia H. Esmail and Mahmoud M. Taha

Mathematics and Theoretical Physics Department, Nuclear Research Center, Atomic Energy Authority, Cairo, P.No. 13759, Egypt.
E-mail: mahmoudmt@hotmail.com

Advection diffusion equation with constant and variable coefficient has a wide range of practical, industrial and environmental applications. Due to the importance of atmospheric dispersion equation, we present this study which deals analytically with the atmospheric dispersion equation. The present model is proposed to estimate the concentration of an air pollutant in an urban area. The model is based on using Differential Transform Method (DTM) to solve the atmospheric dispersion equation. The model assumes 1) the pollutant is released from an elevated continuous point source; 2) there exist an elevated inversion layer; 3) the dispersion coefficients are parameterized as a function of downwind distance in a power law dependence. To test the model accuracy, the model predictions have been applied and compared with the experimental data for the Inshas research reactor (Egypt). The model predictions are shown to be in good agreement with the measurement of field data.

1 Introduction

The advection-diffusion equation of air pollution in the atmosphere is essentially a statement of conservation of the suspended material. The concentration of turbulent fluxes are assumed to be proportional to the mean concentration gradient which is known as Fick-theory.

This assumption, combined with the continuity equation, leads to the steady-state advection-diffusion equation, Blackadar [1]

$$\frac{\partial C}{\partial t}+u\frac{\partial C}{\partial x}+v\frac{\partial C}{\partial y}+w\frac{\partial C}{\partial z}=\frac{\partial}{\partial x}\left(k_x\frac{\partial C}{\partial x}\right)+\\+\frac{\partial}{\partial y}\left(k_y\frac{\partial C}{\partial y}\right)+\frac{\partial}{\partial z}\left(k_z\frac{\partial C}{\partial z}\right) \tag{1}$$

where $C(x,y,z)$ denotes the concentration, k_x, k_y, k_z are the cartesian components of eddy diffusivity and u, v, w are the cartesian components of wind speed, where x, y are cartesian horizontal distance and z is the height above ground surface.

In order to solve (1) we included the following assumptions: the pollutants are inert and have no additional sinks or sources downwind from the point source, the vertical w and lateral v components of the mean flow are assumed to be zero, k_x is neglected, k_y and k_z are functions of downwind distance. The mean horizontal flow is incompressible and horizontally homogeneous (steady state). Then, (1) is simplified to be:

$$u\frac{\partial C}{\partial x}=k_y\left(\frac{\partial^2 C}{\partial y^2}\right)+k_z\left(\frac{\partial^2 C}{\partial z^2}\right). \tag{2}$$

Both z and y are confined in the range $0<z<h$ and $0<y<L_y$ where h is the height of the planetary boundary layer (PBL) and L_y is a cross-wind distance faraway from the source, while the downwind distance $x>0$. The mathematical description of the dispersion problem (2) is completed by the following boundary conditions:

$$u\,C(x,y,z)=Q\,\delta(z)\,\delta(y)\,,\quad \text{at } x=0 \tag{3}$$

$$C(x,y,z)=0\,,\quad \text{at } x,y,z\to\infty \tag{4}$$

$$\frac{\partial C}{\partial y}=0\,,\quad \text{at } y=0,L_y \tag{5}$$

$$C(x,y,z)=R\,,\quad \text{at } y=0 \tag{6}$$

$$\frac{\partial C}{\partial z}=0\,,\quad \text{at } z=h \tag{7}$$

$$k_z\frac{\partial C}{\partial z}=-v_d\,C\,,\quad \text{at } z=0 \tag{8}$$

where v_d is the deposition velocity, Q is the emission rate and $R(x,z)$ is a variable.

The modeling of air pollution dispersion, including dry deposition, was first attempted by modifying the Gaussian plume equation (Chamberlain [2] and Overcamp [3]) and including operative algorithm, as in the surface depletion models (Horst [4, 5]). Ermak [6] found also an analytical solution but with diffusivity and wind as functions of down distance only and Berkowicz and Prahm [7] gave a numerical solution for the dependent time two dimensional equation including dry deposition. The solutions proposed by Smith [8] and Rao [9] also retained the framework of invariant wind speed and eddies with height (as the Gaussian approach). Tsuang [10] proposed a Gaussian model where the dispersion coefficients (the so-said "sigma") are functions of time and height.

Recent analytical solutions of the advection diffusion equation with dry deposition at the ground have utilized height-dependent wind speed and eddy diffusivities (Horst and Slinn [4], Koch [11], Chrysikopoulos et al. [12] and Tirbassi [13]). However, these solutions are restricted to the specific case in which the source is located at the ground level and/or with restrictions to the wind speed and eddy diffusivity vertical pro-

files. It is to be noted that the previous works, Moreira et al. [14, 15] assumed boundary conditions only of the second type (zero flux to the ground) and also Tirabassi et al. [16], but Tirabassi et al. [17] assumed boundary conditions of the the third kind (with deposition to the ground), which encompass the contaminant deposition speed and eddy, where eddy diffusivity profiles are functions in the z direction only.

The differential transform method is used in many fields and many mathematical physical problems such as a system of differential equations [18], a class of time dependent partial differential equations (PDEs) [19], wave, Laplace and heat equations [20], the fractional diffusion equations [21], two-dimensional transient heat flow [22], nonlinear partial differential equations [23], diffusion-convection equation [24], convection-dispersion problem [25], linear transport equation [26], two-dimension transient atmospheric pollutant dispersion [27], Helmholtz equation [28].

The aim of this work is to find the analytical solution developed for concentration of the pollutant released from an elevated source in an inversion layer by using the differential transform method (DTM) [29, 30] with different formulas of dispersion parameters (σ).

The paper is organized as follows. In section 2, we introduce the analytical solution using the differential transform method. In section 3, we apply both the standard method, power law, Briggs formula and other sigma to specific problems in analytical solution.

The validity of the present model is examined by comparing its results with the data for Cs^{137} which were performed around the Atomic Energy Authority (AEA) First Research Reactor in Egypt. The results are tabulated with the observed data and clarified in the conclusion.

2 Analytical solution

Applying DTM for (2) with respect to x, we get:

$$\frac{\partial U_i(x,y)}{\partial x} = k_y \frac{\partial^2 U_i(x,y)}{\partial y^2} + (i+1)(i+2)\, k_z U_{i+2}(x,y) \quad (9)$$

where the inverse of the differential transform is defined as:

$$C(x,y,y,z) = \sum_{i=0} z^i U_i(x,y)\,; \quad (10)$$

from boundary condition (8), we obtain:

$$U_1 = \left(\frac{-v_g}{k_z}\right) U_0\,; \quad (11)$$

from equations (9) and (11), we find that:

$$U_2 = \frac{1}{2k_z}\left(u\,\frac{\partial U_0(x,y)}{\partial x} - k_y \frac{\partial^2 U_0(x,y)}{\partial y^2}\right), \quad (12)$$

$$U_3 = \frac{-v_g}{6k_z}\left[u\,\frac{\partial}{\partial x}\left(\frac{U_0(x,y)}{k_z}\right) - \left(\frac{k_y}{k_z}\right)\frac{\partial^2 U_0(x,y)}{\partial y^2}\right]; \quad (13)$$

from boundary condition (7), we obtain:

$$\left(\frac{-2k_z v_g}{hk_y(2k_z - hv_g)}\right) U_0(x,y) + \left(\frac{u}{k_y}\right)\frac{\partial U_0(x,y)}{\partial x} - \\ -\left(\frac{2hk_z u v_g}{k_y(2k_z - hv_g)}\right) U_0(x,y)\,\frac{\partial}{\partial x}\left(\frac{1}{k_z}\right) = \frac{\partial^2 U_0(x,y)}{\partial y^2}\,. \quad (14)$$

By using separation of variables method for (14), we get:

$$\frac{d^2Y}{dy^2} + \lambda^2 Y = 0 \quad (15)$$

and

$$\frac{dX}{dx} - (A - B)X = 0 \quad (16)$$

where

$$A = \left(\frac{hk_z v_g \frac{\partial}{\partial x}\left(\frac{1}{k_z}\right)}{2k_z - hv_g}\right)$$

and

$$B = \frac{hk_z\lambda^2(2k_z - hv_g) - 2v_g k_z}{hu(2k_z - hv_g)}\,.$$

The solution of (14) becomes:

$$U_0(x,y) = c_1\, e^{\int (A-B)dx}\cos\lambda y \quad (17)$$

where $\lambda = n\pi/l_y$.

For practical application of solutions, we need to find the dispersion parameters σ_y, σ_z and the wind speed u. The dispersion parameters are an important function of downwind distance and stability. The empirical σ_y, σ_z curves suggested by Pasquill [31], Gifford [32] and Turner [33] have often been used and are based on the stability. There are different methods to find these parameters.

The meteorological conditions defining Pasquill turbulence types are

A- Extremely unstable conditions
B- Moderately unstable conditions
C- Slightly unstable conditions
D- Neutral conditions
E- Slightly stable conditions
F- Moderately stable conditions.

Here, we used four methods for estimating dispersion parameters:

1. Standard method: This method is based on a single atmospheric stability. Analytical expressions based on Pasquill-Gifford (P-G) curves used for the dispersion estimates have the forms [34]: .

$$\sigma_y = \frac{rx}{(1 + x/a)^p}, \quad (18)$$

$$\sigma_z = \frac{sx}{(1 + x/a)^q}, \quad (19)$$

where r, s, a, p and q are constants depending on the atmospheric stability. Table 1 shows the values of these constants for different stability classes [35].

Table 1: Meteorological data of the eight convective test runs [35]

Pasquil classes	A	B	C	D	E	F
σ_θ	25°	20°	15°	10°	5°	2.5°
$a(km)$	0.927	0.370	0.283	0.707	1.07	1.17
$s(m/km)$	102.0	96.2	72.2	47.5	33.5	22.0
q	-1.918	-0.101	0.102	0.465	0.624	0.70
$r(m/km)$	250	202	134	78.7	56.6	37.0
p	0.189	0.162	0.134	0.135	0.137	0.134

2. Power law of sigma: In this method σ_z and σ_y can be calculated from:

$$\sigma_y = cx^m \tag{20}$$

$$\sigma_z = dx^n . \tag{21}$$

The parameters c, d, m, n in Smith's (1968) [8] are estimated in table 2.

Table 2: Meteorological data of the eight convective test runs [36]

Pasquil classes	c	m	d	n
A-B	1.46	0.71	0.01	1.54
C	1.52	0.69	0.04	1.17
D	1.36	0.67	0.09	0.95
E-F	0.79	0.70	0.40	0.67

3. Briggs formulas: Formulas had been recommended by Briggs 1973 [37]; they should be used in place of the formulas in Table 3 to estimate σ_z and σ_y.

Table 3: Meteorological data of the eight convective test runs [35,37]

stability class	σ_y	σ_z
A-B	$0.32x\,(1+0.0004x)^{-\frac{1}{2}}$	$0.24x\,(1+0.001x)^{\frac{1}{2}}$
C	$0.22x\,(1+0.0004x)^{-\frac{1}{2}}$	$20x$
D	$0.16x\,(1+0.0004x)^{-\frac{1}{2}}$	$0.14x\,(1+0.0003x)^{-\frac{1}{2}}$
E-F	$0.11x\,(1+0.0004x)^{-\frac{1}{2}}$	$0.08x\,(1+0.00015x)^{-\frac{1}{2}}$

4. Hosker expression: Hosker 1973 [38] well-known analytical "best-fit" expression as:

$$\sigma_z = \left(\frac{\alpha x^\beta}{1+\gamma x^\delta}\right) F(z_0, x) \tag{22}$$

where z_0 is the roughness length, α, β, γ and δ are constants depending on the stability classes in Table 4 and $F(z_0, x)$ is defined as:

$$F(z_0, x) = \ln\left(mx^g\left[1+\left(lx^j\right)^{-1}\right]\right), \quad z_0 \geqslant 0.1m \tag{23}$$

where m, g, l, j are constants depend on the value of the roughness length, where our application z_0 (roughness length) = 0.5, so $l = 18.6$, $m = 5.16$, $j = 0.225$ and $g = 0.098$.

Table 4: The constant values of the roughness length, α, β, γ and δ [38]

Pasquil classes	α	β	γ	δ
A	0.112	1.06	5.0×10^{-4}	0.815
B	0.130	0.950	6.52×10^{-4}	0.750
C	0.112	0.920	9.05×10^{-4}	0.718
D	0.098	0.889	1.35×10^{-3}	0.688
E	0.0609	0.895	1.96×10^{-3}	0.684
F	0.0638	0.783	1.36×10^{-3}	0.672

On the other hand, Briggs 1973 [37] proposed a series of algebraic interpolation formulae based on a wide variety of data sources containing surface and elevated sources with a range of initial buoyancies:

$$\sigma_y = b_1(1+b_2x)^{b_3} . \tag{24}$$

The coefficient values b_1, b_2 and b_3 were derived for both rural and urban terrain and are given in Table 5 [37].

Table 5: The coefficient values b_1, b_2 and b_3 for equation (24) [37]

PG stability	b_1	b_2	b_3
A	0.20	0	–
B	0.12	0	–
C	0.08	0.0002	-0.5
D	0.06	0.0015	-0.5
E	0.03	0.0003	-1
F	0.016	0.0003	-1

3 Results and discussion

Meteorological data provided by Inshas meteorological tower for four months at a smooth flat site (Inshas area, Egypt) for the year (2006) are given in Table 6, [39]. Air samples were collected from 98 m to 186 m around the first and second research reactor in AEA, Egypt. The study area is flat, dominated by sand soil with poor vegetation cover. The study area was divided into 16 sectors (with 22.5° width for each sector), beginning from the north direction. Aerosols were collected at a height of 0.7 m above the ground of 10.3 cm diameter filter paper with a desired collection efficiency (3.4%)

Table 6: Meteorological Data of the nine Convective test runs at Inshas Site

No.	Stability	Down distance x (m)	Mixing height (m)	Emission rate Q (Bq)	Wind speed (m/s)	Initial wind velocity u_0(m/s)
1	A	98	600.85	0.555429	4	3.95
2	A	100	801.13	0.567	4	3.7
3	B	106	973.0	0.023143	6	5.1
4	C	106	888.0	0.254577	4	3.95
5	A	135	921.0	0.266143	4	3.1
6	D	136	443.0	0.277714	4	3.95
7	E	154	1271.0	0.543857	4	3.95
8	C	165	1842.0	0.563529	4	3.1
9	A	186	1642	0.558321	4	3.95

Table 7: Observed and calculated concentrations (Bq/m^3) for nine experiments

Run no.	Observed Con. [39]	Calculated concentrations			
		Standard Model	Power law of sigma	Briggs formulas	Hosker expression
1	0.002	0.0140799	0.0189143	0.013563	0.01379
2	0.004	0.0153392	0.011873	0.01475	0.014014
3	0.005	0.00448	0.00507518	0.004391	0.04422
4	0.007	0.0062904	0.013799	0.00624019	0.00625
5	0.009	0.00859466	0.00870	0.0081565	0.0081117
6	0.007	0.0070497	0.01596	0.0068969	0.00674
7	0.007	0.0137824	0.015015	0.019399	0.013155
8	0.019	0.0177893	0.019194	0.0171672	0.017135
9	0.006	0.0141444	0.01312	0.0132115	0.01311

using a high volume air sampler with 220 V / 50 Hz bias. The air sampler had an air flow rate of approximately 0.7 m^3/min (25 ft^3/min). Sample collective time was 30 min with an air volume of 21.2 m^3 (750 ft^3). This air volume was corrected to standard conditions (25 C^o and 1013 mb) [39].

Table 7 indicates comparison between experimental data of the nine convective test runs at Inshas site and our calculation of concentration by Briggs formula, power law variation, standard method and Hosker's expression, which shows that the power law formula for the dispersion coefficients achieves the best agreement with the experimental results.

3.1 Statistical evaluation

Statistical analysis of the predictions and observations is central to the model performance evaluation. The predicted and the corresponding observed concentrations are treated as pairs in this evaluation.

The statistical index FB indicates weather the predicted quantities underestimate or overestimate the observed ones. The statistical index NMSE represents the quadratic error of the predicted quantities in relation to the observed ones. Best results are indicated by values nearest zero in NMSE, FB, nearest 1 in MG, VG and FAC2 and are factor of two if are greater than 1 and less than 2. The statistical measures chosen to compare performances of the models described here [40]:
(i) Fractional bias FB is defined as:

$$\mathrm{FB} = \frac{\bar{C}_o - \bar{C}_p}{0.5(\bar{C}_o + \bar{C}_p)}$$

where the subscripts o and p refer to the observed and predicted values, respectively, and the overbars indicate mean values. A good model should have FB value close to zero.
(ii) Normalized mean square error (NMSE) is defined as:

$$\mathrm{NMSE} = \frac{\overline{(C_o - C_p)^2}}{\bar{C}_o \bar{C}_p} .$$

This provides information on the overall deviations between predicted and observed concentrations. It is a dimensionless statistic and its value should be as small as possible for a good model.
(iii) The geometric mean bias is defined as:

$$MG = \exp\left(\overline{\ln C_o} - \overline{\ln C_p}\right) .$$

(iv) The geometric variance is defined as:

$$VG = \exp\left(\overline{(\ln C_o - \ln C_p)^2}\right) .$$

Table 8: Comparison between the Standard method, Power law of sigma, Briggs formulas and Hosker expression in terms of FB, FAC2, NMSE, MG and VG

	Standard method	Power law of sigma	Briggs formulas	Hosker expression
FB	-0.42543	-0.65368	-0.445	-0.69646
FAC2	1.540371	1.971068	1.572352	2.068563
NMSE	0.189565	0.478407	0.208342	0.551991
MG	0.605381	0.46362	0.601944	0.490099
VG	1.286468	1.805587	1.293883	1.662927

(v) Fraction within a factor of two (FAC2) is given by:

$$0.5 \leqslant (C_p/C_o) \leqslant 2\,.$$

Statistical evaluation of the models results are given in Table 8, which compares the Standard method, Power law of sigma, Briggs formulas and Hosker expression in terms of FB, FAC2, NMSE, MG and VG.

4 Conclusion

In the present study, an analytical treatment for the dispersion of air pollutant released from point source is formulated. A mathematical solution has been obtained for the steady-state form of the three-dimensional advection-diffusion equation using the Differential Transform Method. Different realistic formulae for the dispersion coefficients as a function of downwind distance have been adopted (namely: Briggs formula, power law variation, standard method and Hosker's expression). In order to validate and verify our model, and for the sake of comparison, we apply our obtained mathematical formulae on the experimental data performed for the release from the first Research Reactor in Egypt. The comparison shows that the power law formula for the dispersion coefficients achieves the best agreement with the experimental results. Finally, the good agreement between the power law variation of the dispersion parameter and the experiential data gives us confidence to extend this work for the case of different sources types, namely, line, area and volume sources. In addition, it is also our intention to perform the mathematical analysis of this method for the case of high penetrated inversion layer (i.e. different stability conditions that permits the pollutant penetration and diffusion through the mixing height).

Submitted on January 25, 2016 / Accepted on February 11, 2016

References

1. Blackadar A. Turbulence and Diffusion in the Atmosphere: Lectures in Environmental Sciences. Springer-Verlag, 1997.
2. Chamberlain A. Aspects of Travel and Deposition of Aerosol and Vapour Clouds. UKAEA Rep. AERE-HP/R-1261, Harwell, Berkshire, United Kingdom, 1953.
3. Overcamp T. A general Gaussian diffusion deposition model for elevated point sources. *J. Appl. Meteor.*, 1976, v. 15, 1167–1171.
4. Horst T., Slinn W. G. A surface depletion model for deposition from a Gaussian plume. *Atmos. Environ.*, 1984, v. 18, 1339–1346.
5. Horst T. The modification of plume models to account for dry deposition. *Bound. Layer Meteor.*, 1984, v. 30, 413–430.
6. Ermak D. An analytical model for air polutant transport and deposition from a point source. *Atmos. Environ.*, 1976, v. 11, 231–237.
7. Berkowicz R. and Prahm L. Pseudospectral simulation of dry deposition from a point source. *Atmos. Environ.*, 1977, v. 12, 379–387.
8. Smith F. The problem of deposition in atmospheric diffusion of particulate matter. *J. Atmos. Sci.*, 1962, v. 19, 429–434.
9. Rao K. Analytical Solutions of a Gradient-transfer Model for Plume Deposition and Sedimentation. NOAA Tech. Memo. ERL ARL-109, Air Resources Laboratories, 1981.
10. Tsuang B. A Gaussian plume trajectory model to quantify the source/receptor relationship of primary pollutants and secondary aerosols: Part I. Theory. *Atmos. Environ.*, 2003, v. 37, 3981–3991.
11. Koch W. A solution of two-dimensional atmospheric diffusion equation with height-dependent diffusion coefficient including ground level absorption. *Atmos. Environ.*, 1989, v. 23, 1729–1732.
12. Chrysikopoulos C., Hildemann L. and Roberts P. A three-dimensional steady-state atmospheric dispersion deposition model for emissions from a ground-level area source. *Atmos. Environ.*, 1992, v. 26A, 747–757.
13. Tirabassi T. Operational advanced air pollution modeling. *PAGEOPH*, 2003, v. 160, 5–16.
14. Moreira D., Rizza U., Vilhena M. and Goulart A. Semi-analytical model for pollution dispersion in the planetary boundary layer. *Atmos. Environ.*, 2005, v. 39, 2689–2697.
15. Moreira D., Vilhena M., Tirabassi T., Costa C. and Bodmann B. Simulation of pollutant dispersion in atmosphere by the Laplace transform: the ADMM approach. *Water Air Soil Pollut.*, 2006, v. 177, 411–439.
16. Tirabassi T., Buske D., Moreira D. and Vilhena M. A two-dimensional solution of the diffusion equation with dry deposition to the ground. *J. Appl. Meteorol. Climatol.*, 2008, v. 47, 2096–2104.
17. Tirabassi T., Moreira D., Vilhena M. and Goulart A. A multi-layer model for pollutants dispersion with dry deposition to the ground. *Atmos. Environ.*, 2010, v. 44, 1859–1865.
18. Thongmoon M. and Pusjuso S. The numerical solutions of differential transform method and the Laplace transform method for a system of differential equations. *Nonlinear Anal. Hybrid Sys.*, 2010, v. 4, 425–431.
19. Kong W. and Xionghua Wu. The Laplace transform and polynomial Trefftz method for a class of time dependent PDEs. *Appl. Math. Model.*, 2009, v. 33, 2226–2233.
20. Eltayeb H. and Kiliçman A. A note on solutions of wave, Laplace's and heat equations with convolution terms by using a double Laplace transform. *Appl. Math. Let.*, 2008, v. 21, 1324–1329.
21. Zamani K. One-dimensional, mass concervative, spatially dependent transport equation: new analytical solution. 12th Pan-American Congress of Applied Mechanics, 2012, 1–6.
22. Mahajerin E. and Burgess G. A Laplace transform-based fundamental collocation method for two-dimensional transient heat flow. *Appl. Therm. Eng.*, 2003, v. 23, 101–111.

23. Abassy T., El-Tawil M. and El-Zoheiry H. Exact solutions of some nonlinear partial differential equations using the variational iteration method linked with Laplace transforms and the Padé technique. *Comput. Math. Appl.*, 2007, v. 54, 940–954.
24. Davis G. A Laplace transform technique for the analytical solution of a diffusion-convection equation over a finite domain. *Appl. Math. Model.*, 1985, v. 9, 69–71.
25. Ren L. and Zhang R. Hybrid Laplace transform finite element method for solving the convection-dispersion problem. *Advan. Wat. Res.*, 1999, v. 23, 229–237.
26. Cardona A. and Vilhena M. A solution of the linear transport equation using Walsh function and Laplace transform. *Annal. Nucl. Energ.*, 1994, v. 21, 495–505.
27. Cassol M., Wortmann S. and Rizza U. Analytic modeling of two-dimensional transient atmospheric pollutant dispersion by double GITT and Laplace Transform techniques. *Environ. Model. Software*, 2009, v. 24, 144–151.
28. Souchet R. Laplace transform method in boundary value problems for the Helmholtz equation. *Mech. Res. Commun.*, 1980, v. 7, 159–164.
29. Raslan K. R., Biswas A. and Abou Sheer Z. F. *Int. J. Phys. Sci.*, 2012, v. 7 (9), 1412–1419.
30. Sohail M. and Mohyud-Din S. T. *Int. J. Modern Eng. Sci.*, 2012, v. 32 (1), 14–22.
31. Pasquill F. *Meteorol. Mag.*, 1961, v. 90, 33–49.
32. Gifford F. A. *Nucl. Safety*, 1961, v. 2 (4), 47–57.
33. Turner D. B. *J. Air Poll. Control Assoc.*, 1979, v. 29, 502–519.
34. Green A. E. S., Singhal R. P. and Venkateswar R. Analytic extensions of the Gaussian plume model. *J. Air Poll. Control Assoc.*, 1980, v. 30 (7), 773–776.
35. Essa K., Mubarak F. and Abu Khadra S. Comparison of some sigma schemes for estimation of air pollutant dispersion in moderate and low winds. *Atmos. Sci. Let.*, 2005, v. 6, 90–96.
36. Panofsky H. A., Dutton J. A. Atmospherical Turbulence Models and Methods for Engineering Applications. John Wiley & Sons, New York, 1984.
37. Briggs G. A. Diffusion Estimation for Small Emissions. ATDL Contribution File No. 79, 1973.
38. Hosker R. Estimates of Dry Deposition and Plume Depletion over Forests and Grasslands in Physical Behaviour or Radioactive Contaminats in the Atmosphere. Symposium Proceedings, Vienna International Atomic Energy Agency, Vienna, 1973, 291–308.
39. Halfa I. K. I. Analysis of the Experimental Data of Air Pollution Using Atmospheric Dispersion Modeling and Rough Set, Doctoral Thesis, Faculty of Science, Tanta University, 2008.
40. Hanna S., Hansen O., and Dharmavaram S. FLACS CFD Air Quality Model Performance Evaluation with Kit Fox, MUST, Prairie Grass, and EMU Observations. *Atmos. Environ.*, 2004, v. 38, 4675–4687.

On the Applicability of Bell's Inequality

Pierre A. Millette
PierreAMillette@alumni.uottawa.ca, Ottawa, Canada

We investigate the applicability of Bell's inequality based on the assumptions used in its derivation. We find that it applies to a specific class of hidden variable theories referred to as Bell theories, but not necessarily to other hidden variable dynamic theories. We consider examples of quantum dynamical processes that cannot be represented by the initial representation defined in Bell's derivation. We highlight two hidden assumptions identified by Jaynes [11] that limit the applicability of Bell's inequality, as derived, to Bell hidden variable theories and that show that there are no superluminal physical influences, only logical inferences.

1 Introduction

Bell's inequality [1–3] sets constraints for the existence of local hidden variable theories in quantum mechanics. Bohr, of the Copenhagen probabilistic school, and Einstein, of the objective reality school, who both contributed to the foundation of quantum mechanics, did not agree on its interpretation – their views and correspondence on the topic are well documented in many books [4–7].

In 1935, Einstein, Podolsky and Rosen published a paper [8] that aimed to show that quantum mechanics was not a complete description of physical reality. Bohr provided a response to the challenge [9], but the EPR paper remained an argument for hidden variables in quantum mechanics. In 1964, Bell [1] published an inequality that imposed constraints for local hidden variable theories to be valid in quantum mechanics. The experiments performed by Aspect *et al* [10] with entangled photons confirmed that Bell's inequality was violated within experimental errors, taken to mean that local hidden variable theories are not valid in quantum mechanics. Only non-local hidden variable theories are possible, based on these results.

In this paper, we investigate the applicability of Bell's inequality, based on the assumptions used in its derivation.

2 Bell's inequality

Bell's derivation [1] considers a pair of spin one-half particles of spin σ_1 and σ_2 respectively, formed in the singlet state, and moving freely in opposite directions. Then $\sigma_1 \cdot \mathbf{a}$ is the measurement of the component of σ_1 along some vector $\mathbf{a}$, and similarly for $\sigma_2 \cdot \mathbf{b}$ along some vector $\mathbf{b}$. Bell then considers the possibility of a more complete description using hidden variable parameters λ.

He writes down the following equation for the expectation value of the product of the two components $\sigma_1 \cdot \mathbf{a}$ and $\sigma_2 \cdot \mathbf{b}$ with parameters λ:

$$P(\mathbf{a}, \mathbf{b}) = \int d\lambda \, \rho(\lambda) \, A(\mathbf{a}, \lambda) \, B(\mathbf{b}, \lambda) \tag{1}$$

where

$$A(\mathbf{a}, \lambda) = \pm 1 \text{ and } B(\mathbf{b}, \lambda) = \pm 1 \tag{2}$$

and $\rho(\lambda)$ is the probability distribution of parameter λ. This should equal the quantum mechanical expectation value

$$<\sigma_1 \cdot \mathbf{a} \; \sigma_2 \cdot \mathbf{b}> = -\mathbf{a} \cdot \mathbf{b} \, . \tag{3}$$

Bell says that it does not matter whether λ is "a single variable or a set, or even a set of functions, and whether the variables are discrete or continuous" [1]. He uses a single continuous parameter described by a probability distribution. In a later paragraph, he states that (1) represents all kinds of possibilities, such as any number of hidden variables, two sets of hidden variables dependent on A and B, or even as initial values of the variables λ at a given time if one wants to assign "dynamical significance and laws of motion" [1] to it. However, it is doubtful that the probability distribution $\rho(\lambda)$ can be used to represent all possible theories of hidden variables.

Indeed, the basic limitation of (1) with its use of a probability distribution $\rho(\lambda)$ is that it imposes a quantum mechanical calculation representation on the analysis. Other quantum level dynamic theories, which we will refer to as hidden variable dynamic theories, could obey totally different dynamic principles, in which case, (1) would not be applicable. Equation (1) is only applicable to a specific class of hidden variable theories that can be represented by that equation, which Jaynes [11] refers to as Bell theories. In the following sections, we consider examples of quantum dynamical processes that cannot be represented by (1) or by the probability distribution $\rho(\lambda)$ used in (1).

3 Measurement limitations and inherent limitations

It is important to note that Bohr's responses to Einstein's *gedanken* experiments were based on measurements arguments, which acted as a barrier to any further analysis beyond that consideration. As pointed out by Jaynes [12], Einstein and Bohr "were both right in the essentials, but just thinking on different levels. Einstein's thinking [was] always on the ontological level traditional in physics; trying to describe the realities of Nature. Bohr's thinking [was] always on the epistemological level, describing not reality but only our information about reality".

As discussed in [13], the Heisenberg Uncertainty Principle arises because x and p form a Fourier transform pair of variables at the quantum level due to the momentum p of a quantum particle being proportional to the de Broglie wave number k of the particle. It is a characteristic of quantum mechanics that conjugate variables are Fourier transform pairs of variables.

It is thus important to differentiate between the measurement limitations that arise from the properties of Fourier transform pairs, and any inherent limitations that may or may not exist at the quantum level for those same variables, independently of the measurement process. Conjugate variable measurement limitations affect how we perceive quantum level events as those can only be perceived by instrumented measurements at that level. However, as shown in [13], conjugate variable measurement limitations affect *only* our perception of the quantum environment, and are *not* inherent limitations of the quantum level.

The Nyquist-Shannon Sampling Theorem of Fourier transform theory allows access to the range of values of variables below the Heisenberg Uncertainty Principle limit under sampling measurement conditions, as demonstrated by the Brillouin zones formulation of solid state physics [13] [14, see p. 21] [15, see p. 100]. Physically this result can be understood from the sampling measurement operation building up the momentum information during the sampling process, up to the Nyquist limit. This shows that there are local hidden variables at the quantum level, independently of the measurement process. The dynamical process in this case is masked by the properties of the Fourier transform.

4 Wave-particle duality in *STCED*

The Elastodynamics of the Spacetime Continuum (*STCED*) [16] has similarities to Bohmian mechanics in that the solutions of the *STCED* wave equations are similar to Louis de Broglie's "double solution" [17, 18]. Bohmian mechanics also known as de Broglie-Bohm theory [19–21] is a theory of quantum physics developed by David Bohm in 1952 [22], based on Louis de Broglie's original work on the *pilot wave*, that provides a causal interpretation of quantum mechanics. It is empirically equivalent to orthodox quantum mechanics, but is free of the conceptual difficulties and the metaphysical aspects that plague the interpretation of quantum theory.

Interestingly, Bell was aware of and a proponent of Bohmian mechanics when he derived his inequality [23]:

> "Bohm showed explicitly how parameters could indeed be introduced, into nonrelativistic wave mechanics, with the help of which the indeterministic description could be transformed into a deterministic one.More importantly, in my opinion, the subjectivity of the orthodox version, the necessary reference to the 'observer,' could be eliminated... I will try to present the essential idea... so compactly, so lucidly, that even some of those who know they will dislike it may go on reading, rather than set the matter aside for another day."

In Bohmian mechanics, a system of particles is described by a combination of the wavefunction from Schrodinger's equation and a guiding equation that specifies the location of the particles. "Thus, in Bohmian mechanics the configuration of a system of particles evolves via a deterministic motion choreographed by the wave function" [21] such as in the two-slit experiment. We will see a similar behavior in the *STCED* wave equations below. Bohmian mechanics is equivalent to a non-local hidden variables theory.

In the Elastodynamics of the Spacetime Continuum, as discussed in [24], energy propagates in the spacetime continuum by longitudinal (*dilatation*) and transverse (*distortion*) wave displacements. This provides a natural explanation for wave-particle duality, with the transverse mode corresponding to the wave aspects of the deformations and the longitudinal mode corresponding to the particle aspects of the deformations.

The displacement u^ν of a deformation from its undeformed state can be decomposed into a longitudinal component $u^\nu_\parallel$ and a transverse component $u^\nu_\perp$. The volume dilatation ε is given by the relation $\varepsilon = u_\parallel{}^\mu{}_{;\mu}$ [16]. The wave equation for $u^\nu_\parallel$ describes the propagation of longitudinal displacements, while the wave equation for $u^\nu_\perp$ describes the propagation of transverse displacements in the spacetime continuum. The u^ν displacement wave equations can be expressed as a longitudinal wave equation for the dilatation ε and a transverse wave equation for the rotation tensor $\omega^{\mu\nu}$ [16].

Particles propagate in the spacetime continuum as longitudinal wave displacements. Mass is proportional to the volume dilatation ε of the longitudinal mode of the deformation [16, see (32)]. This longitudinal mode displacement satisfies a wave equation for ε, different from the transverse mode displacement wave equation for $\omega^{\mu\nu}$. This longitudinal dilatation wave equation for ε is given by [16, see (204)]

$$\nabla^2\varepsilon = -\frac{\bar{k}_0}{2\bar{\mu}_0 + \bar{\lambda}_0}\, u^\nu_\perp \varepsilon_{;\nu} \tag{4}$$

where $\bar{\mu}_0$ and $\bar{\lambda}_0$ are the Lamé constants and $\bar{k}_0$ the elastic volume force constant of the spacetime continuum. It is important to note that the inhomogeneous term on the R.H.S. includes a dot product coupling between the transverse displacement and the volume dilatation for the solution of the longitudinal dilatation wave equation for ε.

The transverse distortion wave equation for $\omega^{\mu\nu}$ [16, see (210)]

$$\nabla^2\omega^{\mu\nu} + \frac{\bar{k}_0}{\bar{\mu}_0}\,\varepsilon(x^\mu)\;\omega^{\mu\nu} = \frac{1}{2}\frac{\bar{k}_0}{\bar{\mu}_0}\,(\varepsilon^{;\mu}u^\nu_\perp - \varepsilon^{;\nu}u^\mu_\perp) \tag{5}$$

also includes a R.H.S. coupling, in this case a cross product, between the transverse displacement and the volume dilatation for the solution of the transverse distortion wave equa-

tion for $\omega^{\mu\nu}$. The transverse distortion wave $\omega^{\mu\nu}$ corresponds to a multi-component wavefunction Ψ.

A deformation propagating in the spacetime continuum consists of a combination of longitudinal and transverse waves. The coupling between $\varepsilon^{,\mu}$ and $u^{\nu}_{\perp}$ on the R.H.S. of both wave equations explains the behavior of electrons in the double slit interference experiment. It shows that even though the transverse wave is the source of the interference pattern in double slit experiments, the longitudinal dilatation wave, which behaves as a particle, follows the interference pattern dictated by the transverse distortion wave as observed experimentally. The longitudinal dilatation wave behaves as a particle and goes through one of the slits, even as it follows the interference pattern dictated by the transverse distortion wave, as observed experimentally [25, see in particular Figure 4] and as seen in the coupling between $\varepsilon^{,\mu}$ and $u^{\nu}_{\perp}$ in (4) and (5) above. This behavior is the same as that in Bohmian mechanics seen above. These results are in agreement with the results of the Jánossy-Naray, Clauser, and Dagenais and Mandel experiments on the self-interference of photons and the neutron interferometry experiments performed by Bonse and Rauch [26, see pp. 73-81].

As mentioned previously, the solutions of the *STCED* wave equations are similar to Louis de Broglie's "double solution". The longitudinal wave is similar to the de Broglie "singularity-wave function" [17]. In *STCED* however, the particle is not a singularity of the wave, but is instead characterized by its mass which arises from the volume dilatation ε propagating as part of the longitudinal wave. There is no need for the collapse of the wavefunction Ψ, as the particle resides in the longitudinal wave, not the transverse one. A measurement of a particle's position is a measurement of the longitudinal wave, not the transverse wave.

In addition, $|\Psi|^2$ represents the physical energy density of the transverse (*distortion*) wave. It corresponds to the transverse field energy of the deformation. It is not the same as the particle, which corresponds to the longitudinal (*dilatation*) wave displacement and is localized within the deformation via the massive volume dilatation. However, $|\Psi|^2$ can be normalized with the system energy and converted into a probability density, thus allowing the use of the existing probabilistic formulation of quantum theory.

The dynamical process, although it has some similarities to Bohmian mechanics, is also different from it as it is centered on longitudinal (particle) and transverse (wavefunction) wave equations derived from the properties of the spacetime continuum of general relativity. It is thus deterministic and causal as is general relativity.

5 Physical influence versus logical inference

We have considered two examples of quantum dynamical processes where the starting equation (1) and the probability distribution $\rho(\lambda)$ used in (1) do not apply to the situation. We now examine in greater details the probabilistic formulation of Bell's inequality derivation of section 2 to better understand its limitations.

Physicist E. T. Jaynes was one of the proponents of the usage of probability theory as an extension of deductive logic. His textbook "Probability Theory: The Logic of Science" [27] published posthumously is an invaluable resource for scientists looking to understand the scientific use of probability theory as opposed to the conventional mathematical measure theory. As he states in [11],

> "Many circumstances seem mysterious or paradoxical to one who thinks that probabilities are real physical properties existing in Nature. But when we adopt the "Bayesian Inference" viewpoint of Harold Jeffreys [28,29], paradoxes often become simple platitudes and we have a more powerful tool for useful calculations."

Jaynes clarifies this approach to probability theory and contrasts it to frequencies as follows [11]:

> "In our system, a probability is a theoretical construct, on the epistemological level, which we assign in order to represent a state of knowledge, or that we calculate from other probabilities according to the rules of probability theory. A frequency is a property of the real world, on the ontological level, that we measure or estimate."

The probability distributions used for inference do not describe a property of the world, only a certain state of information about the world, which provides us with the means to use prior information for analysis as powerfully demonstrated in numerous applications in [11, 12, 27].

The Einstein–Podolsky–Rosen (EPR) paradox and Bell inequality in quantum theory is one of the examples examined by Jaynes in [11]. In quantum mechanics, the belief that probabilities are real physical properties leads to quandaries such as the EPR paradox which lead some to conclude that there is no real world and that physical influences travel faster than the speed of light, or worse ("a spooky kind of action at a distance" as Einstein called it). As Jaynes points out, it is important to note that the EPR article did not question the existence of the correlations, which were expected, but rather the need for a physical causation instead of what he calls "instantaneous psychokinesis", based on experimenter decisions, to control distant events.

Jaynes' analysis of the derivation of Bell's inequality uses the following notation for conditional probabilities which corresponds to Bell's notation as follows:

$$P(AB\,|\,ab) = P(\mathbf{a},\mathbf{b}) \tag{6}$$

$$P(A\,|\,a\lambda) = A(\mathbf{a},\lambda)\,, \tag{7}$$

such that Bell's equation (1) above becomes

$$P(AB\,|\,ab) = \int d\lambda\,\rho(\lambda)\,P(A\,|\,a\lambda)\,P(B\,|\,b\lambda)\,. \tag{8}$$

However, as Jaynes notes, the fundamentally correct relation for $P(AB\,|\,ab)$ according to probability theory should be

$$P(AB\,|\,ab) = \int d\lambda\, P(AB\,|\,ab\lambda)\, P(\lambda\,|\,ab)\,. \qquad (9)$$

Assuming that knowledge of the experimenters' choices gives no information about λ, then one can write

$$P(\lambda\,|\,ab) = \rho(\lambda)\,. \qquad (10)$$

The fundamentally correct factorization of the other probability factor of (9), $P(AB\,|\,ab\lambda)$, is given by [11]

$$P(AB\,|\,ab\lambda) = P(A\,|\,ab\lambda)\, P(B\,|\,Aab\lambda)\,. \qquad (11)$$

However, as Jaynes notes, one could argue as Bell did that EPR demands that A should not influence events at B for space-like intervals. This requirement then leads to the factorization used by Bell to represent the EPR problem

$$P(AB\,|\,ab\lambda) = P(A\,|\,a\lambda)\, P(B\,|\,b\lambda)\,. \qquad (12)$$

Nonetheless, the factorization (12) disagrees with the formalism of quantum mechanics in that the result of the measurement at A must be known before the correlation affects the measurement at B, *i.e.* $P(B\,|\,Aab)$. Hence it is not surprising that Bell's inequality is not satisfied in systems that obey quantum mechanics.

Two additional hidden assumptions are identified by Jaynes in Bell's derivation, in addition to those mentioned above:

1. Bell assumes that a conditional probability $P(X\,|\,Y)$ represents a physical causal influence of Y on X. However, consistency requires that conditional probabilities express logical inferences not physical influences.
2. The class of Bell hidden variable theories mentioned in section 2 does not include all local hidden variable theories. As mentioned in that section, hidden variable theories don't need to satisfy the form of (1) (or alternatively (8)), to reproduce quantum mechanical results, as evidenced in Bohmian mechanics.

Bell's inequality thus applies to the class of hidden variable theories that satisfy his relation (1), *i.e.* Bell hidden variable theories, but not necessarily to other hidden variable dynamic theories.

The superluminal communication implication stems from the first hidden assumption above which shows that what is thought to travel faster than the speed of light is actually a logical inference, not a physical causal influence. As summarized by Jaynes [11],

> "The measurement at A at time t does not change the real physical situation at B; but it changes our state of knowledge about that situation, and therefore it changes the predictions we are able to make about B at some time t'. Since this is a matter of logic rather than physical causation, there is no action at a distance and no difficulty with relativity."

There is simply no superluminal communication, as required by special relativity. Assuming otherwise would be similar to Pauli assuming that the established law of conservation of energy mysteriously fails in weak interactions instead of successfully postulating a new particle (the neutrino).

6 Discussion and conclusion

In this paper, we have investigated the applicability of Bell's inequality, based on the assumptions used in its derivation. We have considered two examples of hidden variable dynamic theories that do not satisfy Bell's initial equation (1) used to derive his inequality, and consequently for which Bell's inequality is not applicable: one based on the Nyquist-Shannon Sampling Theorem of Fourier transform theory and the other based on the wave-particle solutions of the *STCED* wave equations which are similar to Louis de Broglie's "double solution". We highlight two hidden assumptions identified by Jaynes [11] that limit the applicability of Bell's inequality, as derived, to Bell hidden variable theories and that show that there are no superluminal physical influences, only logical inferences.

We close with a quote from Jaynes [27, see p. 328] that captures well the difficulty we are facing:

> "What is done in quantum theory today... when no cause is apparent one simply postulates that no cause exists – ergo, the laws of physics are indeterministic and can be expressed only in probability form."

Thus we encounter paradoxes such as seemingly superluminal physical influences that contradict special relativity, and "spooky action at a distance" is considered as an explanation rather than working to understand the physical root cause of the problem. This paper shows that, in this case, the root cause is due to improper assumptions, specifically the first hidden assumption identified by Jaynes highlighted in section 5 above, that is assuming that a conditional probability represents a physical influence instead of the physically correct logical inference. In summary,

> "He who confuses reality with his knowledge of reality generates needless artificial mysteries." [11]

Submitted on February 19, 2016 / Accepted on February 22, 2016

References

1. Bell J. S. On the Einstein–Podolsky–Rosen Paradox. *Physics*, 1964, v. 1, 195–200. Reprinted in Bell J. S. Speakable and Unspeakable in Quantum Mechanics. Cambridge University Press, Cambridge, 1987, pp. 14–21.
2. Goldstein S., Norsen, T., Tausk D. V., Zanghi, N. Bell's Theorem. *Scholarpedia*, 2011, v. 6 (10), 8378.
3. Gouesbet G. Hidden Worlds in Quantum Physics. Dover Publications, New York, 2013, p. 280–303.
4. Mehra J. Einstein, Physics and Reality. World Scientific Publishing, Singapore, 1999.

5. Whitaker A. Einstein, Bohr and the Quantum Dilemma; From Quantum Theory to Quantum Information, 2nd ed. Cambridge University Press, Cambridge, 2006.
6. Home D., Whitaker A. Einstein's Struggles with Quantum Theory; A Reappraisal. Springer, New York, 2007.
7. Stone A. D. Einstein and the Quantum; The Quest of the Valiant Swabian. Princeton University Press, Princeton, 2013.
8. Einstein A, Podolsky B. and Rosen N. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Phys. Rev.*, 1935, v. 47, 777–780.
9. Bohr N. Can Quantum Mechanical Description of Reality Be Considered Complete? *Phys. Rev.*, 1935, v. 48, 696.
10. Aspect A, Dalibard J. and Roger G. Experimental test of Bell's inequalities using time-varying analyzers. *Phys. Rev. Lett.*, 1982, v. 49, 1804–1807.
11. Jaynes E. T. Clearing Up Mysteries – The Original Goal. In Skilling J., ed. Proceedings Volume, Maximum Entropy and Bayesian Methods. Kluwer Academic Publishers, Dordrecht, 1989, pp. 1–27.
12. Jaynes E. T. Probability in Quantum Theory. In Zurek W. H., ed. Complexity, Entropy and the Physics of Information. Addison Wesley Publishing, Reading, MA, 1990.
13. Millette P. A. The Heisenberg Uncertainty Principle and the Nyquist-Shannon Sampling Theorem. *Progress in Physics*, 2013, v. 9 (3), 9–14. arXiv: quant-ph/1108.3135.
14. Ziman J. M. Principles of the Theory of Solids, 2nd ed. Cambridge University Press, Cambridge, 1979.
15. Chaikin P. M. and Lubensky T. C. Principles of Condensed Matter Physics. Cambridge University Press, Camvridge, 1995.
16. Millette P. A. Elastodynamics of the Spacetime Continuum. *The Abraham Zelmanov Journal*, 2012, v. 5, 221–277.
17. de Broglie L. Non-Linear Wave Mechanics. Elsevier Publishing, Amsterdam, 1960.
18. de Broglie L. Les incertitudes d'Heisenberg et l'interprétation probabiliste de la Mécanique Ondulatoire. Gauthier-Villars, Paris, 1982. Available in English: Heisenberg's Uncertainties and the Probabilistic Interpretation of Wave Mechanics. Kluwer Academic Publishers, Dordrecht, 1990.
19. Holland P. R. The Quantum Theory of Motion; An Account of the de Broglie-Bohm Causal Interpretation of Quantum Mechanics. Cambridge University Press, Cambridge, 1993.
20. Dürr D., Teufel, S. Bohmian Mechanics; The Physics and Mathematics of Quantum Theory. Springer-Verlag, Berlin, 2009.
21. Goldstein S. Bohmian Mechanics. *The Stanford Encyclopedia of Philosophy*, 2013, http://plato.stanford.edu/archives/spr2013/entries/qm-bohm/.
22. Bohm D. A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables I, II. *Physical Review*, 1952, v. 85, 166-–179, 180–193.
23. Bell J. S. On the Impossible Pilot Wave. *Foundations of Physics*, 1982, v. 12, 989–999. Reprinted in Bell J. S. Speakable and Unspeakable in Quantum Mechanics. Cambridge University Press, Cambridge, 1987, pp. 159–168.
24. Millette P. A. Wave-Particle Duality in the Elastodynamics of the Spacetime Continuum (STCED). *Progress in Physics*, 2014, vol. 10 (4), 255–258.
25. Hasselbach F. Recent Contributions of Electron Interferometry to Wave-Particle Duality. In Selleri F., ed. Wave-Particle Duality. Plenum, New York, 1992, 109–125.
26. Selleri F. Quantum Paradoxes and Physical Reality. Kluwer Academic Publishers, Dordrecht, 1990.
27. Jaynes E. T., Bretthorst G. L., ed. Probability Theory; The Logic of Science. Cambridge University Press, Cambridge, 2003.
28. Jeffreys H. Scientific Inference. Cambridge University Press, Cambridge, 1931.
29. Jeffreys H. Theory of Probability. Oxford University Press, Oxford, 1939.

Coincident Down-chirps in GW150914 Betray the Absence of Event Horizons

Robin James Spivey

Biological Sciences, Bangor University, Brambell, Deiniol Road, Bangor, Gwynedd, Great Britain
E-mail: y.gofod@gmail.com

A century has elapsed since gravitational waves were predicted. Their recent detection by the LIGO-Virgo collaboration represents another feather in Einstein's cap and attests to the technological ingenuity of experimentalists. However, the news has been portrayed as affirmation of the existence of black holes, objects whose defining characteristics are event horizons. Whilst a gravitational wave chirp is indicative of coalescing bodies and the inferred masses, $29\pm4M_\odot$ and $36\pm5M_\odot$, rule out neutron stars, a prominent yet overlooked feature in the Hanford and Livingston spectrograms points to a curious mass ejection during the merger process. The spectral bifurcations, beyond which down-chirps are clearly discernible, suggest that a considerable quantity of matter spiralled away from the binary system at the height of the merger. Since accretion disks cannot survive until the latter stages of coalescence, a black hole model seems untenable, and Einstein's expectation that black holes can neither form nor ingest matter in a universe of finite age would appear to be upheld. By virtue of general relativity's logical consistency and the fact that gravity propagates at light speed, gravitational collapse must terminate with the formation of pathology-free temporally suspended objects.

1 The black hole controversy

Einstein realised in 1916 that spacetime could mediate the propagation of energy-transporting gravitational waves travelling at light speed [1]. This entirely theoretical deduction was recently confirmed by the LIGO-Virgo collaboration, demonstrating once again the impeccable physical insights of this great scientist. However, the conclusion drawn on the back of this detection, that coalescing black holes triggered the waves [2], directly contradicts Einstein's published stance [3] regarding the outcome of gravitational collapse.

The first static solution to the field equations of general relativity was found that same year describing the gravitational influence of an idealised, infinite density point mass on asymptotically flat space [4]. Due to the Birkhoff theorem, regions of Schwarzschild's metric accurately represent the gravity external to spherically symmetric bodies such as irrotational stars and planets. However, Einstein appreciated that in the immediate vicinity of Schwarzschild's point mass the solution was physically unrealistic, being unreachable from regions outside the event horizon [3].

Einstein's cogent objection to black holes is easily illustrated by a concrete example. If a ray of light moving directly towards a Schwarzschild black hole can neither arrive at the event horizon nor penetrate it, then no particle can. For a lightlike radial trajectory leading towards the event horizon, the Schwarzschild metric reduces to $(dr/dt)^2 = (2m/r - 1)^2$. Assigning initial coordinates $(r, t) = (r_0, 0)$ to a photon, radius $r_1 < r_0$ is attained at time $t_1 > 0$, which can be readily obtained through integration:

$$t_1 = \int_{r_0}^{r_1} \frac{dr}{2m/r - 1} = \int_{r_0}^{r_1} \left(-1 - \frac{2m}{r - 2m} \right) dr, \qquad (1)$$

$$t_1 = r_0 - r_1 + 2m \ln\left(\frac{r_0 - 2m}{r_1 - 2m} \right). \qquad (2)$$

As the photon nears the horizon, $r_1 \to 2m(1 + \epsilon)$ where $0 < \epsilon \ll 1$. Since ϵ is a factor in the denominator of the logarithm, t_1 grows without limit as $\epsilon \to 0$. Accordingly, even though proper time does not advance for lightlike particles, global relationships within the spacetime impose an insurmountable temporal impediment to their arrival at the event horizon. For timelike particles the situation is much the same. As general relativity is a deterministic theory, this calculation has profound implications, despite its brevity. Even in the most favourable of circumstances a black hole cannot absorb matter and, hence, a universe initially devoid of black holes remains forever devoid of black holes. Since general covariance is integral to this theory, changes of coordinates, as detailed for example in references 5–6 of [2], cannot alter this fundamental conclusion.

Although the stationary black hole metrics satisfy the field equations, they lack a dynamical formation mechanism. It is known that event horizons never quite form during gravitational collapse in a universe of finite age [6–12]. Some theorists claim that infalling matter can arrive at the event horizon of a pre-existing black hole in finite proper time, but in practice this is forbidden by the existence of inviolable temporal relationships that permeate spacetime [13]. In addition, a variety of imprecise arguments commonly advanced for the existence of black holes have been robustly refuted [5].

Some stubborn problems now occupying the time of theoretical physicists are symptomatic of misunderstandings. Belief in black holes has given rise to difficulties such as the information paradox [14], loss of causality within rotating black holes, singularities of infinite mass density and the fact

that when matter is trapped within an event horizon, it has no means of influencing external matter, even gravitationally. Furthermore, tension has arisen between the observed characteristics of certain astrophysical phenomena and popular black hole models. In particular, the finite lifetimes and extreme energetics of quasars and active galactic nuclei (AGN) are difficult to reconcile with nearby galaxy clusters which have only reprocessed around 10% of their primordial gas reserves, yet harbour quiescent galactic nuclei.

The ultrarelativistic emission of charged particles by quasars along biaxial jets alludes to an electromagnetically active central engine of some form. Whereas any charge accruing on a spheroidal black hole would be rapidly neutralised, a gravitationally collapsed object of toroidal topology would be defended by a magnetosphere whose flux lines run locally parallel to its surface [13, 15]. This inference clashes with the "principle of topological censorship", a theorem that is irrelevant if a spacetime has no trapped surfaces [16]. Hence, the characteristics of quasars and AGN offer empirical evidence that gravitational collapse produces "dark holes" lacking event horizons [13]. Quasar extinction would coincide with topological collapse and charge nullification.

Appreciation of the impossibility of event horizon formation inspired the first detailed proposal concerning a future mechanism for dark energy decay. It involves the discharge of vacuum energy via the Unruh effect by intense accelerations exposed within the deepest innards of dark holes [17]. The same work also highlights a novel objection to the existence of black holes relating to their unacceptable influence on the total entropy of the universe. A single supermassive black hole devouring matter could potentially double the entropy of the visible universe in the space of a few seconds, despite poor opportunities for interactions of the captured matter.

2 The dawn of gravitational wave astronomy

Since the announcement that gravitational waves have been detected it has emerged that the GW150914 event closely coincided with a gamma ray burst originating in the same sector of the sky [19]. Suggesting a common source, the binary system must have, as the authors put it, become "unexpectedly active" during coalescence. The possibility that one or more neutron stars were involved can be rejected due to the large masses involved [2,20]. The gamma rays are clearly inconsistent with the no-hair conjecture: any accreting matter should be ejected well before the merger [21]. It is therefore interesting to revisit the gravitational wave data to look for any other evidence of unanticipated peculiarities.

Two such examples draw the eye. In the spectrograms of *both* laser interferometers a down-chirp can be clearly discerned, bifurcating from the somewhat stronger up-chirp during the final crescendo of the merger (see Figure 1). These appear to be comfortably above the noise floor of each detector. The down-chirps are not only present in both spectrograms, they are identically located and share the same characteristics: important hallmarks of a genuine signal.

For a binary dark hole or binary frozen star model, significant mass loss is conceivable during a cataclysmic merger of this kind. The particles held in suspension by time dilation would be strongly perturbed by the gravitational ripples, transporting here a total energy estimated at $3M_{\odot}c^2$ [2]. The combination of this disruption and the violent rotation, particularly during the non-axisymmetric dumbbell phase of coalescence, could plausibly give rise to significant expulsion of matter at the peripheral fringes of the system. The spectral traces are consistent with matter being centrifugally launched with a radial velocity component of approximately $0.04c$. There is also a marked acceleration of the chirp following the shedding of mass, as might be anticipated if the rest mass energy of the ejecta was comparable to the energy radiated in gravitational waves.

From (2), at late times an infalling photon asymptotically approaches the radius $r = 2m$. Why must the photon halt at the exact radius of the event horizon? Why does general relativity only marginally forbid the growth and formation of black holes? Could matters have been any different?

Einstein's theory of gravitation was built upon special relativity which insists that nothing can travel faster than the speed of light in vacuum, prohibiting objects from exerting any form of superluminal influence. As in Newton's theory, gravity has infinite range. This demands that gravitons be massless, with current experimental constraints providing an upper limit of 1.2×10^{-22} eV. Signals from LIGO's geographically separated interferometers support the expectation that gravity travels at the speed of light [2]. Were the speed of gravity any different, the terminal radius of the photon would change, and philosophical problems would ensue.

If photons could only asymptotically approach some radius $r > 2m$, gravitational time dilation could then grow without limit in relatively moderate circumstances, curbing the maximum curvature of spacetime irrespective of Planck-scale limitations. If photons could asymptotically approach some radius $r < 2m$ then event horizons could form, bringing with them all the pathologies associated with black holes. Only if gravity travels at the speed of light can spacetime be arbitrarily warped without fear of event horizon formation, points of infinite mass density, time travel paradoxes or violation of unitarity. Like gravity, electromagnetism has unlimited range. Electric fields are mediated by virtual photons. If black holes did exist then the electric fields of charged particles would vanish upon capture, creating an 'electrical paradox' akin to the very widely acknowledged information paradox. Fortunately, Einstein appears to have formulated a consistent theory of gravitation in which anomalies are avoided but all else is permitted. A strongly curved spacetime may be vital for the timely decay of dark energy [17], a possible requirement for gravity to propagate no slower than light.

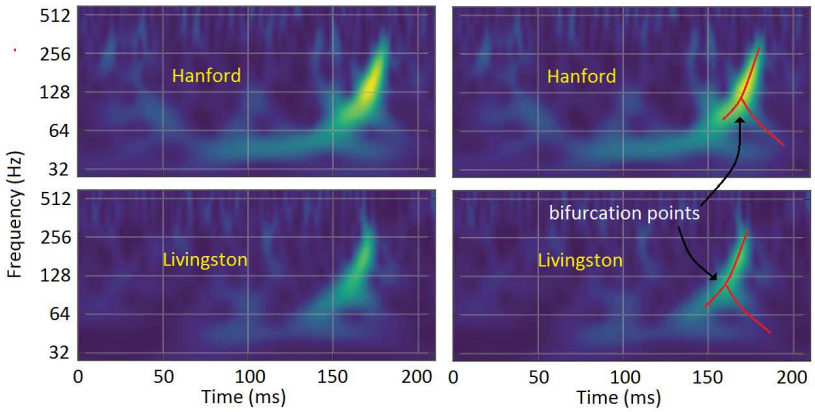

Fig. 1: The gravitational wave spectrograms for the Hanford (top) and Livingston (bottom) Advanced LIGO detectors [2]. Right column: spectral traces have been annotated to show the primary up-chirp and a matching pair of bifurcations beyond which the decline in frequency and amplitude suggests the ejection of mass spiralling away from the merging binary system.

3 Discussion

Gravitational waves have the capability to rectify some long-standing theoretical misconceptions. With improvements in sensitivity already scheduled we shall soon know whether mass-ejections are a generic feature of dark hole coalescence events. If so, we might in time witness some spectacular mergers of supermassive dark holes in the aftermath of galactic mergers within galaxy clusters. For coalescing bodies of large and favourably aligned angular momenta, the resulting gravitational wave signatures could be morphologically very distinct from GW150914 due to the formation of a toroidal dark hole with an unusually lengthy ringdown phase [15, 17]. The publicity and interest surrounding the announcement that gravitational waves have been detected is understandable. However, there has been little or no mention of the fact that the presence of a black hole event horizon cannot be verified even in principle [22] or that Einstein had mathematical grounds for dismissing the notion that black holes exist [3]. Black hole proponents might care to take note that our civilisation still awaits evidence that any of Einstein's predictions concerning gravity were incorrect.

Submitted on February 23, 2016 / Accepted on February 28, 2016

References

1. Einstein A. Approximative integration of the field equations of gravitation. *Sitzungsber. Preuss. Akad. Wiss. Berlin*, 1916, v. 1, 688.
2. Abbott B. P. et al., LIGO & VIRGO collaborations. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 2016, v. 116, 061102.
3. Einstein A. On a stationary system with spherical symmetry consisting of many gravitating masses. *Annals of Mathematics*, 1939, v. 40, 4.
4. Schwarzschild K. Über das gravitationsfeld eines massenpunktes nach der Einsteinschen theorie. *Sitzungsber. Preuss. Akad. Wiss. Berlin*, 1916, v. 1, 189.
5. Weller D. Five fallacies used to link black holes to Einstein's relativistic space-time. *Prog. Phys.*, 2011, v. 7 (1), 93.
6. Oppenheimer J. R. & Snyder H. On continued gravitational contraction. *Phys. Rev.*, 1939, v. 56 (5), 455.
7. Vachaspati T., Stojkovic D. & Krauss L. M. Observation of incipient black holes and the information loss problem. *Phys. Rev. D*, 2007, v. 76 (2), 024005.
8. Kiselev V. V., Logunov A. A. & Mestvirishvili M. A. Black holes: theoretical prediction or fantasy? *Physics of Particles and Nuclei*, 2006, v. 37 (3), 317–320.
9. Kiselev, V. V. E., Logunov A. A. & Mestvirishvili M. A. The physical inconsistency of the Schwarzschild and Kerr solutions. *Theor. & Math. Physics*, 2010, v. 164 (1), 972–975.
10. Shiekh A. Y. Approaching the event horizon of a black hole. *Adv. Stud. Theor. Phys.*, 2012, v. 6 (23), 1147–1152.
11. Chafin C. E. Globally Causal Solutions for Gravitational Collapse. arXiv: gr-qc/1402.1524.
12. Piñol M. A model of dust-like spherically symmetric gravitational collapse without event horizon formation. *Prog. Phys.*, 2015, v. 11 (4), 331.
13. Spivey R. J. Dispelling black hole pathologies through theory and observation. *Prog. Phys.*, 2015, v. 11, 321–329.
14. Mathur S. D. The information paradox: a pedagogical introduction. *Class. Quant. Grav.*, 2009, v. 26 (22), 224001.
15. Spivey R. J. Quasars: a supermassive rotating toroidal black hole interpretation. *MNRAS*, 2000, v. 316 (4), 856–874.
16. Friedman J. L., Schleich K. & Witt D. M. Topological censorship. *Phys. Rev. Lett.*, 1993, v. 71 (10), 1486.
17. Spivey R. J. A non-anthropic solution to the cosmological constant problem. *Prog. Phys.*, 2016, v. 12 (1), 72–84.
18. Spivey R. J. A cosmological hypothesis potentially resolving the mystery of extraterrestrial silence with falsifiable implications for neutrinos. *Physics Essays*, 2015, v. 28 (2), 254–264.
19. Connaughton V. et al. Fermi GBM observations of LIGO gravitational wave event GW150914. arXiv: astro-ph.HE/1602.03920v3.
20. Oppenheimer J. R. & Volkoff G. M. On massive neutron cores. *Physical Review*, 1939, v. 55 (4), 374.
21. Loeb A. Electromagnetic counterparts to black hole mergers detected by LIGO. arXiv: astro-ph.HE/1602.04735v1.
22. Abramowicz M. A., Kluzniak W. & Lasota J. P. No observational proof of the black-hole event-horizon. *Astronomy & Astrophysics*, 2002, v. 396 (3), L31–L34.

Repulsive Gravity in the Oppenheimer-Snyder Collapsar

Trevor W. Marshall
Buckingham Centre for Astrobiology, The University of Buckingham, Buckingham MK18 1EG, UK
E-mail: trevnat@talktalk.net

The Oppenheimer-Snyder metric for a collapsing dust ball has a well defined equilibrium state when the time coordinate goes to plus infinity. The entire ball is contained within the gravitational radius r_0, but half of its content lies within a thin shell between r_0 and $0.94r_0$. This state has the acausal property that no light ray escapes from it, but if one boundary condition at the surface, which Oppenheimer and Snyder imposed without justification, is removed, then all points in the interior remain in causal contact by null geodesics with the exterior. This modification causes the half shell's interior radius to increase to $0.97r_0$. Together with the results of a previous article on the density inside a spherosymmetric neutron star, the present results indicate that, in contrast with the universal attraction of Newtonian gravity, General Relativity gives gravitational repulsion at high density.

1 Introduction

The modern concept of black hole originates with Chandrasekhar's [1] discovery of an upper bound for the mass of a Newtonian white dwarf; it has been claimed (see, for example [2] section 11.3) that the replacement of Newtonian gravitation by General Relativity (GR) makes no significant difference. Using GR, Oppenheimer and Volkoff [3] (OV) found a similar result for neutron stars, the upper bound being somewhat lower than in the white-dwarf case. The OV article, in its footnote 10, did indicate that the GR field equations allow for a stable solution having zero density at the origin in place of the maximum density there of the Newtonian solution, but gave no further attention to this possibility; there seems to have been no serious attempt to return to it since, though a well known text ([5] after equation 23.20) has described it as "unphysical". We showed [4] that solutions of the OV-footnote variety may easily be obtained. The only new feature of such solutions which could conceivably qualify for the "unphysical" label is that the metric has a simple-pole singularity at the origin. This singularity is curiously similar to that now very widely used to describe a black hole, but with the crucial difference that its residue is positive, so that instead of infinite density there we find zero density.

In our previous article we advocated a field, rather than the geometric interpretation of GR, constructing a field energy tensor to explain why the stellar material is concentrated in a spherical shell and not at the origin. Here we shall use an exclusively geometric description, but will nevetheless be able to demonstrate, by studying the particle geodesics inside the shell, that the picture which emerges almost demands that we accept there is gravitational repulsion in the interior of the shell. We conclude that the black hole is a Newtonian concept, superseded by GR.

Our geometric investigation is based on what seems to be the only time-dependent study of a collapsar, namely that of Oppenheimer and Snyder [6] (OS). In an early stage of black-hole theory this article's conclusion was seriously misquoted by Penrose [7] who stated:

"The general situation with regard to a spherically symmetrical body is well known [6]. For a sufficiently great mass, there is *no final equilibrium state* (our emphasis). When sufficient thermal energy has been radiated away, the body contracts and continues to contract until a physical singularity is encountered at $r = 0$."

OS did not say anything resembling this assertion of Penrose. Indeed we shall show below that the OS density distribution approaches a stationary distribution, whose diameter is twice the gravitational radius, as the time goes to plus infinity. It is true that OS also found that in this limit there is a region inside the collapsar from which light may not be emitted, but we shall show below that this is not a real property of the model, and that it may be easily repaired so that all points of the physical space, exterior and interior, remain causally connected at all times. Nobody has demonstrated that any real collapse situation leads to the "trapped surfaces" of the Penrose article, and I would argue that such surfaces would violate the kind of causality described in Weinberg's text ([2] section 7.5). This conclusion was also stated recently by Chafin [8].

2 The OS metric

OS used the comoving coordinates (τ, R, θ, ϕ) with the metric

$$ds^2 = d\tau^2 - \frac{8m^3R}{r}\,dR^2 - r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right), \qquad (1)$$

$$r = 2m\left(R^{3/2} - \frac{3\tau}{4m}\right)^{2/3},$$

in the exterior region $R > 1$ and

$$ds^2 = d\tau^2 - \frac{r^2}{R^2}\,dR^2 - r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right), \qquad (2)$$

$$r = 2mR\left(1 - \frac{3\tau}{4m}\right)^{2/3},$$

in the interior region $0 < R < 1$. By the transformation

$$t = \frac{4m}{3} R^{3/2} - \frac{2}{3}\sqrt{\frac{r^3}{2m}} - 2\sqrt{2mr} + 2m \ln \frac{\sqrt{r} + \sqrt{2m}}{\sqrt{r} - \sqrt{2m}}, \quad (3)$$

the exterior metric converts to the Schwarzschild form

$$ds^2 = \frac{r - 2m}{r} dt^2 - \frac{r}{r - 2m} dr^2 - r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right). \quad (4)$$

We note that, since R is a comoving coordinate, $R = const.$ is a freefall geodesic and in particular the surface $r_1(t)$, that is $R = 1$, satisfies

$$t = \frac{4m}{3} - \frac{2}{3}\sqrt{\frac{r_1^3}{2m}} - 2\sqrt{2mr_1} + 2m \ln \frac{\sqrt{r_1} + \sqrt{2m}}{\sqrt{r_1} - \sqrt{2m}}, \quad (5)$$

and also any such geodesic, for $R > 1$, has its speed v increasing up to a maximum $v = 2c/(3\sqrt{3})$ and then decreasing asymptotically to zero as r approaches $2m$. This confirms the OS statement [6] "...an external observer sees the star asymptotically shrinking to its gravitational radius".

For $0 < R < 1$, OS identified an "internal time" t by defining a *cotime* y as

$$t = \frac{4m}{3} - \frac{4m}{3}\sqrt{y^3} - 4m\sqrt{y} + 2m \ln \frac{\sqrt{y} + 1}{\sqrt{y} - 1}, \quad (6)$$

and then putting

$$y = \frac{r}{2mR} + \frac{R^2 - 1}{2}. \quad (7)$$

This not only gives a continuous match for the internal and external t at $R = 1$, but also the metric in $0 < R < 1$ is

$$ds^2 = \frac{2mr^2(y-1)^2}{Ry^3(r - 2mR^3)} dt^2 - \frac{r}{r - 2mR^3} dr^2 - \\ - r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right), \quad (8)$$

which is continuous with (4) at $R = 1$.

From (6) and (7) we now see that the equilibrium state of the OS model is given by

$$r = mR(3 - R^2) \quad (0 < R < 1), \quad (9)$$

which contradicts the conclusion stated by Penrose and quoted in the previous section of this article. The density ρ is obtained from the curvature tensor of (1)

$$\rho = \frac{mR^3}{4\pi r^3}, \quad (10)$$

and since

$$\sqrt{-g} = \frac{r^3}{R}\sin\theta, \quad (11)$$

it integrates over the volume of the collapsar to give

$$\int_{R<1} \rho\sqrt{-g}\, dR\, d\theta\, d\phi = m. \quad (12)$$

In the remote past, when $r \sim yR, y \to \infty$, the dust particles are distributed uniformly over the sphere's interior, but as collapse proceeds their trajectories, $R = const$, crowd near the surface. This may be shown by considering that in the remote past half of the particles are contained within a shell between $R = 2^{-1/3} = 0.7937$ and $R = 1$, and that their final positions are $r = mR(3 - R^2)$, so that they end up between $r = 1.881m$ and $r = 2m$.

3 A problem with causality

At no time does the entire content of the collapsar go inside the sphere $r = 2m$, so Figure 1 of Penrose [7] is an incorrect picture of the OS collapsar, as is the discussion about *trapped surfaces* on which the figure is based. There is, however a causal anomaly in the OS model, in that, for any $R < 1$, there is a value of t beyond which no light signal emerges.

For the region $R < 1$ we introduce the coordinates (x, R, θ, ϕ), where $x = r/(2mR)$. The metric is

$$\frac{1}{4m^2} ds^2 = xdx^2 - x^2 dR^2 - x^2R^2\left(d\theta^2 + \sin^2\theta d\phi^2\right). \quad (13)$$

A radial light wave or radial null geodesic (RNG) satisfies

$$\frac{dR}{dx} = -\frac{1}{\sqrt{x}}, \quad (14)$$

that is

$$R = 2\sqrt{x(0)} - 2\sqrt{x} = 2\sqrt{x_0} - 2\sqrt{x}. \quad (15)$$

In order to reach the surface at $R = 1$ we need $x(1) > 1$ and therefore $x_0 > 9/4$, but from (7) we find that the minimum value of x_0 is 3/2, reached at cotime $y = 1$, that is when t is plus infinity. It follows that, for $y > 7/4$, an RNG from the origin cannot escape.

It is a simple matter to repair this flaw in the OS model; we replace (7) by

$$y = 1 + x - \frac{(R-3)^2}{4}, \quad (16)$$

so that all RNGs for $y > 1$ escape and causality is preserved. I have established [9] that the metric tensor with these coordinates is again continuous at $R = 1$. It differs from (8) in that the tensor component g_{rt} is not zero in $R < 1$, but only at $R = 1$. It was the unjustified imposition of the condition $g_{rt} = 0$ which led OS to claim that the connection (7) between the (x, R) and the (t, r) coordinates is unique. Our amendment of the OS metric leads to a more concentrated shell, because the equilibrium state is now specified by

$$r = \frac{mR(3-R)^2}{2}, \quad (17)$$

which, putting $R = 2^{-1/3}$, leaves half of the original dust matter in a shell between $r = 1.932m$ and $r = 2m$.

4 Gravity becomes repulsive at high densities

We have all believed since 1687 that gravity is universally attractive, so it requires some effort to adjust to the idea that gravity may repel; even the new mode of thought which came with GR did not change the paradigm of attractive gravity. We have attempted to show elsewhere [4] how a full appreciation of the gravitational field may cause us to change our intuition. However, for the present article we shall stay within the geometric presentation of GR, merely pointing the way towards an understanding of repulsive gravity.

We consider the motion of a foreign dust particle of small mass which crashes radially into the surface $R = 1$ at time t, that is at the point $r = r_1(t)$ given by (5), with a speed greater than that at which the surface itself is moving. We ignore the gravitational force exerted by this foreign particle, so it moves along a radial geodesic of the metric (13). The coordinate R is cyclic, so we have a conservation equation

$$x^2 \frac{dR}{ds} = -C\,, \quad (C > 0)\,, \tag{18}$$

and it then follows that

$$\frac{dR}{dx} = \frac{C}{\sqrt{C^2 x + x^3}}\,. \tag{19}$$

This equation, when integrated with initial conditions $(x, R) = (r_1/2m, 1)$, leads to a relation between the final values x_∞ and R_∞ at t equal to plus infinity

$$R_\infty = 1 - \int_{x_\infty}^{r_1/2m} \frac{C}{\sqrt{C^2 x + x^3}}\, dx\,. \tag{20}$$

Now substituting $y = 1$ in (16) provides a second such relation, so eliminating x_∞ we obtain R_∞ in terms of r_1 and C. This is not a difficult process numerically, but in the limiting ultrarelativistic case $C \to \infty$ – effectively a null geodesic – it becomes especially simple

$$R_\infty = 2 - \sqrt{r_1/(2m)}\,, \quad (r_1 < 8m)\,. \tag{21}$$

If $r_1 > 8m$ such a particle passes through the centre and exits at the opposite end of the diameter. A particle which crashes into the collapsar when the latter is close to its final state – r_1 close to 1 – does not penetrate it beyond the surface shell described in the previous section.

As long as we stay within the constraints of the geometric interpretation of GR, we are not able to draw inferences about what causes such a dramatic deceleration; we could, for example [10], continue to insist that it results from time dilation of the metric. I suggest, however, that a return to the language of field theory offers us, at the very least, an attractive alternative; we may claim that the force of repulsive gravity which decelerates the incident particle is the very same as the one which compresses the particles of the collapsar into a thin shell. In the context of a collapsar having a more realistic equation of state we pursued this point of view in our previous article [4].

Submitted on March 2, 2016 / Accepted on March 6, 2016

References

1. Chandrasekhar S. Stellar Structure. Dover, New York, 1939, Chapter IV.
2. Weinberg S. Gravitation and Cosmology. John Wiley, New York, 1972.
3. Oppenheimer J. R. and Volkoff G. M. *Phys. Rev.*, 1939, v. 54, 540.
4. Wallis M. K. and Marshall T. W. Energy in General Relativity — the case of the neutron star. *Proc. Physical Interpretations of Relativity Theory*, 2015, Bauman State Technical University, Moscow, 544–556.
5. Misner C. W., Thorne K. S. and Wheeler J. A. Gravitation. Freeman, San Francisco, 1970.
6. Oppenheimer J. R. and Snyder H. *Phys. Rev.*, 1939, v. 56, 455.
7. Penrose R. *Phys. Rev. Lett.*, 1965, v. 14, 57.
8. Chafin C. E. Globally Causal Solutions for Gravitational Collapse. arxiv: gr-qc/1402.1524.
9. Marshall T. W. *Astrophys. Space Sci.*, 2012, v. 342, 329–332.
10. Thorne K. S. Black Holes and Time Warps. Norton, New York, 1994, p. 417.

The Dirac-Electron Vacuum Wave

William C. Daywitt

National Institute for Standards and Technology (retired), Boulder, Colorado. E-mail: wcdaywitt@me.com

This paper argues that the Dirac equation can be interpreted as an interaction between the electron core and the Planck vacuum state, where the positive and negative solutions represent respectively the dynamics of the electron core and a vacuum wave propagating within the vacuum state. Results show that the nonrelativistic positive solution reduces to the Schrödinger wave equation.

1 Introduction

In its rest frame the massive electron core $(-e_*, m)$ exerts the two-term coupling force [1, Sec.7-8]

$$F(r) = \frac{e_*^2}{r^2} - \frac{mc^2}{r} = \frac{(-e_*)(-e_*)}{r^2} - \frac{mm_*G}{r_* r} \qquad (1)$$

on the PV quasi-continuum, where e_* is the massless bare charge with its derived electron mass m, and $G\,(= e_*^2/m_*^2)$ is Newton's gravitational constant. The first $(-e_*)$ in (1) belongs to the electron and the second to the separate Planck particles making up the degenerate PV state. The two terms in (1) represent respectively the Coulomb repulsion between the electron charge and the separate PV charges, and their mutual gravitational attraction.

The particle/PV coupling force (1) vanishes at the electron Compton radius $r_c\,(= e_*^2/mc^2)$. In addition, the vanishing of $F(r_c)$ is a Lorentz invariant constant [2] that leads to the important Compton-(de Broglie) relations

$$r_c \cdot mc^2 = r_d \cdot cp = r_L \cdot E = r_* \cdot m_* c^2 = e_*^2 \;(= c\hbar) \qquad (2)$$

where $r_d = r_c/\beta_0\gamma_0$ and $r_L = r_c/\gamma_0$, and $r_*\,(= e_*^2/m_*c^2)$ and m_* are the Compton radius and mass of the Planck particles within the PV. The ratio of the electron speed v to the speed of light c is β_0 and $\gamma_0 = 1/(1-\beta_0^2)^{1/2}$. The relativistic momentum and energy following from the invariance of $F(r_c) = 0$ are $p\,(= m\gamma_0 v)$ and $E\,(= m\gamma_0 c^2)$, from which $E = (m^2c^4 + c^2p^2)^{1/2}$ is the relativistically important energy-momentum relationship.

The results of the previous paragraph show that the important relativistic energy E and momentum p (or its vector counterpart $\mathbf{p}$) are determined at the basic core-PV interaction level. Furthermore, since the core is many orders-of-magnitude smaller than the electron Compton radius, it is reasonable to assume that this point-core picks up its wave-particle nature (including its Compton radius and its energy and momentum operators) from its coupling to the PV continuum.

2 Dirac equation

The Dirac equation [3, p.79]

$$i\hbar\frac{\partial}{\partial t}\begin{pmatrix}\phi\\ \chi\end{pmatrix} = \begin{pmatrix}c\vec{\sigma}\cdot\widehat{\mathbf{p}}\chi\\ c\vec{\sigma}\cdot\widehat{\mathbf{p}}\phi\end{pmatrix} + mc^2\begin{pmatrix}\phi\\ -\chi\end{pmatrix} \qquad (3)$$

where $\widehat{\mathbf{p}}\;(=-i\hbar\nabla)$ is the momentum operator and $\hbar$ is the reduced Planck constant, can be expressed using (2) as

$$ir_c\frac{\partial}{c\partial t}\begin{pmatrix}\phi\\ \chi\end{pmatrix} + \begin{pmatrix}\vec{\sigma}\cdot ir_c\nabla\chi\\ \vec{\sigma}\cdot ir_c\nabla\phi\end{pmatrix} = \begin{pmatrix}\phi\\ -\chi\end{pmatrix} \qquad (4)$$

where the solutions ϕ and χ for this electron-vacuum system are 2x1 Dirac spinors, and $\vec{\sigma}$ is the Pauli 2x2 vector matrix derived from the three 2x2 Pauli spin matrices σ_k $(k = 1, 2, 3)$ [3, p.12]

$$\sigma_1 = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\quad \sigma_2 = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix},\quad \sigma_3 = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}. \qquad (5)$$

The gradient operators $\partial/c\partial t$ and ∇ in (4) are normalized by the electron Compton radius r_c

$$\frac{\partial}{c\partial t/r_c} \qquad \text{and} \qquad \frac{\partial}{\partial x_k/r_c} \qquad (6)$$

whose denominators can be looked upon as the normalized line elements cdt/r_c and dx/r_c of a spacetime [4, p.27] perturbed by the electron core $(-e_*, m)$ in (1). Following from this viewpoint is the concept that *the 2x1 spinors ϕ and χ represent, respectively, the PV response to the electron core $(-e_*, m)$ and some type of vacuum wave.* Furthermore, the vacuum wave cannot be a Planck-particle wave, since the PV is a degenerate state (where the vacuum eigenstates are fully occupied). Thus the wave must be of the nature of a percussion wave, analogous to a wave traveling on the head of a kettle drum.

3 Dirac-Schrödinger reduction

The solution χ in the two simultaneous equations of (4) is assumed in the PV theory to represent a relativistic vacuum wave propagating within the PV state. What follows derives the nonrelativistic version of that wave to add more credence and understanding the vacuum wave idea.

The Dirac-to-Schrödinger reduction [3, p.79] of (4) begins with the elimination of its mass related, high-frequency, components by assuming

$$\begin{pmatrix}\phi\\ \chi\end{pmatrix} = \begin{pmatrix}\phi_0\\ \chi_0\end{pmatrix}e^{-i(mc^2t/\hbar)} = \begin{pmatrix}\phi_0\\ \chi_0\end{pmatrix}e^{-i(ct/r_c)} \qquad (7)$$

where ϕ_0 and χ_0 are slowly varying functions of time compared to the exponential. This result implies that the frequency $\omega_c = c/r_c \gg \omega_0$ for any ω_0 associated with ϕ_0 or χ_0. Inserting (7) into (4) gives

$$ir_c \frac{\partial}{c\partial t}\begin{pmatrix}\phi_0\\ \chi_0\end{pmatrix} + \begin{pmatrix}\overrightarrow{\sigma}\cdot ir_c\nabla\chi_0\\ \overrightarrow{\sigma}\cdot ir_c\nabla\phi_0\end{pmatrix} = \begin{pmatrix}0\\ -2\chi_0\end{pmatrix} \tag{8}$$

where the 0 on the right is a 2x1 null spinor. This zero spinor indicates that the mass energy of the free electron core is being ignored, while the effective negative mass-energy of the vacuum wave has doubled ($-2\chi_0$). In effect, mass energy for the core-vacuum system has been conserved by shifting the mass energy from the free relativistic core to the vacuum wave.

The lower of the two simultaneous equations in (8) can be reduced from three to two terms by the assumption

$$\left| ir_c \frac{\partial\chi_0}{c\partial t} \right| \ll |-2\chi_0| \tag{9}$$

if the kinetic energy of the vacuum wave is significantly less than its effective mass energy. Inserting (9) into (8) then yields

$$ir_c \frac{\partial}{c\partial t}\begin{pmatrix}\phi_0\\ 0\end{pmatrix} + \begin{pmatrix}\overrightarrow{\sigma}\cdot ir_c\nabla\chi_0\\ \overrightarrow{\sigma}\cdot ir_c\nabla\phi_0\end{pmatrix} = \begin{pmatrix}0\\ -2\chi_0\end{pmatrix} \tag{10}$$

as the nonrelativistic version of (4). The mass energy of the free core, and the kinetic energy of the vacuum wave (associated with the lower-left null spinor), are discarded in this nonrelativistic approximation to (4).

Separating the two equations in (10) produces

$$ir_c \frac{\partial\phi_0}{c\partial t} + \overrightarrow{\sigma}\cdot ir_c\nabla\chi_0 = 0 \tag{11}$$

and

$$\overrightarrow{\sigma}\cdot ir_c\nabla\phi_0 = -2\chi_0 \tag{12}$$

where the second term in (11) and the first term in (12) represent the connection between the free-space core dynamics (ϕ_0) and the vacuum wave (χ_0). Inserting (12) into (11) then leads to [3, p.80]

$$ir_c \frac{\partial\phi_0}{c\partial t} - \frac{(\overrightarrow{\sigma}\cdot ir_c\nabla)^2}{2}\phi_0 = 0\,. \tag{13}$$

Finally, using the Pauli-matrix identity [3, p.12]

$$(\overrightarrow{\sigma}\cdot\nabla)^2 = I\,(\nabla)^2 \tag{14}$$

in (13) yields the free-core Schrödinger equation

$$ir_c \frac{\partial\phi_0}{c\partial t} = \frac{(ir_c\nabla)^2}{2}\phi_0 \quad \text{or} \quad i\hbar\frac{\partial\phi_0}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\phi_0 \tag{15}$$

where the two spin components in ϕ_0 are ignored in this nonrelativistic approximation; so ϕ_0 becomes a simple scaler wavefunction rather than a 2x1 spinor.

4 Conclusions and comments

Although the spin components are missing from the standard version of the Schrödinger equation [5, p.20], the solutions to (11) and (12) indicate that those components are still meaningful.

Using $r_c\,(= e_*^2/mc^2 = \hbar/mc)$ from (2) in (11) and (12) yields

$$ir_c \frac{\partial\phi_0}{c\partial t} = \overrightarrow{\sigma}\cdot(\widehat{\mathbf{p}}/mc)\,\chi_0 \tag{16}$$

and

$$\chi_0 = \overrightarrow{\sigma}\cdot(\widehat{\mathbf{p}}/2)\,\phi_0 \tag{17}$$

where $\widehat{\mathbf{p}}\,(= -i\hbar\nabla)$ is the vector momentum operator.

Equations (16) and (17) from the perturbed spacetime can be understood as follows: the free-space energy from ϕ_0 in the first term of (16) drives the vacuum energy associated with the second term; this χ_0 energy of the second term in (17) then feeds back into the ϕ_0 term in (16), leading to a circular simultaneity between the two equations that represent the coupled nonrelativistic behavior of the core-PV system. Furthermore, the fact that there is no kinetic-energy term in (17) suggests that the localized energy in the PV travels as a percussion wave through that vacuum state. This scenario represents the PV view of the Dirac electron equation (4): that is, the dynamics of the free-space electron core $(-e_*, m)$ lead to a vacuum wave propagating within the PV state, in step with the free electron core.

Dedication

This paper is dedicated to the memory Dr. Petr Beckmann [6], Professor Emeritus Electrical Engineering, the University of Colorado, Boulder, Colorado.

Submitted on March 18, 2016 / Accepted on March 20, 2016

References

1. Daywitt W. C. The trouble with the equations of modern fundamental physics. *American Journal of Modern Physics*, Special Issue: *Physics Without Higgs and Without Supersymmetry*, 2016, v. 5 (1–1), 22. See also www:planckvacuum.com.
2. Daywitt W. C. The de Broglie relations derived from the electron and proton coupling to the Planck vacuum state. *Progress in Physics*, 2015, v. 11 (2), 189.
3. Gingrich D.M. Practical Quantum Electrodynamics. CRC, The Taylor & Francis Group, Boca Raton, 2006.
4. Leighton R.B. Principles of Modern Physics. McGraw-Hill Book Co., New York, 1959.
5. Schiff L.I. Quantum Mechanics, 2nd. McGraw Book Co., Inc., New York, 1955.
6. Petr Beckmann. Article from Wikipedia.

Mass of a Charged Particle with Complex Structure in Zeropoint Field

Kundeti Muralidhar

Physics Department, National Defence Academy, Khadakwasla, Pune-411023, India. kundetimuralidhar@gmail.com

A charged particle immersed in the fluctuating zeropoint field may be visualized as an oscillator and such an oscillating particle is considered to possess an extended structure with center of mass and center of charge separated by radius of rotation in a complex vector space. Considering stochastic electrodynamics with spin, the zeropoint energy absorbed by the particle due to its internal motion has been derived. One may initially assume a massless charged particle with complex structure and after interaction with zeropoint field, the absorbed energy of the particle may correspond to the particle mass. This gives an idea that an elementary particle may acquire mass from the interaction of zeropoint field. When the particle moves as a whole, there appears to be a small energy correction of the order of fine structure constant and it may be attributed to the mass correction due to particle motion in the zeropoint field.

1 Introduction

The Dirac electron executes rapid oscillations superimposed on its normal average translational motion and this oscillatory motion is known as zitterbewegung and it was first shown by Schrödinger. In the zitterbewegung motion, the electron appears vibrating rapidly with a very high frequency equal to $2mc^2\hbar^{-1}$ and with internal velocity equal to the velocity of light. These oscillations are confined to a region of the order of Compton wavelength of the particle. It has been shown by several authors over decades that the center of charge and center of mass of charged particle are not one and the same but they are separated by a distance of the order of Compton wavelength of the particle. The approach of extended particle structure was developed by Wyssenhoff and Raabe [1], Barut and Zhanghi [2], Salesi and Recami [3] and others. The list of references connected with the validity of the extended or internal structure of charged particle are too many and some of them are mentioned in the reference [4]. Thus the structure of an elementary charged particle is not definitely a point particle with charge and mass or a spherical rigid body with charge distribution. The structure of electron may be visualized as the point charge in a circular motion with spin angular momentum. The frequency of rotation is equal to the zitterbewegung frequency and the radius of rotation is equal to half the average Compton wavelength. The circular motion is observed from the rest frame positioned at the centre of rotation which is the centre of mass point. Thus the centre of mass point and the centre of charge point are separated by the radius of rotation. The electron spin generated from the circular motion of zitterbewegung was advocated by several researchers. Holten [5] discussed the classical and quantum electrodynamics of spinning particles. In the Holten theory, the spinning particle emerges as a modification of relativistic time dilation by a spin dependent term and the zitterbewegung appears as a circular motion and the angular momentum of such circular motion represents the spin. In the Hestene model of Dirac electron [6], the spin was considered as a dynamical property of the electron motion. In the approach of geometric algebra, using multivector valued Lagrangian, the angular momentum of this internal rotation represents particle spin and it has been explicitly shown as a bivector quantity representing the orientation of the plane of rotation [7, 8]. In quantum theories, the internal oscillations of the particle are attributed due to vacuum fluctuations. However, in stochastic electrodynamics, the internal oscillatory motion of the particle is attributed to the presence of zeropoint field throughout space [9]. The mass of the particle is seen as the energy of oscillations confined to a region of space of dimensions of the order of Compton wavelength [10].

The classical concept of space is an infinite void and featureless. However, it has been replaced by the vacuum field or the zeropoint random electromagnetic field when the quantum oscillator energy was found to contain certain zeropoint energy and with the substitution of the quantum oscillator energy into the Planck's radiation formula yields the energy density of zeropoint field at absolute zero temperature [11]. In a classical approach to the radiation problem, Einstein and Stern obtained blackbody radiation spectrum and suggested that a dipole oscillator possessed zeropoint energy. In 1916, Nernst proposed that the universe might actually contain ubiquitous zeropoint field without any presence of external electromagnetic sources [12, 13]. Thus the origin of zeropoint field is presumed to be purely a quantum mechanical effect and considered to be uniformly present throughout space in the form of stochastic fluctuating electromagnetic field. The zeropoint radiation is found to be homogeneous and isotropic in space. The spectral density of zeropoint radiation is proportional to ω^3 and it is therefore Lorentz invariant. The electromagnetic zeropoint field consists of fluctuating radiation that can be expressed as a superposition of polarised plane waves. Because of the random impulses from fluctuating zeropoint field, a free particle cannot remain at rest but oscillates about its equilibrium position.

The Planck's idea of zeropoint radiation field was revis-

ited by Marshall and explicitly showed that the equivalence between classical and quantum oscillators in the ground state [14]. This has inspired interesting modifications to classical electrodynamics and the developed subject is called stochastic electrodynamics. Stochastic electrodynamics deals with the movement of charged particles in the classical electromagnetic fluctuating zeropint field. The presence of classical, isotropic, homogeneous and Lorentz invariant zeropoint field in the universe is an important constituent of stochastic electrodynamics. The stochastic electrodynamics approach was used to explain classically several important fundamental results and problems of quantum mechanics [15–20]. Boyer [15] showed that for a harmonic oscillator, the fluctuations produced by zeropint field are exactly in agreement with the quantum theory and as a consequence the Heisenberg minimum uncertainty relation is satisfied for the oscillator immersed in the zeropoint field. Stochastic electrodynamics was used to explain the long standing problems of quantum mechanics, namely the stability of an atom, Van der Walls force between molecules [16], Casimir force [17], etc. All these studies reveal the fact that the conventional concept of space has been changed by the emergence of zeropoint field. A detailed account of stochastic electrodynamics as a real classical electromagnetic field and a phenomenological stochastic approach to the fundamental aspects of quantum mechanics was given by de La Pena *et al.,* [13, 21]. In the stochastic electrodynamics, if the upper cut-off frequency to the spectrum of zeropoint field is not imposed, the energy of the oscillator would be divergent. Despite of its success in explaining several quantum phenomena, the results obtained in the stochastic electrodynamics have certain drawbacks [20]; it neglects Lorentz force due to zeropoint magnetic field, it fails in the case of nonlinear forces, explanation of sharp spectral lines is not possible, diffraction of electrons cannot be explained and further the Schrödinger equation can be derived in particular cases only.

A charged point particle immersed in the fluctuating electromagnetic zeropoint field is considered as an oscillator. In the stochastic electrodynamics approach, the equation of motion of the charged particle in the zeropoint field is known as Brafford-Marshall equation [13] which is simply the Abraham-Lorentz [22] equation of motion of a charged particle of mass m and charge e and it is given by

$$m\ddot{\mathbf{x}} - \Gamma_a m\ddot{\mathbf{v}} + m\omega_0^2\mathbf{x} = e\mathbf{E}_z(\mathbf{x}, t), \tag{1}$$

where $\Gamma_a = 2e^2/3mc^3$, ω_0 is the frequency of oscillations of the particle, $\mathbf{v}$ is the velocity of the particle, c is the velocity of light, $\mathbf{E}_z(\mathbf{x}, t)$ is the external electric zeropoint field and an over dot denotes differentiation with respect to time. In the above equation, the force term contains three parts; the binding force $m\omega_0^2\mathbf{x}$, damping force $\Gamma_a m\ddot{\mathbf{v}}$ and external electric zeropoint field force $e\mathbf{E}_z(\mathbf{x}, t)$. In the case of point particles, the strength of these forces follows the relation

$$m\omega_0^2\mathbf{x} < \Gamma_a m\ddot{\mathbf{v}} < e\mathbf{E}_z(\mathbf{x}, t). \tag{2}$$

The energy absorbed by the particle oscillator in the zeropoint field was given by several authors by introducing certain approximations. There are two main approaches found in the literature; one is due to Boyer [6] and the other is due to Rueda [19]. In addition to these main approaches, recently Cavalleri *et al.,* [20] introduced stochastic electrodynamics with spin and explained several interesting phenomena for example, stability of elliptical orbits in an atom, the origin of special relativity and the explanation for diffraction of electrons. It has been shown that the drawbacks of stochastic electrodynamics can be removed with the introduction of spin into the problem. The particle has a natural cut-off frequency equal to the spin frequency which is the maximum frequency radiated by the electron in the zitterbewegung interpretation. This eliminates the problem of divergence in stochastic electrodynamics. These recent advancements in the field of stochastic electrodynamics fully support the assumption that the stochastic electromagnetic field represents the zeropoint field and renew the interest in studying the fundamental aspects of quantum systems and in particular the charged particle oscillator in zeropoint fields.

In Boyer's extensive studies, the harmonic oscillator was developed under the dipole approximation and the charged particle was considered as a point particle without any internal structure. The point particle limit is endowed with two assumptions; i) when the particle size tends to zero, $\omega_c\tau \ll 1$, where ω_c is the cut-off frequency and τ is the characteristic time and ii) when the radiation damping term is very small compared to the external force, $\Gamma_a\omega_c \ll 1$. In Boyer's process of finding the zeropoint energy associated with the charged particle, an integral under narrow line width approximation was solved and finally the zeropoint energy per mode of the oscillator was obtained [16]. This energy has been shown to be equal to the zeropoint energy of the quantum oscillator.

In Rueda's approach, the classical particle was considered as a homogeneously charged rigid sphere and to find the energy absorbed by the particle, the radiation damping and binding terms were neglected when compared to the force term in the Lorentz Abraham equation of motion. The integration was performed over a range 0 to τ, where τ is the characteristic time taken by the electromagnetic wave to traverse a distance equal to the diameter of the particle. The main difference from Boyer's approach is that Rueda assumed $\omega_c\tau \gg 1$ and this condition means the cut-off wavelength is much smaller than the particle size. Further, Rueda introduced a convergence factor $\eta(\omega)$ in the zeropoint energy of the particle oscillator. The average zeropint energy of the oscillator is given by [19]

$$\langle E_0\rangle = \frac{\Gamma_a\hbar\omega_c^2}{\pi}\,\eta(\omega_c). \tag{3}$$

In the later studies, Haitch, Rueda and Puthoff [23] studied an accelerated charged particle under the influence of zeropint field and obtained a relation for inertial mass of a charged particle which is similar to (3). Recently, Haitch *et al.* [9] suggested that the radiation damping constant in the zeropoint field as Γ_z which is not necessarily equal to the damping constant Γ_a of Larmor formula for power radiated by an accelerated charged particle. If we set $\eta(\omega)\Gamma_z\omega_c \sim 1$, the ground state energy of the particle oscillator in the zeropoint field is written as $(\hbar\omega_c)/\pi$. In the case the cut-off frequency is simlar to the resonant frequency of the particle oscillator in the electromagnetic zeropoint field, the ground state energy is equal to the zitterbewegung energy of the Dirac electron. Here, the frequency ω_c is not generally equal to the frequency of oscillation of the particle and it differs by a fraction of fine structure constant. However, the reason for assuming Γ_a as Γ_z is obscure. It may be understood that the energy in (3) corresponds only to a mass correction but not to the mass of the charged particle.

In the stochastic electrodynamics with spin, the particle is considered to possess an extended internal structure and the particle spin is sensitive to the zeropoint frequency that is equal to the frequency of gyration. The particle gyration motion explains the spin properties and refers to a circular motion at the speed of light [20]. The velocity of the particle is not the real velocity of gyrating particle, but centre of mass point around which the particle revolves. The special relativity is not present at the particle level and arises mainly because of the helical motion of the particle when observed from an arbitrary inertial frame of reference [24]. The centre of circular motion responds only to the force parallel to the spin direction. The equation of motion of centre of mass point can be expressed by (1) provided the external force is parallel to the spin direction.

Clifford algebra or Geometric algebra has been considered to be a superior mathematical tool to express many of the physical concepts and proved to provide simpler and straightforward description to the mathematical and physical problems. The geometric algebra was rediscovered by Hestenes [25] in 1960's and it is being used by a growing number of physicists today. In Geometric algebra, a complex vector is defined as a sum of a vector and a bivector. In the complex vector algebra, the oscillations of a charged particle immersed in zeropoint field have been studied recently by the author [26]. The oscillations of the particle in the zeropoint field may be considered as complex rotations in complex vector space. The local particle harmonic oscillator is analysed in the complex vector formalism considering the algebra of complex null vectors. It has been shown that the average zeropoint energy of the particle is proportional to particle bivector spin and the mass of the particle may be interpreted as a local spatial complex rotation in the rest frame.

In the electromagnetic world, the particle mass originates from the electromagnetic field and it is purely electromagnetic in nature [27]. In the classical Lorentz theory of electron, the self-energy is closely connected to the electromagnetic mass of the electron. The self-energy problem in classical theory or quantum theory is essentially connected to the structure of electron and it may not be correct to assign the structure to the electron as a form factor [28]. Further the classical electromagnetic field may be only responsible for the interaction and gives the particle mass as purely electromagnetic in nature. In quantum field theories, the energy, momentum and charge of a particle appear as a consequence of field quantisation and leads to natural classification of particles depending on their spin values. In the renormalization procedure of quantum field theory with finite cut-off for the radiatively induced mass, it has been shown that mass depends on particle spin in the limit when the bare mass tends to zero [29]. However, in the quantum electrodynamics it is well known that the sum of bare mass and the mass correction equals the electron mass and the mass correction is due to the interaction of the particle with vacuum fluctuations [11]. Recently, Pollock interpreted particle mass (fermion or boson) arising from the zeropoint vacuum oscillations by introducing a matrix mass term in the Dirac equation [30]. The standard model deals with the fundamental particles through interaction of bosons, and at a deeper level one may consider the particles as field excitations. Though the vacuum fluctuations have been treated in a different manner in quantum theory and in quantum electrodynamics, the particle oscillations considered either in the vacuum field or in the classical stochastic electrodynamics with spin, are attributed to the fluctuations of the zeropoint field. The idea that the mass arises from the external electromagnetic interaction may lead to the conclusion that charge retains intrinsic masslessness [31]. It has been argued that for there to be correspondence with the particle mass, perhaps at pre-quantum level, inertial mass must originate from external electromagnetic interaction [32].

The aim of this article is to find the energy absorbed by the particle due to its intrinsic motion in the presence of zeropoint field and to discuss the possible origin of mass generation. In section 2, we have explained the modalities of the extended structure of the charged particle in the complex vector algebra. In the present extended particle structure, since we have considered the center of mass point and center of charge separated by radius of rotation in the complex plane, the equation of motion of the particle as a whole is considered as a combination of equation of motion of center of charge and the equation of motion of center of mass. These equations of motion of center of charge and center of mass are derived in section 3. Considering the equation of motion of center of charge in the zeropoint field, the energy absorbed by an extended charged particle is obtained in section 4, and the possible origin of mass generation is discussed in section 5. Finally, conclusions are presented in section 6. Throughout this article a charged particle implies a particle like electron.

2 The complex structure of a charged particle

In the extended particle structure, the centre of mass and the centre of charge positions are considered as separate. Denoting the centre of local complex rotations by the position vector $\mathbf{x}$ and the radius of rotation by the vector ξ , a complex vector connected with both the motion of the centre of mass point and internal complex rotation is expressed as [26]

$$X(t) = \mathbf{x}(t) + \mathbf{i}\,\xi(t). \tag{4}$$

In the geometric algebra, a bivector represents an oriented plane and $\mathbf{i}$ is a pseudoscalar which represents an oriented volume [33]. Differentiating (4) with respect to time gives the velocity complex vector.

$$U(t) = \mathbf{v}(t) + \mathbf{i}\,\mathbf{u}(t). \tag{5}$$

Here, the velocity of centre of mass point is $\mathbf{v}$ and the internal particle velocity is $\mathbf{u}$. A reversion operation on U gives $\bar{U} = \mathbf{v} - \mathbf{i}\,\mathbf{u}$ and the product

$$U\bar{U} = v^2 + u^2. \tag{6}$$

In the particle rest frame $\mathbf{v} = 0$ and $U\bar{U} = u^2$. Since the particle internal velocity in the particle rest frame $u = c$ the velocity of light, $|U| = u = c$. However, when the particle is observed from an arbitrary frame different from the rest frame of the particle centre of mass, as the centre of mass moves with velocity $\mathbf{v}$, the particle motion contains both translational and internal rotational motion of the particle. Then the particle internal velocity can be seen as

$$u^2 = c^2 - v^2 \tag{7}$$

or

$$u = c(1 - \beta^2)^{1/2} = c\gamma^{-1}, \tag{8}$$

where $\boldsymbol{\beta}= \mathbf{v}/c$ and the factor γ is the usual Lorentz factor. The angular frequency of rotation of the particle internal motion is equal to the ratio between the velocity c and radius of rotation ξ, $\omega_s = c/\xi$. When observed from an arbitrary frame, the angular frequency ω would be equal to the ratio between u and ξ

$$\omega = \frac{u}{\xi} = \omega_s\gamma^{-1}. \tag{9}$$

Thus the angular frequency of rotation decreases when observed from an arbitrary frame and the decrease depends on the velocity of the centre of mass. Considering the helical motion of the particle, this method of calculation for time dilation was first shown in a simple manner by Cavelleri [24]. The above analysis shows that the basic reason for the relativistic effects that we observe is due to the internal rotation which is a consequence of fluctuating zeropoint field and elucidates a deeper understanding of relativity at particle level in addition to the constancy of velocity of light postulate. The difference between ω and ω_s corresponds to the particle velocity. In other words, when the particle moves with velocity $\mathbf{v}$, an important consequence is that the particle itself induces certain modification in the field to take place at a lower frequency ω_B. Thus the motion of a free particle is conveniently visualized as a superposition of frequencies ω_0 and ω_B such that the particle motion as observed from an arbitrary frame appears to be a modulated wave containing internal high frequency ω_0 and an envelope frequency ω_B. The ratio between the envelop frequency and the internal frequency is then expressed as

$$\frac{\omega_B}{\omega_0} = \frac{v}{c}. \tag{10}$$

This result is simply a consequence of superposition of internal complex rotations on translational motion of the particle. The relativistic momentum of the center of mass point can be expressed as $\mathbf{p} = \gamma m\mathbf{v}$ and in the complex vector formalism momentum complex vector is given by [26]

$$P = \mathbf{p} + \mathbf{i}\,\pi, \tag{11}$$

where $\pi = m\mathbf{u}$. The total energy of the particle is now expressed as

$$E^2 = P\bar{P}c^2 = (\mathbf{p} + \mathbf{i}\,\pi)(\mathbf{p} - \mathbf{i}\,\pi) = p^2c^2 + m^2c^4. \tag{12}$$

However, in the presence of external electromagnetic field we normally replace the momentum by $\mathbf{p} - e\mathbf{A}/c$ in the minimal coupling prescription. Now, using $\mathbf{p} \rightarrow \mathbf{p} - e\mathbf{A}/c$ in (12) and equating the scalar parts, the total energy of the particle becomes

$$E^2 = p^2c^2 - 2ec\mathbf{p}.\mathbf{A} + e^2A^2 + m^2c^4. \tag{13}$$

Here, $\mathbf{A}$ represents the zeropoint electromagnetic field vector potential. In the rest frame of the particle, i.e., when the velocity $\mathbf{v} = 0$, the above expression reduces to

$$E_0 \sim mc^2 + \frac{e^2A^2}{2mc^2}, \tag{14}$$

where the higher order terms are neglected. Thus, under the influence of zeropoint field, the term $e^2A^2/2mc^2$ in the above equation gives a correction to mass. Expanding the vector potential in terms of its creation and annihilation operators and averaging in the standard form, it can be shown that the correction term [23]

$$\frac{e^2}{2mc^2}\langle A^2\rangle = \frac{\alpha}{2\pi}\frac{(\hbar\omega_c)^2}{mc^2}, \tag{15}$$

where ω_c is the cut-off frequency and α is the fine structure constant. When the cut-off frequency is equal to the frequency of oscillation of the particle, $\omega_c = \omega_0$ and using Einstein-de Broglie formula $\hbar\omega_0 = mc^2$, the mass correction can be expressed in the following form

$$\langle\Delta E_0\rangle = \delta mc^2 = \frac{\alpha}{2\pi}\,mc^2. \tag{16}$$

Thus in the presence of zeropoint field, the vector potential term in (15) gives the mass correction and it was obtained by Schwinger in quantum electrodynamics. The particle mass which arises due to local complex rotations in the zeropoint field is regarded as the so called bare mass and when the particle is observed from an arbitrary frame, the particle mass has some mass correction due to the presence of external zeropoint field.

3 Equation of motion of the particle with complex structure

It should be noted that, (1) contains the so called runaway and causal problems. In the Landau approximation, the damping term is written as a derivative of external force. In this case, the runaway and causal problems are eliminated and the exact equation of motion of a charged particle was recently given by Rohlrich [34] and Yaghjian [35]. In the equation of motion of the charged particle, centre of mass appears as if the total charge is at that point. In other words, there is no distinction between centre of mass and centre of charge points. In the case of extended particle structure, it has been clarified in the previous sections that the external zeropoint field must be responsible for the internal complex rotations and at the same time for the deviations in the path of the particle when it is moving with certain velocity. The external zeropoint field is then expressed as a function of complex vector X, $\mathbf{E}_z = \mathbf{E}_z(X,t)$ and expanding it gives

$$\mathbf{E}_z(X,t) = \mathbf{E}_z(\mathbf{x},t) + \mathbf{i}\,\xi \left|\frac{\partial \mathbf{E}_z(\mathbf{x},t)}{\partial \mathbf{x}}\right|_{\mathbf{x}\to 0} + O(\xi^2). \tag{17}$$

The second term on right hand side of the above equation is independent of x and it is a function of ξ only. Neglecting higher order terms in (17) and representing the second term on right by $\mathbf{i}\,\mathbf{E}_z(\xi,t)$, the external zeropoint field $\mathbf{E}_z(X,t)$ can be decomposed into a vector and a bivector parts

$$\mathbf{E}_z(X,t) = \mathbf{E}_z(\mathbf{x},t) + \mathbf{i}\,\mathbf{E}_z(\xi,t). \tag{18}$$

The random fluctuations produce kicks in all directions and leads to random fluctuations of the centre of mass point and at the same time random fluctuations also produce internal complex oscillations or rotations. Thus the force acting on the charged particle can be decomposed into two terms, the force acting on the centre of mass and the force acting on the centre of charge. For the field acting on the centre of mass, the particle mass and charge appear as if they are at the centre of mass point and we treat the equation of motion of the particle in the point particle limit. However, for the field acting on the centre of charge, the effective mass seen by the zeropoint field is the mass due to the potential $U_z \sim e^2/2R \sim m_z c^2$. The magnitude of R is of the order of Compton wavelength. Then the effective mass m_z in the zeropoint field is approximately equal to the electromagnetic mass which is proportional to the electromagnetic potential due to charge e at the center of mass position. Replacing the position vector $\mathbf{x}$ by the complex vector X and $\mathbf{E}_z(x,t)$ by the complex field vector $\mathbf{E}_z(X,t)$ in (1) and separating vector and bivector parts gives the equations of motion of the centre of mass and the centre of charge respectively. The equation of motion of center of mass is the Abraham-Lorentz equation of motion of a charged point particle in the external electromagnetic zeropoint field given by (1) and the motion of the centre of mass of the particle is observed from an arbitrary frame of reference. In the rest frame of the particle, the equation of motion represents the equation of motion of center of charge

$$m_z\ddot{\xi} - \Gamma_z m_z \ddot{\mathbf{u}} + m_z\omega_0^2\xi = e\mathbf{E}_z(\xi,t). \tag{19}$$

The terms $\Gamma_z m_z\ddot{\mathbf{u}}$ and $m_z\omega_0^2\xi$ are radiation damping and binding terms respectively. The damping constant in the above equation is defined as $\Gamma_z = (2e^2)/(3m_z c^3)$.

4 Average zeropoint energy associated with the particle in its rest frame

The zeropoint field and particle interaction takes place at resonance and the particle oscillates at resonant frequency ω_0. In other words, the particle oscillator absorbs energy from the zeropoint field at a single frequency which is the characteristic frequency of oscillation. Since, both radiation damping and binding terms are much smaller than the force term in (19) one can neglect these terms and integrating with respect to time t gives the internal velocity of rotation of the particle

$$\mathbf{u}(t) = \frac{e}{m_z}\int_0^\tau \mathbf{E}_z(\xi,t)dt. \tag{20}$$

Here, the upper limit of integration is chosen as the characteristic time τ required by the electromagnetic wave to traverse a distance equal to the size of the particle. The electric field vector $\mathbf{E}_z(\xi,t)$ is expressed in the same form as that of Rueda [19],

$$\mathbf{E}_z(\xi,t) = \sum_{\lambda=1}^{2}\int d^3k\,\epsilon(\mathbf{k},\lambda)\frac{H(\omega)}{2}\times \left[a e^{i(\mathbf{k}.\xi-\omega t)} + a^* e^{i(\mathbf{k}.\xi-\omega t)}\right], \tag{21}$$

where $a = \exp(-i\theta(\mathbf{k},\lambda))$, $a^* = \exp(i\theta(\mathbf{k},\lambda))$ and $\epsilon(\mathbf{k},\lambda)$ is the polarization vector and the normalization constant is set equal to unity. The phase angle $\theta(\mathbf{k},\lambda)$ is a set of random variables uniformly distributed between 0 and 2π and are mutually independent for each choice of wave vector $\mathbf{k}$ and λ. The stochastic nature of the field lies in these phase angles and a statistical average of these phase angles gives an effective value of the field. For point particles, because the size is zero, we find the spectral divergence of zeropoint field. However, for particles with extended structure, one can discern a natural cut-off wavelength associated with the particle size. The convergence factor gives an upper bound to the energy available

from the electromagnetic zeropoint field and it is associated with the characteristic function $H(\omega)$ of the zeropoint field. The function $H(\omega)$ is given by $2\pi^2 H^2(\omega) = \eta(\omega)\hbar\omega$. In (21), integrating the electric field vector with respect to time gives

$$I = \sum_{\lambda=1}^{2} \int_0^\infty d^3k \, \frac{H(\omega)}{2} \left[\epsilon(\mathbf{k}, \lambda)\, a e^{i\mathbf{k}.\xi} \left(\frac{e^{-i\,\omega\tau} - 1}{-i\,\omega} \right) + \epsilon(\mathbf{k}, \lambda)\, a e^{i\mathbf{k}.\xi} \left(\frac{e^{-i\,\omega\tau} - 1}{-i\,\omega} \right) \right].$$

The charge current in the rest frame of the particle is the charge times the internal velocity of the particle. The interaction energy of the charged particle with the zeropoint field is expressed as the charge current times the vector potential of the zeropoint field. However, one can express the vector potential as the integral of the zeropoint electric field vector. Then the average zeropoint energy acquired by the particle is expressed as

$$\langle E_0 \rangle = \frac{e^2}{m} \langle II^* \rangle. \tag{22}$$

The averages of random phase and the polarization vector are expressed as follows

$$\langle aa^* \rangle = \delta(\lambda - \lambda')\, \delta^3(k - k')\,;\ \langle aa \rangle = 0\,;\ \langle a^* a^* \rangle = 0$$
$$\langle \epsilon(\mathbf{k}, \lambda) \epsilon^*(\mathbf{k}, \lambda) \rangle = \delta_{ij} - \tfrac{k_1}{k^2}$$
$$\textstyle\sum_{\lambda=1}^{2} \int d^3k \, \langle \epsilon(\mathbf{k}, \lambda) \epsilon^*(\mathbf{k}, \lambda) \rangle = \tfrac{8\pi}{3} \int \omega^2 \, d\omega \,.$$

Using these stochastic averages, replacing the convergence factor by $\eta(\omega_0)$ and setting the upperlimit of integration to the frequency of oscillations in (22) gives

$$\langle E_0 \rangle = \frac{4e^2\hbar}{3\pi m_z c^3}\, \eta(\omega_0) \int_0^{\omega_0} \omega(1 - \cos\omega\tau)\, d\omega. \tag{23}$$

For an extended particle structure $\omega_0\tau = 2\pi$ and the above equation after integration reduces to

$$\langle E_0 \rangle = \eta(\omega_0)\, \frac{\Gamma_z \hbar \omega_0^2}{\pi}. \tag{24}$$

This result is similar to the result obtained by Reuda [19] and Puthoff [36]. However, the difference is that the damping constant is now replaced by Γ_z and cut-off frequency ω_c is replaced by the resonant internal frequency of oscillation of the particle. In (24), both the values for m_z and $\eta(\omega_0)$ are not known exactly and must be approximated. Instead, one can approximate $\eta(\omega_0)\Gamma_z\omega_0 \sim 1$ for the particle with extended structure. Then the average zeropoint energy acquired by the particle in its rest frame is

$$\langle E_0 \rangle = \frac{\hbar\omega_0}{\pi}. \tag{25}$$

This energy is similar to the zitterbewegung energy of Dirac electron in quantum mechanics.

5 Equation of motion of the particle with complex structure

In the above procedure, initially we have considered the charged particle without any mass. Such particle interacting with zeropoint field acquires mass due to particle resonant oscillations and gains energy from the electromagnetic zeropoint field. This average zeropoint energy of the particle appears as the mass of the particle. In the complex vector formalism of internal harmonic oscillator in zeropoint field, it has been shown by the author that the average energy $\langle E_0 \rangle$ is related to the mass through particle spin and represents the mass generated from the local complex rotations produced by the interaction of zeropoint field with the particle. The relation between average zeropoint energy and particle spin is given by the expression [26]

$$\langle E_0 \rangle - \omega_0 \langle \mathbf{s} \rangle = 0. \tag{26}$$

Let us denote $\omega_s = 2\omega_0$ and write the angular velocity bivector as $\Omega_s = -\mathbf{i}\,\sigma_s\omega_s$, where σ_s is a unit vector along the direction of spin. The average value of spin is obtained by taking the average over a half cycle, $\langle \mathbf{s} \rangle = \frac{2}{\pi}\mathbf{s}$. Substituting this average value of spin and $\langle E_0 \rangle$ from (25) in (26) gives the relation between particle mass and spin

$$mc^2 = \sigma_s \Omega_s . S, \tag{27}$$

where the relation $\hbar\omega_0 = mc^2$ is used and the bivector spin $S = \mathbf{i}\,\sigma_s\hbar/2$. The unit vector σ_s acting on an idempotent $\mathcal{J}_+ = (1 + \sigma_S)/2$ gives an eigenvalue +1. This statement is represented by an equation $\sigma_s\mathcal{J}_+ = +1\mathcal{J}_+$. When (27) is multiplied from right by an idempotent $\mathcal{J}_+$ on both sides the unit vector is absorbed by the idempotent and equating the scalar parts gives

$$mc^2 = \Omega_s . S. \tag{28}$$

Thus the mass of the particle turns out to be the local internal rotational energy given by the term $\Omega_s . S$. Since, the magnitude of spin and velocity of light are constants, the value of particle mass depends on the frequency of spin rotation and the different particles may have different frequencies of spin rotation. The above analysis shows that the internal complex rotation is responsible for the existence of particle mass. Then, one may initially consider a massless charged particle and it may acquire mass from zeropoint field through a local complex rotation.

When the particle is observed from an arbitrary frame of reference, the center of mass point moves with velocity $\mathbf{v}$. The equation of motion of centre of mass point is given by (1) and solving it by assuming the radiation damping and binding terms as small when compared to the force term, one can obtain the zeropoint energy absorbed by the point particle and it is given by (3). The cut-off frequency ω_c is the limiting frequency in the integration. When we assume the cut-off

frequency $\omega_c = \omega_0$ [37, 38] and after introducing the convergence factor $\eta(\omega_0) \sim 3/4$ in (3), the average energy representing the mass correction of the particle in the zeropoint field can be expressed as

$$\delta m = \frac{\alpha}{2\pi} mc^2. \quad (29)$$

This mass correction is too small and found to be similar to the expression found in quantum electrodynamics to the first order in the fine structure constant α.

6 Conclusions

In the stochastic electrodynamics with spin, it has been shown that the average zeropoint energy absorbed by the particle due to its internal motion gives the particle mass. When the particle center of mass point moves with certain velocity, we find the average energy absorbed by the particle gives the mass correction. In deriving both particle mass and mass correction, a convergence factor has been introduced for an extended particle. To understand the mechanism of mass generation of an elementary particle, one may initially assume a massless charged particle with complex structure and such a particle can be visualized as an oscillator in the fluctuating zeropoint field. Then the average energy absorbed by the oscillator refers to the particle mass. Finally, we conceive the idea that an elementary particle acquires mass from the interaction of ubiquitous zeropoint field.

Submitted on March 18, 2016 / Accepted on March 21, 2016

References

1. Weyssenhoff J., Raabbe A. Relativistic dynamics of spin fluids and spin particles. *Acta. Phys. Pol.*, 1947, v. 9, 7.
2. Barut A. O., Zanghi A. J. Classical model of the Dirac electron. *Phys. Rev. Lett.*, 1984, v. 52, 2009–2012.
3. Salesi G., Recami E. A velocity field and operator for spinning particles in (nonrelativistic) quantum mechanics. *Found. Phys.*, 1998, v. 28, 763–773.
4. Pavsic M., Recami E., Rodrigues W. A., Maccarrone G. D., Raciti F., Saleci G. Spin and electron structure. *Phys. Lett. B.*, 1993, v. 318, 481.
5. van Holten J. W. On the electrodynamics of spinning particles. *Nuclear Phys. B.*, 1991, v. 356, 3–26.
6. Hestenes D. Zitterbewegung in quantum mechanics. *Found. Phys.*, 2010, v. 40, 1–54.
7. Muralidhar K. Classical origin of quantum spin. *Apeiron*, 2011, v. 18, 146.
8. Muralidhar K. The spin bivector and zeropoint energy in geometric algebra. *Adv. Studies Theor. Phys.*, 2012, v. 6, 675–686.
9. Haisch B., Rueda A., Dobyns Y. Inertial mass and quantum vacuum fields. *Ann der Physik.*, 2001, v. 10, 393–414.
10. Sidharth B. G. Revisiting zitterbewegung. *Intl. J. Theor. Phys.*, 2009, v. 48, 497–506.
11. Milonni P. W. The Quantum Vacuum: An Introduction to Quantum Electrodynamics. Academic Press, Boston, 1994.
12. Boyer T. H. The classical Vacuum. *Sci. Am.*, 1985, v. 253, 70.
13. de La Pena L., Cetto A. M. The Quantum Dice – An Introduction to Stochastic Electrodynamics. Kluwer Academic Publishers, Dordrecht, 1996.
14. Marshall T. W. Random electrodynamics. *Proc. Roy. Soc. A.*, 1963, v. 276, 475–491.
15. Boyer T. H. Random electrodynamics – The theory of classical electrodynamics with classical electromagnetic zero point radiation. *Phys. Rev. D.*, 1975, v. 11, 790.
16. Boyer T. H. Unretarded London-van der Waals forces derived from classical electrodynamics with classical electromagnetic zeropoint radiation. *Phys. Rev.*, 1972, v. 6, 314.
17. Boyer T. H. Quantum zeropoint energy and long range forces. *Annals of Phys.*, 1970, v. 56, 474.
18. Boyer T. H. Connection between the adiabatic hypothesis of old quantum theory and classical electrodynamics with classical electromagnetic zero-point radiation. *Phys. Rev. A.*, 1978, v. 18, 1238.
19. Rueda A. Behaviour of classical particles immersed in electromagnetic zero-pint field. *Phys. Rev. A.*, 1981, v. 23, 2020.
20. Cavalleri G., Barbero F., Bertazzi G., Cesaroni E., Tonni E., Bosi L., Spavieri G., Gillies G. T. A qualitative assessment of stochastic electrodynamics with spin (SEDS): Physical principles and novel applications. *Front. Phys. China.*, 2010, v. 5, 107–122.
21. de la Pena L., Cetto A. M., Hernandez A. V. The Emerging Quantum: The Physics Behind Quantum Mechanics. Springer, Cham, 2015.
22. Lorentz H. A. The Theory of Electrons and its Applications to the Phenomena of Light and Radiant Heat. G. E. Stechert and Co., New York, 1916.
23. Haitch B., Reuda A., Puthoff H. E. Inertia as a zero point field Lorentz force. *Phys. Rev. A.*, 1994, v. 49, 678–894.
24. Cavalleri G. $\hbar$ derived from cosmology and origin of special relativity and QM. *Nuovo Cimento. B.*, 1997, v. 112, 1193–1205.
25. Hestenes D. Oersted Medal Lecture 2002: Reforming the Mathematical Language of Physics. *Am. J. Phys.*, 2003, v. 71, 104–121.
26. Muralidhar K. Complex vector formalism of harmonic oscillator in geometric algebra: Particle mass, spin and dynamics in complex vector space. *Found. Phys.*, 2014, v. 44, 265–295.
27. Jammer M. Concepts of Mass in Contemporary Physics and Philosophy. Princeton University Press, Princeton, New Jersey, 2000.
28. Wentzel G. Quantum Theory of Fields. Dover Publications Inc., Mineola, New York, 2003.
29. Modanese G. Inertial mass and vacuum fluctuations in quantum field theory. *Found. Phys. Lett.*, 2003, v. 16, 135–141.
30. Pollock M. D. On vacuum fluctuations and particle masses. *Found. Phys.*, 2012, v. 42, 1300–1338.
31. Ibison M. Massless classical electrodynamics. *Fizika. A.*, 2003, v. 12, 55–74.
32. Ibison M. A Massless Classical Electron. In: Chubykalo A., Espinoza A., Smirnov-Rueda R., Onoochin V., eds. Has the Last Word Been Said on Classical Electrodynamics? – New Horizons. Rinton Press, New Jersey, 2003.
33. Doran C., Lasenby A. Geometric Algebra for Physicists. Cambridge University Press, Cambridge, 2003.
34. Rohrlich F. Classical Charged Particles. World Scientific, Singapore, 2007.
35. Yaghjian A. D. Relativistic Dynamics of a Charged Sphere. Springer, New York, 2006.
36. Puthoff H. C. Gravity as zeropoint fluctuation force. *Phys. Rev. A.*, 1989, v. 39, 2333–2342.
37. Masperi L. Qualitative description of electron's anomalous magnetic moment. *Revista Brasillera de Fisica.*, 1989, v. 19, 215.
38. Bjorken J. D., Drell S. D. Relativistic Quantum Mechanics. McGraw-Hill, New York, 1964.

The Relationship Between the Possibility of a Hidden Variable in Time and the Uncertainty Principle

Eyal Brodet
E-mail: eyalbrodet@hotmail.com

In this paper we will discuss the relationship between the possibility of a hidden variable in time and the uncertainty principle. The discussion consists in a fundamental look at the decay time processes of unstable elementary particles. As will be argued, the hidden variable in time possibility may result in a possible way to bypass the energy-time uncertainty principle. Therefore energy and time information may be known simultaneously in the decay time process. A fundamental and general experimental way to test the above is suggested.

1 Introduction

In this paper we investigate the connection between the possibility of a hidden variable in time and the uncertainty principle. In [1] the process of $e^+e^- \to \mu^+\mu^-$ was discussed and it was questioned how come under what appears to be identical local initial conditions we get a distribution of decay time values for the μ^+ and the μ^-. In[1] there was no discussion about the muon mass width that in fact means that the muons are in principle not completely identical to each other and therefore the local initial conditions are not completely identical between the different events. In [1] it was assumed that the muon mass width could not explain (at least not by itself) the exponential decay time distribution of the muons. Therefore the suggestion was that there exists another internal property within the muons that is responsible for generating the muon decay time distribution.

However this could not be the complete explanation as the muons do have a narrow distribution of mass values and therefore there are slightly different local initial conditions in this process between different events. This fact has to be taken into account in a complete explanation for the decay time distribution in this process. As the muon mass width is part of the uncertainty principle, in this paper we will discuss the connection of the uncertainty principle to the hidden variable in time possibility and attempt to incorporate the two. Moreover we will discuss how the existence of a hidden variable in time could help to bypass the energy-time aspect of the uncertainty principle. Finally an experimental way to test the above is discussed.

This paper is organized as follows. Section 2 discusses the theoretical background. Section 3 describes a possible experimental way to bypass the energy-time uncertainty principle in case a hidden variable in time do exist. The conclusions are presented in section 4.

2 Theoretical discussion

2.1 Background

In this paper we investigate the connection between the possibility of a hidden variable in time and the uncertainty principle. The hidden variable in time possibility first presented in [1] gives a deterministic approach that attempts to explain the distribution in decay time as a result of a compatible distribution in an additional internal property within the particles. The suggestion was that this additional internal property is related to the frequency of the virtual boson emission and absorption and therefore as it is related to time and affects the decay time of particles it was termed a hidden variable in time (f_r). However even if the above is correct this could not be a complete explanation as we have to take into account a known distribution in the initial decaying particles which is the distribution in their mass.

In [1] it was assumed that this mass distribution of for example the muon particles can not solely explain the muon exponential decay time distribution. More specifically it was assumed that the Breit-Wigner distribution of the mass value could not be translated in a deterministic, unique and logical way, using the Standard Model, into the exponential decay time distribution that we observe. One could convince oneself intuitively that this is the case by considering the peaks of the two distributions which are at $m = M_{mean}$ for the Breit-Wigner case and $t = 0$ for the exponential decay case and also the tails which are for the mass Breit-Wigner case at $m = 0$, and at $m = \infty$ (two tails) and for decay time case at $t = \infty$.

Therefore one can not get a logical connection between the two distributions because one could not associate the process initial and final condition logically considering what we know from the Standard Model. That is if we start from the two peaks as the most common and popular initial conditions where most of the events are then we get two different initial conditions for the mass value at the tails ($m = 0,\ m = \infty$) that give a single final condition which is the tail of the exponential at $t = \infty$.

This does not give a logical and deterministic explanation as logically under different initial conditions we should get different final conditions considering the dependence of the decay time on the mass as described in the Standard Model and the experimental decay time results, i.e the higher the mass is the shorter the decay time is. Therefore the mass distribution could not generate deterministically and logically

using the Standard Model, the observed muon decay time distribution, and we need an alternative explanation. Perhaps in the form of the hidden variable f_r.

The standard model does however, give a general link between the Breit-Wigner shape and the exponential decay time shape for a given particle which narrows down the uncertainty by telling us the favorite mass value of the particles is M_{mean} and that an enhanced fraction of them will decay almost instantly after they are born. For example if 40% of the masses are at a bin around M_{mean} then $40 * 40 = 16\%$ of them will decay in the first decay time bin in the decay time distribution.

That is if we measure a specific mass to be on the mean value then we know that there is a 16% probability that the particle would decay in the first decay time bin. This is compared to what we know from the uncertainty principle, where knowing exactly the mass value yields a complete uncertainty on the decay time. Therefore the Standard Model reduces the uncertainty with respect to the uncertainty principle by allowing us to calculate the Breit-Wigner and the exponential decay time distributions.

2.2 $\Delta m, \Delta t$ and f_r

The distribution of Δm is known from the Breit-Wigner but the distribution of Δt is experimentally unknown (we do not know how to deduce it from the exponential decay distribution), we only know the maximum value of it from the knowledge of Δm and the boundaries given by the uncertainty principle.

According to [1], if one knows the true particle decay $T1$, time then one may know f_{r1} from Fig. 1. In this case this particular f_{r1} has two possible mass value $M1, M2$ as shown in Fig. 2. These two mass values may have the same value of f_{r1} but with very different mean lifetimes given for example in the muon case, from the known Standard Model formula:

$$\tau_{\mu(1,2)} = \frac{192\,\pi^3 h^7}{G_f M(1,2)_\mu^5\, c^4}\,. \qquad (1)$$

where G_f is the Fermi coupling constant. In the case when f_r exists, one may have a deterministic link between $\Delta m, \Delta t, f_r$ and t which may cancel the uncertainty limitations as will be discussed later on. Without f_r there is no deterministic link between a specific mass $M1$ and a specific decay time $T1$.

If f_r exists the exponential shape is the slope of the f_r depending on the mean lifetime which gives a deterministic description for a specific event using extra information in the form of f_r.

2.3 Mathematical relationship between f_r and the exponential and Breit-Wigner distributions

Putting the above into a mathematical form gives us two expressions for f_r:

$$f_{r(i)} = f(m_i)\,A_i = f(m_i)\,\exp\left(-\frac{t_i}{\tau_i}\right) \qquad (2)$$

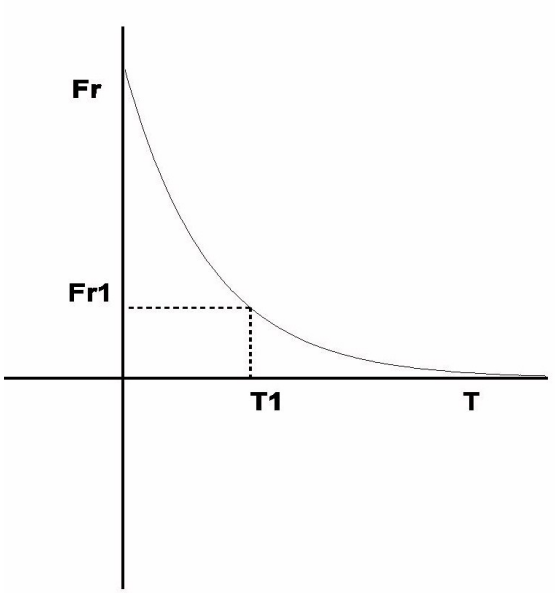

Fig. 1: f_r versus the distribution of the particles decay time.

$$f_{r(i)} = f(m_i)\,f(E_i) = f(m_i)\,\frac{k_1}{(E_i^2 - M_{mean}^2)^2 + M_{mean}^2\,\Gamma_i^2} \qquad (3)$$

where $f(m_i)$ is the mass amplitude and $E_i = m_i\,c^2$. One possibility for $f(m_i)$ may be: $f(m_i) = M_{mean} + (m_i - M_{mean})\,k_2$ where k_1 and k_2 are parameters with yet unknown values. The above relationship suggests the following:

For mass measured close to the muon mean mass values Δm is small and Δt is large (but we cannot be sure at this stage how large as we do not know how to deduce the uncertainty Δt from the exponential muon decay time distribution). We can only get the maximum value for Δt for small Δm using the uncertainty principle. From Fig. 2 we can see that small Δm values correspond to high f_r values and therefore, as can be seen from Fig. 1 to short decay time values.

For mass measured far from the mean muon mass value (lower or bigger), we know that Δm is large and Δt is small. Again we cannot be sure for a particular event how small is Δt, however we only know that it has to be small as Δm is large in order to satisfy the uncertainty principle. From Fig. 2 we can see that large Δm corresponds to low f_r values and therefore, as can be seen from Fig. 1, to long decay time values.

Therefore the effect of the uncertainty principle assuming the existence of f_r, on the decay process is that it associates a particular and different uncertainty on each decay time and mass values. This is where masses around M_{mean} are assumed to have short decay times and have small Δm and large Δt, and masses that are away from M_{mean} (smaller or greater) are assumed to have larger decay times and larger Δm and smaller Δt.

This is where the limitation associated by the uncertainty principle of knowing simultaneously the exact mass and de-

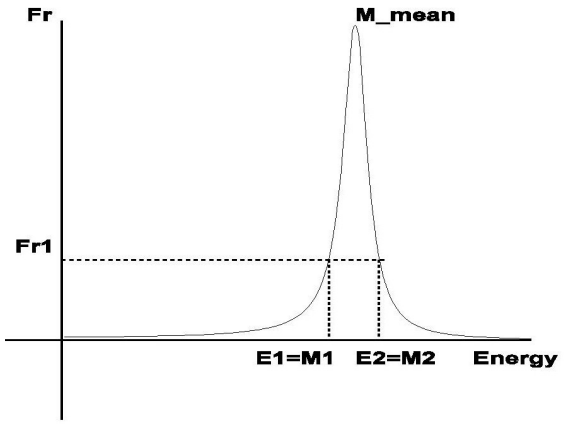

Fig. 2: f_r versus the distributions in the particles mass.

cay time of a particle still remains. In the next section we will discuss how this limitation could be bypassed.

3 Possible bypass of the energy-time uncertainty principle

The possibility of a hidden variable in time opens up a new way to fundamentally bypass the above limitation on the simultaneous knowledge of the muon mass and decay time values. This could be expressed by the following measurement that may be done by known detectors [2]:

As measuring the muon mass exactly is experimentally difficult due to the missing energy of the neutrino involved, we may turn to measuring exactly its decay time.

Therefore, if we measure a specific muon decay time $t1$, we know its f_r from Fig. 1. Therefore we could know its two associated masses $m1$ and $m2$ from Fig. 2 and its mass uncertainty Δm. From the uncertainty principle we could then also know its maximum Δt uncertainty. Therefore this gives us three possible decay time $t1$, $t1 - \Delta t$, $t1 + \Delta t$ and six possible masses that are associated to these three decay times. Now we need to decide which pair of decay time and mass values is the correct one for that particular event. We can attempt to do that by measuring exactly the muon electric charge Q in that event from:

$$\frac{M V^2}{r} = QVB \qquad (4)$$

where B is the external magnetic field, M is the muon mass and we can measure the momentum from the curvature r and the velocity V from the Cherenkov detector. Therefore after knowing the charge we may deduce the factor $A = \exp(-t/\tau)$ according to [3]. This A value corresponds to a part of the particles f_r in that particular event where $f_{r(i)} = f(m_i) A_i$ as shown in (2). Now all we need to do is to see which pair of mass and decay time values is closest in value to the measurement of the factor A and find the correct initial mass and final decay time in that particular event, thereby bypassing the uncertainty principle.

4 Conclusion

The implication of a hidden variable in time on the energy-time uncertainty principle was discussed. A fundamental way was presented to bypass the uncertainty principle through measuring the decay time and charge value in a specific $e^+e^- \rightarrow \mu^+\mu^-$ event, thereby knowing the exact value of the initial muon mass and decay time.

Submitted on April 2, 2016 / Accepted on April 6, 2016

References

1. Brodet E. The possibility of a hidden variable in time. *Physics Essays*, 2010, v. 23 (4), 613–617.
2. The DELPHI detector at LEP. http://delphiwww.cern.ch/. Accessed Dec 2015.
3. Brodet E. Incorporating the possibility of a hidden variable in time in the Standard Model and the experimental implications. *Physics Essays*, 2016, v. 29 (1), 57–61.

A Modern Interpretation of the Dirac-Electron Continuity Equation

William C. Daywitt

National Institute for Standards and Technology (retired), Boulder, Colorado. E-mail: wcdaywitt@me.com

This paper re-derives the Dirac continuity equation for the electron from the viewpoint of the Planck vacuum (PV) theory. Results show the equation to be a spacetime equation (whose line elements are cdt and dx^k) that equates the normalized ct-gradient of the probability density ($\psi^\dagger\psi$) to the normalized negative divergence of the quantity ($\psi^\dagger\boldsymbol{\alpha}\psi$).

1 Introduction

The Dirac equation that defines the free-electron spinor field $\psi = \psi(\mathbf{r}, t)$ [1, p.74]

$$ic\hbar\frac{\partial\psi}{c\partial t} = (c\boldsymbol{\alpha}\cdot\widehat{\mathbf{p}} + mc^2\beta)\psi \tag{1}$$

where $\widehat{\mathbf{p}}$ ($=-i\hbar\nabla$) is the vector momentum operator, can be expressed as

$$ic\hbar\left(\frac{\partial}{c\partial t} + \boldsymbol{\alpha}\cdot\nabla\right)\psi = mc^2\beta\psi \tag{2}$$

where c is the speed of light, $\hbar$ is the reduced Planck constant, and m is the electron mass. The spinor field ψ is the 4×1 column vector

$$\psi = \begin{pmatrix}\psi_1\\ \psi_2\\ \psi_3\\ \psi_4\end{pmatrix}. \tag{3}$$

The two 4x4 matrices in (1) and (2) are defined by

$$\alpha_k = \begin{pmatrix}0 & \sigma_k\\ \sigma_k & 0\end{pmatrix} \quad\text{and}\quad \beta = \begin{pmatrix}I & 0\\ 0 & -I\end{pmatrix} \tag{4}$$

where $k = (1, 2, 3)$ and I is the 2×2 unit matrix. The three 2×2 Pauli spin matrices σ_k are [1, p. 12]

$$\sigma_1 = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\quad \sigma_2 = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix},\quad \sigma_3 = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \tag{5}$$

and the operator on the left side of (2) reduces to

$$\left(\frac{\partial}{c\partial t} + \boldsymbol{\alpha}\cdot\nabla\right) = \left(\frac{\partial}{c\partial t} + \sum_{k=1}^{3}\alpha_k\frac{\partial}{\partial x^k}\right). \tag{6}$$

In its rest frame the massive electron core $(-e_*, m)$, with its zero-point derived mass m [2], exerts the two-term coupling force [3, Sec. 7-8]

$$F(r) = \frac{e_*^2}{r^2} - \frac{mc^2}{r} = \frac{(-e_*)(-e_*)}{r^2} - \frac{mm_*G}{r_*r} \tag{7}$$

on the PV quasi-continuum, where e_* is the massless bare charge and $G\,(=\,e_*^2/m_*^2)$ is Newton's gravitational constant. The first $(-e_*)$ in (7) belongs to the electron and the second to the separate Planck particles making up the degenerate PV state. The two terms in (7) represent respectively the Coulomb repulsion between the electron charge and the separate PV charges, and the second their mutual gravitational attraction.

The particle/PV coupling force (7) vanishes at the electron Compton radius $r_c\,(=\,e_*^2/mc^2)$. In addition, the vanishing of $F(r_c)$ is a Lorentz invariant constant [4] that leads to the important Compton-(de Broglie) relations

$$r_c\cdot mc^2 = r_d\cdot cp = r_L\cdot E = r_*\cdot m_*c^2 = e_*^2 \quad (=c\hbar) \tag{8}$$

where $r_d = r_c/\beta_0\gamma_0$ and $r_L = r_c/\gamma_0$, and $r_*\,(=\,e_*^2/m_*c^2)$ and m_* are the Compton radius and mass of the Planck particles within the PV state. The ratio of the electron speed v to the speed of light c is β_0 and $\gamma_0 = 1/(1-\beta_0^2)^{1/2}$. The relativistic momentum and energy following from the invariance of $F(r_c) = 0$ are $p\,(=\,m\gamma_0 v)$ and $E\,(=\,m\gamma_0c^2)$, from which $E = (m^2c^4 + c^2p^2)^{1/2}$ is the relativistically important energy-momentum relationship.

Using (8), (2) can be expressed as

$$ie_*^2\left(\frac{\partial}{c\partial t} + \boldsymbol{\alpha}\cdot\nabla\right)\psi = mc^2\beta\psi \tag{9}$$

or

$$ir_c\left(\frac{\partial}{c\partial t} + \boldsymbol{\alpha}\cdot\nabla\right)\psi = \beta\psi \tag{10}$$

where the partial derivatives within the parentheses are normalized by the Compton radius r_c. The spinor field that is the hermitian conjugate of ψ is the 1×4 row vector $\psi^\dagger = (\psi_1^\dagger, \psi_2^\dagger, \psi_3^\dagger, \psi_4^\dagger)$. Then, pre-multiplying (10) by $\psi^\dagger$ leads to

$$ir_c\psi^\dagger\left(\frac{\partial}{c\partial t} + \boldsymbol{\alpha}\cdot\nabla\right)\psi = \psi^\dagger\beta\psi\,. \tag{11}$$

Taking the hermitian conjugate of (10), post-multiplying by ψ, then yields [1, p. 76]

$$-ir_c\left(\frac{\partial}{c\partial t} + \boldsymbol{\alpha}\cdot\nabla\right)\psi^\dagger\psi = \psi^\dagger\beta\psi\,. \tag{12}$$

Subtracting (12) from (11) finally leads to the continuity equations [1, p. 76]

$$ir_c\left[\frac{\partial(\psi^\dagger\psi)}{c\partial t} + \nabla\cdot(\psi^\dagger\boldsymbol{\alpha}\psi)\right] = 0 \tag{13}$$

or

$$\frac{\partial(\psi^\dagger\psi)}{c\partial t/r_c} + \sum_{k=1}^{3} \frac{\partial(\psi^\dagger\alpha_k\psi)}{\partial x^k/r_c} = 0 \qquad (14)$$

for the electron. From (8), the presence of r_c in these two equations connects the electron core dynamics to a wave traveling within the vacuum state [5].

2 Comments and Conclusions

Dividing (13) by ir_c yields the equation

$$\frac{\partial(\psi^\dagger\psi)}{\partial t} + \nabla\cdot(\psi^\dagger c\boldsymbol{\alpha}\psi) = 0 \qquad (15)$$

where the 4×4 matrix $c\boldsymbol{\alpha}$ looks like a velocity operator because of the speed of light c. This observation then leads intuitively to the standard continuity equation [1, p. 76]

$$\frac{\partial\rho}{\partial t} + \nabla\cdot\mathbf{j} = 0 \qquad (16)$$

where $\rho = \psi^\dagger\psi$ is the probability density and $j^k = \psi^\dagger c\alpha_k\psi$ is the kth component of the probability current density. Integrating (16) over the volume V (assumed to contain the electron core $(-e_*, m)$), and using the divergence theorem, leads to [1, p. 77]

$$\frac{\partial}{\partial t}\int_V d\rho\, d^3x + \int_S \mathbf{j}\cdot d\vec{S} = 0\,. \qquad (17)$$

where the surface S surrounds the volume V.

So far, so good. But there is a problem: treating $c\boldsymbol{\alpha}$ as a free-space matrix velocity leads to a tortured interpretation of that operator that cries out for a better explanation. From the PV perspective, that explanation is apparent from equation (14)

$$\frac{\partial(\psi^\dagger\psi)}{c\partial t/r_c} + \sum_{k=1}^{3} \frac{\partial(\psi^\dagger\alpha_k\psi)}{\partial x^k/r_c} = 0$$

where the Minkowski-like line elements, cdt and dx^k associated with the partial derivatives, are normalized by the electron Compton radius r_c. The form of this equation suggests that it is associated with a distorted spacetime [6, p. 27] (the distortion coming from the r_c and the α_k), rather than a free-space velocity dynamic. Furthermore, the absence of the dynamical electron parameters p and E from (8), and the fact that $c\boldsymbol{\alpha}$ is not a recognizable free-space operator, suggest that (14) refers to a PV substructure dynamic [7] (driven by the electron core dynamic), where the normalized ct-gradient of $(\psi^\dagger\psi)$ equals the normalized negative divergence of $(\psi^\dagger\boldsymbol{\alpha}\psi)$.

Finally, the assumption that the PV is a degenerate state implies that the Planck-particle energy eigenstates are full. So if there is a current wave propagating within the PV, it cannot involve a Planck particle current (because the Planck particles are not free to move macroscopically). Thus $c\boldsymbol{\alpha}$ must refer, in part, to a localized percussion-like spinor wave within that vacuum state, analogous to a wave traveling on the surface of a kettle drum.

Equations (13) and (14) and the previous two paragraphs represent the PV view of the Dirac-electron continuity equation.

Submitted on April 15, 2016 / Accepted on April 30, 2016

References

1. Gingrich D.M. Practical Quantum Electrodynamics. CRC, The Taylor & Francis Group, Boca Raton, London, New York, 2006.
2. Daywitt W.C. Why the Proton is Smaller and Heavier than the Electron. *Progress in Physics*, 2014, v. 10, 175.
3. Daywitt W.C. The Trouble with the Equations of Modern Fundamental Physics. *American Journal of Modern Physics, Special Issue: Physics Without Higgs and Without Supersymmetry*, 2016, v. 5, no. 1-1, 22.
4. Daywitt W.C. The de Broglie Relations Derived from the Electron and Proton Coupling to the Planck Vacuum State. *Progress in Physics*, 2015, v. 11(2), 189.
5. Daywitt W.C. The Dirac-Electron Vacuum Wave. *Progress in Physics*, 2016, v. 12(3), 222.
6. Leighton R.B. Principles of Modern Physics. McGraw-Hill Book Co., New York, Toranto, London, 1959.
7. Daywitt W.C. The Strong and Weak Forces and their Relationship to the Dirac Particles and the Vacuum State. *Progress in Physics*, 2015, v. 11(1), 18.

Optics of the Event Horizon Telescope

Trevor W. Marshall
Buckingham Centre for Astrobiology, The University of Buckingham, Buckingham MK18 1EG, UK
E-mail: trevnat@talktalk.net

It is suggested in this article that part of the signal in the 1.3 mm range from Sagittarius A* originates inside the central collapsar, rather than coming entirely from its accretion disc. The suggestion has its origin in the discovery that the classic article of Oppenheimer and Snyder contains a basic error in its assertion that the light, from a collapsing object lying entirely within its own photonsphere, is progressively cut off as the object shrinks towards its gravitational radius, where a large part of the Oppenheimer-Snyder collapsar's material is concentrated. The signal from the collapsar has certain features which may make it possible to distinguish its image from that of the accretion disc.

1 Introduction

At the centre of our galaxy, 8 kp distant from us, there is an object named Sagittarius A* whose mass is 4.1 megasuns. It is popularly classified as a black hole, with a spherical* region of radius 1.2×10^7 km around it bounded by an "event horizon"; according to black-hole theory no light from Sagittarius A* can cross this horizon.

In two recent articles [1, 2] it was shown that there is a solution of the field equations of General Relativity for such a supermassive object, which has no singularity at $r = 0$, and which allows light signals to cross the horizon. The latter property of the solution was demonstrated for the case of rays which are normal to the event horizon, and the present article demonstrates that it may be extended to all orientations. In addition we consider the range of angles for which light originating at the surface of such a collapsar crosses the photonsphere, at 1.5 times the gravitational radius, and consequently may reach a terrestrial telescope. There is currently a project called the Event-Horizon Telescope [3] (EHT) designed to look at the signal from the neighbourhood of Sagittarius A* in the 1.3 mm range.

Central to the widespread belief in the validity of black-hole theory is the article of Oppenheimer and Snyder (OS) [4]. This reported, without giving details, an investigation of the light signal from a supermassive object, arriving at the following conclusion

> All energy from the surface of the star will be reduced very much in escaping ... by the gravitational deflection of light which will prevent the escape of radiation except through a cone about the outward normal of progressively shrinking aperture as the star contracts. The star thus tends to close itself off from any communication with a distant observer.

The property of the OS metric claimed by Penrose, which he needed as a prerequisite for his singularity theorem [5], was the stronger one known as the *trapped surface*. The publications cited above show that neither of these properties in fact holds for the OS metric.

In the following two sections we shall use precisely the OS metric to show that the progressively shrinking aperture of the emission cone has no effect on the size of the image of the collapsing object, and only a marginal effect on its total luminosity. This result leads us to suggest that the signal from Sagittarius A* comes partly from the surface of the collapsar itself, and not entirely from the accretion disc, as is assumed in most current analyses. The accretion disc may well have a substantially higher temperature than the collapsar, but that is probably offset by the vastly greater area of the latter. Note also that the millimetre range of wavelength investigated by the EHT corresponds to the maximum of a Planck spectrum of just a few degrees Kelvin; to support our analysis, the collapsar must retain only the merest relic of its thermal energy.

The OS article reached another conclusion, stated in their Abstract, namely

> ... an external observer sees the star asymptotically shrinking to its gravitational radius.

This result contradicts directly Penrose's description of the OS results and was verified by me in [1]. The point is that OS showed that there is a common system of coordinates applicable to both the exterior and interior of the collapsar. My article [1] demonstrated that the density distribution of the OS "dust cloud" becomes concentrated near the surface as it shrinks to the gravitational radius; no exotic process like the modern black-hole one of "spaghettification" [6] occurs when a notional spaceship crosses the event horizon. OS should be considered responsible for the notion that further shrinkage occurs within the gravitational radius only in so far as they gave their article the misleading title "On continued gravitational contraction".

It should be noted that in the exterior, and hence in what should now be recognized as the universal, time frame the collapsar's shrinkage to the gravitational radius takes an infinite lapse of time. We shall show in the following section that in the limit there is an underlying infinite red shift, which

*For the purposes of this article we ignore its spin.

causes not only the surface itself, but also all light signals approaching it, to be infinitely slowed down. This is the real significance of the event horizon, but it is my contention that a real collapsar, with an internal pressure resulting from the intervention of forces other than gravitational, stops shrinking before it reaches the gravitational radius. For example, we have investigated [7] a collapsar whose equation of state is an idealized form of neutron fluid*, and for which, above a certain mass, its maximum density lies between the event horizon and the photonsphere.

2 The exterior light orbits

Darwin [9, 10] described the null geodesics of the Schwarzschild metric

$$ds^2 = \frac{r-2m}{r}\,dt^2 - \frac{r}{r-2m}\,dr^2 - r^2 d\theta^2 - r^2\sin^2\theta\, d\phi^2, \quad (1)$$

where $2m$ is the gravitational radius. He extended the standard theory of light deflection, his method being equivalent to minimising the action integral for a light ray with small impact parameter starting from infinity; for orbits in the plane $\theta = \pi/2$,

$$\delta\int L d\phi = 0, \quad (2)$$

with the lagrangian

$$L = \left[\frac{r-2m}{r}\,t'^2 - \frac{r}{r-2m}\,r'^2 - r^2\right]^{1/2}, \quad (3)$$

where a prime denotes differentiation with respect to ϕ. The Lagrange equation for the cyclic coordinate t is

$$\left[\frac{d}{d\phi} - \frac{L'}{L}\right]\frac{r-2m}{r}\,t' = 0. \quad (4)$$

The corresponding conservation integral for ϕ enables us to put $L'/L = 2r'/r$, so we obtain

$$t' = \frac{r^3}{p(r-2m)}, \quad (5)$$

the constant p being the impact parameter

$$p = \lim_{r\to\infty} r^2\frac{d\phi}{dt}. \quad (6)$$

The ray orbit is then obtained by substituting for t' and then putting $L = 0$, that is

$$r'^2 = \frac{r^4}{p^2} - r^2 + 2mr. \quad (7)$$

Darwin deduced that a ray with impact parameter p greater than $3m\sqrt{3}$ returns to $r = \infty$; the deflection angle may be many multiples of 2π as p approaches $3m\sqrt{3}$, and in the limiting case $p = 3m\sqrt{3}$ the ray circles indefinitely at $r = 3m$, which is nowadays called the *photonsphere*. For p less than this, the ray is captured, and it goes to what Darwin termed the "barrier", nowadays called the *event horizon*, at $r = 2m$. He also repeated the point previously made by OS, that the journey from $r = 3m$ to $r = 2m$ takes an infinite time. When the collapse is incomplete, the surface being at $r = r_1 > 2m$, a ray arrives there making an angle with the normal of

$$\xi = \tan^{-1}\left[\left(\frac{r_1^2}{p^2} - 1 + \frac{2m}{r_1}\right)^{-1/2}\right], \quad (8)$$

and in the limiting case $r_1 = 2m$ this becomes

$$\xi = \tan^{-1}\left(\frac{p}{2m}\right). \quad (9)$$

We may deduce directly the orbits of rays exiting from the barrier; those falling within a cone of semiangle $\tan^{-1}(3\sqrt{3}/2) = 68.9$ degrees go to our telescope at "infinity", forming an image of parallax $6m\sqrt{3}$. Any collapsar with $2m < r_1 < 3m$ has this same parallax, but at $3m$ the cone has opened up fully to 90 degrees. A collapsar bigger than $3m$ has a parallax bigger than $6m\sqrt{3}$, while for much larger collapsars, like white dwarfs of solar mass, light deflection is negligible, and the parallax is then simply twice the surface radius. In Figure 1 a number of rays have been plotted, leaving various points in the surface, when $r_1 = 2.2m$, and going towards our telescope; we note that the rays going to the edge of the image come from points on the "invisible face" of the collapsar.

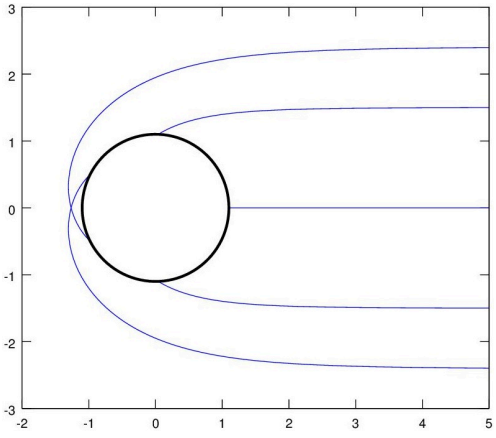

Fig. 1: The light rays issuing from the surface of a collapsar at 1.1 times the gravitational radius, collimated towards a distant telescope. The outer rays are close to the edge of the image, which has a diameter of 5.2 times the gravitational radius; these rays have their sources on what, in the absence of gravitational lensing, would be the invisible part of the surface. The unit of distance is the gravitational radius.

*This model is simply that of Oppenheimer and Volkoff [8] with a different boundary condition at the origin.

For Sagittarius A* the minimum parallax, according to the above analysis, and with the distance of EHT from the galactic centre equal to 2.4×10^{17} km, is 52 arc microseconds, which exceeds the best available current value [3] by about 50 percent. The image profile, that is its intensity $C(p)$ as p goes from zero to $3m\sqrt{3}$, is given by

$$C(p) = \left| \frac{r_1 \sin\phi_0 \cos\xi}{p} \frac{d\phi_0}{dp} \right| , \qquad (10)$$

where ξ is given by (8), that is

$$\cos\xi = \sqrt{\frac{r_1^3 - p^2 r_1 + 2p^2 m}{r_1^3 + 2p^2 m}} , \qquad (11)$$

and ϕ_0 is the angle between the outward normal at the surface and the ray's final direction, that is

$$\phi_0 = \int_{r_1}^{\infty} \frac{p dr}{\sqrt{r^4 - p^2 r^2 + 2mp^2 r}} , \qquad (12)$$

leading to

$$\frac{d\phi_0}{dp} = \int_{r_1}^{\infty} \frac{r^4 dr}{(r^4 - p^2 r^2 + 2mp^2 r)^{3/2}} . \qquad (13)$$

Note that, for $r_1 \gg 2m$, $\xi = \phi_0$, $p = r_1 \sin\phi_0$, and $C(p) = 1/r_1 = \text{const.}$ with ϕ_0 going from $-\pi/2$ to $\pi/2$, giving a uniform circular image of radius r_1; for our case ϕ_0 takes all real values. In Figure 2 the image profile $C(p)$ is plotted. The edge of the image is at $p = 3m\sqrt{3} = 2.598\, r_0$, where r_0 is the gravitational radius. Note that, though $C(p)$ drops to zero at $p = 2.388 r_0$, there is a bright fringe between that value and $p = 2.588\, r_0$; though not shown in the Figure, there is a series of narrower fringes between the latter value and the edge of the image at $p = 2.598\, r_0$. The fringes result from light rays circling close to the photonsphere before finally escaping to reach the telescope, their minima occurring at p-values for which ϕ_0 are integer multiples of π.

A ray which leaves the surface in a direction falling outside the limiting cone, that is with an orbit described by $p > 3m\sqrt{3}$, turns round before reaching the photonsphere, and returns to the barrier after an infinite time.

None of this accords with the OS description, in which the cone closes down to zero at $r = 2m$.

3 The interior light orbits

According to the OS [4] model, the surface of the collapsar completely contracts to the barrier only at $t = \infty$; in the words of that article

> ... an external observer sees the star asymptotically shrinking to its gravitational radius.

Specifically $r_1(t)$ is given by

$$t = -\frac{2}{3}\sqrt{\frac{r_1^3}{2m}} - 2\sqrt{2mr_1} + 2m \ln \frac{\sqrt{r_1} + \sqrt{2m}}{\sqrt{r_1} - \sqrt{2m}} . \qquad (14)$$

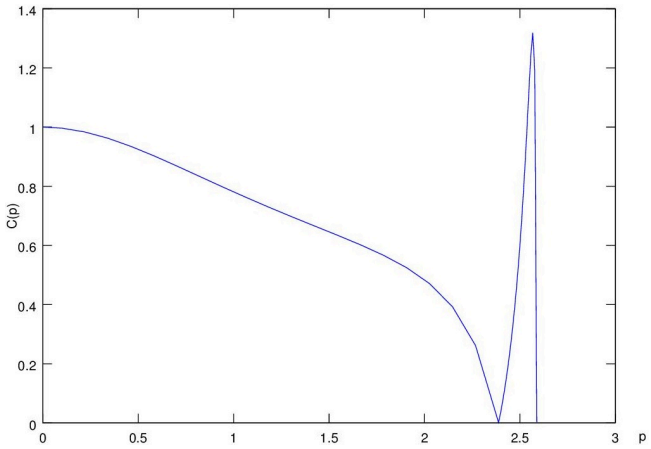

Fig. 2: The image profile $C(p)$ formed by the rays in Figure 1. Again the gravitational radius is the distance unit on the horizontal axis p, and $C(p)$ is normalized to $C(0) = 1$.

For $r < r_1$ the OS metric is

$$ds^2 = \frac{r^3}{2mR^3}\left(\frac{dr}{r} - \frac{dR}{R}\right)^2 - \frac{r^2}{R^2} dR^2 - \\ - r^2 d\theta^2 - r^2 \sin^2\theta\, d\phi^2 , \qquad (15)$$

where the coordinate R lies between 0 and 1, and is related to t in a manner to be determined by matching conditions imposed at the surface.

The interior null geodesics in the plane $\theta = \pi/2$ are constructed from the Lagrangian

$$L = \left[\frac{r^3}{2mR^3}\left(\frac{r'}{r} - \frac{R'}{R}\right)^2 - \frac{r^2}{R^2} R'^2 - r^2\right]^{1/2} . \qquad (16)$$

The Lagrange equations for r and R are, putting $L'/L = 2r'/r$ as in the exterior case,

$$\frac{2rr''}{R^3} - \frac{2r^2R''}{R^4} - \frac{3r'^2}{R^3} - \frac{2rr'R'}{R^4} + \frac{5r^2R'^2}{R^5} + \\ + \frac{4mrR'^2}{R^2} + 4mr = 0 \qquad (17)$$

and

$$\frac{2r^2r''}{R^4} - \frac{2r^3R''}{R^5} - \frac{3rr'^2}{R^4} - \frac{2r^2r'R'}{R^5} + \frac{5r^3R'^2}{R^6} + \\ + \frac{4mr^2R''}{R^2} - \frac{4mr^2R'^2}{R^3} = 0 . \qquad (18)$$

Combining these to eliminate r'', we obtain

$$R'' = R + \frac{2R'^2}{R} , \qquad (19)$$

for which a sufficiently general solution, for $0 < R < 1$, is

$$R = \sin\phi_0 \csc\phi \quad (\phi_0 < \phi < \pi - \phi_0) . \qquad (20)$$

Then, as in the exterior case, we obtain a first order equation for r by substituting this in L and putting $L = 0$, namely

$$r' = \sqrt{\frac{2mr\sin^3\phi_0}{\sin^5\phi}} - r\cot\phi\,, \tag{21}$$

with the solution

$$r = \frac{m\sin\phi_0}{2\sin\phi}\,(A - \sin\phi_0\cot\phi)^2. \tag{22}$$

A ray which arrives at $R = 1$, that is $\phi = \phi_0$, with $r = r_1$ has $A = 2\sqrt{r_1/(2m)} + \cos\phi_0$; the special case $\phi_0 = 0$ was given in eq (14) of [1]. At this point the ray has gradient

$$r' = \csc\phi_0\,\sqrt{2mr_1} - r_1\cot\phi_0\,. \tag{23}$$

4 Matching at the surface

OS [4] matched their metric with the exterior (1) by defining the *cotime* $y(r, R)$ related to t by

$$\frac{t}{2m} = -\frac{2}{3}\,y^{3/2} - 2\sqrt{y} + \ln\frac{\sqrt{y}+1}{\sqrt{y}-1}\,; \tag{24}$$

this they required to satisfy $y(r_1, 1) = r_1/(2m)$. To match the two metric tensors at $R = 1$ they then put

$$y = \frac{r}{2Rm} + \frac{R^2-1}{2}\,. \tag{25}$$

With the metrics matched at the surface, that means the refractive indices are also matched, so the corresponding light rays should join smoothly there. Eq (23) for the interior ray gives, at $r = r_1$,

$$r'^2 + r^2 - 2mr = \left(r_1\csc\phi_0 - \sqrt{2mr_1}\cot\phi_0\right)^2, \tag{26}$$

so the value

$$p = \frac{r_1^2\sin\phi_0}{r_1 - \sqrt{2mr_1}\cos\phi_0} \tag{27}$$

gives a smooth connection between the interior (23) and exterior (7) rays at $r = r_1$. Differentiating (24) and (25), we then find that the values of t' also match at r_1, which confirms that the light speed r'/t' is continuous there.

It may now be seen that, as r_1 approaches $2m$, the speed of light at the surface goes to zero, which generalizes the particular case treated in [1], where the light ray was normal to the surface. Such behaviour may be understood as resulting from the infinite "dust" density there (see below). This behaviour will be modified by the intervention of nongravitational forces; in particular we have studied the effect of the Fermi degeneracy pressure in a neutron star [7], for which the density has a finite maximum well separated from both the surface and $r = 0$. Thus, for a collapsar made of real stellar matter, it makes sense to consider a state of equilibrium whose radius exceeds the gravitational, and for which light leaves the surface with a finite speed; this was the situation depicted in Figure 1.

I add that the matching relation (25) is not unique, though OS stated that it was. In my previous articles [1, 2] the alternative

$$y = \frac{r}{2Rm} - \frac{(1-R)(5-R)}{4} \tag{28}$$

was given. This is part of a wider family of matching relations, and, for this particular choice, has certain advantages in respect of causality.

The infinite surface density of the OS final state may be seen in their calculation of the scalar density ρ, namely

$$\rho = \frac{3R^3}{8\pi r^3}\,. \tag{29}$$

Multiplying this by their three-volume element, we obtain

$$\rho\sqrt{-g}\,dR\,d\theta\,d\phi = \frac{3R^2\sin\theta}{8\pi}\,dR\,d\theta\,d\phi\,, \tag{30}$$

which, in terms of r, gives the density

$$\rho\sqrt{-g}\,dr\,d\theta\,d\phi = \frac{3R^2\sin\theta}{8\pi}\left(\frac{\partial R}{\partial r}\right)_t dr\,d\theta\,d\phi\,. \tag{31}$$

The partial derivative is given, at cotime $y = 1$, by

$$\left(\frac{\partial R}{\partial r}\right)_t = \left(\frac{\partial R}{\partial r}\right)_y = \frac{1}{3m(1-R^2)}\,, \tag{32}$$

giving infinite density at $R = 1$.

Actually we have found that the density in the shell just inside r_1 is very much reduced for a supermassive object like Sagittarius A*, and I propose that the material there is an electron gas with a nearly stationary nucleonic background* which should have broadly similar optical properties to both the OS dust cloud and the neutron star. In these cases the light speed will still be considerably reduced near the surface, but will remain finite.

5 Discussion

The suggestion about the origin of the EHT image of Sagittarius A*, namely that part of the light we receive comes from the collapsar itself, has implications for the direction future observations with the telescope should take. A central problem is to explain the present-day value of the parallax, which is 37μas as opposed to the 52μas we obtained in Section 2. We note that the size of this image is not at all well defined, because of the need to separate the signal from the background noise of nearby objects; this is reflected by the wide error bar in the above parallax. It should be noted also that the image of the accretion disc has almost the same diameter as the one

This entails classifying Sagittarius A as a *supermassive white giant*.

described in Section 2 if the disc is inside the photonsphere, and that its image is larger if it lies outside the photonsphere.

The fringes of the image, described in Section 2, do not seem to have been noticed previously, though they are surely present also in the image of the accretion disc. To distinguish between the two images, arising, as they do, from two superimposed sources of almost the same diameter, will require further analysis along the lines of Section 2; the principal difference is the three-dimensional form of the collapsar, as opposed to the flat, effectively two-dimensional form of the disc. Some progress, both in image enhancement and in theoretical modelling, would help to clarify matters.

The classic article of Oppenheimer and Snyder [4], based in turn on the equally classic one of Tolman [11], was essential for the construction of the matched orbits. In particular these articles (see also [12]) enable us to identify the comoving coordinate R used in Section 3. But the step required to describe fully the orbits of particles of "dust", that is the stellar material, and of light rays near the surface, is the identification of the time coordinate $t(r, R)$ made in our earlier article [1] and in Section 4 of this one.

Acknowledgement

I gratefully acknowledge the help of my colleagues Clifford Chafin, Charles Kinniburgh, Robin Spivey and Max Wallis.

Submitted on May 1, 2016 / Accepted on May 4, 2016

References

1. Marshall T.W. *Astrophys. Space Sci.*, 2012, v.342, 329–332.
2. Marshall T.W. *Progress in Physics*, 2016, v.12, issue 3, 219–221.
3. Doeleman S.S., Weintroub J.W., Rogers A.E.E., Plambeck R., Freund R., Tilanus R., Ziurys L., Moran J., Corey B., Young K.H., Smythe D.L., Titus M., Marrone D.P., Cappallo R.J., Bock D.C.J., Bower G.C., Chamberlin R., Maness H., Niell A.E., Roy A., Strittmatter P., Werhimmer D., Whitney A.R. and Woody D. *Nature*, 2008, v.455, 78.
4. Oppenheimer J.R. and Snyder H. *Phys. Rev.*, 1939, v.56, 455.
5. Penrose R. *Phys. Rev. Letters*, 1965, v.14, 57.
6. Thorne K.S. Black Holes and Time Warps. Norton, New York, 1994, see *Prologue*.
7. Wallis M.K. and Marshall T.W. Energy in General Relativity – the case of the neutron star. *Physical Interpretations of Relativity Theory*, Proceedings of International Meeting, Bauman State Technical University, Moscow, 2015 pp. 544–556 (2015). See also http://tinyurl.com/PIRT-Moscow-15-MarshallWallis
8. Oppenheimer J.R. and Volkoff G.M. *Phys. Rev.*, 1939, v.54, 540.
9. Darwin, Sir Charles. *Proc. Roy. Soc. A*, 1958, v.249, 180–194.
10. Darwin, Sir Charles. *Proc. Roy. Soc. A*, 1961, v.263, 39–50.
11. Tolman R.C., *Proc. Nat. Acad. Sci. (USA)*, 1934, v.20, 169–176.
12. Weinberg S. Gravitation and Cosmology. John Wiley, New York, 1972, pp. 338–341.

The Dual Behavior of Quantum Fields and the Big Bang

Malik Matwi

Damascus University, Damascus, Syria. E-mail: malik.matwi@hotmail.com

We modify the propagator of the quantum fields for the quarks and gluons. With that we have finite results (without ultraviolet divergence) in the perturbation theory. Then we search for $a^2p^2 \to 0$ and $a^2k^2 \to 0$ with fixing the Lagrangian parameters Z_i, therefore we can ignore our modification. We find the situation $a^2p^2 \to 0$ and $a^2k^2 \to 0$ associates with the free particles situation $g \to 0$ (g is the coupling constant) and the situation $a \neq 0$ associates with the perturbation breaking. We try to give the modification terms $a^2p^2/(1+a^2p^2)$ and $a^2k^2/(1+a^2k^2)$ physical aspects, for that we find the corresponding terms in the Lagrangian. To do that we find the role of those terms in the Feynman diagrams, in self energies, quarks gluons vertex, ... We see we can relate the propagator modification to fields dual behavior, pairing particle-antiparticle appears as scalar particles with mass $1/a$. For the quarks we can interrupt these particles as pions with charges $(-1, 0, +1)$. If we used the propagator modification for deriving the quarks static potential $U(r)$ of exchanged gluons and pions we find $U(0) \sim 1/a$ if we compare this with the Coulomb potential we find the length a equivalent to the smallest distance between the interacting quarks. We use the static potential in quarks plasma study. We find the free and confinement quarks phases. We suggest a nuclear compression. We find there is a decrease in the global pressure due to the nuclear condensation. We use this decrease in the Friedman equations solutions, we find we can control the dark matter and dark energy, we can cancel them.

1 Quarks and gluons propagator modification

To remove the ultraviolet (UV) divergences in the quarks and gluons perturbed interaction, we modify the propagator like:

$$\overline{\Delta}^{ab}_{\mu\nu}(k^2) = \frac{g_{\mu\nu}\,\delta^{ab}}{k^2 - i\varepsilon}\left(1 - \frac{a^2k^2}{1+a^2k^2}\right) \quad \text{for gluons} \tag{1.1}$$

$$\bar{S}_{ij}(\not p) = \frac{-\not p\,\delta_{ij}}{p^2 - i\varepsilon}\left(1 - \frac{a^2p^2}{1+a^2p^2}\right) \quad \text{for quarks} \tag{1.2}$$

the indexes a and b are gluons indexes, i and j color indexes and a is critical length, $\hbar = c = 1$. We use this modification in calculating the quarks self-energy for the perturbation interaction with the gluons, then we renormalize the interaction and search for the condition $a^2p^2 \to 0$ and $a^2k^2 \to 0$. We have

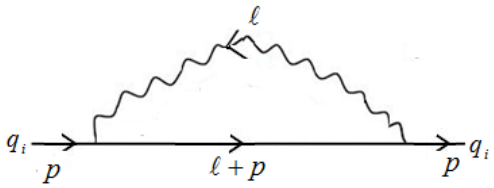

Fig. 1: The quarks self energy in strong interaction.

$$i\Sigma_{ij}(\not p) = \int \frac{d^4\ell}{(2\pi)^4}\left[ig_s\gamma^\mu T^a_{ik}\frac{\overline{S}_{kl}(\not p + \not\ell)}{i}ig_s\gamma^\nu T^b_{lj}\right]\frac{\overline{\Delta}^{ab}_{\mu\nu}(\ell^2)}{i}$$
$$= g_s^2 T^a_{ik}T^b_{lj}\int \frac{d^4\ell}{(2\pi)^4}\left[\gamma^\mu\frac{(-\not p - \not\ell)\delta_{kl}}{(p+\ell)^2}\gamma^\nu\right]\frac{g_{\mu\nu}\delta^{ab}}{\ell^2}\,.$$

So

$$i\Sigma_{ij}(\not p) = g_s^2 T^a_{ik}T^a_{kj}\int \frac{d^4\ell}{(2\pi)^4}\left[\gamma^\mu\frac{(-\not p - \not\ell)}{(p+\ell)^2}\gamma^\nu\right]\frac{g_{\mu\nu}}{\ell^2}$$
$$= g_s^2 C(R)\delta_{ij}\int \frac{d^4\ell}{(2\pi)^4}\left[\gamma^\mu\frac{(-\not p - \not\ell)}{(p+\ell)^2}\gamma_\mu\right]\frac{1}{\ell^2}$$

using $\gamma^\mu(-\not p - \not\ell)\gamma_\mu = 2(-\not p - \not\ell)$, it becomes

$$i\Sigma_{ij}(\not p) = 2g_s^2 C(R)\delta_{ij}\int \frac{d^4\ell}{(2\pi)^4}\frac{(-\not p - \not\ell)}{(p+\ell)^2}\frac{1}{\ell^2}\,.$$

Now we use the gluon modified propagator

$$\overline{\Delta}^{ab}_{\mu\nu}(k^2) = \frac{g_{\mu\nu}\delta^{ab}}{k^2 - i\varepsilon}\left(1 - \frac{a^2k^2}{1+a^2k^2}\right)$$

we get

$$i\Sigma_{ij}(\not p) = 2g_s^2 C(R)\delta_{ij}$$
$$\int \frac{d^4\ell}{(2\pi)^4}\frac{(-\not p - \not\ell)}{(p+\ell)^2}\frac{1}{\ell^2}\left(1 - \frac{a^2\ell^2}{1+a^2\ell^2}\right) \tag{1.3}$$
$$= 2g_s^2 C(R)\delta_{ij}\int \frac{d^4\ell}{(2\pi)^4}\frac{(-\not p - \not\ell)}{(p+\ell)^2}\frac{1}{\ell^2}\frac{1}{1+a^2\ell^2}\,. \tag{1.4}$$

For massive quarks, the self-energy becomes:

$$i\Sigma_{ij}(\not p) = g_s^2 C(R)\delta_{ij}\int \frac{d^4\ell}{(2\pi)^4}\frac{N}{(p+\ell)^2 + m_q^2}\frac{1}{\ell^2 + m_\gamma^2}\frac{1}{1+a^2\ell^2}$$

with $N = \gamma^\mu(-\not{p} - \not{\ell} + m)\gamma_\mu$, using the Feynman formula:

$$\frac{1}{((p+\ell)^2+m^2)\cdot(\ell^2+m_\gamma^2)\cdot(1/a^2+\ell^2)} = \int dF_3 \frac{1}{\left[((p+\ell)^2+m^2)x_1 + (\ell^2+m_\gamma^2)x_2 + (1/a^2+\ell^2)x_3\right]^3}$$

with $\int dF_3 = 2\int_0^1 dx_1 dx_2 dx_3\, \delta(x_1+x_2+x_3-1)$ and setting the transformation $q = \ell + x_1 p$ with changing the integral to be over q and making transformation to Euclidean space, the self-energy becomes [2]

$$i\Sigma_{ij}(\not{p}) = g_s^2 C(R)\delta_{ij} i \int \frac{d^4\bar{q}}{(2\pi)^4}\frac{1}{a^2}\int dF_3 \frac{N}{[\bar{q}^2+D]^3}$$

with $D = -x_1^2 p^2 + x_1 p^2 + x_1 m^2 + x_2 + m_\gamma^2 + (1-x_1-x_2)1/a^2$. The linear term in q integrates to zero, using $q = \ell + x_1 p$, N is replaced with [2]

$$N \to -2(1-x_1)\not{p} - 4m\,.$$

Using the relation

$$\int \frac{d^d\bar{q}}{(2\pi)^d}\frac{(\bar{q}^2)^a}{(\bar{q}^2+D)^{\mathrm{b}}} = \frac{\Gamma(\mathrm{b}-a-\frac{d}{2})\Gamma(a+\frac{d}{2})}{(4\pi)^{\frac{d}{2}}\Gamma(\mathrm{b})\Gamma(\frac{d}{2})} D^{-(\mathrm{b}-a-\frac{d}{2})}\,,$$

the integral over q in Euclidean space becomes:

$$\Sigma_{ij}(\not{p}) = g_s^2 C(R)\delta_{ij}\frac{1}{a^2}\int dF_3 N \frac{\Gamma(3-2)\Gamma(2)}{(4\pi)^2\Gamma(3)\Gamma(2)} D^{-(3-2)} = g_s^2 C(R)\delta_{ij}\frac{1}{a^2}\int dF_3 \frac{N}{16\pi^2\times 2} D^{-1}\,.$$

The self-energy becomes

$$\Sigma_{ij}(\not{p}) = g_s^2 C(R)\delta_{ij}\frac{1}{a^2}\int_0^1 dx_1\int_0^{1-x_1} dx_2 \frac{N}{16\pi^2}\frac{1}{D} = \frac{g_s^2 C(R)\delta_{ij}}{16\pi^2}\int_0^1 dx_1\int_0^{1-x_1} dx_2 \frac{-2(1-x_1)\not{p} - 4m}{a^2\left[-x_1^2 p^2 + x_1 p^2 + x_1 m^2 + x_2 m_\gamma^2 + (1-x_1-x_2)/a^2\right]}\,.$$

We write

$$\Sigma_{ij}(\not{p}) = C(R)\delta_{ij}\frac{g_s^2}{8\pi^2}\int_0^1 dx_1\int_0^{1-x_1} dx_2 \frac{-(1-x_1)\not{p} - 2m}{[a^2 f + (1-x_1-x_2)]} \tag{1.5}$$

with $f = -x_1^2 p^2 + x_1 p^2 + x_1 m^2 + x_2 m_\gamma^2$ this is a finite result (without divergences).

Now we renormalize the fermions propagator to give the real states and let $a \to 0$. The interacting quarks propagator becomes [2]:

$$\overline{S}(\not{p})^{-1} = \not{p} + m - \Sigma(\not{p})\,. \tag{1.6}$$

To renormalize the interacting field, we write it as

$$\overline{S}(\not{p})^{-1} = \not{p} + m - \Sigma(\not{p}) = Z_2\not{p} + Z_m m\,. \tag{1.7}$$

The parameters Z_2 and Z_m are the renormalization parameters, later we try to make them constants. For the interacting field ψ we have:

$$\langle 0|\,\psi(\not{p})\bar{\psi}(-\not{p})\,|0\rangle = \frac{1}{i}\frac{1}{\not{p}+m-\Sigma(\not{p})} = \frac{1}{i}\frac{1}{Z_2\not{p}+Z_m m} = \frac{1}{iZ_2}\frac{1}{\not{p}+Z_2^{-1}Z_m m}\,.$$

We can rewrite as

$$\langle 0|\,\sqrt{Z_2}\psi(\not{p})\sqrt{Z_2}\bar{\psi}(-\not{p})\,|0\rangle = \frac{1}{i}\frac{1}{\not{p}+Z_2^{-1}Z_m m}$$

and make $m_0 = Z_2^{-1}Z_m m$ and $\psi_0 = \sqrt{Z_2}\psi$ with that we have bare fields ψ_0 that are like the free fields and like the classical fields, so we can make them independent of the interaction, so $\partial\psi_0/\partial p^2 = \partial m_0/\partial p^2 = 0$ for $a \to 0$ and by that we renormalize the interaction. We make ψ the interacting field with mass m the physical mass, but we have to make $\Re\,[\Sigma(-m)] = 0$ in (1.6) but with $m_\gamma^2 < 0$. From (1.5) and (1.7) we have

$$Z_2 = 1 + C(R)\frac{g_s^2}{8\pi^2}\int_0^1 dx_1\int_0^{1-x_1} dx_2 \frac{1-x_1}{[a^2 f + (1-x_1-x_2)]}$$

$$Z_m = 1 + C(R)\frac{g_s^2}{8\pi^2}\int_0^1 dx_1\int_0^{1-x_1} dx_2 \frac{2}{[a^2 f + (1-x_1-x_2)]}$$

and $f = -x_1^2 p^2 + x_1 p^2 + x_1 m_q^2 + x_2 m_\gamma^2$.

By that we remove the self-energy of the interacting quark and make the mass variable. For easiness we ignore m_q and m_γ so

$$Z_2 = 1 + C(R)\frac{\alpha_s}{2\pi}\int_0^1 (1-x)\ln\left(1+\frac{1}{a^2p^2x}\right)dx = 1 + \frac{C(R)\alpha_s}{4\pi(a^2p^2)^2}\left[(a^2p^2)^2\ln\left(1+\frac{1}{a^2p^2}\right) - a^2p^2 + (2a^2p^2+1)\ln(a^2p^2+1)\right].$$

Now we fix $Z_2 = constant$ and search for the situations $-a^2p^2 \to 0$ for timelike and $a^2p^2 \to 0$ for spacelike, we have

$$\frac{\alpha_s}{(a^2p^2)^2}\left[(a^2p^2)^2\ln\left(1+\frac{1}{a^2p^2}\right) - a^2p^2 + (2a^2p^2+1)\ln\left(a^2p^2+1\right)\right] = \mathrm{c}\,.$$

For spacelike $p^2 > 0$, we have Fig. 2. According to this figure, we have $a^2p^2 = \exp(-c/\alpha_s) \to 0$ when $\alpha_s \to 0$ this is the

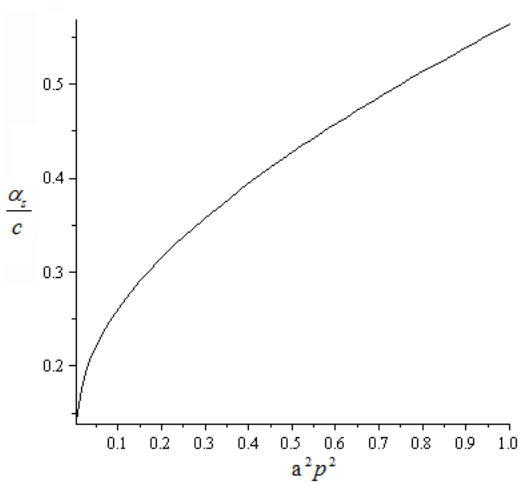

Fig. 2: The behavior of the length a with fixing Z_2.

decoupling; $p^2 \gg \Lambda^2_{QCD}$. It is the free quarks and gluons situation; $\alpha_s \to 0$ occurs at high energy for the free quarks phase. Because $ap \to 0$ so $p \ll 1/a$ this gives $r \gg a \to 0$, therefore the propagator modification is ignored. So the behavior of the length a is like the behavior of the coupling constant α_s and the modification terms are removed $ap \ll 1$ at high energy (free quarks phase).

For the limited low energy we fix $\alpha_s/a^2 = constant \times \sigma$, σ is string tension that appears in the low energy static potential $U(r)$ as we will see, for $a \to 0$ we have

$$Z_2 = 1+C(R)\frac{\alpha_s}{4\pi}\left(\frac{3}{2} - \ln\left(p^2a^2\right) + O(p^2a^2)\right) \to 1, \text{when } a \to 0$$

$$Z_m = 1 + C(R)\frac{\alpha_s}{\pi}\left(1 - \ln\left(p^2a^2\right) + O(p^2a^2)\right) \to 1, \text{when } a \to 0$$

We know the strong interaction coupling constant α_s increases extremely at low limited energy, therefore, according to the figure, we can't let $a \to 0$, so we assume when the perturbation breaks down the length a could not be removed and takes non-zero value, let it be a_0, so the propagator modification takes place.

1.1 The confinement situation

According to Fig. 2 it is possible to have $ap > 1$ (the coupling constant α_s increases extremely at low energy), therefore $p > 1/a \to r < a$ which is the quarks confinement phase at low energy.

To study the quarks confinement, we use the modified gluons propagator in deriving the static potential of the quark-quark gluons exchange. We define this potential in momentum space using M matrix elements for quark-quark (gluons exchange) interaction, with $\omega_0 = k_0 = 0$ (like the Born approximation to the scattering amplitude in non-relativistic quantum mechanics [1])

$$iM = -i\tilde{V}(k)J^\mu(p'_2, p_2)J_\mu(p'_1, p_1)$$

with the transferred current $J^\mu(p', p) = \bar{u}(p')\gamma^\mu u(p)$ where spinor states $u(p)$ include the helicity states.

We find M matrix elements using the Feynman diagrams for quark-quark gluons exchange using color representation for one quark like

$$u(p)_{color\otimes spinor} = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}u(p)_{spinor}\,.$$

For distinguishable quarks (only one diagram), we have

$$iM = \bar{u}^i(p'_2)ig_s\gamma^\mu(T^a)^j_i u_j(p_2)\frac{\Delta^{ab}_{\mu\nu}(k^2)}{i}\bar{u}^k(p'_1)ig_s\gamma^\nu(T^b)^\ell_k u_\ell(p_1)$$

with $k = p'_2 - p_2 = p_1 - p'_1$.

Using Gell-Mann matrices, we consider the matrices $T^a = \lambda^a; \lambda_1, \ldots, \lambda_8$ as $SU(3)$ generators, and using the modified gluons propagator we have

$$iM = \sum_{ijk\ell} ig_s^2\bar{u}^i(p'_2)\gamma^\mu(T^a)^j_i u_j(p_2)\,\frac{g_{\mu\nu}\delta^{ab}}{k^2}\left(1 - \frac{a^2k^2}{1+a^2k^2}\right)$$
$$\bar{u}^k(p'_1)\gamma^\nu(T^b)^\ell_k u_\ell(p_1)$$

to sum over the color indexes i, j with the color representation like above and over gluon index a we write

$$\sum_{ij}\bar{u}^i(p'_2)\gamma^\mu(T^a)^j_i u_j(p_2)$$
$$= \bar{u}(p'_2)\gamma^\mu\,\frac{1}{\sqrt{3}}\begin{pmatrix}1 & 1 & 1\end{pmatrix}(T^a)\,\frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}u(p_2)$$

and

$$\frac{1}{\sqrt{3}}\begin{pmatrix}1 & 1 & 1\end{pmatrix}(T^a)\,\frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix} = \frac{1}{3}\sum_{ij}(T^a)^j_i.$$

Therefore the M matrix elements become

$$M = \frac{1}{9}\sum_a\left(\sum_{ij}(T^a)^j_i\right)^2 g_s^2\bar{u}(p'_2)\gamma^\mu u(p_2)\,\frac{1}{k^2}\left(1 - \frac{a^2k^2}{1+a^2k^2}\right)$$
$$\bar{u}(p'_1)\gamma_\mu u(p_1)\,.$$

The Gell-Mann matrices with nonzero sum of the elements are

$$\lambda_1 = \begin{pmatrix}0&1&0\\1&0&0\\0&0&0\end{pmatrix},\ \lambda_4 = \begin{pmatrix}0&0&1\\0&0&0\\1&0&0\end{pmatrix} \text{ and } \lambda_6 = \begin{pmatrix}0&0&0\\0&0&1\\0&1&0\end{pmatrix}.$$

So

$$\sum_a \left(\sum_{ij} (T^a)_i^j \right)^2 = 3\,(2)^2 = 12.$$

Therefore we have

$$M = \frac{12 g_s^2}{9} \frac{1}{k^2} \left(1 - \frac{a^2 k^2}{1 + a^2 k^2} \right) \bar{u}(p_2') \gamma^\mu u(p_2) \bar{u}(p_1') \gamma_\mu u(p_1)\,.$$

We have the potential $\tilde{V}(k)$ in momentum space as we defined

$$\begin{aligned} iM &= -i\tilde{V}(k) J^\mu(p_2', p_2) J_\mu(p_1', p_1) \\ &= i \frac{12 g_s^2}{9} g_s^2 \bar{u}(p_2') \gamma^\mu u(p_2) \frac{1}{k^2} \left(1 - \frac{k^2}{k^2 + 1/a^2} \right) \bar{u}(p_1') \gamma_\mu u(p_1) \end{aligned}$$

with the transferred currents $J^\mu(p_2', p_2) = \bar{u}(p_2') \gamma^\mu u(p_2)$ and $J^\mu(p_1', p_1) = \bar{u}(p_1') \gamma^\mu u(p_1)$. So we have

$$\tilde{V}(k) = -\frac{4 g_s^2}{3} \frac{1}{k^2} \left(1 - \frac{k^2}{k^2 + 1/a^2} \right).$$

Making the Fourier transformation to the space XYZ, we have the static potential $U(x)$ ($k_0 = 0$) like the electric potential [1]

$$\begin{aligned} U(x) &= \int \frac{d^3k}{(2\pi)^3} \tilde{V}(k)\, e^{ik\cdot x} \\ &= -\frac{4 g_s^2}{3} \int \frac{d^3k}{(2\pi)^3} \frac{1}{k^2} \left(1 - \frac{k^2}{k^2 + 1/a^2} \right) e^{ik\cdot x} \\ &= -\frac{4 g_s^2}{3 \times 4\pi r} \left(1 - \exp\left(-\frac{r}{a} \right) \right) \quad \text{with } r = \sqrt{x^2 + y^2 + z^2}\,. \end{aligned}$$

For low limited energy we have $ap > 1$ (Fig. 2) so $r < a$, the static potential becomes

$$U(r) = -\frac{4 g_s^2}{3 \times 4\pi r} \left[1 - \exp\left(-\frac{r}{a} \right) \right] = -u_0 + a_1 r - a_2 r^2 + \ldots$$

with

$$\begin{aligned} u_0 &= \frac{4}{3} \frac{g_s^2}{4\pi a} = \frac{4\alpha_s}{3a}, \\ a_1 &= \sigma = \frac{g_s^2}{3 \times 2\pi a^2} = \frac{2\alpha_s}{3a^2}, \\ a_2 &= \frac{4\alpha_s}{3 \times 6a^3}. \end{aligned}$$

To fix $u_0 = 4\alpha_s/3a$ we write it as

$$u_0 = \frac{4\alpha_s}{3a} = \frac{4\alpha_s}{3a^2} a = 2\sigma a$$

fixing the string tension σ and the length $a \to a_0$ at low energy.

This potential appears at low limited energy and prevents the quarks from spreading away, $r < a$ so it holds the quarks inside the hadrons. But starting from the high energies $a \to 0$, although the quarks masses are small but they are created only at high energies where they are free and by dropping the energy the situation $r < a$ appears, the length a would run and becomes higher at low energies, so have $-a^2k^2 > 1$ for $r < a$ which is the confinement. The confinement (at low limited energy) means when $r \to a$ the two interacting quarks kinetic energy becomes zero (ignore the quark mass), therefore the highest kinetic energy that the quark can get equals σa which relates to the potential $U(r) = -u_0 + \sigma r + \ldots$ for $r < a$.

We can make $U(r)$ the potential for all quarks in $r < a$ so $\sigma \to \sum\sigma$ and consider r as average distance between the interacting quarks, so the energy σa becomes the highest kinetic energy of all quarks. When $r \to a$ the potential becomes $U(0) = -u_0 = -4\alpha_s/3a = -\sigma a < 0$ therefore the total quarks energy becomes negative.

In this situation the free quarks disappear, they become condensed in the hadrons. So the role of the potential is reducing the number of free quarks. Therefore the potential $u_0 = \sigma a$ leads to decrease of the free quarks chemical potential μ_0, and we have

$$\begin{aligned} \mu_0 \to \mu_0 + U(r) &= \mu_0 - \frac{\alpha_s}{r} \left(1 - e^{-r/a} \right) = \mu(r) \\ &\approx \mu_0 - u_0 + \sigma r \quad \text{for } r < a \end{aligned}$$

where we replaced $4\alpha_s/3$ with α_s. We renormalize this step at high energy for the free quarks, quarks plasma.

2 The quarks field dual behavior

To have finite results in the perturbation interaction, we modified the propagator like

$$\overline{\Delta}_{\mu\nu}^{ab}(k^2) = \frac{g_{\mu\nu}\delta^{ab}}{k^2 - i\varepsilon} \left(1 - \frac{a^2 k^2}{1 + a^2 k^2} \right) \text{ for gluons}$$

$$\bar{S}_{ij}(\not{p}) = \frac{-\not{p}\delta_{ij}}{p^2 - i\varepsilon} \left(1 - \frac{a^2 p^2}{1 + a^2 p^2} \right) \text{ for quarks}\,.$$

We saw we can ignore the modification terms $a^2p^2/(1 + a^2p^2)$ and $a^2k^2/(1 + a^2k^2)$ at high energy, but when the energy drops down to limited energy, those terms take place, we can give them a physical meaning, for that we search for the corresponding terms in the Lagrangian.

To do this, we find the role of those terms in the Feynman diagrams, in self energies, quarks-gluons vertex, ... We find that the terms $a^2p^2/(1 + a^2p^2)$ and $a^2k^2/(1 + a^2k^2)$ can be related to pairing quark-antiquark that appear as scalar particles with mass $1/a$ and charges $(-1, 0, +1)$ and we can interpret these particles as pions.

That appears in the particles-antiparticles composition in Feynman diagrams which mean for the fields, there is fields dual behavior, free fields and composite fields, this behavior leads to the possibility of separating the particles and possibility for their composition, so the dual behavior of the fields is elementary behavior. In general, for any particle A and its

antiparticle $\overline{A}$, in pertubation interaction, they pair and have a scalar particle $\overline{A}A$, this leads to reduce the currents (charges) of particles and antiparticles.

That is, for each outcoming particle, in Feynman diagrams, there is incoming antiparticle with positive energy and negative mass, depending on the coupling constant behavior (this is at high energy for the electromagnetic interaction and at low energy for the strong interaction, quarks and gluons). Therefore reducing their interactions with the charges in a way leads to finite results in the perturbation results.

Using the gluons modified propagator, the quark self-energy becomes (1.3)

$$i\Sigma_{ij}(\not p) = 2g_s^2 C(R)\delta_{ij}\int \frac{d^4\ell}{(2\pi)^4}\,\frac{(-\not p-\not\ell)}{(p+\ell)^2}\,\frac{1}{\ell^2}\left(1-\frac{a^2\ell^2}{1+a^2\ell^2}\right).$$

We can separate it into two parts

1. Quark−gluon part:

$$i\Sigma_{ij}(\not p) = 2g_s^2 C(R)\delta_{ij}\int \frac{d^4\ell}{(2\pi)^4}\,\frac{(-\not p-\not\ell)}{(p+\ell)^2}\,\frac{1}{\ell^2}\,;$$

2. pairing quarks part:

$$i\Sigma_{ij}(\not p) = 2g_s^2 C(R)\int \frac{d^4\ell}{(2\pi)^4}\,\frac{(-\not p-\not\ell)\delta_{ij}}{(p+\ell)^2}\,\frac{1}{\ell^2}\left(-\frac{a^2\ell^2}{1+a^2\ell^2}\right).$$

It appears that in the pairing part there is a scalar field φ propagator:

$$\frac{1}{i}\,\frac{1}{\ell^2+1/a^2}$$

which is real scalar particles field propagator with mass $1/a$, to preserve the charges, spin, ..., this particle must be condensed of quark-antiquark $|\overline{q}q\rangle$ (particle-antiparticle in general) so we have new diagram (Fig. 3), we rewrite

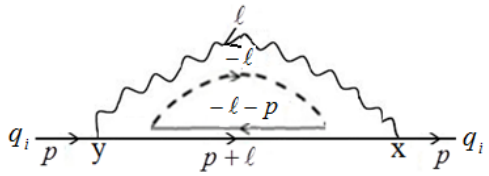

Fig. 3: Representation the dual behavior, joined particle−antiparticle with opposite momentum−energy.

$$i\Sigma_{ij}(\not p) = 2(-g_s)^2 C(R)\int \frac{d^4\ell}{(2\pi)^4}\,\frac{(-\not p-\not\ell)\delta_{ij}}{i(p+\ell)^2}\,\frac{-i}{(\ell)^2+1/a^2}\,.$$

Therefore we must add new interaction terms to the quarks Lagrangian, the possible terms are:

$$\Delta L = -ig_{\varphi q}\varphi\bar{Q}Q \quad \text{with} \quad g_{\varphi q} = g_s\sqrt{2C(R)}$$

or

$$\Delta L = g_{\varphi q}\varphi\bar{Q}\gamma_5 Q\,.$$

We expect the pairing particles-antiparticles preserve the flavor symmetry, so the real scalar field φ becomes $|\overline{q_i}q_j\rangle$. For two flavors q_i and q_j we write the quarks field like $Q = \begin{pmatrix} q_i & q_j \end{pmatrix}^T$ so

$$\Delta L = -ig_{\varphi q}\varphi^a\bar{Q}T_2^a Q \quad \text{or} \quad \Delta L = g_{\varphi q}\varphi^a\bar{Q}T_2^a\gamma_5 Q\,.$$

The real scalar fields φ^a could interact with itself and have real non-zero ground value υ then $\langle\varphi\rangle = \upsilon$ so we can renormalize it like

$$\varphi^a T_2^a \to \nu - i\nu\pi^a T_2^a + ...$$

then we have

$$\begin{aligned}\Delta L &= -ig_{\varphi q}\bar{Q}\left(\nu - i\nu\pi^a T_2^a + ...\right)Q\\ &= -ig_{\varphi q}\nu\bar{Q}Q - g_{\varphi\pi}\pi\bar{Q}Q + ...\ \text{Chiral symmetry breaking}\end{aligned}$$

or

$$\Delta L = g_{\varphi q}\bar{Q}(\nu - i\nu\pi^a T_2^a + ...)\gamma_5 Q \to g_{\pi q}\bar{Q}\gamma_5 Q - ig_{\pi q}\pi\bar{Q}\gamma_5 Q + ...$$

Here the particles $\pi^a T_2^a \to \pi = (\pi^0, \pi^-, \pi^+)$ are the pions. The unusual terms $-ig_{\varphi a}\nu\bar{Q}Q$ and $g_{\pi q}\bar{Q}\gamma_5 Q$ are not hermitian and violate the symmetries, so they let the quarks disappear, damping at low energy $r < a$:

$$\begin{aligned}e^{i\Delta Et}\,|Q\rangle &= e^{-i\Delta Lt}\,|Q\rangle = e^{-g_{\varphi q}\nu\bar{q}qt}\,|Q\rangle\\ &= \sum_n e^{-g_{\varphi q}\nu(\bar{q}q)t}\,|E_n\rangle\langle E_n\,|Q\rangle \to |0\rangle\langle 0\,|Q\rangle\,.\end{aligned}$$

E_n is the energy of the quarks in state $|n\rangle$ and $e^{i\hat{H}t}\,|Q\rangle$ is the eigenstate of the quarks field operator $\hat{Q}(t)$ in Heisenberg picture, $\hat{Q}(t) = e^{i\hat{H}t}\hat{Q}e^{-i\hat{H}t}$.

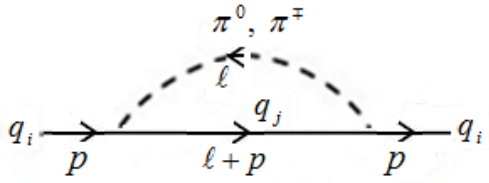

Fig. 4: The quarks interaction with pions as a result of dual behavior.

That damping in the states is because of the pairing quark-antiquark at low energy $a \neq 0$, this pairing reduces the charges (currents) of free quarks (Fig. 5). We can see that if we relate the minus sign in $-a^2\ell^2/(1+a^2\ell^2)$ to the fermions propagator:

$$S(x-y) = \int \frac{d^4p}{(2\pi)^4}\,\frac{-\not p}{p^2}e^{ip(x-y)}\ \text{(propagator from } y \text{ to } x)$$

so

$$-S(x-y) = -\int \frac{d^4p}{(2\pi)^4}\,\frac{-\not p}{p^2}e^{ip(x-y)} = \int \frac{d^4p}{(2\pi)^4}\,\frac{+\not p}{p^2}e^{ip(x-y)}$$

change $p \to -p$ (propagator from x to y)

$$-S(x-y) = \int \frac{d^4p}{(2\pi)^4} \frac{-\not{p}}{p^2} e^{-ip(x-y)} = \int \frac{d^4p}{(2\pi)^4} \frac{-\not{p}}{p^2} e^{ip(y-x)} .$$

So it is equivalent to invert the propagator $y \to x$ to $x \to y$ with positive energy and negative mass. Therefore it reduces the charges, currents, energies, ... of the particles and antiparticles, we have

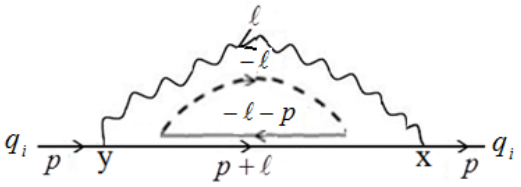

Fig. 5: Omitting the distance x-y from the propagator.

$$(p+\ell) + (-p-\ell) = 0 \text{ and } (-\ell) + (\ell) = 0$$

so incoming with p and outcoming with p, it is like to say the particles jump from y to x, in other words the distance $y - x$ is removed from the interaction. We expect the fields dual behavior takes place in negative potential. If there is no negative potential the paired particles would not survive (never condense). For the quarks, the case $0 < r < a$ must associate with negative potential u and $E + u < 0$. Because the behavior of the strong interaction coupling constant at low energy α_s is high, we expect negative potential at low energy $E + u < 0$ $(E > 0, u < 0)$, so the quarks condense.

Because of the dual behavior of the quarks field which leads to quarks composite in scalar charged particles like the pions (π^-, π^0, π^+) and because of their quantized charges $(-1, 0, +1)$ we expect the hadrons charges to be also quantized $(-Q, -Q+1, \ldots 0, +1, \ldots, +Q)$ this quantization relates to the dual behavior of the quarks field in different hadrons, pairing quarks of different hadrons, so these condensed quarks; pions, kaons, ... are shared between the hadrons, so we put them together with the hadrons in groups, like the pions $(-1, 0, +1)$ which can be inserted in $SU(2)$ generators which can represent the proton-neutron pairing. Therefore the protons and neutrons Lagrangian contains the terms $-ig_{\pi N}\pi^\alpha \bar{N} T_2^\alpha N$ with the nucleon field $N = \begin{pmatrix} p \\ n \end{pmatrix}$.

3 The quarks plasma

We tried before to explain how the quarks are confined, for the strong interaction, we have the condition $r < a \neq 0$ at low limited energy and the condition $r > a \to 0$ at high energies for free quarks where the length a is removed from the propagators. But it appears to be fixed at low limited energy. In the last section we showed there is dual behavior for the quarks field, but when the length a is fixed, the result is scalar particles (pions) with mass $1/a_0$ at low limited energy and the result is the chiral symmetry breaking. We found the length a appears in the quark-quark strong interaction (gluons exchanging) potential $U(r)_{r<a} < 0$, so it relates to interaction strength. That is because the behavior of the length a is like the behavior of the coupling constant α_s. The confinement (at low energy $r < a$) means when $r \to a$ the two interacting quarks kinetic energy becomes zero (ignore the quark mass), therefore the highest kinetic energy the quark can get equals σa which relates to the potential $U(r) = -u_0 + \sigma r + \ldots$ for $r < a$ (at low limited energy). When $r \to a$ the potential becomes $U(0) = -u_0 = -4\alpha_s/3a < 0$ therefore the total quarks energy becomes negative. In this situation the free quarks disappear ($\mu_0 \to 0$), they become condensed in the hadrons.

We try here to use statistical thermodynamics to show how the free quarks disappear at low energies (low temperatures) where the length a becomes fixed, so the chiral symmetry breaking and the quarks condensation. One of the results is that the confinement phase (3.14) not necessarily associates with chiral symmetry breaking, that is, the chiral symmetry breaking appears at the end of the cooling process when the expanding and cooling are ended and the length a becomes fixed, therefore the chiral symmetry breaking occurs and the pions become massive $m = 1/a_0$.

We start with the massless quarks, their energy in volume V is

$$E = c \int_{a^3} d^3r \int_0^\infty d\varepsilon\, g(\varepsilon)\, \varepsilon \frac{1}{e^{\beta(\varepsilon-\mu(r))}+1} \quad : g(\varepsilon) = g_q \frac{V}{2\pi^2} \varepsilon^2 \tag{3.1}$$

where $\mu(r) = \mu_0 + u(r)$ with $u(r) = -\frac{4\alpha_s}{3r}\left(1 - e^{-r/a}\right)$.

Here we inserted the quark-quark strong interaction potential $U(r)$ in the chemical potential (for decreasing the free quarks energy, as we think, the quarks potential reduces the free quarks chemical potential and make them condense at low energy) and because $r < a$ we integrate over the volume a^3: r is the distance between the interacting quarks. We can replace $4\alpha_s/3 \to \alpha_s$.

The constant c is determined by comparing with free quarks high energy where the potential $U(r) \to 0$ and $\alpha_s \to 0$ (decoupling) at high energies, so the length $a \to 0$ that is as we said before, the behavior of the length a is like the behavior of the coupling constant g_s therefore the quarks become free at high energies.

By integrating over the energy (Maple program) we have:

$$\begin{aligned} E &= c g_q \frac{V}{2\pi^2} \int_{a^3} d^3r \int_0^\infty d\varepsilon \frac{\varepsilon^3}{e^{\beta(\varepsilon-\mu(r))}+1} \\ &= c g_q \frac{V}{2\pi^2\beta^4} \int_{a^3} d^3r \left[\frac{7\pi^4}{60} + \frac{\pi^2}{2} u_0(r)^2 + \frac{1}{4} u_0(r)^4 + \right. \\ &\qquad \left. + 6 \sum_{k=1}^{\infty} \frac{(-1)^k e^{-k\beta\mu(r)}}{k^4} \right] \end{aligned}$$

with $u_0(r) = \beta\mu(r) = \beta(\mu_0 + u(r))$. By integrating over r (the distance between the interacting quarks) we have

$$E = cg_q \frac{2Va^3}{\pi x^4}\Bigg[3.78 + 2\,(\beta\mu_0)^2\Bigg(0.82 - 1.16\,\frac{\alpha_s}{a\mu_0} + \\ + 0.41\left(\frac{\alpha_s}{a\mu_0}\right)^2\Bigg) + (\beta\mu_0)^4\Bigg(0.08 - 0.23\,\frac{\alpha_s}{a\mu_0} + \\ + 0.25\left(\frac{\alpha_s}{a\mu_0}\right)^2 - 0.12\left(\frac{\alpha_s}{a\mu_0}\right)^3 + 0.02\left(\frac{\alpha_s}{a\mu_0}\right)^4\Bigg) + \\ + 6\sum_{k=1}^{\infty}\int_0^1 x^2dx\,\frac{(-1)^k e^{-k\beta\mu(x)}}{k^4}\Bigg].$$

g_q is the quarks degeneracy number and $x = \beta\mu_0$. For easiness we write $\alpha_s/a\mu_0 = 2\sigma a/\mu_0 = y$ in the energy relation. So it becomes

$$E = cg_q \frac{2Va^3}{\pi x^4}\Big[3.78 + 2(\beta\mu_0)^2\left(0.82 - 1.16y + 0.41y^2\right) \\ + (\beta\mu_0)^4\left(0.08 - 0.23y + 0.25y^2 - 0.12y^3 + 0.02y^4\right) \\ + 6\sum_{k=1}^{\infty}\int_0^1 x^2dx\,\frac{(-1)^k e^{-k\beta\mu(x)}}{k^4}\Big]. \qquad (3.2)$$

at high energy: $x = \beta\mu_0 = \mu_0/T \to 0$. To find the constant c we compare with quarks high energy where they are free massless particles:

$$E_{high} = g_q V\,\frac{7\pi^2}{240}\,T^4\,.$$

When T is high, $x = (\mu_0/T) \to 0$ and $y \to 0$ therefore $\beta\mu(x) \to 0$ so we expand $e^{-k\beta\mu(x)}$ near $\beta\mu(x) = 0$, we have:

$$E_{high} = cg_q\frac{2a^3V}{\pi x^4}\,[3.78 - 1.88 + O(x,y)] \to cg_q\frac{2a^3V}{\pi x^4}1.9$$

$$\to g_q\frac{7\pi^2V}{240}T^4 = cg_q\frac{2a^3V}{\pi x^4}1.9 \to c = \frac{\pi}{2a^3 1.9}\frac{7\pi^2}{240}\mu_0^4 \qquad (3.3)$$

The energy becomes:

$$E = \frac{1}{1.9}\,\frac{7\pi^2}{240}\,\mu_0^4 g_q\,\frac{V}{(\beta\mu_0)^4}\Big[3.78 + 2(\beta\mu_0)^2(0.82 - 1.16y + \\ + 0.41y^2) + (\beta\mu_0)^4(0.08 - 0.23y + 0.25y^2 - 0.12y^3 + \\ + 0.02y^4) + 6\sum_{k=1}^{\infty}\int_0^1 x^2dx\,\frac{(-1)^k e^{-ku_0(x)}}{k^4}\Big].$$

Now we see the effects of the length a on the energy, at high energy, by fixing $x = \mu_0/T$ and varying $y = \sigma a/2\mu_0 < 1$:

$$E_{high} = \frac{1}{1.9}\frac{7\pi^2}{240}\,g_q V\mu_0^4\,x^{-4}\Big[1.9 + x(1.8 - 1.24y) + \\ + x^2(0.82 - 1.18y + 0.42y^2) + x^3(0.23 - 0.47y + \\ + 0.33y^2 - 0.08y^3) + x^4(0.04 - 0.12y + \\ + 0.13y^2 - 0.07y^3 + 0.01y^4) + \ldots\Big]_{x=\beta\mu_0\to 0}\,. \qquad (3.4)$$

We expanded $e^{-k\beta\mu(x)}$ near $\beta\mu(x) = 0$ and fixed the tension σ as we assumed before, so we have Fig. 6.

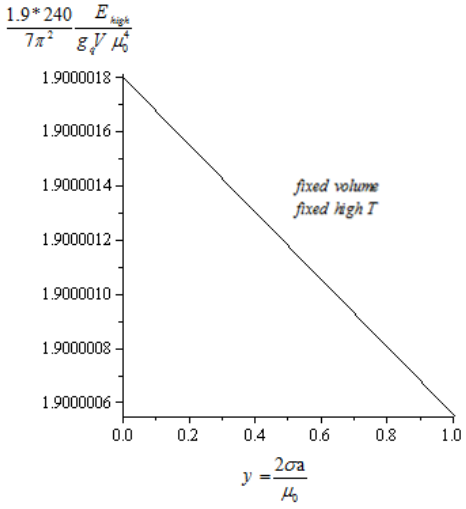

Fig. 6: Decreasing the high energy with increasing y.

It appears in the figure that the high energy quarks lose energy when the length a increases although the temperature is fixed. That means, when the length a increases the number of the excited quarks decreases. That is because of the attractive linear potential $\sigma r\ldots$ between the quarks, that potential absorbs an energy ($r < a$ confinement, section 1), so the quarks are cooled faster by the expansion. As we said before, the behavior of length a is like the behavior of the coupling constant α_s so when the energy dropped to lowest energy, the length a increased extremely and this is fast cooling (extreme cooling). That occurs when the particles spread away, the length a, as a distance between the quarks, increases.

To determine the end, we search for the balance situations, such as zero pressure, confinement condition, ... First we find the high energy pressure including the effects of the potential σa. Starting from the general pressure relation:

$$p = -\frac{\partial}{\partial V}F \text{ where } F = -T\ln Z = -\frac{1}{\beta}\,\ln Z$$

here we use the relation:

$$\ln Z = c\int_{a^3} d^3r\int_0^{\infty} d\varepsilon\, g(\varepsilon)\ln\left(e^{-\beta(\varepsilon-\mu(r))}+1\right)$$

$$: \; g(\varepsilon) = g_q\,\frac{V}{2\pi^2}\,\varepsilon^2$$

and the pressure becomes

$$P = \frac{1}{3}\,\frac{\partial}{\partial V}E$$

so for high energy $x = \beta\mu_0 \to 0$ we have the pressure:

$$\begin{aligned} P_{high} &= \frac{1}{3}\frac{\partial}{\partial V}E_{high} \\ &= \frac{\partial}{\partial V}\frac{1}{3\times 1.9}\frac{7\pi^2}{240}\,g_q V\mu_0^4\,x^{-4}\Big[1.9 + x(1.8 - 1.24y) + \\ &+ x^2(0.82 - 1.18y + 0.42y^2) + x^3(0.23 - 0.47y + \\ &+ 0.33y^2 - 0.08y^3) + x^4(0.04 - 0.12y + \\ &+ 0.13y^2 - 0.07y^3 + 0.01y^4) + \dots\Big]. \end{aligned}$$

Now the key point is, we want to include the potential effect on the pressure so we replace the volume V with the volume $a^3 \sim y^3$ so

$$\begin{aligned} P_{high} \to \frac{\partial}{\partial y^3}y^3\frac{1}{3\times 1.9}\frac{7\pi^2}{240}\,g_q\,\mu_0^4\,x^{-4}\Big[1.9 + x(1.8 - 1.24y) + \\ + x^2(0.82 - 1.18y + 0.42y^2) + x^3(0.23 - 0.47y + \\ + 0.33y^2 - 0.08y^3) + x^4(0.04 - 0.12y + \\ + 0.13y^2 - 0.07y^3 + 0.01y^4) + \dots\Big] \end{aligned} \tag{3.5}$$

which is represented in Fig. 7, without conditions on y or on the length a.

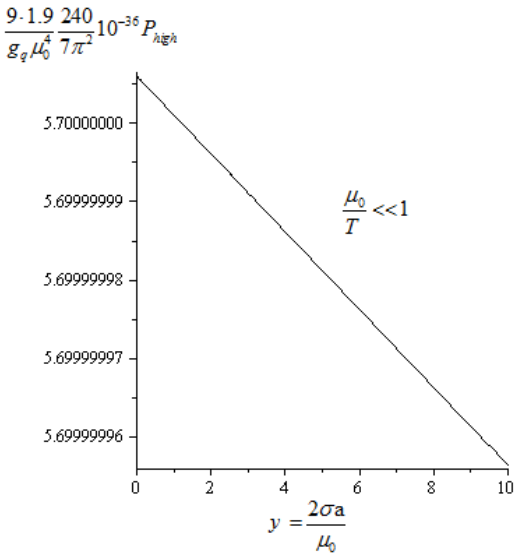

Fig. 7: The effects of potential σa on the pressure.

It is clear (without conditions on y) the pressure decreases with increasing the length a (decreasing the quarks energy $-p^2$) until it becomes zero, then negative. That becomes clear at low energy where there are conditions on y and so on the length a.

For the low energy quarks, $T \to 0$ so $\beta\mu(x) \to \infty$ so $e^{-k\beta\mu(x)} \to 0$. The energy becomes:

$$\begin{aligned} E_{low} = \frac{1}{1.9}\frac{7\pi^2}{240}\,\mu_0^4\,g_q\,\frac{V}{(\beta\mu_0)^4}\Big[3.78 + \\ + 2\,(\beta\mu_0)^2\left(0.82 - 1.16y + 0.41y^2\right) + \\ + (\beta\mu_0)^4\left(0.08 - 0.23y + 0.25y^2 - 0.12y^3 + 0.02y^4\right)\Big]. \end{aligned} \tag{3.6}$$

Making $x = T/\mu_0$ so

$$\begin{aligned} E_{low} = \frac{1}{1.9}\frac{7\pi^2}{240}\,\mu_0^4\,g_q V x^4\Big[3.78 + 2x^{-2}\left(0.82 - 1.16y + 0.41y^2\right) \\ + x^{-4}\left(0.08 - 0.23y + 0.25y^2 - 0.12y^3 + 0.02y^4\right)\Big]. \end{aligned}$$

Now the key point, we want to show the effect of the potential σa on the energy so we see the behavior of the energy in the volume a^3 with respect to $y = 2\sigma a/\mu_0$ the diagram is given in Fig. 8. That is extreme behavior after $y = 0.6$ where the

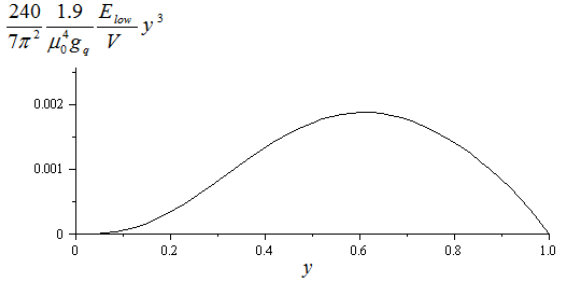

Fig. 8: The extremely decreasing in quarks low energy in the strong interaction.

energy $(E/V)\,a^3$ decreases when the volume a^3 increases, the end in $y = 1$ where the free quarks disappear for $y > 1$.

Now we can distinguish between the confinement and the chiral symmetry breaking, when $y > 0.6$ there is confinement: extreme cooling, negative pressure. But when reach $y = 1$ there is chiral symmetry breaking where the length a becomes fixed, and from the quarks field dual behavior there are scalar charged particles with mass $1/a$ appear when the length a is fixed with non-zero value a_0. Here the evidence for fixing the length a is the lowest limited quarks energy, that is as we said before, the behavior of the length a is like the behavior of the coupling constant α_s so when the quarks energy dropped (extreme cooling) the length a increases extremely to reach the highest value when $y = 1$ which is equivalent to smallest energy $E = 0$ (the cooling end). Another evidence for fixing the length a (chiral symmetry breaking) is the low energy pressure:

$$P_{low} = \frac{1}{3}\frac{\partial}{\partial V}E_{low} \to \frac{1}{3}\frac{\partial}{\partial y^3}\frac{E_{low}}{V}\,y^3\,.$$

To include the potential effect we study the pressure using the

volume $a^3 \sim y^3$ therefore

$$P_{low} \to \frac{1}{3}\frac{\partial}{\partial y^3}\frac{1}{1.9}\frac{7\pi^2}{240} g_q \mu_0^4 y^3 x^4 \left[3.78 + \right.$$
$$+ 2x^{-2}\left(0.82 - 1.16y + 0.41y^2\right)$$
$$\left. + x^{-4}\left(0.08 - 0.23y + 0.25y^2 - 0.12y^3 + 0.02y^4\right)\right]$$

and therefore

$$\frac{P_{low}}{\mu_0^4} = \frac{1}{9\times 1.9}\frac{7\pi^2}{240} g_q \left[3\times 3.78\, x^4 + \right.$$
$$+ 3\times 2\times x^2\left(0.82 - 1.16y + 0.41y^2\right) \qquad (3.7)$$
$$+ 3\times(0.08 - 0.23y + 0.25y^2 - 0.12y^3 + 0.02y^4) +$$
$$+ 2yx^2(-1.16 + 0.82y) +$$
$$\left. + y\left(-0.23 + 0.5y - 0.36y^2 + 0.08y^3\right)\right].$$

We see its behavior in Fig. 9 below

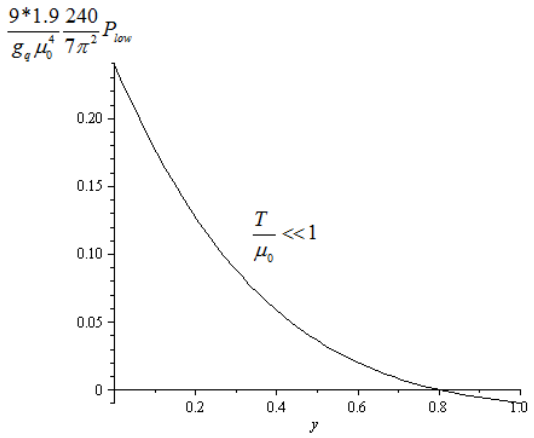

Fig. 9: The extremely decreasing in the pressure at low energy.

It is clear from the figure, when $y > 0.6$ the quarks pressure becomes negative. We expect the condensed quarks phase (confinement quarks) has positive pressure, so the preferred phase is the condensed quarks phase. So when $y > 0.6$ the quarks condense until $y = 1 : a \to a_0 \approx 1/(135 - 140\,\text{Mev})$ the quarks disappear, the scalar charged particles (pions) appear instead of them, that is because of the quarks dual behavior (free-condensed quarks), but at low limited energy the condensed phase has a big chance instead the free phase.

3.1 The confinement phase

In this paper we study two quarks (up and down) condensation in the pions (π^0, π^+, π^-) and baryons (n, p^+, p^-), so the degeneracy number is $g_q = 2_{flavor}\times 2_{charge}\times 2_{spin}\times 3_{color} = 24$.

We need more clarifying for determining if the quarks could stay free particles or they condense in hadrons. We can think they could be free if their energy is enough for covering the strong interaction potential and stay free particles with least possible energy (at 0 temperature). Unless they condense in the hadrons.

To cover the strong interaction potential means to lose an energy E_u which is transferred to the exchanged static gluons and pions which are created between the low energy quarks. So the remaining energy in the volume $4\pi a^3/3$ is

$$\frac{E_{q,low}}{V}\frac{4\pi}{3}a^3 - \frac{E_u}{V}\frac{4\pi}{3}a^3 . \qquad (3.8)$$

This energy must be enough for the least possible free quarks. Therefore we must determine the chemical potential μ_0 of the free quarks with smallest possible density at 0 temperature.

According to the quarks confinement $r < a$ at low limited energy, which means the highest possible distance between the two interacting quarks is a, we expect the least quarks density is two quarks in the volume $4\pi\,(a/2)^3/3$.

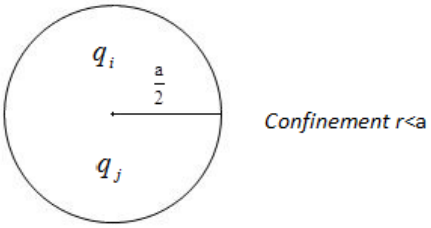

Fig. 10: The quarks confinement at low energy.

From this view we can calculate the least quarks chemical potential μ_0 of free quarks:

$$2\left(\frac{4\pi}{3}\left(\frac{a}{2}\right)^3\right)^{-1} = \frac{1}{V}\int_0^{\mu_0} g(\varepsilon)\,d\varepsilon = g_q\frac{\mu_0^3}{6\pi^2}$$
$$\to \left(\mu_0\frac{a}{2}\right)^3 = \frac{9\pi}{g_q}$$
$$\to (\mu_0 a)^3 = \frac{8\times 9\pi}{g_q}$$

$1/a$ is the pion mass when $a \to a_0$ in the end of free quarks phase so $1/a \to (135 - 140)\,\text{Mev}$. So the least free quarks energy density in 0 temperature is

$$\frac{\varepsilon_{free}}{V} = \frac{1}{V}\int_0^{\mu_0} g(\varepsilon)\,\varepsilon\,d\varepsilon = g_q\frac{\mu_0^4}{4\times 2\pi^2} .$$

The smallest energy of the free quarks in the volume $4\pi a^3/3$ is

$$\varepsilon_{free,a^3} = g_q\frac{\mu_0^4}{4\times 2\pi^2}\frac{4\pi a^3}{3} = \frac{4\pi}{3}g_q\frac{\mu_0}{4\times 2\pi^2}(\mu_0 a)^3$$
$$= \frac{4\pi}{3}g_q\frac{\mu_0}{4\times 2\pi^2}\frac{8\times 9\pi}{g_q} = \frac{4\pi}{3}\frac{9}{\pi}\mu_0 \qquad (3.9)$$

therefore

$$\frac{\varepsilon_{free,a^3}}{2\mu_0} = \frac{4\pi}{3}\frac{9}{2\pi} = \frac{4\pi}{3}\times 1.43\,.$$

Because the chemical potential $\mu \sim 1/a$ and $\mu \to \mu_0$ when $a \to a_0$ and because $y \sim a$ so we modified $\mu_0 \to \mu_0/y$ so

$$\frac{4\pi}{3}\times 1.43 \to \frac{4\pi}{3}\times\frac{1.43}{y}\,. \tag{3.10}$$

Now we find the least energy E_u which is transferred to the static exchanged gluons and pions according to the potential

$$u(r) = -\frac{4\alpha_s}{3r}\left(1-e^{-r/a}\right) \approx -u_0 + \sigma r : r < a\,.$$

We absorbed $4/3$ to α_s so and made $\alpha_s/a\mu_0 = 2\sigma a/\mu_0 = y$ the constant σ is the string tension. This potential is inserted to reduce the chemical potential μ_0 and the energy is renormalized at high energy. So we have $\mu_0 \to \mu_0 + u(r)$:

$$\mu(r) = \mu_0 - \frac{\alpha_s}{r}\left(1-e^{-r/a}\right) \approx \mu_0 - u_0 + \sigma r : r < a\,.$$

Therefore we can calculate the least absorbed energy by this potential, by calculating the changes on the energy density at 0 temperature

$$\begin{aligned}\frac{\varepsilon(\alpha_s/a)}{V} = \frac{\varepsilon(y)}{V} &= \frac{1}{V}\,c\int_0^a 4\pi r^2 dr\int_0^{\mu(r)} g(\varepsilon)\,\varepsilon\,d\varepsilon\\ &= c\int_0^a 4\pi r^2 dr\, g_q\,\frac{\mu(r)^4}{4\times 2\pi^2}\,.\end{aligned}$$

The constant c is determined

$$c = \frac{\pi}{2a^3 1.9}\,\frac{7\pi^2}{240}$$

so the interaction energy is

$$\begin{aligned}\frac{\varepsilon(\alpha_s/a)}{V} = \frac{\varepsilon(y)}{V} &= g_q\,\frac{\pi}{2a^3 1.9}\,\frac{7\pi^2}{240}\,\frac{4\pi}{4\times 2\pi^2}\int_0^a r^2 dr\mu(r)^4\\ &= g_q\,\frac{7\pi^2}{4\times 1.9\times 240a^3}\int_0^a r^2 dr\mu(r)^4\,.\end{aligned}$$

This becomes

$$\begin{aligned}\frac{\varepsilon(y)}{V} &= g_q\,\frac{7\pi^2}{4\times 1.9\times 240a^3}\int_0^a r^2 dr\left[\mu_0 - \frac{\alpha_s}{r}\left(1-e^{-r/a}\right)\right]^4\\ &= g_q\,\frac{7\pi^2}{4\times 1.9\times 240a^3}(\mu_0)^4\int_0^a r^2 dr\left[1-\frac{\alpha_s}{\mu_0 r}\left(1-e^{-r/a}\right)\right]^4\end{aligned}$$

Using the change $r = ax$ so

$$\begin{aligned}\frac{\varepsilon(y)}{V} =& g_q\,\frac{7\pi^2}{4\times 1.9\times 240a^3}\,(\mu_0)^4\\ &\int_0^1 a^3x^2dx\left[1-\frac{\alpha_s}{\mu_0 ax}\,(1-e^{-x})\right]^4\end{aligned}$$

therefore

$$\frac{\varepsilon(y)}{V} = g_q\,\frac{7\pi^2}{4\times 1.9\times 240}\,(\mu_0)^4\int_0^1 x^2dx\left[1-\frac{y}{x}\,(1-e^{-x})\right]^4\,.$$

The spent energy for the interaction in the volume $4\pi a^3/3$ is

$$\begin{aligned}\varepsilon_{u,a^3} =& \frac{\varepsilon(1)-\varepsilon(0)}{V}\,\frac{4\pi a^3}{3}\\ =&\frac{4\pi}{3}\,g_q\,\frac{7\pi^2(\mu_0 a)^3\mu_0}{4\times 1.9\times 240}\\ &\left(\int_0^1 x^2dx\left[1-\frac{1}{x}\,(1-e^{-x})\right]^4 - \int_0^1 x^2dx\right)\end{aligned} \tag{3.11}$$

and it becomes

$$\begin{aligned}\varepsilon_{u,a^3} &= -\frac{4\pi}{3}\,g_q\,\frac{7\pi^2}{4\times 1.9\times 240}\,(\mu_0 a)^3\mu_0\,0.33\\ &= -\frac{4\pi}{3}\,g_q\,\frac{7\pi^2}{4\times 1.9\times 240}\,\frac{8\times 9\pi}{g_q}\,\mu_0\times 0.33\,.\end{aligned}$$

Therefore

$$\begin{aligned}\varepsilon_{u,a^3} &= -\frac{4\pi}{3}\,g_q\,\frac{7\pi^2}{4\times 1.9\times 240}\,(\mu_0 a)^3\mu_0\times 0.33\\ &= -\frac{4\pi}{3}\,\frac{7\times 8\times 9\times 0.33\pi^3}{4\times 1.9\times 240}\,\mu_0 = -\frac{4\pi}{3}\times 2.82\,\mu_0\,.\end{aligned} \tag{3.12}$$

So we have

$$\frac{\varepsilon_{u,a^3}}{2\mu_0} = -\frac{4\pi}{3}\times 1.41\,.$$

As for E_{free} we replace

$$\frac{4\pi}{3}\times 1.41 \to \frac{4\pi}{3}\times\frac{1.41}{y}\,.$$

Now we find the confinement condition at any temperature, if the quarks energy is not enough to cover the interaction energy E_u and give free quarks with smallest density, at 0 temperature, then they become confinement ($r < a$), so the confinement condition

$$E(T,y) - \varepsilon_u - \varepsilon_{free} < 0\,. \tag{3.13}$$

Then

$$\frac{E(T,y)}{V}\,\frac{4\pi a^3}{3} - \frac{\varepsilon_u}{V}\,\frac{4\pi a^3}{3} - \frac{\varepsilon_{free}}{V}\,\frac{4\pi a^3}{3} < 0$$

or

$$\frac{E(T,y)}{2\mu_0 V}\,\frac{4\pi a^3}{3} - \frac{\varepsilon_u}{2\mu_0 V}\,\frac{4\pi a^3}{3} - \frac{\varepsilon_{free}}{2\mu_0 V}\,\frac{4\pi a^3}{3} < 0\,.$$

We consider

$$\sigma_{a^3}a = \frac{\varepsilon_{free}}{V}\,\frac{4\pi a^3}{3}$$

as critical energy of free quarks for lowest energy, the tension σ_{a^3} here is the volume tension. Therefore this critical energy

is transferred to the produced hadrons and photons. Using the quarks low energy

$$E_{low} = \frac{1}{1.9}\frac{7\pi^2}{240} g_q \mu_0^4 V x^4 \left[3.78 + \right.$$
$$+ 2\, x^{-2}\left(0.82 - 1.16y + 0.41y^2\right) +$$
$$\left. + x^{-4}\left(0.08 - 0.23y + 0.25y^2 - 0.12y^3 + 0.02y^4\right)\right].$$

With $x = T/\mu_0 \ll 1$, the confinement condition becomes

$$\frac{1}{2}\frac{1}{1.9}\frac{7\pi^2}{240}\mu_0^3 g_q\, x^4\, \frac{4\pi a^3}{3}\left[3.78 + \right.$$
$$+ 2x^{-2}\left(0.82 - 1.16y + 0.41y^2\right) +$$
$$\left. + x^{-4}\left(0.08 - 0.23y + 0.25y^2 - 0.12y^3 + 0.02y^4\right)\right] -$$
$$- \frac{\varepsilon_u}{2\mu_0 V}\frac{4\pi a^3}{3} - \frac{\varepsilon_{free}}{2\mu_0 V}\frac{4\pi a^3}{3} < 0\,.$$

It becomes

$$\frac{1}{2}\frac{1}{1.9}\frac{7\pi^2}{240} g_q \frac{4\pi(\mu_0 a)^3}{3}\left[3.78x^4 + \right.$$
$$+ 2x^2\left(0.82 - 1.16y + 0.41y^2\right) +$$
$$\left. + \left(0.08 - 0.23y + 0.25y^2 - 0.12y^3 + 0.02y^4\right)\right] -$$
$$- \frac{4\pi}{3}\times\frac{1.41}{y} - \frac{4\pi}{3}\times\frac{1.43}{y} < 0\,.$$

We had the relation

$$(\mu_0 a)^3 = \frac{8\times 9\pi}{g_q}\,,$$

therefore, the condition becomes

$$\frac{1}{2}\frac{1}{1.9}\frac{7\pi^2}{240}\frac{4\times 8\times 9\pi^2}{3}\left[3.78x^4 + \right.$$
$$+ 2\,x^2\left(0.82 - 1.16y + 0.41y^2\right)$$
$$\left. + \left(0.08 - 0.23y + 0.25y^2 - 0.12y^3 + 0.02y^4\right)\right] -$$
$$- \frac{4\pi}{3}\times\frac{1.41}{y} - \frac{4\pi}{3}\times\frac{1.43}{y} < 0\,.$$

It becomes

$$3.78x^4 + 2x^2\left(0.82 - 1.16y + 0.41y^2\right) +$$
$$+ \left(0.08 - 0.23y + 0.25y^2 - 0.12y^3 + 0.02y^4\right) - \quad (3.14)$$
$$- 0.16y^{-1} < 0$$

with the curve of Fig. 11.

The critical situation x_c with $y \to 1$ (the end of the extreme cooling)

$$3.78x_c^4 + 2\times 0.07x_c^2 - 0.16574 = 0 \to x_c = 0.438\,.$$

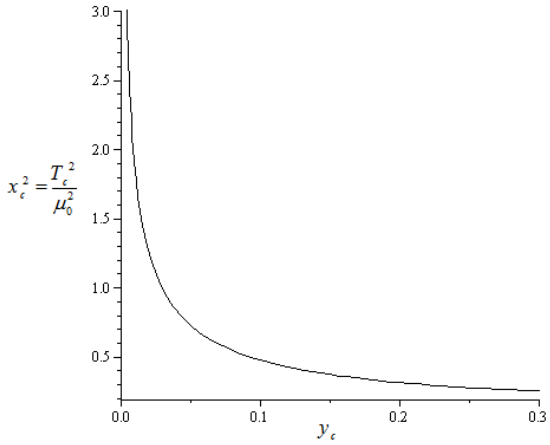

Fig. 11: The critical $x_c^2 y_c$ curve separates the free and confinement quarks phases.

So the critical temperature of the confinement condition when $y \to 1$ from $x_c = T_c/\mu_0$ is $T_c = 0.438\mu_0$. We determine μ_0 from

$$(\mu_0 a)^3 = \frac{8\times 9\pi}{g_q}$$

when $y \to 1$ so $a \to a_0$ we set $1/a_0$ = pion mass = (135 − 140) Mev so therefore, the condition becomes

$$\mu_0 = \frac{1}{a}\left(\frac{8\times 9\pi}{g_q}\right)^{\frac{1}{3}} \to 135\left(\frac{8\times 9\pi}{24}\right)^{\frac{1}{3}}$$
$$= 285.15\,\text{Mev} \quad \text{for}\ \frac{1}{a_0} = 135\,\text{Mev}. \quad (3.15)$$

Hence the critical temperature is $T_c = 0.438 \times 285.15 = 124.9$ Mev.

Now we try to find the produced hadrons, after covering the potential (3.12), the quarks critical energy (possible smallest energy) E_{free} (3.9) is transferred to the produced hadrons and photons. The key idea here is: because the cooling is an extreme cooling, it is expanding $a : 0 \to a_0 = 1/(135 - 140\,\text{Mev})$ so this process is thermally isolated from the other fields (adiabatic change), therefore the produced particles are in $T_c = 124.9$ Mev. We assume that the produced particles are hadrons (fermions and bosons) and photons. When $a : 0 \to a_0 : y \to 1$ the pions become massive $m = 1/a_0$ so we expect the other hadrons become massive at this stage, we assume that is in $T \to T_c$.

Therefore we assume when $T > T_c$ massless hadrons and $T < T_c$ massive hadrons. Anyway in $x_c y_c$ curve we find the confinement is possible at high energy ($T \gg T_c : a \to 0$). First we write using (3.9)

$$\frac{\varepsilon_{free}}{V} = \frac{\varepsilon_{free,a^3}}{4\pi a^3/3} = g_q\frac{\mu_0^4}{4\cdot 2\pi^2} \quad (3.16)$$
$$= \frac{\sigma_{a^3}\, a}{4\pi a^3/3} \to \frac{E_{hadrons} + E_{photons}}{V}\ \text{below}\ x_c y_c\ \text{curve}$$

or

$$\frac{\sigma_{a^3}a}{4\pi a^3/3} = g_q\frac{\mu_0^4}{4\cdot 2\pi^2} \to \varepsilon_f + \varepsilon_b + \varepsilon_{ph}\,.$$

With the densities

$$\varepsilon_f = \frac{E_f}{V}\,,\ \varepsilon_b = \frac{E_b}{V} \text{ and } \varepsilon_{ph} = \frac{E_{ph}}{V}$$

for spin 1/2 hadrons (fermions p^+, p^-, n), spin 0 hadrons (bosons π^0, π^-, π^+) and photons densities. For massless phase $T \gg T_c$ and $y_c \approx 0$ ignoring the chemical potential we have

$$\begin{aligned} n_f &= \frac{N_f}{V} = g_f\,\frac{3\zeta(3)}{4\pi^2}\,T^3\,,\\ n_b &= \frac{N_b}{V} = g_b\,\frac{\zeta(3)}{\pi^2}\,T^3 \text{ and}\\ \varepsilon_{ph} &= \frac{E_{ph}}{V} = g_{ph}\,\frac{\pi^2}{30}\,T^4\,. \end{aligned} \tag{3.17}$$

Now the key point, because the cooling is extreme cooling, to take all the particles (quarks) from high temperature and put them at low temperature, so the same structure at high energy will be at low energies, like the charges ratios, energy distribution over the particles, spins, ... At $T \to T_c$ and $y_c = 1$ the hadrons become massive, we approximate: for bosons (pions with mass $1/a_0 = 135 - 140$ Mev) the energy density becomes:

$$\begin{aligned} &\varepsilon_b = g_b\,\frac{\pi^2}{30}\,T^4 \to \varepsilon_b = g_b\,\frac{\pi^2}{30}\,T^4 + m_{pion}n_b\\ &\text{with } n_b = \frac{N_b}{V} = g_b\,\frac{\zeta(3)}{\pi^2}\,T^3 \text{ and } m_{pion} = \frac{1}{a_0}\,. \end{aligned}$$

And for fermions (let them be p^+, p^-, n) we approximate (ignoring the chemical potential)

$$\begin{aligned} &\varepsilon_f = g_f\,\frac{7}{8}\,\frac{\pi^2}{30}\,T^4 \to \varepsilon_f = g_f\,\frac{7}{8}\,\frac{\pi^2}{30}\,T^4 + m_f n_f\\ &\text{with } n_f = \frac{N_f}{V} = g_f\,\frac{3\zeta(3)}{4\pi^2}\,T^3\,. \end{aligned}$$

So (3.15) becomes

$$\begin{aligned} &\frac{\sigma_{a^3}a}{4\pi a^3/3} = g_q\,\frac{\mu_0^4}{4\cdot 2\pi^2} = \varepsilon_f + \varepsilon_b + \varepsilon_{ph}\\ &= g_f\,\frac{7}{8}\,\frac{\pi^2}{30}T_c^4 + m_f n_f + g_b\,\frac{\pi^2}{30}T_c^4 + \frac{1}{a_0}n_b + g_{ph}\,\frac{\pi^2}{30}T_c^4\,. \end{aligned} \tag{3.18}$$

with $g_{quarks} = 2_{flavor}\times 2_{charge}\times 2_{spin}\times 3_{color}$, $g_f = 3_{charge}\times 2_{spin}$, $g_b = 3_{charge}$ and $g_{ph} = 2_{polarization}$.

Now we calculate (3.17) for $1/a_0 = 135\,\text{Mev}\,(\pi^0)$, $\mu_0 = 285.15$ Mev, and $T_c = 124.9$ Mev we have

$$\begin{aligned} &2.0096\times 10^9\,\text{Mev}^4 = 6\times\frac{7}{8}\,\frac{\pi^2}{30}(124.9)^4 +\\ &+\ m_f 6\times\frac{3\zeta(3)}{4\pi^2}(124.9)^3 + 3\times\frac{\pi^2}{30}(124.9)^4 +\\ &+ 135\times 3\times\frac{\zeta(3)}{\pi^2}(124.9)^3 + 2\times\frac{\pi^2}{30}(124.9)^4\,. \end{aligned}$$

Its solution is $m_f = 1023$ Mev. We keep 2.0096×10^9 Mev^4 as smallest possible energy density.

For $1/a_0 = 140$ Mev, $\mu_0 = 295.7$ Mev so $T_c = 129.5$ Mev the mass m_f becomes $m_f = 798.4$ Mev. Therefore it must be $135\,\text{Mev} < 1/a_0 < 140\,\text{Mev}$.

For $1/a_0 = 136.8$ Mev we have $T_c = 126.56$ Mev then the mass m_f becomes $m_f \approx 938$ Mev so the fermions (hadrons) are the baryons (p^+, p^-, n).

Therefore we fix it $1/a_0 = 136.8$ Mev, we use it to cancel the dark matter. Maybe there is an external pressure $-P_{ex}$ so the lost energy is $P_{ex}\,4\pi a^3/3$.

Now we try to calculate the ratio N_q/N_h. From the condensation relation

$$N_q\delta\mu_q + N_h\delta\mu_h = 0$$

N_h is the hadrons (consider only the fermions) and μ_h is their chemical potential.

We assumed before the relation for the quarks chemical potential

$$\mu(r) = \mu_0 + u(r)\ \text{ with } u(r) = -\frac{\alpha_s}{r}\left(1 - e^{-r/a}\right)$$

$$\text{so } \delta\mu_q(r) = u(r) = -\frac{\alpha_s}{r}\left(1 - e^{-r/a}\right)\,.$$

The effect of this changing appeared in $y = \alpha_s/a\mu_0$ in the results. For the hadrons we have

$$\delta\mu_h = -\frac{N_q}{N_h}\,\delta\mu_q = -\frac{N_q}{N_h}\,u(r)\,.$$

That is right if we consider the hadrons are massless, that is when $T \gg T_c$ and $y \ll 1$ (in the condensation phase, below the curve $x_c y_c$) so we have the chemical potential for the hadrons

$$\mu_h(r) = \mu_{0h} - u(r)\ \text{ with } u(r) = -\frac{\alpha_s}{r}\left(1 - e^{-r/a}\right)$$

therefore we replace $y \to (-N_q\mu_{0q}/N_h\mu_{0h})\,y$ in the quarks energy to get the hadrons energy. The energy of the hadrons becomes

$$\begin{aligned} &E_{H,low} = \frac{1}{1.9}\,\frac{7\pi^2}{240}\,\mu_{0h}^4\,g_h\,V\,x^4\Bigg[3.78 +\\ &+ 2x^{-2}\left(0.82 + 1.16\left(\frac{N_q\mu_{0q}}{N_h\mu_{0h}}\right)y + 0.41\left(\frac{N_q\mu_{0q}}{N_h\mu_{0h}}\right)^2 y^2\right) +\\ &+ x^{-4}\left(0.08 + 0.23\left(\frac{N_q\mu_{0q}}{N_h\mu_{0h}}\right)y + 0.25\left(\frac{N_q\mu_{0q}}{N_h\mu_{0h}}\right)^2 y^2 +\right.\\ &\left.+ 0.12\left(\frac{N_q\mu_{0q}}{N_h\mu_{0h}}\right)^3 y^3 + 0.02\left(\frac{N_q\mu_{0q}}{N_h\mu_{0h}}\right)^4 y^4\right)\Bigg]. \end{aligned}$$

Assume $\mu_{0h} = \mu_{0q}$ and $y = 1$ so

$$E_{H,low} = \frac{1}{1.9}\,\frac{7\pi^2}{240}\,\mu_{0q}^4\, g_h\, V\left[3.78x^4 + 2x^2\left(0.82 + \right.\right.$$
$$+1.16\left(\frac{N_q}{N_h}\right) + 0.41\left(\frac{N_q}{N_h}\right)^2\Big) + \left(0.08 + 0.23\left(\frac{N_q}{N_h}\right) + \right.$$
$$\left.\left. + 0.25\left(\frac{N_q}{N_h}\right)^2 + 0.12\left(\frac{N_q}{N_h}\right)^3 + 0.02\left(\frac{N_q}{N_h}\right)^4\Big)\right].$$

So the chemical potential μ_h of the hadrons becomes

$$\mu_h^4 = \mu_{0q}^4\left(1 + \frac{0.23}{0.08}\left(\frac{N_q}{N_h}\right) + \frac{0.25}{0.08}\left(\frac{N_q}{N_h}\right)^2 + \right.$$
$$\left. + \frac{0.12}{0.08}\left(\frac{N_q}{N_h}\right)^3 + \frac{0.02}{0.08}\left(\frac{N_q}{N_h}\right)^4\right).$$

When $T < T_c$ the hadrons become massive, as we assumed before, so for massive hadrons with $m_f = 938$ Mev we expect $\mu_h = m_f = 938$ Mev when they cooled with small densities. Therefore

$$(938)^4 = (285.15)^4\left(1 + \frac{0.23}{0.08}\left(\frac{N_q}{N_h}\right) + \frac{0.25}{0.08}\left(\frac{N_q}{N_h}\right)^2 + \right.$$
$$\left. + \frac{0.12}{0.08}\left(\frac{N_q}{N_h}\right)^3 + \frac{0.02}{0.08}\left(\frac{N_q}{N_h}\right)^4\right).$$

Its positive solution is $N_q/N_h = 3.1$ so they are the baryons (fermions with three quarks). For 0 temperature fermions the chemical potential is approximated by

$$\mu_0^2 = m^2 + \left(\frac{N}{V}\frac{6\pi^2}{g_f}\right)^{2/3}.$$

For low hadrons density we ignored the term

$$\left(\frac{N}{V}\frac{6\pi^2}{g_f}\right)^{2/3}.$$

4 The nuclear compression

The cooled hadrons have high density, so there is hidden high pressure, that pressure makes influence δa so δy near $y = 1$ or it makes $y = 1 + \delta y$: $\delta y \approx 0.005$ so the cooled quarks inside the hadrons fluctuate, this depends on the energy, if the energy is high then there are new hadrons. These processes let the interacting hadrons lose kinetic energy and form the pions.

Because the number of quarks increases although the hadrons are fixed, therefore the hadrons energy decreases and they cannot spread away. We can see how the chemical potential of the interacting hadrons changes under the fluctuation $\delta y \sim \delta a$ (due to the quarks interaction) from the condensation relation $N_q\delta\mu_q + N_h\delta\mu_h = 0$ we have $\delta\mu_h = -N_q\delta\mu_q/N_h$ for the fluctuation δy we have

$$\delta\mu_h = -\frac{N_q}{N_h}\frac{\partial\mu_q}{\partial y}\delta y$$

from quarks chemical potential (4.4), we find

$$\frac{\partial\mu_q}{\partial y} < 0 \quad \text{so} \quad -\frac{\partial\mu_q}{\partial y} > 0$$

therefore we have

$$\delta\mu_h = \frac{N_q}{N_h}\left(-\frac{\partial\mu_q}{\partial y}\right)\delta y < 0 \quad \text{when} \quad \delta y < 0$$

which is the quarks compressing, when the hadrons collide together this leads to $\delta y < 0$ (compression) so the hadrons lose energy and new hadrons are created. And when they try to extend (spread away) $\delta y > 0$ so $\delta\mu_h > 0$, there will be a negative potential.

For the interacting hadrons pressure we have the phase changing relation $V_q\delta P_q + V_h\delta P_h = 0 : V$ volume, we have

$$\delta P_h = -\frac{V_q}{V_h}\,\delta P_q = -\frac{V_q}{V_h}\,\frac{\partial P_q}{\partial y}\,\delta y$$

because $\partial P_q/\partial y < 0 \to -\partial P_q/\partial y > 0$ therefore when the hadrons collide together $\delta y < 0$ so their pressure decreases, they lose energy, so new hadrons are created.

We have

$$\delta y = \left(-\frac{V_q}{V_h}\frac{\partial P_q}{\partial y}\right)^{-1}\delta P_h \quad \text{at} \quad y = 1\,.$$

So the hadrons chemical potential becomes

$$\delta\mu_h = \frac{N_q}{N_h}\left(-\frac{\partial\mu_q}{\partial y}\right)\left(-\frac{V_q}{V_h}\frac{\partial P_q}{\partial y}\right)^{-1}\delta P_h \;:\; y = 1\,.$$

It becomes

$$\delta\mu_h = \frac{N_q V_h}{N_h V_q}\left(\frac{\partial\mu_q}{\partial y}\right)\left(\frac{\partial P_q}{\partial y}\right)^{-1}\delta P_h \;:\; y = 1\,. \tag{4.1}$$

We can relate this changing to a constant nuclear potential. Like to write

$$\delta\mu_h = -V_0\,. \tag{4.2}$$

V_0 is the potential for each hadron.

So when the hadron (fermions, like protons or neutrons) join, their density increases $\delta\mu_h > 0$ so their pressure rises $\delta P_h > 0$, therefore there is a negative potential $V_0 < 0$. At low energies this potential prevents them from spreading away.

Now we calculate

$$\delta\mu_h = \frac{N_q V_h}{N_h V_q}\left(\frac{\partial\mu_q}{\partial y}\right)\left(\frac{\partial P_q}{\partial y}\right)^{-1}\delta P_h \;:\; y = 1\,.$$

We use the pressure at low energy (3.7)

$$P_{low}/\mu_0^4 = (9\times 1.9\times 240)^{-1}\,7\pi^2 g_q\left[3\times 3.78x^4+\right.$$
$$+3\times 2\times x^2\left(0.82-1.16y+0.41y^2\right)+3\,(0.08-0.23y+$$
$$+0.25y^2-0.12y^3+0.02y^4)+2yx^2\,(-1.16+0.82y)+$$
$$\left.+y\left(-0.23+0.5y-0.36y^2+0.08y^3\right)\right]$$

and we get

$$\frac{\partial P_q}{\partial y} = -\frac{0.076\times 7\times\pi^2\times g_q\mu_0^4}{240\times 3\times 1.9} : x_c = 0.438\,,\ y=1\,. \quad (4.3)$$

Using the relation

$$\mu_0 = \frac{1}{a_0}\left(\frac{8\times 9\pi}{g_q}\right)^{\frac{1}{3}}$$

we have

$$\mu_0 = 285.15\,\text{Mev}\ \text{ for } g_q = 24 \text{ and } 1/a_0 = 135\,\text{Mev}\ \text{ for } \pi^0$$
$$\mu_0 = 295.7\,\text{Mev}\ \text{ for } 1/a_0 = 140\,\text{Mev}\ \text{ for } \pi^- \text{ and } \pi^+\,.$$

So the chemical potential μ_0 is in the range from 285.15 Mev to 295.7 Mev therefore

$$\frac{\partial P_q}{\partial y} = -6.06\times 10^8\,\text{Mev}^4\ \text{ for } \mu_0 = 285.15\,\text{Mev}$$
$$\frac{\partial P_q}{\partial y} = -7.01\times 10^8\,\text{Mev}^4\ \text{ for } \mu_0 = 295.7\,\text{Mev}\,.$$

Now we try to calculate $\partial\mu_q/\partial y$, according to low energy

$$E_{low} = (1.9\times 240)^{-1}\,7\pi^2\,\mu_0^4\,g_q\,V\,x^4\left[3.78+\right.$$
$$+2x^{-2}\left(0.82-1.16y+0.41y^2\right)$$
$$\left.+x^{-4}\left(0.08-0.23y+0.25y^2-0.12y^3+0.02y^4\right)\right]$$

we can equivalence

$$\mu^4 = \mu_0^4\left(1-\frac{0.23}{0.08}y+\frac{0.25}{0.08}y^2-\frac{0.12}{0.08}y^3+\frac{0.02}{0.08}y^4\right). \quad (4.4)$$

But $\partial\mu/\partial y\to\infty$ when $y\to 1$ so we replace

$$\frac{\partial\mu_q}{\partial y}\to\frac{\mu_{y=1}-\mu_{y=0}}{1-0} = \frac{0-\mu_{y=0}}{1-0} = -\mu_{y=0} = -\mu_0\,.$$

Therefore we have

$$\delta\mu_h = \frac{N_q V_h}{N_h V_q}\left(\frac{\partial\mu_q}{\partial y}\right)\left(\frac{\partial P_q}{\partial y}\right)^{-1}\delta P_h$$
$$= \frac{N_q V_h}{N_h V_q}\mu_0\left(0.09\mu_0^4\right)^{-1}\delta P_h$$
$$= \frac{N_q V_h}{N_h V_q}\left(0.09\mu_0^3\right)^{-1}\delta P_h\,.$$

So we have

$$\delta\mu_h = \frac{N_q V_h}{N_h V_q}\left(0.09\,\mu_0^3\right)^{-1}\delta P_h$$
$$= 4.7\times 10^{-7}\frac{N_q V_h}{N_h V_q}\,\delta P_h\ \text{ for } \mu_0 = 285.15\,\text{Mev}$$

and

$$\delta\mu_h = 4.2\times 10^{-7}\frac{N_q V_h}{N_h V_q}\,\delta P_h\ \text{ for } \mu_0 = 295.7\,\text{Mev}\,. \quad (4.5)$$

We use them to cancel the dark matter and dark energy.

5 The Big Bang

We assume there were two universal phases, high energies massless particles phase (let them be the quarks plasma) and then the massive low energies particles (let them be the hadrons).

The first phase associated with high energy density (drops from infinity to finite), the time of that stage is $\tau: 0\to a_0 = 1/(135-140)$ Mev then the massive hadrons phase begins (the time $t: 0\to\infty$).

In both stages the highest universal expansion must not exceed the light speed, for the first phase, high energies massless quarks phase, the density of the energy is the same in all space points so the universal expansion is the same in every point in the space, we let the speed of that expansion equal the light speed, therefore the Hubble parameter $H(t<a_0)$ of this stage $t<a_0$ is given by (5.2).

To find the Hubble parameter for the massive hadrons phase $H(t>a_0)$, we suggest the geometry transformation (5.3) in which the time $\tau: 0\to a_0$ for the quarks corresponds to the time $t: 0\to\infty$ for the massive hadrons phase. We can relate that change in the geometry to the high differences in the energy densities of the two phases. The phase $\tau: 0\to a_0$ high quarks energy, uniform high energy density, massless, ... The phase $t: 0\to\infty$ the massive hadrons, low energy density, separated particles, ...

Now we try to explain how the universe exploded and expanded, we start from our assumptions we made before and find the Hubble parameter and try to find the dark energy and matter. We found that the quarks expand to the length $a_0 = 1/(135-140)$ Mev then the hadrons appear instead.

We assume that the universe was created in every point in two dimensional space XY then the explosion in the Z direction. That is by the quarks, in each point in the XY plane the quarks were created and then they expanded in each point XY to the length a_0 then the explosion in the Z direction, the result is the universe in the space XYZ. There was no universal explosion in the XY plane, the universal explosion was only in the Z direction, in the plane XY there was extension due to the quarks expanding from $r=0$ to $r=a_0 = 1/(135-140)$ Mev the plane XY was infinity before the quarks expansion and it is infinity after that expansion, what happened is an increase

in the number of the XY points, then the explosion in the Z direction. We assume both expansion (XY and Z) occurred with the light speed c.

To find the lost matter, dark matter and dark energy, we use the relation (4.5) we found before:

$$\delta\mu_h = \frac{N_q V_h}{N_h V_q}\left(0.09\,\mu_0^3\right)^{-1}\delta P_h$$
$$= 4.7\times 10^{-7}\,\frac{N_q V_h}{N_h V_q}\,\delta P_h \ \text{ for } \mu_0 = 285.15\,\text{Mev}$$

and

$$\delta\mu_h = 4.2\times 10^{-7}\,\frac{N_q V_h}{N_h V_q}\,\delta P_h \ \text{ for } \mu_0 = 295.7\,\text{Mev}\,.$$

Here we relate this changing in the pressure δP (independent of time) to the hadrons condensation process to form the nucleuses, where the global pressure $\delta P = \delta P_h$ dropped extremely due to the nuclear attractive potential (make it the nuclear binding energy) $V_0 = (-7-8)$ Mev [3]. This pressure δP_h remains contained in the nucleuses, but globally is not visible.

So there is hidden global pressure δP_h and we have to include that problem in the Friedman equations solutions, we notice that the nuclear attractive potential leads to increasing in the cooled hadrons densities. Therefore the decreasing in the hadrons pressure associated with the increasing of their densities (inside the nucleuses). The result is excess in the local energy density, that effects appear in the equations, that is, the matter density appears to be larger than the right energy density. So there is neither dark matter nor dark energy, it is just global and local densities.

We start from the definition of the scale parameter $R(t)$ for the universe expansion, we write [6]

$$ds^2 = -dt^2 + R^2(t)\left(\frac{dr^2}{1-kr^2} + r^2 d\Omega^2\right). \qquad (5.1)$$

We set $k = 0$ flat Universe. Now we try to find the Hubble parameter

$$H(t) = \frac{1}{R(t)}\,\frac{dR(t)}{dt} = \frac{\dot R(t)}{R(t)}\,.$$

There are two phases $t < a_0$ free quarks phase and $t > a_0$ hadrons phase which is the expansion in the Z direction. That means there are two different spacetime geometry, $t < a_0$ and $t > a_0$.

In the first phase $\tau = t < a_0$ the expansion is the same in all space points, so the expansion velocity

$$\frac{dR_1}{dt} = \dot R(t)\,r$$

is the same in all space points and equals the light speed $c = \hbar = 1$ here, so

$$1 = \dot R(t) r : \ t < a = a_0\,.$$

Therefore

$$\dot R(t) = \frac{1}{r} : \ t < a = a_0\,.$$

So we can write

$$R(t) = \frac{t}{r} : \ t < a = a_0\,.$$

So the Hubble parameter becomes

$$H(t) = \frac{\dot R(t)}{R(t)} = \frac{1/r}{t/r} = \frac{1}{t} : \ t < a = a_0\,. \qquad (5.2)$$

Now we want to find the Hubble parameter in the phase $t > a_0$ low energy phase. Actually when the quarks expand from $r = 0$ to $r = a \to a_0$ there will be infinity points expanding, so infinity expanding distance in XY space, but the expansion cannot exceed the light speed $c = 1$ therefore an explosion occurs in the Z direction, so the universal explosion. Therefore the time $t = \tau : 0 \to a_0$ for the free quarks phase will associate with $t : 0 \to \infty$ for the universal expansion, so we make the geometry transformation

$$t = \frac{-c_0}{\tau - a_0} : \ \tau < a_0\,. \qquad (5.3)$$

c_0 is constant, we can relate that relation to a difference in spacetime geometry. That means if the quarks space $r < a_0 = 1/(135-140)$ Mev is flat, so the hadrons space is not, it is curved space, where we live. It is convenient to consider the quarks space ($r < a_0$ large energy density) is curved not our space (low energy density).

Now we can find the Hubble parameter for the universe $t : 0 \to \infty$. We can find the Hubble parameter $H(t > a_0)$ for the geometry $t : 0 \to \infty$ from $H(t < a_0)$:

$$H(\tau < a) = \frac{1}{R(\tau < a)}\,\frac{dR(\tau < a)}{d\tau}$$
$$= \frac{1}{f(r,\theta,\varphi)R(t > a)}\,\frac{d}{d\tau} f(r,\theta,\varphi)R(t > a)\,.$$

We set the geometry transformation

$$R(\tau < a) = f(r,\theta,\varphi)\,R(t > a)$$

so

$$\frac{1}{\tau} = \frac{1}{f(r,\theta,\varphi)R(t > a)}\,\frac{d}{d\tau} f(r,\theta,\varphi)R(t > a)$$
$$= \frac{1}{R(t > a)}\,\frac{d}{d\tau}R(t > a) \qquad (5.4)$$

or

$$\frac{1}{\tau} = \frac{1}{R(t > a)}\,\frac{dt}{d\tau}\,\frac{d}{dt}R(t > a)$$
$$= \frac{dt}{d\tau}\,\frac{1}{R(t > a)}\,\frac{d}{dt}R(t > a) = \frac{dt}{d\tau}H(t > a)\,.$$

Using the geometry transformation

$$t = \frac{-c_0}{\tau - a_0} : \tau < a \to a_0 ,$$

we have the Hubble parameter of the low energy density of the cold Universe

$$H(t > a_0) = \frac{1}{R(t > a_0)} \frac{d}{dt} R(t > a_0)$$
$$= \frac{c_0}{t(a_0 t - c_0)} = \frac{1}{t\left(\frac{a_0}{c_0} t - 1\right)} = \frac{1}{t\left(c'_0 t - 1\right)}$$

where c'_0 is constant.

The Friedman equations can be written, for $k = 0$, like [6]

$$3\frac{\dot{R}^2(t)}{R^2(t)} = 8\pi G_N \rho + \Lambda \qquad (1)$$
$$-\frac{\ddot{R}(t)}{R(t)} + \frac{\dot{R}^2(t)}{R^2(t)} = 4\pi G_N(\rho + p) \qquad (2) \qquad (5.5)$$
$$\frac{d}{dt}(\rho + \delta p) = -3(\rho + p)\frac{\dot{R}(t)}{R(t)} . \qquad (3)$$

To control (or cancel) the dark matter and energy, we make the transformations in the Friedman equations which keep the Hubble parameter unchanged

$$3\frac{\dot{R}^2(t)}{R^2(t)} = 8\pi G_N(\rho + \delta P) + \Lambda - 8\pi G_N \delta P \qquad (1')$$
$$-\frac{\ddot{R}(t)}{R(t)} + \frac{\dot{R}^2(t)}{R^2(t)} = 4\pi G_N(\rho + \delta P + p - \delta P) \qquad (2') \qquad (5.6)$$
$$\frac{d}{dt}(\rho + \delta p) = -3(\rho + \delta P + p - \delta P)\frac{\dot{R}(t)}{R(t)} . \qquad (3')$$

So we have (for same Hubble parameter we had before)

$$\rho' = \rho + \delta p_h$$
$$p' = p - \delta p_h$$
$$\Lambda' = \Lambda - 8\pi G_N \delta p_h = 0 .$$

For the universal nuclear condensation, we assume the universal change $\delta\rho = \delta P = \delta p_h > 0$ is independent of the time.

We can say ρ', p' and $\Lambda' = 0, P' = 0$ are for the located matter, when the hadrons are cooled, they condense and locate in small volumes with high matter density, because of the strong nuclear attractive interaction, so their pressure decreases extremely $P' \approx 0$. That pressure is contained (hidden) in the nucleus. It is like to condense a gas with certain mass m and fixed volume V, the density m/V is the same before and after the condensation, but the real density of the produced liquid is not. Like that we consider ρ the right matter ρ_{matter} and the problems; the increasing $\rho' = \rho + \delta p_h$ and $\Lambda \neq 0$ are because of the phase changing.

We set $\rho' = \rho(t)$ and solve the two equations:

$$-\frac{\ddot{R}(t)}{R(t)} + \frac{\dot{R}^2(t)}{R^2(t)} = 4\pi G_N \rho(t) \qquad (2')$$
$$\frac{d}{dt}(\rho + \delta p) = \dot{\rho}(t) = -3\rho(t)\frac{\dot{R}(t)}{R(t)} \qquad (3')$$

using the Hubble parameter

$$H(t) = \frac{1}{R}\frac{dR}{dt} = \frac{1}{t\left(c'_0 t - 1\right)} \qquad : \quad t > a_0 .$$

From (3') we have

$$-\frac{1}{3}\frac{R(t)}{\dot{R}(t)}\dot{\rho}(t) = \rho(t)$$

so (2') becomes

$$-\frac{\ddot{R}(t)}{R(t)} + \frac{\dot{R}^2(t)}{R^2(t)} = -\frac{4\pi G_N}{3}\frac{R(t)}{\dot{R}(t)}\dot{\rho}(t) .$$

This equation becomes

$$\frac{\dot{R}(t)}{R(t)}\left(-\frac{\ddot{R}(t)}{R(t)} + \frac{\dot{R}^2(t)}{R^2(t)}\right) = -\frac{4\pi G_N}{3}\dot{\rho}(t)$$

or

$$H(t)\left(-\frac{\ddot{R}(t)}{R(t)} + H^2(t)\right) = -\frac{4\pi G_N}{3}\dot{\rho}(t) .$$

Using

$$\frac{d}{dt}\frac{\dot{R}(t)}{R(t)} = \frac{\ddot{R}(t)}{R(t)} - \frac{\dot{R}^2(t)}{R^2(t)}$$

we get

$$H(t)\frac{d}{dt}H(t) = \frac{4\pi G_N}{3}\dot{\rho}(t) \to \frac{1}{2}H(t)^2 = \frac{4\pi G_N}{3}(\rho(t) - \rho_0) .$$

For finite results we put $\rho_0 = 0$ so

$$\frac{1}{2}H(t)^2 = \frac{4\pi G_N}{3}\rho(t) .$$

Now we calculate the contributions of the vacuum energy to the total energy using the cosmological constant Λ' from (1')

$$\Omega_{\Lambda'} = \frac{\rho'_\Lambda}{\rho_c} = \frac{\Lambda'}{3H^2} = \frac{3H^2 - 8\pi G_N \rho(t)}{3H^2}$$
$$= 1 - 2\frac{4\pi G_N}{3H^2}\rho(t) = 1 - 2\frac{1}{H^2}\frac{1}{2}H(t)^2 = 0$$

with the critical energy density

$$\rho_c = \frac{3H^2}{8\pi G_N} .$$

So the vacuum energy density is canceled, and the total energy is the matter energy $\Omega_{matter} = 1$ so $\rho(t)/\rho_c = 1$. Here

$\rho(t) = \rho_c$ is $\rho(t) = \rho' = \rho_{matter} + \delta p_h$, so $\rho(t)$ is higher than the right matter ρ_{matter}.

Now we see if this relation is satisfied or not. We use the global change on the pressure $\delta p = \delta p_h > 0$ which we derived in (4.5):

$$\delta\mu_h = \frac{N_q V_h}{N_h V_q}\left(0.09\mu_0^3\right)^{-1}\delta P_h$$
$$= 4.7\times 10^{-7}\,\frac{N_q V_h}{N_h V_q}\,\delta P_h \ \text{ for } \mu_0 = 285.15\,\text{Mev}$$

and

$$\delta\mu_h = 4.2\times\ 10^{-7}\,\frac{N_q V_h}{N_h V_q}\,\delta P_h \ \text{ for } \mu_0 = 295.7\,\text{Mev}\,.$$

Now we try to find V_q/V_h the quarks volume $V_q = S\,d_q$ and the hadrons volume $V_h = S\,d_h$ as shown in Fig. 12 where the

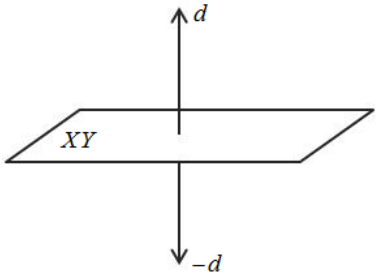

Fig. 12: The universal explosion in Z direction starting from XY flat.

universal explosion is in the $Z = d$ direction. If we assume the explosion speed is the same for both hadrons and quarks, light speed $v = c = 1$, so for the quarks

$$H_q(t) = \frac{\dot R(t)_q}{R(t)_q} = \frac{v_q}{d_q} = \frac{1}{d_q}\,.$$

For the hadrons

$$H_h(t) = \frac{\dot R(t)_h}{R(t)_h} = \frac{v_h}{d_h} = \frac{1}{d_h}$$

therefore

$$\frac{V_q}{V_h} = \frac{S\,d_q}{S\,d_h} = \frac{d_q}{d_h} = \frac{H_h}{H_q}\,,$$

H_h is the universal Hubble parameter, today is

$$H = 71\,\text{km/s/mpc} = 2.3\times 10^{-18}\,s^{-1}$$
$$= 2.3\times 10^{-18}\times 6.58\times 10^{-22}\,\text{Mev} = 151.34\times 10^{-41}\,\text{Mev}\,.$$

The quarks Hubble parameter $H_q = 1/\tau\to 1/a_0 = (135 - 140)$ Mev. So we have (for 135 Mev)

$$\frac{H_h}{H_q} = \frac{151.34\times 10^{-41}\,\text{Mev}}{135\,\text{Mev}} = 1.127\times 10^{-41}\,.$$

Therefore

$$\frac{V_q}{V_h} = \frac{H_h}{H_q} = 1.127\times 10^{-41}\,.$$

We set $\delta\mu_h = -V_0 = (7 - 8)$ Mev the nuclear potential (nucleon binding energy). Therefore, from (4.5), we have

$$\delta\rho = \delta P_h = -\frac{N_h}{N_q}\times 1.127\times 10^{-41}\times\frac{-V_0}{47}\times 10^8\,\text{Mev}^4$$
$$\text{for}\ \frac{1}{a_0} = 135\,\text{Mev} : \mu_0 = 285.15\,\text{Mev}$$

and

$$\delta\rho = \delta P_h = -\frac{N_h}{N_q}\times 1.087\times 10^{-41}\times\frac{-V_0}{42}\times 10^8\,\text{Mev}^4$$
$$\text{for}\ \frac{1}{a_0} = 140\,\text{Mev} : \mu_0 = 295.7\,\text{Mev}\,.$$

For $N_h/N_q = 1/5$, like the interaction $P^+ + \pi^- \to n$ the neutron n appears to have five quarks, that is acceptable according to the fields dual behavior. Therefore

$$\delta\rho = \delta P_h = -\frac{1}{5}\times 1.127\times 10^{-41}\times\frac{-7}{47}\times 10^8\,\text{Mev}^4$$
$$= 335.7\times 10^{-37}\,\text{Mev}^4$$
$$\text{for}\ \mu_0 = 285.15\,\text{Mev and}\ V_0 = -7\,\text{Mev}$$

and

$$\delta\rho = \delta P_h = -\frac{1}{5}\times 1.087\times 10^{-41}\times\frac{-8}{42}\times 10^8\,\text{Mev}^4$$
$$= 414\times 10^{-37}\,\text{Mev}^4$$
$$\text{for}\ \mu_0 = 295.7\,\text{Mev and}\ V_0 = -8\,\text{Mev}.$$

So the change $\delta\rho = \delta P_h$ is in the range:

$$\text{from}\ 335.7\times 10^{-37}\,\text{Mev}^4\ \text{to}\ 414\times 10^{-37}\,\text{Mev}^4\,.$$

Therefore the visible matter is in the range

$$\text{from}\ \rho_{matter} = \rho_c - \delta p_h = 335.7\times 10^{-37}\,\text{Mev}^4$$
$$\text{to}\ \rho_{matter} = \rho_c - \delta p_h = 414\times 10^{-37}\,\text{Mev}^4\,.$$

For the critical energy $\rho_c = 406\times 10^{-37}$ Mev^4 the visible matter is in the range

$$\text{from}\ \rho_{matter} = 0\ \text{to}\ \rho_{matter} = 70\times 10^{-37}\,\text{Mev}^4\,.$$

The right baryonic matter energy density is

$$\rho_b = 4.19\times 10^{-31}\,\text{g/cm}^3 \approx 17.97\times 10^{-37}\,\text{Mev}^4$$

which belongs to the range 0 to 70×10^{-37} Mev^4. We can control this and have

$$\rho_{matter} = \rho_c - \delta P_h = 406\times 10^{-37} - \delta P_h = 17.97\times 10^{-37}\,\text{Mev}^4$$

by finding r:

$$140r + 135(1 - r) = \frac{1}{a_0}$$

where $1/a_0$ satisfies

$$406 \times 10^{-37}\,\text{Mev}^4 - \delta P_h = 17.97 \times 10^{-37}\,\text{Mev}^4.$$

For $1/a_0 = 136.8\,\text{Mev}$ (we used it in (3.17) to have $m_f \approx 938\,\text{Mev}$), the chemical potential becomes $\mu_0 = 288.95\,\text{Mev}$. And with $V_0 = 7.776\,\text{Mev}$ we get

$$\begin{aligned}\delta\rho = \delta P_h &= 335.7 \times 10^{-37} \times \left(\frac{288.95}{285.15}\right)^3 \times \frac{7.776}{7}\,\text{Mev}^4 \\ &= 388 \times 10^{-37}\,\text{Mev}^4\,.\end{aligned}$$

The matter density becomes

$$\begin{aligned}\rho_{matter} &= 406 \times 10^{-37}\,\text{Mev}^4 - 388 \times 10^{-37}\,\text{Mev}^4 \\ &= 17.9 \times 10^{-37}\,\text{Mev}^4\,.\end{aligned}$$

which is the right matter (global visible matter density). Therefore we can control the dark matter and dark energy. We can cancel them.

Note that not all of those ideas are contained in the references.

Submitted on March 16, 2016 / Accepted on March 20, 2016

References

1. Peskin M. E. and Schroeder D. V. An Introduction to Quantum Field Theory. Westview, Boulder, 1995.
2. Srednicki M. Quantum Field Theory. Cambridge University Press, Cambridge, 2006.
3. Martin B. R. Nuclear and Particle Physics. University College, London, 2006.
4. Brown L. S. Quantum Field Theory. Cambridge University Press, Cambridge, 1994.
5. Hooft G. Introduction to String Theory, version 14-05-04. Institute for Theoretical Physics, Utrecht University and Spinoza Institute, Netherlands, 2003–2004.
6. Blau M. Lecture Notes on General Relativity. Institut für Theoretische Physik, Universität Bern, CH-3012 Bern, Switzerland, 2012. http://www.blau.itp.unibe.ch/Lecturenotes.html.

On Quantization and the Resonance Paths

Jacques Consiglio

52 Chemin de Labarthe, 31600 Labastidette, France
E-mail: Jacques.Consiglio@gmail.com

We use a mass-resonance equation to analyze the known elementary particles mass spectrum; we first show that masses and charges are quantized together and all couplings are geometry of movement. Next, the long-expected connection between gravitation and the rest of physics appears as we deduce and compute from the equation parameters the resonance corresponding to the reduced Planck mass. In this way, quantum fields and general relativity can be emergent theories where the natural law is unique.

It is in the admission of ignorance and the admission of uncertainty that there is a hope for the continuous motion of human beings in some direction that doesn't get confined, permanently blocked, as it has so many times before in various periods in the history of man. R.P. Feynman.

1 Introduction

In a celebrated paper, Dirac [8] showed that the existence of magnetic poles and quantum mechanics imply symmetrical quantization of magnetic and electric charges. This is the very first attempt to explain the observation of a universal charge quantum. Since then other theories were produced in which the magnetic charge differs. But even though charges have definite symmetry nothing imposes the charge ratio; namely the fine structure constant α.

It is often believed that the standard model (SM) of particles physics is part of a wider theory in which its free parameters are calculable — but possibly free in essence or accepting multiple solutions. One can see the seeds of this line of thoughts in Dirac's quantization: once the idea is extended to all fields, it may structure the logical constraints in such a manner that the full set of equations can be solved. Such result is expected in super-symmetry and string theory.

However, we must remind that we discuss *the parameters of a theory, not a-priori of nature*. At the other extreme, assume quantum theory incomplete or not fully understood, a possibility exists that all *known* parameters are already calculable from known physics. If so, it may be possible to decode some field characteristics directly from known data. At present time, the only rich group of parameters is the elementary particles mass spectrum as we know 12 samples, and it may be enough to understand its underlying structure.

In short, and in a general manner:

- Assume the 12 known masses correspond to solutions of a set of unknown equations.
- In the most favorable case, if no other mass exists (or close enough) all degrees of freedom are used.
- Hence it may be possible to find or approach the equations and the structure of the solution.

The approach is subtler and a lot more risky than any other since instead of building on theoretical knowledge we assume ignorance — and we do not know what we do ignore.

The object of this paper is to prove the existence of a solution, probably unique, and one of the equations in which the solution is visible. One can infer its validity in two manners; firstly by its agreement with phenomenology, and secondly, by its logical coherence, compactness and simplicity.

In a suite of papers [3, 4], we showed how the mass spectrum is structured. We found firstly that the elementary particles mass obey a simple equation, which is geometrical and based on integral resonances; secondly, two coupling constants (including α) are used in the equation while we find no specific couplings related to the SM weak and Higgs fields as they use only specific geometrical degrees of freedom; thirdly, all calculi and equations are compatible with a simple form of compositeness. On this basis, we showed [5] that the electron and muon magnetic moment anomalies can be computed from the equation parameters with no use of QED.

In the next sections, we first repeat the main demonstrations, fix some errors, and then discuss the results and implications; since the mass equation is geometrical, its use of coupling constants and the manner they combine imply that they are also geometrical; we deduce that they correspond to resonance paths and find or approach the related equations. In this way, the field is geometrically self-quantized and has no free parameter related to energy. The same applies to gravitation since, using Wheeler-Feynman absorber equations, we deduce and compute its coupling (and the reduced Planck mass) from the constants and integral resonances used in the mass equation. In this way this mass-resonance theory is linked to gravitation and cosmology; it needs no dark matter and no big bang but comes with a constant linear expansion and energy creation.

We shall use measurement data and constants from CODATA 2014 or the Particle Data Group 2014 except where mentioned. The point is of importance considering the precision reached with leptons masses, anomalies, and α. The reader should keep in mind that the initial study used older values which imposed no difference to the model.

2 Deriving a mass equation

De Broglie [2] imagined a stationary wave of length $h\,c/E$ which relativistic transformation gives a phase wave of length h/p. This is the origin of the wave equations of quantum mechanics. The question of the nature of those waves is still open; in this section, we imagine how a stationary wave can be born *and ring*; then we predict some characteristics of the resonances that we shall later use as verification.

We assume that the wave is the physical exchange at the origin of mass. Energy exchange is momentum, and it gives a pressure field that "cages" the particle charges and some associated self-energy. The initial idea is similar to the Poincaré stress [11] though not identical as we split the particle.

Roughly speaking, we cage a permanent photon-like current in a box also made of currents and we guess that the box and the charge quantize each other. Assume the box size universal, it is sufficient to use a length 1. In the one dimensional case, the pressure is a simple force, and resonance implies an integral number M such that we have:

$$m = \mu + X\,M,$$

where m is the particle mass and X is a universal constant. The quantity μ represents a massless self-energy that necessarily propagates, and it implies a double resonance. Hence the resonance corresponds to a product $M = N\,P$:

$$m = \mu + X\,N\,P.$$

In the 1-dimensional case, we should have $N = P$ corresponding to identical inbound and outbound currents, but we shall need a more general equation and then we use a product. In a wave representation, it represents the number of times the inbound and outbound wave crests hit each other in a universal period of time or within a definite length.

Caging a massless particle requires symmetry, a force that opposes the particle charge to the pressure field, that is precisely the resonance N P and the self-energy μ. There must be a residual distance $d \neq 0$ between the first resonance wall and the current μ at which the force applies. It gives:

$$m = \mu + \frac{X}{d + \dfrac{1}{N\,P}}.$$

Now the distance d should also depend on N and P because energy comes from the distance (d + *1/NP*) which is equivalent to a potential. A potential is quantized and $1/NP$ is already quantized as it comes from $XNP = XM$. Then we use $d = KD$, with K an integral number and D a length. Last, in three dimensions we get a cube:

$$m = \mu + \frac{X}{\left(K\,D + \dfrac{1}{N\,P}\right)^3}. \tag{2.1}$$

The equation has 6 degrees of freedom that can be reduced to 5 by division by X or μ and give unit-less quantities.

Now let us discuss the equation geometry; contrary to the one-dimensional case, we have more degrees of freedom in the resonance and the paths associated to N and P can be radial or circular; here we can use group theory arguments:

— Case 1: A double radial resonance. It needs identical inbound and outbound waves, then $N = P$, giving a stationary wave. Except for the cube, it is identical to the 1-dimensional case then it should address leptons and U(1), and also the Poincare stress in which case we should have $KD > 0$, with K increasing with mass as $1/NP$ reduces since the leptons charges are identical.

— Case 2: A double circular resonance: The resonance geometry is conserved when we invert rotation axis; hence it must be identified to SU(2) and by symmetry $N = P$. But we must change (2.1) with $X \to X/k\,\pi$ with k a constant integral number; this is because compared to the first case even though the resonance is circular the pressure is still applied to its geometrical center. The equation becomes:

$$m = \mu + \frac{X}{k\,\pi\left(K\,D + \dfrac{1}{N\,P}\right)^3}.$$

It addresses massive bosons, which role in nature is to carry interactions. They are similar to a photon and we must integrate to X the term μ (that would be an intrinsic mass). Therefore we will compute their masses (index b) comparatively to the full electron mass (index e) as follows:

$$\frac{m_b}{m_e} = \frac{\left(\dfrac{1}{N_e P_e} + K_e D_e\right)^3}{k\,\pi\left(\dfrac{1}{N_b P_b} + K_b D_b\right)^3}. \tag{2.2}$$

- Case 3: A mixed resonance. It includes both symmetries U(1) and SU(2), it is then SU(3) and this case addresses quarks. If D is related to the strong force and asymptotic freedom ($\approx$ inverse to the Poincare stress) we should have $KD < 0$, ideally constant. It implies $N \neq P$ with a geometrical constraint between π, N and P since a phase lock between the two paths must exist; it requires to squaring a circle, then logically we should get approximate relations like:

$$N\,P\,\pi \approx \textit{an integral number}, \tag{2.3}$$

If the logic above is valid, it follows that particles distant interactions are a manifestation of the resonance; hence we should find relations between the resonance numbers (N, P) and the known symmetries, and also between some coupling constants and the non-integral values of D, X, and μ. *De facto, and most importantly, we cannot understand mass and charge quantization separately*.

3 Massive elementary particles resonances

In this section, we shall fit the equation parameters to all known elementary particles masses; since the equation is related to symmetry, the natural strategy is to proceed by groups (leptons, quarks, massive bosons). We shall assume X universal and μ specific to leptons (where enough precision exists) and, since D addresses forces, it must be group-dependent.

Recall also that a number of relations must be verified by the fit; they can be used as verification of the geometrical constraints imposed by symmetry and by the equation.

3.1 Leptons

The Table 1 shows charged leptons resonances. It uses very small numbers, we get $N = P$ as expected. The equation parameters are given hereafter:

$$\mu = 241.67661953\,\text{eV},$$
$$D_e = 0.0008532218937, \tag{3.1}$$
$$X = 8.1451213299073\,\text{KeV}.$$

Table 1: Electron, muon, tau in MeV/c^2.

–	P = N	K	Computed	Measured
e	2	2	0.510 998 9461	0.510 998 9461(31)
μ	5	3	105.658 3752	105.658 3745(24)
τ	9	5	1 776.84	1 776.82(16)

Using α, the fine structure constant, we define a new constant that will be used later:

$$A_S = D_e/\alpha \approx 0.11692, \tag{3.2}$$

which name A_S is chosen for its value is reminiscent of the strong force coupling.

The values in (3.1) can be tuned so that all masses match exactly regardless of uncertainty; instead those values have been chosen to compute exactly the electron mass and magnetic moment anomaly (assuming the related equations developed later are good-enough for such precision).

3.2 Quarks

Using X and μ constant from (3.1) the quarks resonances are shown Table 2 (masses in the natural scheme) where a regular pattern is obvious.

As expected, the parameter D is slightly different from (3.1) to compute those masses:

$$D_q = D_e(1 + \alpha) = A_S\,(\alpha + \alpha^2). \tag{3.3}$$

Using D_e like for leptons gives the top mass out of range $\approx$ 167 GeV, and then a difference with leptons exists. Quarks masses are no more published in the natural scheme; the estimates used in Table 2 are dated 2011 except for the top [18], see also [19].

We get $N \neq P$ as expected; P and K are constant which is surprisingly simple. The constancy of $K = -6 < 0$ is reminiscent of asymptotic freedom and then also agrees with a connection between D_e and α_s. Note that varying K by ± 1 gives computed quarks masses out of uncertainty range for the four heaviest.

Table 2: Quarks resonances in MeV/c^2.

–	P	N	K	Computed	Estimate
u	3	2	–6	1.93	1.7 – 3.1
d	3	19/7	–6	5.00	4.1 – 5.7
s	3	7	–6	106.4	80 – 130
c	3	14	–6	1,255	1,180 – 1,340
b	3	19	–6	4,285	4,130 – 4,370
t	3	38	–6	172,380	172,040 ±190 ± 750

The approximate relations with $N\,P\,\pi$ (2.3) are verified for the second and third generations; they are:

$$c, s : 7 \times 3\pi \approx 65.97 \approx 66/1.0004025,$$
$$t, b : 19 \times 3\pi \approx 179.07 \approx 179 \times 1.0003954.$$

We also notice that between 1 and 19 no other integral numbers come close to verifying (2.3).

It is interesting that the multiplication of N by 2 in the second and third generations corresponds to the difference in electric charges $(1/3, 2/3)$ as it links mass and charge quantization. For the first generation the down quark needs a fraction $N = 19/7$ which is barely acceptable, and we notice that the relations with (2.3) match with $2\,\pi$ for the d and also indirectly for the u instead of $3\,\pi$ for the four heavier quarks.

Those particularities may relate to quarks mixing, which we see in the fraction $19/7 = 38/14$, and the same logic for u also holds since $2 = 38/19 = 14/7$.

$$u : 2 \times 3\pi \approx 19/1.008,$$
$$d : (19/7) \times 2\pi \approx 17 \times 1.0032.$$

Hence something unique happens to the u and d.

3.3 Massive Bosons

We assume that the $W^{\pm}$, Z^0 and H^0 acquire their masses from the same geometry; recall that we only have three geometries (or mechanisms) and then we cannot address the weak force bosons and the H^0 separately. Using (2.2), it corresponds to the same resonance, that is on the circular path we must have $N = P = constant$, and only the radial K varies (though this is not exact since we shall later find a slight difference).

A factor $k\,\pi$ at the denominator of (2.2) is needed since the resonance is supposed circular, but we do not find a perfect fit with k integral. We need a factor $k \approx 1$; it seems at first that we add a degree of freedom but we shall show that it is a geometrical constraint.

The analysis of those masses is iterative and leads to important reasoning which is repeated hereafter in details. In practice:

- The empirical fit gives the resonances, which are $N = P = 12$, and $K = -2, -7, -19$ for the $W^{\pm}$, Z^0 and H^0 respectively. The weak force bosons come in range but the error on the H^0 is 1 GeV. Those numbers immediately suggests the same underlying geometry as quarks and maybe leptons, then the same field combining potentials expressed by D_e and α.
- The empirical value of D for massive bosons is first approximated as $D_b \approx \alpha^2(1+A_S/2-A_S^2/6)$; it suggests an interaction term that depends on α and D_e; the former is known and the later estimated with precision.
- The expression $[D_e(1+\alpha)]^2 = \alpha^2(1+2\,A_S+A_S^2)$ is similar and may give $D_b \approx \alpha^2(1+A_S/2-A_S^2/6)$ depending on the effective algebra. (Doubling the forces divides the distance, then $2A_S \rightarrow A_S/2$, and the term $-A_S^2/6$ fits with the $K = -6$ in table 2.) .

On this basis we may have enough information to model the interaction; the equations (3.2 – 3.3) suggest:

- Two types of charges corresponding to the mass $\mu : E$ and C ($\approx$ electric and color) on which D depends.
- A free field (charges X), and the pressure is given by interactions: $X\times X$, $E\times X$, and $C\times X$, hence D_b includes 3 terms, but its expression is incomplete as we do not yet compute all masses with precision.

Now we shall complete the reasoning, compute the predicted bosons masses, and compare to experimental data.

Classification and immediate identification gives Table 3. It shows that each individual interaction adds a piece of coefficient in D_b — like simple potentials adding or subtracting. But we can only compute a radial distance (which gives a radial strength), not the orientation of the force which can be symmetry-dependent as we discuss rotations.

Table 3: Classification and minimal interpretation of the coefficients.

–	D	Coeff	Interaction	Interpretation/logic
1	D_e	αA_S	$X\times E$	*Leptons*
2	D_q	αA_S	$X\times E$	*Leptons* $\rightarrow$ *Quarks*
3	D_q	$\alpha(\alpha A_S)$	$X\times C$	*Quarks Charge*
4	D_b	α^2	$X\times X$	–
5	D_b	$\alpha(\alpha A_S)/2$	$X\times C$	*Quarks* $\rightarrow$ *Bosons*
6	D_b	$(\alpha A_S)^2/6$	$(X\times E)^2$	*Leptons* $\rightarrow$ *Bosons*

The important point in this table is that quarks charges resume to $X\times C = X\times(X\times E)$, and the coefficient 1/2 line 5 implies two distinct charges (augmenting the force and then reducing the distance). Interpretation details are given hereafter (referring to the line of the Table 3) and lead to understanding.

Leptons — Line 1; charge E.

- $X\times E \rightarrow \alpha A_S$: There is only one elementary interaction; it just gives us its coefficient.

Quarks — Lines 2 and 3; charges E and C.

- $X\times E \rightarrow \alpha A_S$: Same as electrons, and independent of the quark electric charge.
- $X\times C = X\times(X\times E) \rightarrow \alpha(\alpha A_S)$: This is a different interaction; it is not a new kind of charge but it has the same nature and quantum as X.

Massive Bosons — Lines 4, 5, and 6: charges E and C.

We found the same coefficients for the $W^{\pm}$ and the Z^0. One is electrically neutral but not the other. Still, we find coefficients related to electricity and color charge, and then those bosons are made of two fractional electric charges and their two color charges (as we shall see the term charge is abusive here). Then it is:

- $X\times X \rightarrow \alpha^2$: The interaction of two charges X gives a distance α^2. This is the main force on the circular path that other interactions will impact — they are secondary forces or loops impacting this path.
- $X\times C = X\times(X\times E) \rightarrow \alpha(\alpha A_S)/2$: The coefficient $\alpha(\alpha A_S)$ comes with quarks color charge; it also shows that the charges of a weak force boson are equivalent to that of two quarks, and different of that of a lepton. Increasing the force by a factor 2 reduces the length proportionally; thus the factor 1/2.
- $(X\times E)\times(X\times E) \rightarrow -(\alpha A_S)^2/6$: This coefficient corresponds to the effect of the main resonance on separate electric charges. We recognize $D_e = \alpha A_S$ from leptons, but 1/6 is new; it is only associated to D_e^2 and this interaction is not present in Tables 1 and 2.

At this point, we understand how the interaction works and we can logically deduce all missing terms in the expression of D_b using α and A_S. For this, we need to complete the series of interaction loops with the field X:

$X\times X\times X \rightarrow -\alpha^4$: Since $X\times X \rightarrow \alpha^2$ positive, and $K < 0$, the force in $X\times X$ is compressive and then this coefficient is scalar (and positive), it increases the compression and then reduces the length: the coefficient is then negative $-\alpha^4$. The next coefficient is positive as it reduces $-\alpha^4$. Similarly, we must add loops indefinitely ($X\times X\times X\times X$ etc.); it gives a simple series converging to $\alpha^2/(1+\alpha^2)$.

Last, each interaction must be augmented with any number of X where the corresponding length is modified depending on its sign; then the coefficient $-A_S^2/6$ is multiplied by $1/(1+\alpha^2)$ and the coefficient $A_S/2$ by $1/(1-\alpha^2)$. The series make a small difference in D_b which is far from negligible when it comes to computing masses. The coefficient D_b for the $W^{\pm}$ and Z^0 is then:

$$D_{WZ} = \alpha^2 \left(\frac{1}{1+\alpha^2} + \frac{A_S}{2(1-\alpha^2)} - \frac{A_S^2}{6(1+\alpha^2)} \right),$$

$$D_{WZ} = 5.62404904 \times 10^{-5}. \tag{3.4}$$

It also reads:

$$D_{WZ} = \frac{\alpha^2}{1+\alpha^2} + \frac{D_e}{2(1-\alpha^2)} - \frac{D_e^2}{6(1+\alpha^2)}.$$

But it cannot be identical for the H^0, firstly because its spin is not 1. Assuming it holds four charges organized in a tetrahedral manner, a tetrahedron has 6 lines of forces, and the last interaction term is six times stronger:

$$D_H = \alpha^2 \left(\frac{1}{1+\alpha^2} + \frac{A_S}{2(1-\alpha^2)} - \frac{A_S^2}{1+\alpha^2} \right),$$

$$D_H = 5.56338664 \times 10^{-5}. \tag{3.5}$$

Or, alternately,

$$D_H = \frac{\alpha^2}{1+\alpha^2} + \frac{D_e}{2(1-\alpha^2)} - \frac{D_e^2}{1+\alpha^2}.$$

It may also include additional loops thru the tetrahedron. The strength of a line linking two charges is $1/6$, it gives the first term A_S^2 in (3.5), but for the H^0 it propagates thru 6 lines of a tetrahedron and it gives $6\,A_S^4$. But it is not a free field, and then it may not need an infinite number of loops. We shall use a one-loop approximation since additional loops makes a small difference ($\approx -10\,MeV$):

$$D_H = \alpha^2 \left(\frac{1}{1+\alpha^2} + \frac{A_S}{2(1-\alpha^2)} - \frac{A_S^2\,(1+6A_S^2)}{1+\alpha^2} \right),$$

$$D_H = 5.55741566 \times 10^{-5}. \tag{3.6}$$

This expression is the only reason here for A_S to be physical since all others uses of this coefficient reduce to D_e.

Now let us come back to the coefficient k in (2.2). In Table 4, we have $N = P$, and then those two resonances have the same orientation with opposite paths, but we find K in $(-2, -7, -19)$ the same numbers as for the quarks N which resonance is mixed.

Consequently, there is, like for quarks (2.3), a geometrical constraint which here is between the length D_b and the circular path π/NP. Taking only the circular path into account and keeping the constraint coming from the radius, D_b should be a divisor of $\pi/NP = \pi/144$, a division that must hold with any K in $-2, -7, -19$. Since all Ks are primes numbers the constraint applies to their product. In this simplified picture (that cannot hold yet) we should have:

$$(\pi/144)/D_b = 2 \times 7 \times 19 \rightarrow \pi/144 = 266\,D_b$$

Now D_b is radial and a 3-sphere volume depends on the cube of its radius. Then we must use $D_b\pi^{1/3}$ on the right hand side; it gives a modified equation that is close to hold:

$$\pi/144 = 266\,D_b\pi^{1/3}.$$

This equation is equivalent to squaring the circle, then we miss the coefficient k which is now a logical geometrical constraint related to phase lock. In (2.2), π is multiplied by k and this equation addresses a volume; hence we must use its cube on the left hand side, and reduce π accordingly on the right-hand side; in this way we get comparable quantities and it gives the geometrical resonance constraint:

$$k^3\,\pi/144 = 266\,D_b\,(\pi/k)^{1/3}. \tag{3.7}$$

Here the interaction term D_b constrains k thru geometry. The two sides of (3.7) represent lengths, and then taking their cube we get volumes verifying:

$$(266\,D_b)^3 = k^{10}\,\pi^2\,(1/144)^3. \tag{3.8}$$

It equates the volume of a 3-cube of edge $266\,D_b$ on the left hand-side to that of a 4-ball ($V^4 = \pi^2\,R^4/2$) divided by half its radius on the right-hand side, where a correction k is needed for cubing the sphere. Here D_b is an interaction term in 4D, k a geometrical wave coherence constraint, and (3.8) links a radial and a circular path in 4D. Now compute from (3.8):

$$(3.4) \rightarrow k_{WZ} = 1.00128565, \tag{3.9.1}$$

$$(3.5) \rightarrow k_H = 0.998033312, \tag{3.9.2}$$

$$(3.6) \rightarrow k_H = 0.997711845. \tag{3.9.3}$$

Using the coefficients above and (2.2), gives the masses in Table 4, where precision is impressive.

Table 4: Bosons resonances in MeV/c^2, H^0 mass in [17].

–	P = N	K	Computed	Measured
$W^{\pm}$	12	−2	80,384.9	80,385 ± 15
Z^0	12	−7	91,187.56	91,187.6 ± 2.1
H^0	12	−19	125,206	125.090 ± 240
H^0	12	−19	125,094	125.090 ± 240

After modeling the interaction we compute the weak force bosons masses in perfect agreement with measurement and it and confirms the validity of our reasoning. We get an effective

unified theory of resonances where the forces compositeness decays from leptons and quarks and this is truly unexpected.

In this table, the last two lines correspond to the equations (3.5 – 3.9.2) and (3.6 – 3.9.3) respectively for D_H and k_H.

We can now better analyze the resonance in Table 4. Consider the length $2 \times 7 \times 19 = 266$. A phase lock between the radial and circular paths and the $K = -7$ and -19 imply two circular path lengths which are $L1 = 2\pi(1 - 7/266)$, and $L2 = 2\pi(1 - 19/266)$. Those are compatible if and only if $A\,L1 = B\,L2$, with A and B integral numbers. We must solve the following equation which solution is trivial:

$$\frac{A \times 2\pi(266-7)}{266} = \frac{B \times 2\pi(266-19)}{266},$$

$$A = 266-19 = 247, B = 266-7 = 259, B - A = 12. \quad (3.10)$$

The resonance number, 12, appears on the left hand side of (3.10); it comes from phase coherence between the circular path and the spots on the radius and we naturally get $N = P = A - B = 12$ which then depends only on K (we use only -7 and -19, but $K = -2$ is not a problem since 12 is even).

Finally all numbers and parameters used in Table 4 appear constrained; the specific degree of freedom used here is just geometry. We have two forces coefficients (α and D_e) and no specific coupling in this sector which is then emergent; this result disagrees with the SM concept and requires unification from below (as opposed to distinct fields).

3.4 Bosons widths

The expression (2.2) is a resonance equation and the computed masses correspond to the poles of the resonances. Then it should be possible to compute widths and then lifetimes; at best, the widths are the size of some working resonance "spots"; it would show that this theory gives the SM weak field. For this we have to understand the phase coherence between multiple paths. Recall that the bosons charges are found interacting and organized in a minimal manner; in 3D, it is a tetrahedron for the H^0 and a simple straight line for the Z^0 and $W^\pm$. For the weak force bosons:

With two circular phases the symmetry is loose, it has some freedom, and on the circular path it suffices that N and P hold on $1/2$ phase to stabilize the resonance. It authorizes a circular phase shift $\pm\pi/12$ which extends or reduces the sphere; with two charges, it gives on the radial part $\Delta K = (\pm 1/2)(1/12) = \pm 1/24$.

In the radial direction, we have 266 slots, and the same reasoning applies; it adds $\Delta K = \pm 1$.

For the H^0, with 4 charges, the symmetry is fully constrained in 3D; N and P hold together: $\Delta K = 1/144$. A tetrahedron has 6 lines of force that can break; hence the width is reduced accordingly $\Delta K = 1/144/6$. Other loops add nothing since a tetrahedron is fully constrained in 3D.

On this basis, the resonance width is the difference in mass ΔM given by (2.2) with respect to the pole in Table 4 when we use $K + \Delta K$ in (2.2) to compute the particle mass $M + \Delta M$. We get:

$W^\pm \to \Delta K = (1 + 1/24) \to \Gamma_W = 2.0857$ GeV, a perfect match with experiment (2.085 ± 0.042 GeV).

$Z^0 \to \Delta K = (1 + 1/24) \to \Gamma_Z = 2.468$ GeV, 1% less than expected (2.4952 ± 0.0023 GeV).

$H^0 \to \Delta K = 1/(144 \times 6) \to \Gamma_H = 4.10$ MeV, which agrees with the SM prediction at 125.09 GeV.

Hence, the widths come straightforwardly from geometry. But the Z^0 width is out of range and this can only be due to the difference in charges with the $W^\pm$ that we have ignored. Reasoning simply:

$W^\pm$: The charges e/3 and 2e/3 (or opposite) repel each other with a force coefficient $2e^2/9$.

Z^0: The charges e/3 and -e/3 (or 2e/3 and -2e/3) attract each other, the force coefficient is $e^2/9$ or $4e^2/9$.

The difference in inner charges between the Z^0 and the $W\pm$ gives a difference in forces which is:

$$\frac{2e^2}{9} + \frac{e^2}{9} = \frac{e^2}{3} \text{ Or : } \frac{2e^2}{9} + \frac{4e^2}{9} = \frac{2e^2}{3}.$$

It implies that the forces cannot be balanced in the same manner for the two bosons. Assuming the $W^\pm$ width computed value is exact, we need an additional term to compute the Z^0 width. Since the forces in the calculus of D_b depend on charges, from the equations above the missing coefficient is $1.5/137$ or $1.5\,\alpha$. It gives:

$Z^0 \to \Delta K = (1 + 1/24 + 1.5/137) \to \Gamma_Z = 2.4946$ GeV,

which agrees with the SM prediction and experimental data. However the experimental precision for the $W^\pm$ and Z^0 widths differ by one order of magnitude; hence this reasoning, which is differential, is risky and non conclusive.

3.5 Resonance terms, analysis and reduction

The resonance terms found in the previous tables (all N and P) reduce to 2, 3, 7, and 19 in the following manner:

Leptons: 2, 7 – 2, and 7 + 2.

Quarks: 3, 7, 2×7, 19, and 2×19, if we omit the u and d where we know from the CKM matrix that mixing is large as compared to the other angles.

Massive bosons: 12 = 19 – 7.

It is remarkable that $7 = 2^3 - 1^3$, and $19 = 3^3 - 2^3$; here it reduces to the three "symmetry numbers" of U(1), SU(2) and SU(3), and their cubes differences. Moreover for all quarks we get $P = 3$, including the u and d, where the polarity appears, meanwhile for leptons it seems that we have the polarity 2 in a mixed manner. In this way the radial paths are based on 2 and 3, while 7 and 19 only come with circular paths (and unstable or mixing particles).

Moreover, the difference in resonance between the electron and the muon and tau relate to the $K = -7$ of the Z^0, while the heavy quarks decays include a factor 2 in charge and resonance which fit the $K = -2$ of the $W^\pm$.

Therefore we get the strong impression that the equation relates to an intricate resonance scheme based only on the SM symmetries — or something close. The simplicity of the reasoning and numerical results suggest that the mass spectrum may be unavoidable, and since it relies on charges it also suggests the absence of free parameters in nature.

3.6 Charges ratios

The results in this section suggest a single field "below" and it is interesting to estimate charges ratios, but we can only compute their radial effect, not the forces orientation; from the analysis of Table 3 the distances building the Ds are in reverse proportions of charges. Then for the electron, $K = 2$, and for quarks, $K = -6$, and $D_e \approx D_q$.

From Table 3 and the different parameters D, taking into account the differences in K, since we have $X \times E \to 2 \times D_e$ for the electron and for quarks $X \times C \to -6\,\alpha\,D_e$; we estimate:

$$\frac{C}{E} = \frac{2\,D_e}{6\,\alpha\,D_e} \to C = \frac{E}{3\,\alpha},$$

which has a clear scent of monopole; importantly, it does not depend on the quark electric charge since the coefficient 3 (from $K = -6$) is constant in Table 2.

In Table 3, we also have $X \times X \to \alpha^2$ and $X \times C \to 6\,\alpha\,D_e$, and then we estimate:

$$\frac{C}{X} = \frac{\alpha^2}{6\,\alpha\,D_e} = 1.4254503 \approx \sqrt{2},$$

which is in the range of 1 and then the same type of charges ($\approx \sqrt{2}$ suggests geometry of the force orientation).

4 Coupling constants

4.1 Introduction

We found two real constants in the expression of the parameters D which represents a length in the equation. In the expressions (3.4 – 3.5 – 3.6) used for D_b those two constants stand on equal grounds. Hence since α is the coupling constant of QED, then D_e is also a coupling constant. It then relates directly to the strong and weak forces couplings (recall that we also have $A_S = D_e/\alpha$ in the range of $\alpha_{S(MZ)}$) and since K appears constant for quarks, D_q should be related to asymptotic freedom. Therefore it seems that the equation addresses a field below with two and only two couplings (neglecting gravitation for now).

Since all resonances are integral (N, P, K) and reduce to a few numbers, it is minimal and elegant to generalize the concept and assume that the field is entirely self-quantizing (or self-constraining) and that quantization is entirely based on geometry and integral numbers; in this way, those two coupling correspond to some counter-resonances ($1/N \to N$ or $N \to N$) and then to constant path lengths (or relative path lengths).

In practice the only known constant integral path length is that of photons for which $r^2 - c^2t^2 = 0$. At the opposite, in special relativity, massive particles obey $r^2 - c^2t^2 = const \neq 0$ which we write $r^2 - c^2t^2 - const = 0$. But now the paths of the resonance define the massive particles — *we mean entirely*; it is a repeat pattern that fits into this equation and it first implies that the path includes a rotation which is around the time axis. Then we guess that D_e (as a length) must be computed from a pseudo-norm like expression of the form:

$$n^2 + m\,\pi^2 - p^2 = D_e^{-2}$$

where the central term introduces a rotation and n, m, and p are expressions based on the resonance terms. Now of course, α must obey a similar pattern and, since α and D_e have distinct but complimentary roles, the expressions giving D_e and α should use resonance terms in a complimentary manner. Last, the bosons resonances are based on 4-dimensional paths; then n, m, and p must be seen as the coordinates of a 4-path which projection on 3-dimensional space gives real numbers.

Because of 4D resonances, we shall suppose that there is no punctual particle or 1D string and that the field is entirely fluid. It implies that some currents propagating in a direction orthogonal to the observable 3-space (possibly back and/or forth in time) are preserving and propagating the characteristics of the particle and we shall abusively denote those "time-currents". In this way the electric field of the electron is seen similar to the effect of a magnetic current propagating forward and/or backward in time with respect to the present. Here the present is seen as the surface of an expanding 4-sphere, but 4D space is assumed preexisting and permanent.

It results in an interesting minimal model where all known massive particles are composites of time-currents:

Leptons:

- e^- : $[\uparrow^-\downarrow^+]$,
- μ^- : $[\uparrow^-\downarrow^+\downarrow^-\downarrow^+]$,
- τ^- : $[\uparrow^-\downarrow^+\uparrow^-\uparrow^+]$.

Quarks:

- t^+ : $[\uparrow^+\downarrow^-\downarrow^+\uparrow^-\uparrow^+]$,
- b^- : $[\downarrow^+\uparrow^-\uparrow^+]$,
- $c+$: $[\uparrow^+\downarrow^-\downarrow^+]$,
- s^- : $[\downarrow^+]$.

Bosons:

- Z^0: $[\downarrow^+\downarrow^-]$,
- $W^\pm$: $[\uparrow^+\downarrow^-]$ and $[\uparrow^-\downarrow^+]$,
- H^0 : $[\uparrow^+\uparrow^-\downarrow^+\downarrow^-]$.

where the notations are trivial for up-time and down-time currents sign and directions (the sign is the current, not the electric charge which, by convention, is inverted for down currents); the apparent electric charge is $2/3$ for an up-time cur-

rent and 1/3 for a down-current (still by convention). Several aspects of the model are of interest:

- The model is based on 4 dimensions of space; it is then coherent with the calculus of the coefficient k used for bosons, but it is also reminiscent of QCD where quarks live in 4 dimensions.
- The difference between the H^0 and the weak bosons is consistent with the calculus of D_{WZ} and D_H.
- All quarks decays consist in a separation of currents where the sum of the produced $W^{\pm}$ boson's current and quark is equal to the currents of the original quark (and of course the picture is reversible).
- The same is valid for leptons decay, but with a Z^0.
- There is no room to make a d quark except by mixing (and the d comes with resonances ratios).
- The notion of time-currents removes the need for particles "inhabiting space". In this way, the concept is minimalist and particularly elegant since, eventually, it must result in self-quantizing movement where we do not need to distinguish space and matter.
- All particles include a down-type current (taking this as strict rule implies mixing for the u and d, and the absence of FCNC). The model agree with Cramer's interpretation of quantum mechanics — though in an almost classical 4-dimensional manner. All particles are connected to and can send information to their past or receive some from their future because a communication channel exists which is the particle itself.

4.2 Coincidences

In this sub-section we discuss three numerical coincidences involving the numerical values found in section 3. In this way, we seek coherence with known but older theory.

4.2.1 Lamb shift, Bethe's equation

Bethe [1] computes the hydrogen Lamb shift; he gets:

$$\Delta E = \frac{\alpha^5 m_e c^2}{6\pi} \ln\left(\frac{m_e^2 c^2}{8.9\,\alpha^2 m_e^2 c^2}\right), \tag{4.1}$$

where m_e is the electron mass; the expression in the logarithm depends on the cutoff and gives a ratio between the electron absorption and self-interaction and then in our model μ and $(m_e - \mu)$ respectively (though according to the mass-equation, self-interaction and absorption may be reversed with respect to QED,) we find:

$$\frac{(m_e - \mu)}{\mu} = \frac{1}{8.8857\,\alpha^2}. \tag{4.2}$$

The relative difference with respect to Bethe's result is 1.6×10^{-3} (or 2×10^{-4} for ΔE) and then μ seems relevant with respect to Bethe's analysis. We notice a similar coincidence:

$$\frac{(m_e - \mu)}{\mu} \approx \frac{\sqrt{2}}{4\pi\,\alpha^2}. \tag{4.3}$$

The relative error in (4.3) is $\approx 1.25\times10^{-5}$. Consequently, since Bethe's paper is seen as the very first step to QED, X and μ should be fundamental quantities directly linked to QED.

4.2.2 The electron mass and spin, rough analysis of the coincidences

A physical action is a product of charges or currents; then we analyze action and not energy. Accordingly, the electron mass comes as a repeated action ($E = h\nu$).

Action is a product that we first write in complex form:

$$\left(G + \frac{i\,e}{2}\right)\left(G - \frac{i\,e}{2}\right) = G^2 + \frac{e^2}{4} \to m_e, \tag{4.4}$$

where e/2 represents the currents, not the apparent charges, and G the resonant component. Now we write (4.4) in quaternion form:

$$\left(G + \frac{i\,e}{2}\right)\left(G + \frac{k\,e}{2}\right) = G^2 - \frac{j\,e^2}{4} + (k+i)\frac{e\,G}{2}. \tag{4.5}$$

Those equations may approach the natural algebra, but the result seems wrong. Still, assume the algebra is broken, (4.4) gives the mass and (4.5) angular momentum:

$$G^2 + \frac{e^2}{4} \to m_e;\ (k\ +\ i)\frac{e\,G}{2} \to \textit{angular momenta}. \tag{4.6}$$

The angular momentum splits into two components on orthogonal axis — which agrees with the idea of time-currents. Then one is the magnetic moment and the other is along the time axis; we will denote the latter "spin". Now we identify the squared charges in (4.6) with the masses in (4.3); it gives:

$$4\pi\,\alpha^2 G^2 \approx e^2\frac{\sqrt{2}}{4}.$$

Substituting G with a Dirac charge, we get $1 \approx \sqrt{(2)}/4\pi$; now multiply each side of this ridiculous result by the Planck constant we get the following correspondence:

$$h \leftrightarrow \sqrt{2}\,\frac{\hbar}{2} = \left|(k+i)\frac{e\,G}{2}\right|, \tag{4.7}$$

which interpretation is obvious: a repeated action h is energy ($E = h\nu$) and it makes the leptons spin and magnetic moment.

4.2.3 The Dirac condition and the parameters X and μ

Dirac [8] analyzes the possibility of existence of magnetic monopoles using quantum mechanics. Based on the mathematical properties of the electron wave function interpreted

as a density of probability of presence, he shows that a magnetic monopole is compatible with the existence of quantum mechanics in Hamiltonian form if and only if the so called Dirac condition is respected:

$$e\,g = \frac{(n\,\hbar\,c)}{2} \rightarrow g = \frac{n\,e}{2\,\alpha}. \tag{4.8}$$

It results in the elegant idea that the existence of magnetic poles fixes the electric charge and conversely.

Now let us assume that the electron wave is a magnetic current; since Dirac's demonstration is based on the "fields of force" acting on the electron wave then magnetic currents acting on electric charges must obey the same condition. But in our model e is an apparent charge (say e_e) and also a sum of time-currents (say e_m) and its monopole (denoted g_m).

Both must be taken into account in the condition as part of the total current; then the condition is:

$$e_e(g_m + e_m) = \frac{(n\,\hbar\,c)}{2}. \tag{4.9}$$

Now compare with our data and use $e_m = e_e$. The fundamental resonance in equation (2.1) corresponds to a theoretical half electron, that is $N = P = 1, K = 0$, and a self-energy $\mu/2$ that we shall ignore. It gives, as per (1 – 3.3):

$$m = X/1 = 8.1451213299073\,\mathrm{keV/c^2}. \tag{4.10}$$

This mass must be compared to μ as it comes from the interaction of the time-currents (not the apparent charges) and then, for an electron, as the product $e^2/4$. The rest of the electron mass (N = P = K = 2) is given by the resonance; then in (4.10) the numbers (N = P = 1) correspond to a hypothetical particle where a current G is interacting with $e/2$ which mass is given by an action corresponding to $Ge/2$.

Now we analyze how action comes as a product of currents, but not energy for which we rely on resonances. In the hypothetical resonance above, it corresponds to the products $e\,G$ and $e^2/4$, where G^2 is absent. It leads to a correspondence between action and energy:

$$\frac{e\,G}{2} \leftrightarrow m; \quad \frac{e^2}{4} \leftrightarrow \mu. \tag{4.11}$$

We divide the two expressions in (4.11) and in light of (4.9) we add $\mu/2$ that we initially ignored; we find:

$$\frac{2\,G}{e} = \frac{m}{\mu} \rightarrow 4\,G + e = 68.4051246306057\,e \approx \frac{e}{2\,\alpha}. \tag{4.12}$$

We *want* to recognize here the modified Dirac condition in (4.5), because the fine structure constant appears linked to the equation parameters.

But the result seems approximate; at first the relative discrepancy (-1.65×10^{-3}) seems acceptable since we analyze a hypothetical particle but we shall see that this numerical value holds precisely.

There is a second aspect related to the Dirac condition which comes from the time-currents model and the apparent electric charges $e/3$ and $2e/3$ going respectively down and up the time; assume their individual self interactions are squared charges. Once again, we can link action and energy:

$$(e/3)^2 + (2e/3)^2 \rightarrow \mu(1/3)^2 + \mu(2/3)^2 = 5\mu/9. \tag{4.13}$$

Now from (4.10):

$$4(m + 5\mu/9)/\mu = 137.032471483434 \approx 1/\alpha \tag{4.14}$$

The relative discrepancy with respect to α is $\approx 2.26\times10^{-5}$. The coincidence can, at first sight, be seen redundant with the equation (4.12) as it is almost identical, but it comes from a different interaction and we shall see now that this value also holds.

4.3 Leptons magnetic moment anomaly

We assumed that the resonances in the previous section "construct" the leptons waves; unlike the classical wave equations the geometrical construction is not unique but lepton-dependent. Thus, even for the electron it seems hardly possible to make an exact link with the Dirac equation which, according to (2.1), should be too general; consequently we go back to de Broglie's thesis which is fully relativistic.

4.3.1 De Broglie wave geometry

In his thesis, de Broglie uses a standing wave, that we will denote the Compton wave and finds a phase wave as a result of the relativistic transformation of the former. The agreement of the stationary wave assumption with the results in Table 1 is straightforward since we get N = P for all leptons.

The change in phase of the de Broglie wave over the first Bohr orbit of a hydrogen atom is $2\,\pi$, while the Compton wavelength change in phase over this orbit is $2\,\pi/\alpha$. Then over any number of Compton wavelengths, we have:

$$\Delta\phi_D = \alpha\,\Delta\phi_C, \tag{4.15}$$

where $\Delta\phi_D$ and $\Delta\phi_C$ are the changes in phase of the de Broglie and Compton waves over any length. On the n^{th} orbit we find:

$$\Delta\phi_D = \frac{\alpha\,\Delta\phi_C}{n}, \tag{4.16}$$

There are n de Broglie wavelengths around the n^{th} Bohr orbit and we get a constant angular differential term α. The same reasoning applies in the case of a nucleus of charge $Z\,e$ and gives the same value. Hence, considering that the de Broglie wave defines the motion of the electron this term is universal in the Bohr model. As a result, and taking into account simultaneously the motion of the electron and the phase velocity of the de Broglie wave going around the proton, the

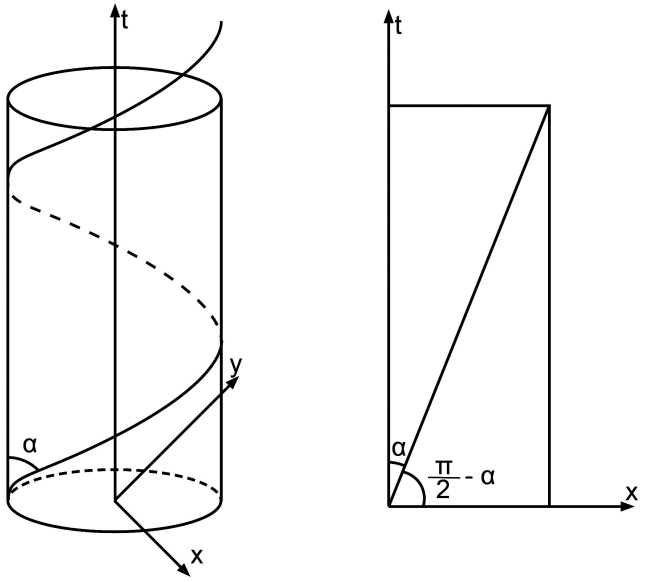

Fig. 1: Left, the electron classical Bohr orbit; right, the same cylinder unfolded (the angle is $\approx \alpha$).

phases of the two waves at any location of the electron classical trajectory are permanently identical.

Assume α is a path length based on integral and geometrical numbers. On the cylinder Figure 1, and using a system of unit where the radius of the cylinder is 1, the length of the unfolded tangent is approximated with $L \approx \sqrt{137^2 + (2\pi)^2}$. Now we know that the electron spin is 1/2, and then the rotation of the resonance is reduced to π when the electron runs one turn; we get the well-known $\sqrt{137^2 + \pi^2} \approx \alpha$.

Consider now the de Broglie wave as a shortcut permanently joining the electron with itself, but one (or n) Compton wavelength later, with an action 1/137/2 (taking again the spin 1/2 into account), it gives:

$$\alpha^{-1} \approx \sqrt{137^2 + \pi^2 - \frac{1}{137} \times \frac{1}{2}}$$

which holds with a relative precision $\approx 3 \times 10^{-8}$. Last, consider that the electron progresses in time, but that its waves are composed of two currents going up and down. If the up-time part of the waves gives a factor 1/2, the down-time part sees the electron with a charge twice lesser since in the case of quarks the down-time and up-time currents manifest fields 1/3 and 2/3 respectively. It must be augmented with a resonance length dependent on the time-velocity of the electron; twice longer for the same reason (charges 1/3, 2/3); finally it gives a factor 1/8 for the down-time part and we get:

$$\alpha^{-1} = \sqrt{137^2 + \pi^2 - \frac{1}{137}\left(\frac{1}{2} + \frac{1}{8}\right)}$$

$$\rightarrow \alpha = 72\,973\,525\,698 \times 10^{-13}$$

which is *exactly* the value of α given in CODATA 2012! Considering precision together with the simplicity of this geometry, it looks pretty much like time-currents exist.

In special relativity, one would consider the so called rapidity of the electron defined as a hyperbolic angle. However, the path length α can also be seen as a simple angle in the Euclidean coordinates (x, y, z, ict) as originally used by Minkowski. Moreover, one must consider this angle universal, and it implies a complimentary angle $\pi/2 - \alpha$. At first the existence of those angles can be checked numerically as it must also correspond to the coincidence (4.3); after appropriate replacements of α^2 by two coefficients corresponding to the two angles α and $(\pi/2 - \alpha)$, the equation (4.3) gives:

$$4\pi\,(m_e - \mu)\,\sin(\alpha)\left[\left(\frac{\pi}{2} - \alpha\right)\sin\left(\frac{\alpha}{\pi/2 - \alpha}\right)\right] = \mu\sqrt{2},$$

which holds with a relative precision of 2.9×10^{-8} instead of 1.25×10^{-5} for (4.3).

4.3.2 Other resonance coefficients and action

When the electron is on the first orbit there is a rotation of the time-current of a hyperbolic angle α which ratio to the space current changes in proportion of the hyperbolic tangent of this angle. As stated, the impact is a phase differential and considering resonances, a simple angle gives $tan(\alpha)$; it runs around the full Bohr orbit and then the instantaneous action term is $tan(\alpha)/2\pi$. The action given by $tan(\alpha)$ is that of a resonance going around the full orbit.

It must cycle on 1/2 quantum; hence the first correction term to the electron magnetic moment anomaly is:

$$a_0^e = \frac{tan(\alpha)}{2\pi} \approx \frac{g-2}{2} \tag{4.17}$$

where we denote a_0 and g the correction and the g-factor respectively. Compare to the first order QED correction by Schwinger [12], the well known $\alpha/2\pi$. The difference comes from a different manner to taking into account relativistic effects. Here it suggests that taking into account together the particle resonances and special relativity in the original Minkowski manner could give an analytic solution. In facts, the difference is that we consider the electron as a 4D gyroscope which axis is bent by velocity. This axis is shown with the orientation of the resonances N, P, K in Figure 2.

Therefore in (2.1) the resonance $N\,P$ corresponds to G^2 in (4.13) while K corresponds to $e^2/4$. The product $N\,P$ makes and "absorbs" the spin and the full space-resonance cycle is then $(N\,P\,-\,2)\,K$ which is a product G^2e^2 while the spin is given by $G\,e$. Action depends on the number of currents C (which, according to the model, is lepton-dependent) while the mass μ is constant; then we divide this coefficient by the number of currents.

We get a spin-dependent coefficient where the spin relates to the interaction of the G-currents and the apparent electric charges — which is logical. It is:

$$E = \sqrt{\frac{NP-2}{C}}K. \tag{4.18}$$

In the direction of time (K in Figure 2), the same reasoning gives NK^2 for a product $e^2/4$. But we get a spin independent coefficient which relates only to the currents and does not need a square root; it is:

$$F = \frac{N\,K^2}{4}. \tag{4.19}$$

The coefficients above are valid for an electron but for the muon and tau the coefficient a_0 corresponding to the time current rotation is not α like in (4.17), it depends on the resonance numbers. The electron is the special case because all resonance numbers are identical and even (N = P = K = 2) and then all phases are identical.

For the muon and the tau, N = P and K are odd and prime with each other, and then the action cycle is $N\,K$. Using (4.18) for an electron, the cycle uses N = K = 2 and its angle should be written $2\,\alpha/2$. Then for a muon and a tau the corresponding coefficient is:

$$\phi = \frac{tan(N\,K\,\alpha/2)}{NK/2}, \qquad a_0 = \frac{\phi}{2\,\pi}. \tag{4.20}$$

The expression mixes angles and resonance and fits with the interaction of current where action is angle-dependent; it will be the geometric form used in this section. We introduce $\alpha/2$ which we now consider as the physical angle of each time-current — it gives α for two currents of opposite directions taken together.

4.3.3 The electron

Now we want to compute the anomaly from the following picture: the electron is seen as a 4D rotation which (in all cases) has the following mathematical property: two orthogonal planes exist which are conserved by the rotation. The identifications are then obvious; the angles in the previous section define the two planes rotations and correspond to the resonances. The rotation is said double since we find distinct angles α and $(\pi/2-\alpha)$. The planes intersect at a single point (a mathematical property of any 4D rotation) where the resonances apply, and it defines the punctual particle — but we do not need to introduce anything material at this place (no particle). The planes intersection point also moves in space and in the direction of time defining a classical trajectory.

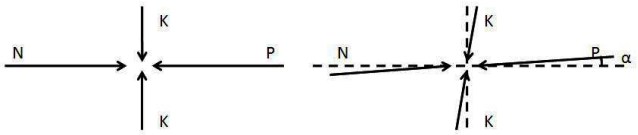

Fig. 2: Resonance geometry on N, P, and K. Left: an electron seen at rest, K on the time axis, N and P in 3-space. Right: an angle $\approx \alpha$ appears as a relativistic shift on the first Bohr orbit where axes are bent by velocity.

One plane is orthogonal to the time axis and hosts the leptons resonances N = P, and K is on the other one which includes the "time translation" of the particle. Finally those two planes are lepton-independent and then their translation and the associated angles define entirely the seemingly anomalous values in (4.8 – 4.10) as they are also lepton-independent. Consequently, the lepton-dependent resonances imply different magnetic moment anomalies. Therefore we can reverse-compute the anomaly from those two quantities. In this way, we define:

From (4.8): $4(X/\mu + 1/2) = \beta_1^{-1} = 136.810249261211$,

From (4.10): $4(X/\mu + 5/9) = \beta_2^{-1} = 137.032471483434$.

The Dirac equation gives $g = 2$ and it is known that the correction is entirely related to relativistic shifts. The quantities above correspond to distinct interactions and then distinct types of charges; hence the correction is a product $a_T = a_0\,a_1\,a_2$ where a_0 is geometrical and corresponds to the angle α in (4.17) or ϕ in (4.20), a_1 to the action of the apparent electric charges (4.10), and a_2 to the action of (magnetic) currents (4.8).

Since β_1 and β_2 are deduced from the leptons masses, they are related to the tangent of some angles part of the resonance geometry (in the same manner as $tan(\alpha)/2\pi$). The anomaly is angular and differential and then a_1 and a_2 must be computed as ratios involving α and the arctangents of some angles involving respectively β_2 or β_1, and resonance numbers. The electron correction term a_1^e is then given by an expression of following form:

$$\frac{\tan(\alpha)Y}{\tan^{-1}(\beta_2 Y)} \to a_1^e.$$

It links an action given by the angle α and another one given by β_2 and the anomaly relates to their ratio. Now β_2 relates to the apparent electric charges giving the spin; then $Y = E$ as defined in (4.18). The angle $\alpha/2$ also impacts the coefficient and subtracts from K.

Then we write:

$$E \to \sqrt{\frac{NP-2}{C}\left(K+\frac{\alpha}{2}\right)} \tag{4.21}$$

$$a_1^e = \frac{\tan(\alpha)\,\sqrt{2+\alpha/2}}{\tan^{-1}\left(\beta_2\,\sqrt{2+\alpha/2}\right)} \tag{4.21.1}$$

Now β_1 comes from the time-currents of the electron; we must make a similar reasoning involving F defined in (4.19).

Naturally, this correction will be similar in form to the equation above. The logic is:

- The first order effect is null; it is second order where the cross-products cancel.
- The angle must be α instead of $\alpha/2$ since the two angles $\alpha/2$ on the axis of K sum up.

It gives, for an electron:

$$a_2^e = \frac{\tan(\alpha) F\left(1-\alpha^2\right)}{\tan^{-1}\left(\beta_1 F\left(1-\alpha^2\right)\right)}, \tag{4.22}$$

$$a_2^e = \frac{\tan(\alpha)\left(2-2\alpha^2\right)}{\tan^{-1}\left(\beta_1\left(2-2\alpha^2\right)\right)}. \tag{4.22.1}$$

Note that in the equations (4.21 – 4.22) the angle $\alpha/2$ affects K and $-\alpha^2$ affects K^2; it is the same geometry where only K is impacted. Now from (4.17 – 4.21.1 – 4.22.1) and using the value of α in CODATA 1014 we find:

$$g_T^e/2 = 1 + a_0\, a_1\, a_2 = 1.00115965218091. \tag{4.23}$$

The values of X and μ in (3.1) were tuned to fit with CODATA 2014 which gives:

$$g^e/2 = 1.00115965218091\,(26). \tag{4.24}$$

The relative error on g_T^e in (4.23) with respect to (4.24) is less than 10^{-14}, but it can be down to $\approx 10^{-8} - 10^{-9}$ without ad-hoc tuning and keeping all leptons masses within uncertainty — the result would still be very significant.

4.3.4 The muon and tau

We get the equations needed to compute the muon anomaly in the same manner as for the electron but using (4.20) and including in (4.21) the four currents given by the model, and the resonance numbers in Table 1. We get:

$$g_T^\mu/2 = 1.00116592081. \tag{4.25}$$

The CODATA 2014 experimental value is:

$$g^\mu/2 = 1.00116592089\,(63). \tag{4.26}$$

The result is well within experimental uncertainty and independent of the adjustments of (3.1) since the precision is in the range 10^{-9}. The SM prediction disagrees with a $2-4\sigma$ discrepancy. Typically:

$$a^\mu_{SM} - a^\mu_{experiment} = (2.8 \pm 0.8) \times 10^{-9}. \tag{4.27}$$

The very short lifetime of the tau makes impossible at present to measure its $(g-2)$. The SM prediction is:

$$g^\tau_{SM}/2 = 1.00117721\,(5). \tag{4.28}$$

Using the tau resonances in Table 1 we get:

$$g_T^\tau/2 = 1.00125789. \tag{4.29}$$

But on the other hand, in the tau resonance, N = P = 9 is not a prime number, it is a square and then, perhaps, we should use 3 instead of 9 in the equations to compute its anomaly (we find a second reason later). It gives:

$$g_T^\tau/2 = 1.00117037, \tag{4.30}$$

where the difference with the SM prediction is more coherent with that of muons.

4.4 The fine structure constant

We made a first calculus of α as a simple path length. Now we shall first show that the shortcuts in this path length, namely 1/2 and 1/8, also defines the leptons resonances, and then find an immediate origin to the number 137.

4.4.1 A second view on leptons resonances

Our analysis of the resonances in Table 1 fits with the supposed geometry, and complimentary angles α and $(\pi/2-\alpha)$. It is a quasi-symmetrical picture that suggests the existence of a second view on the leptons resonances agreeing with the equation (2.1). In this equation we use three resonance terms (N, P, and K), but the rotation is in 4 dimensions; then the resonance terms correspond to one rotation plane used completely (N, and P), while K lives in the other plane but we only use an axis (not the full plane). The second view should split oppositely; it cannot hold with $N = P$ but it must with $P = K$ because of phase coherence. Then using angular ratios, we should have a different mass: $\mu' \approx \mu\,\pi/2 \approx 380$ eV/c^2. Starting with this value, imposing P = K, and using the equation (2.1), an empirical fit to the same decimal as shown in Table 1 gives Table 5 and the coefficients in (4.31).

Table 5: Second view on electron, muon, tau in MeV/c^2.

–	P=K	N	Computed	Measured
e	2	2	0.510 998 9461	0.510 998 9461(31)
μ	3	8	105.658 3752	105.658 3745(24)
τ	4	16	1 776.84	1 776.82(16)

$$\begin{aligned} \mu' &= 385.6750521055\,\text{eV/c}^2, \\ D' &= 0.0002255984538, \\ X' &= 8.02160795579\,\text{keV/c}^2. \end{aligned} \tag{4.31}$$

$P = K$ is verified, and we can estimate:

$$\mu' = \mu\left(\frac{\pi}{2} + \frac{\pi}{137} + \left(\frac{2\pi}{137}\right)^2\right),$$

which was used to compute (4.31); it uses 1/137 and no simple fit was found with α.

The remarkable point in Table 5 is that we find for N the numbers 2 and 8, and their product 16 for the tau. Those numbers show that, in the EM field, the resonance is tachyonic and the shortcuts can ring independently or in a combined manner. The product 16 also justifies our doubts for the tau $(g-2)$ in (4.29).

4.4.2 Alpha and 137

Following the first equation giving α, assuming time-currents exist and correspond to $e/2 \to 1/274$ we find an empirical fit compatible with CODATA 2014:

$$\alpha^{-1} = \sqrt{137^2 + \pi^2 + \frac{1}{274^2} - \frac{1}{137}\left(\frac{1}{2} + \frac{1}{8}\right)}$$

$$\to \alpha = 72\,973\,525\,672 \times 10^{-13}, \qquad (4.32)$$

where the difference with CODATA 2014 is about half the standard deviation:

$$\alpha_{CODATA\,2014} = 72\,973\,525\,664\,(17) \times 10^{-13}.$$

But now, why 137? A straightforward calculus gives a possible origin; taking all integral N and P from all tables, we get a seemingly absurd suite of numbers that sums to:

$$\Sigma_{NP} = 2 + 3 + 4 + 5 + 7 + 8 + 9 + 12 + 14 + 16 + 19 + 38$$

$$= 137. \qquad (4.33)$$

Is that a coincidence, or rather the signature of a discrete wave packet? If one thinks of exponentiation, each term of the sum corresponds to a different piece of the phase of a unique signal which includes all symmetries and all the manners they combine, interact and condense (or ring). Since N and P are space currents, Σ_{NP} defines a universal oscillator. With respect to field theory, it is straightforward that such a wave includes or represents all virtual particles fields.

A complimentary result on $K \to 274$ seems doubtful; however, taking 266 from bosons instead of (-2, -7, -19), and the distinct values of K from leptons and quarks, we notice:

$$\Sigma_K = (2 \times 7 \times 19) + 2 + 3 + 4 + 5 - 6 = 274. \qquad (4.34)$$

The interpretation is less obvious and the link with known theory is nil, because this quantity addresses the effect in space of vibrations or rotations along the time axis and their participation to particles mass and interactions; there is no such concept in known theory.

In any case, those relations are complimentary to each other and provide with numerical coherence linked to the concepts developed before.

4.4.3 Splitting D_e and D'

Now, α is a 4D path length as seen in 3+1D, then the couplings D' (4.31) and D_e (3.1) should have a similar form but in a complimentary manner with respect to the resonance terms; hence they should also be expressed with similar expressions but using 3, 7, and 19 (the resonances of quarks) and $\Sigma_K = 274$; we find the following empirical fit which terms show an obvious symmetry:

$$D_e^{-1} = \sqrt{((7-3) \times (274+19))^2 + 7\pi^2 - \frac{19\pi}{19-1}}, \qquad (4.35.1)$$

$$D'^{-1} = \sqrt{((19-3)(274+3))^2 + 2^2 \times 3 \times 7\pi^2 - \frac{3}{3-1}}. \qquad (4.35.2)$$

Those expressions were used to compute the values in (3.1 – 4.31) and then all masses.

Several aspects are remarkable in those expressions:

- We notice that 274 + 3 = 277 and 274 + 19 = 293 are also prime numbers; hence those are not reducible. Their difference is 16 which is also $(7-3)^2$ in D_e and $(19-3)$ in D'.
- The rotation term $7\pi^2$ in (4.36.1) is a perfect fit with the μ and τ resonances (5 = 7 – 2, and 9 = 7 + 2), where 7 was inferred a rotation.
- D' includes a factor 2, which can be inserted in K in Table 5, but not in P; then P and K act on the time and magnetic moment axis respectively and it must be identical to the classical g-factor = 2. This is necessary since Table 5 is in the symmetry of QED.

Importantly, the expressions above are obtained by simple divisions based on the initial empirical fit of the D_e and D'. The left term is the closest square to the empirical value of D^{-2} from which it is subtracted; the middle integral term is the division of the rest by π^2 that gives a small residual term. Then we search to express all terms with integral numbers — preferably those we expect.

5 Gravitation, the keystone

The mass equation and the time-current model are coherent with Cramer's transactional interpretation of quantum mechanics which fills the gap of non-locality (the true signature of quantum physics) but without spooky action in 3-space.

Since the reasoning to the mass equation (thru N and P) and Cramer's interpretation are relevant in absorber theory and uses a pressure field, gravity must be analyzed in a shielding manner using Wheeler-Feynman equations [13, 14]; in this way, it was shown compatible with gravitation in a recent paper [6]. It does not require the existence of dark matter to explain the observations at the origin of this hypothesis and it also explains the cosmos energy densities (visible, dark, and visible + dark). In this section, we shall not restate the piece of theory in [6] but only the logic and main results.

The absorber free energy equivalent mass M_A is given by symmetry of the absorber process in gravitation; we first write the energetic part of the Schwarzshild metric:

$$c^2\,d\tau^2 = (c^2 - 2\,G\,m/r)\,dt^2 - \frac{c^2}{c^2 - 2\,G\,m/r}(dx^2 + dy^2 + dz^2)$$

Then, in the spirit of absorber theory, we symmetrize the equation in geometry and mass terms:

$$\frac{2\,G\,m}{r} = \frac{m\,R_u}{M_A\,r} \rightarrow \frac{R_U}{2\,M_A} = \frac{G}{c^2}, \tag{5.1}$$

$$\rightarrow M_A = \frac{R_U\,c^2}{2\,G} = \frac{P_p\,T}{2\,c^2} = 9.790 \times 10^{52}\ \text{kg} \tag{5.2}$$

where $R_U = cT$, $T = 1/H$ is the age of the event horizon while H is the Hubble factor and P_p is the Planck power.

Concerning visible energies $M_V c^2$, the ratio M_V/M_A is a geometrical constant. This constant links a 4-volume and a linear interaction in 3-space; the surface of a 4-sphere is $2\,\pi^2\,R^3$, and then the factor 2 in (5.2) becomes $4\,\pi^2$ in $3 + 1D$ where visible energies interact thru the light cone. It gives:

$$\frac{M_A}{M_V} = 2\pi^2 \rightarrow M_V = 4.453 \times 10^{51}\ \text{kg}. \tag{5.3}$$

Summing, we get the total energy M_U of the visible universe:

$$M_U = M_A + M_V = 9.236 \times 10^{52}\ \text{kg}. \tag{5.4}$$

It gives to a total density $\rho = 9.91 \times 10^{-27}$ kg/m^3 and the visible energy (5.3) is 4.82% of the total. The benchmark at this time is the Planck mission results [20] which gives $\rho = 9.90\,(6) \times 10^{-27}$ kg/m^3 and 4.86 (10)% of visible matter. Hence according to the standard model of cosmology we get valid quantities. The equation (5.1) also means that the rate of dark energy creation (M_A) since the initial bang is constant and half the non-reduced Planck power: the universe energy is identical to its expansion and we do not find a big bang but a permanent process. Next, using the Wheeler-Feynman equations or Newtonian gravity this creation gives an acceleration excess up to $H\,c$ at the galaxy borders, meaning the absence of dark matter.

But now what is the relation with our analysis of mass? According to (4.34 – 4.35), the numbers 137 and 266 address space and time respectively. They interfere at the point of origin which is visible thru the solid angle 4π, and we should find there the reduced Planck mass giving the Planck power:

$$M_p = \sqrt{\frac{hc}{G}} \times \frac{1}{4\pi} = 2.43536\,(6) \times 10^{18}\ \text{GeV/c}^2.$$

Using the mass equation (2.1) with the parameters in (3.1) and taking $N = P = 137^2$, and $K = +1/266^2$ gives a mass:

$$M = 2.464 \times 10^{18}\ \text{GeV/c}^2,$$

which is very close to Mp.

Looking at (4.36.1), we find $7\pi^2$ in the expression of D_e while 19 has a role similar to 7 in the case of quarks (N, Table 2) and bosons (K, and N = P = 19 – 7 in Table 4); then in order to symmetrize the equation we take:

$$N = P = 137^2 - 19\,\pi^2;\ K = +1/266^2,$$

$$M = 2.43526 \times 10^{18}\ \text{GeV/c}^2.$$

Finally, the next two decimals are given by addition of $\approx 2/3$ to $N = P$; a small empirical term which is expected as it makes this expression homogeneous to coupling:

$$N = P = 137^2 - 19\,\pi^2 + 2/3;\ K = +1/266^2, \tag{5.5}$$

$$M = 2.43536 \times 10^{18}\ \text{GeV/c}^2. \tag{5.6}$$

Since $1/(N\,P) < K\,D$ this resonance is not permitted in 3+1D.

Considering that we now discuss reconciliation of quantum theory and general relativity through a common origin this result is keystone on top of the study. It shows that the same field also leads gravitation.

Here we can define a unit-less quantum gravitational coupling constant which reads:

$$\alpha_G = \frac{X^2}{M_p^2} = \left(\frac{1}{(137^2 - 19\pi^2 + 2/3)^2} + \frac{D_e}{266^2}\right)^6, \tag{5.7}$$

$$\alpha_G = 1.1186 \times 10^{-47},$$

where we see that the rest of quantization lives in and from a single oscillator defining gravity; it is "below" quantum theory and it does not need the existence of a graviton particle.

Unlike the classical definitions of α_G, since X is universal and represents the pressure field, (5.7) is unique and does not depend on an arbitrary choice of mass.

But now the ratio of the electron mass to the Planck mass is constant, which seems a contradiction with (5.1). On the other hand, the observable cosmos has constant atomic physic and chemistry and then its laws use relative constants varying in time and not absolute ones. Thus, only unit-less quantities are constant; since G is used with constant masses in classical theories, then hc and G vary together in the same manner as $\alpha = e^2/\hbar c$ is constant.

Therefore, here is the big picture, the minimal interpretation of all results in this paper (no doubt it can be made more complex and elegant):

- A Planck particle exists at the origin; it emits a wave of Planck length and time. This resonance exists in 4 dimensions, it is not energy but its wave defines the quantum of action.
- This wave interacts thru the light cone (and gives 137 in α), and thru a radial line (giving 266 and 274). In a symmetric absorber concept, it means that the universe and its origin are quantizing each other.

- The emission is constant and corresponds to the Planck power; it builds M_A, and the visible energies field M_V is the absorber. It creates a deficit which is gravitation (see the absorber equations in [6]).
- For complete quantization thru time-currents, 137 is the sum of all resonances in space, and 266 is the product of the bottom space-type resonances, radial or circular (2, 7, 19).
- Increasing masses and the constancy of $e^2/\hbar c$ and hc/G are equivalent to and interpreted as time dilation. It denotes the emergence of the observable time in a frame where it does not exist. The observable time is seen as a radial progression in 4D space.

6 Discussion

Firstly, what have we been discussing all along? Essentially, the reasoning to the mass equations (2.1 – 2.2) is based on the existence of a stationary wave in a universe where:

a) everything propagates,

b) mass and charges do not have a proper existence,

c) the field is self-quantizing, and consequently a unique field and mechanism exists.

We end-up with a wide picture where all (free) parameters of the SM related to energy are self-quantizing geometry of movement (at this stage, and taking all results above, only the SM parameters expressed as phases or angles are not computed); the same is valid in gravitation and cosmology. *Hence we discuss the very nature of energy*, of its forms and formation on top of a unique field; something looking like the natural reductionist path of science.

Secondly, what does it means with respect to the standard theories? In its present form QFT neither considers *definite* rotations nor signals going up and down the time. Therefore no true comparison with our results is possible. Still, we find a number of connections like coupling constants and other aspects which will be discussed in the next paragraph.

In cosmology and using general relativity, a permanent energy creation is not even envisioned. Still, energy conservation comes from time-translation invariance and Neuter's theorem; but we know that the background (R_U) increases and then there is no mathematical reason that energy is conserved in cosmology.

The third point to discuss is the possibility of a different universe (a fashion question). But it seems unlikely because, as shown before, all resonances decay from $2, 3, 7 = 2^3 - 1$, and $19 = 3^3 - 2^3$; then probably only 2 and 3. It leads to conjecturing further the role of symmetry in the mass equation; essentially how do we get 2, 3, 7 and 19, and what is the limiting factor if any? Now let us reason on this aspect.

In the mass equations, the resonances N and P should come straight from the equation geometry and group theory. We shall use 1, 2, 3 to denote U(1), SU(2) and SU(3) respectively and discuss field polarization in the resonance equation. With field polarization p we mean dipoles or tri-poles where summing p charges makes a neutral. In the following, one must just keep in mind that $U(1) \subset SU(2) \subset SU(3)$.

- At the core of a particle resonance, time currents give a charge Q constant; its polarity is p (in 2, 3). In any sphere centered on Q the sum of charges is Q. Then except for Q, the total charge separation in a scale-independent 3-sphere depends on a cube, say n^3 (since the resonance radius is arbitrary) and it is neutral.
- In the radial case, with resonance P, on each layer of the resonance the radial action is layer independent, then the radial coefficient of polarization in $1/n^2$ for each layer, with $1 \leqslant n \leqslant P$; then $P = n$. The polarity of Q is p and defines the interaction of the particle which is also radial, then on the radial path $n = p = P$. Here P defines simultaneously a radial exchange of action and polarity (the symmetry). This is immediately verified for quarks ($P = 3$), and for leptons if we decompose P as shown before.
- On a circular path, a resonance N gives N circular sectors with identical action and action coefficients. Then $N = n^3$ on this path. Since this number does not define the radial interaction of the particle, any subgroup of p is acceptable, then $1 \leqslant n \leqslant p$.

We get the following suites of numbers:

- On a radial path the polarity is p, and $P = p = 2$ or 3;
- On a circular path the polarity is n with $1 \leqslant n \leqslant p \rightarrow N = n^3$; limited to $2 \rightarrow 8, 3 \rightarrow 27$.

But the latter is a rotation, not a resonance as needed, and we need to complete the reasoning.

With geometry and currents (and nothing else), the logical manner is to combine symmetries. Say in the resonance volume we have two symmetries at work; a structural point of equilibrium needs a transformation. Therefore, on the circular path a resonance is seen as a transformer in $n \leq p$ and the subgroups of n, where coefficients are the same for n and its subgroup. Hence, on circular paths we get cubes differences 7 and 19; those come like transformer of charges or currents between a group and its sub-group. That is to say that the field polarization $n \rightarrow n^3$ is always balanced by $(n-1) \rightarrow (n-1)^3$. Importantly, there is nothing in this reasoning preventing more complex oscillators, for instance $19 - 7 = 12$.

This discussion leads to introduce U(1) which is a very special case; since $1^3 - 0^3 = 1$ it seems to be a massless field with any oscillator; the same reasoning on 0 suggests a continuous current — an amplitude according to which masses and then the observable lengths and the rate of time vary in reverse proportions.

Now why only 0, 1, 2, and 3? Within the logic above, the first mathematical explanation is Hurwitz theorem [10]. Consider two charges or currents x and y, we may need to

compute the impact of y on the self-interaction of x; it is equal to the action on x of the interaction of x and y (conservation and symmetry), then:

$$(x\,x)\,y = x\,(x\,y).$$

This is the definition of alternative algebra; according to Hurwitz theorem [10], only four exist which are R real numbers, C complex numbers ~ U(1), H quaternions ~ SU(2), and O octonions ~ SU(3). One can consider this as limiting either the symmetry spectrum, or just our ability to model with particles and charges — or both.

A peculiar case arises with $x\,X = 1$ or unitary; the impact of $x\,X$ on any other quantity of the same group does not change its amplitude. Then $x\,X$ addresses structural conservation and we find simultaneously 137, $1/137$, and 274, $1/274$ in the expression of couplings.

In this way those quantities are related to the monopole as quantized rotational 4D paths, like α and D_e, where only couplings can be measured in 3+1D as seemingly arbitrary real numbers.

7 Conclusions

The breakthrough to wave equations was the assumption of a stationary wave pervading all space. But how can such a wave exist in relativity without a mass of its own? How could it be distinct from the mass of the particle or system it describes? Then how could it be distinct from gravitation?

Those naive but unsolved questions are almost a century old as they address the nature of the wave, wave-particle duality, the completeness of quantum mechanics, and the physical link between gravitation and quantum physics.

The novelty here is that those questions are justified by the existence of a solution to the free parameters problem, including and linking particles physics, gravitation, and cosmology, not only by conceptual disagreements or theoretical incompatibilities.

As stated in introduction, we do not solve any equation; the existence of a solution is first seen when the mass equation is fit to phenomenology, and then extended to couplings. We find logical coherence, a reductionist concept and fantastic precision. Of course it does not look like the usual manner in modern physics where theory and principles reign; but, considering the difficulty of solving this problem from theory, it might be the only practicable way — at least at present time.

As a matter of conclusion, it looks as though the solution shown here can be found only as a whole and provided that we do not build on existing concepts (and maybe even principles); but one must first recognize the existence of a problem together with its ramifications. This situation is fantastic and terrible; if *that* solution exists, physics could remain stuck endlessly in its present conceptual state *because of this conceptual state*: whatever new particles discovered in collision machines modeled with ad-hoc SM extensions, its framework may never be contradicted by experiment.

8 Addendum

As for the 750 GeV resonance possibly detected at CERN [21], since it decays to two photons we assume the same equation and parameters as the H^0 and only K can be fit; it gives $K = -133/2$ which is immediately remarkable. However, since K is not integral the width must be reconsidered, logically to $\Delta K = 1/4$, giving from (2.2 – 3.6 – 3.9.3):

$$N = P = 12, K = -133/2 \rightarrow m \approx 744.9\ \mathrm{GeV/c^2},$$

$$\Delta K = 1/4 \rightarrow \Gamma \approx 9.6\ \mathrm{GeV/c^2}.$$

Using (3.9.2) instead of (3.9.3) adds +3 $\mathrm{GeV/c^2}$ to the mass. The other candidate with $\Delta K \approx 1$ gives $\Gamma \approx 40\ \mathrm{GeV}/c^2$.

At this scale, the equation (2.2) is very sensitive to D and the model in time-currents must be identical to the H^0 otherwise the computed mass is far from the estimate. It would be very similar, but it leads to remark that there are two manners to put four distinct charges at the corners of a tetrahedron; there may be a chiral difference with the H^0, justifying distinct masses and a probable impact on the particle decays.

Last, the number 133/2 verifies (2.3) like 7 and 19, but with $P = 1$ instead of $P = 3$, since $133\,\pi/2 \approx 209/1.0004$. It is even doubly remarkable since 209 is multiple of 19.

Hence the existence of this particle, if confirmed, should not change the values of Σ_{NP} and Σ_K; it fits well and naturally with the logic and results in this paper.

Submitted on April 3, 2016 / Accepted on April 30, 2016

References

1. Bethe H.A. The electromagnetic shift of energy levels. *Phys. Rev.*, 1947, v. 72, 339.
2. De Broglie L. Recherches sur la théorie des quanta. *Annales de Physique* — 10e sérié — Tome III — Janvier-Février 1925.
3. Consiglio J. Below the Standard Model? *Applied Physics Research*, v. 6(2), 19–27, 2014.
4. Consiglio J. On the Field Below SM, Analysis and Predictions *Applied Physics Research*, v. 6(5), 1–17, 2014.
5. Consiglio J. Hacking the Fine Structure Constant in Leptons Geometry *Applied Physics Research*, v. 7(6), 30–46, 2015.
6. Consiglio J. On the Absorber in Gravitation. *Progress in Physics*, v. 12(1), 20–25, 2016.
7. Cramer J. The transactional interpretation of quantum mechanics. *Rev. Mod. Phys.*, 1986, v. 58(3).
8. Dirac P.A.M. Quantized singularities in the Electromagnetic Field, *Proc. Roy. Soc. A*, 1931, v. 133, 60.
9. Dirac P.A.M. (1948). The Theory of Magnetic Poles. *Phys. Rev.*, 1948, v. 74, 817.
10. Dickson L.E. (1919). On Quaternions and Their Generalization and the History of the Eight Square Theorem. *Ann. Math.*, 1919, v. 20, 155.
11. Poincaré H. Sur la dynamique de l'électron. *Rendiconti del Circolo Matematico di Palermo*, 1903, v. 21, 129–176.
12. Schwinger J. On Quantum-Electrodynamics and the Magnetic Moment of the Electron. *Physical Review*, 1948, v. 73(4), 416.
13. Wheeler J.A., Feynman R.P. Interaction with the Absorber as the Mechanism of Radiation. *Reviews of Modern Physics*, 1945, v. 17(2–3), 157–161.

14. Wheeler J.A., Feynman R.P. Classical Electrodynamics in Terms of Direct Interparticle Action. *Reviews of Modern Physics*, 1949, v. 21(3), 425–433.

15. ATLAS collaboration. Measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4l$ channels in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector. *(*Phys. Rev. D), 2014, v. 90(5), 052004.

16. CMS collaboration. Precise determination of the mass of the Higgs boson and studies of the compatibility of its couplings with the standard model (Technical report). 2014, CMS-PAS-HIG-14-009. CERN. https://cds.cern.ch/record/1728249?ln=en

17. ATLAS and CMS Collaborations. Combined Measurement of the Higgs Boson Mass in pp Collisions at sqrt(s)=7 and 8 TeV with the ATLAS and CMS Experiments. 2015. arXiv: 1503.07589.

18. The ATLAS, CDF, CMS and D0 Collaborations. First combination of Tevatron and LHC measurements of the top-quark mass. 2014. arXiv, 1403.4427

19. The CMS Collaboration. Measurement of the top-quark mass in top-quark pair events with lepton+jets final states in pp collisions at sqrt(s)=8 TeV. 2014. CMS Physics Analysis Summary TOP-14-001. https://cds.cern.ch/record/1690093?ln=en

20. The Planck Collaboration. Planck 2015 results. I. Overview of products and scientific results. arXiv: 1502.01582

21. ATLAS note, ATLAS-CONF-2015-081, Search for resonances decaying to photon pairs in 3.2 fb-1 of pp collisions at sqrt(s) = 13 TeV with the ATLAS detector. CMS note, CMS PAS EXO-15-004 Search for new physics in high mass diphoton events in proton-proton collisions at 13 TeV.

On the Nature of Ball Lightning

Anatoly V. Belyakov
E-mail: belyakov.lih@gmail.com

The author proposes a model of ball lightning based on a mechanistic interpretation of John Wheeler's ideas. It is assumed that ball lightning is a quasi-particle that has the Planck mass and consists of closed contours, which in turn are based on the magnetic and gravity force balance. These contours are hard packed in a small volume of ball lightning, forming a multilayer capacitor containing a substantial charge and electrostatic energy. This paper provides calculations of characteristic parameters of ball lightning, which are well consistent with its phenomenology.

1 Introduction

There are many theories about ball lightning. However, the nature of this mysterious phenomenon remains unclear. This paper proposes a model of ball lightning based on a mechanistic interpretation of John Wheeler's concept. Previously, such an approach has been successfully applied to construct both micro-world and space models (see [1–4], etc.).

To some extent, the model proposed is similar to the quantum model by Geert Dijkhuis, Professor at Eindhoven University of Technology and Secretary of the International Committee on Ball Lightning. His model suggests that ball lightning is a macroscopic quantum object. Earlier, a similar hypothesis was proposed by Boris Ignatov [4]. Nevertheless, it must be noted that there is no complete understanding of the nature of such objects. It is assumed that a quasiparticle cannot carry a substance. It only carries energy, pulse, and momentum, while the electrons inside such an object are completely coherent and make up a single wave function. Energy of such a quasiparticle is gradually dissipating, in the visible range in particular. Therefore, ball lightning can be observed as an optical object.

2 Presuppositions

Recall that (according to Wheeler) there are original primary elements of space and matter, which have different names — *wormholes, appendices, current tubes, threads or force lines of a field*. If they are real objects and not just mathematical abstractions, in physical terms, they must be some kind of vortex structures resting on the phase boundary (surface). In particular, Wheeler treats charges as singular points on the surface source-drain connected by current tubes in an additional dimension forming a closed *contour*.

Paper [1] shows that, from a purely mechanistic point of view, the *charge* is proportional to its momentum about the contour of the vortical tube and reflects the extent of non-equilibrium of physical vacuum; *spin* is proportional to the angular momentum relative to the longitudinal axis of the contour, respectively; and the *magnetic interaction* between the conductors is similar to forces existing between the current tubes. It is customary that a single element of such a tube is an element with the size of a classical electron radius r_e and its mass m_e.

The model of ball lightning was only built using the ratio of fundamental interactions as in the above-mentioned papers of the author. The mechanistic interpretation of Wheeler's ideas makes it possible to record formulae for electric and magnetic forces in a *Coulomb-free* form, where the charge is replaced with the ultimate electron momentum. In this case, the electric and magnetic constants, ε_0 and μ_0, are as follows:

$$\varepsilon_0 = \frac{m_e}{r_e} = 3.33 \times 10^{-16}\ \text{kg/m}, \tag{1}$$

$$\mu_0 = \frac{1}{\varepsilon_0 c^2} = 0.0344\ \text{N}^{-1}, \tag{2}$$

where the electric constant becomes linear density of the vortical tube, and the reciprocal of the magnetic constant is the centrifugal force produced by rotation of the vortical tube element with m_e mass with the velocity of light c along r_e radius. This value is also equivalent to the force existing between two elementary charges at the given radius.

For the purpose of mutual comparison of interactions, formulae for the electric, magnetic, gravity, and inertial forces are written in a dimensionless form with a single dimension factor of force $1/\mu_0$. With (1) and (2) in mind, we have the following:

$$F_e = \frac{1}{\mu_0}\left(\frac{r_e}{r_0}\right)^2 z_{e_1} z_{e_2}\,, \tag{3}$$

$$F_m = \frac{1}{\mu_0}\left(\frac{l}{2\pi r_0}\right)\left(\frac{r_e}{c \times [\text{sec}]}\right)^2 z_{e_1} z_{e_2}\,, \tag{4}$$

$$F_g = \frac{1}{\mu_0}\frac{1}{f}\left(\frac{r_e}{r_0}\right)^2 z_{g_1} z_{g_2}\,, \tag{5}$$

$$F_i = \frac{1}{\mu_0}\frac{r_e}{r_0}\left(\frac{V_0}{c}\right)^2 z_g\,, \tag{6}$$

where V_0, r_0, l, z_e, z_g, and f stand for circumferential velocity, circumferential radius or distance between vortical tubes, length of the vortex tube (thread) or contour, relative values of the charge and mass of the electron charge and mass, and the

electric-gravity force ratio, respectively, with the latter having the following formula with the same designations:

$$f = \frac{c^2}{\varepsilon_0 \gamma} = 4.16 \times 10^{42}, \tag{7}$$

where γ is the gravitational constant.

3 Calculation of characteristic parameters of ball lightning

Ball lightning often originates from streak one. Imagine streak lightning as a bundle of vortex threads, which (under certain conditions) form vortical current tubes. The latter, in turn, are closed into contours. It is obvious that there must be balances of some pairs of interactions for ball lightning to exist. They are the following:

1. Ball lightning mass M satisfies the condition of equality of *electric and gravity forces*, so with unit charges we have the following:

$$M = f^{1/2} m_e = 1.86 \times 10^{-9} \text{ kg}, \tag{8}$$

which is in agreement with the Plank mass by order of magnitude.

2. Closed contour branches with opposed currents satisfy the balance of *magnetic and gravity forces* resulting in a linear geometric mean dimension of the contour:

$$l_k = (r_0\, l)^{1/2} = \left(\frac{z_{g_1} z_{g_2}}{z_{e_1} z_{e_2}}\right)^{1/2} \left(\frac{2\pi}{f}\right)^{1/2} c \times [\text{sec}], \tag{9}$$

where the ratio of the product $\varepsilon = (z_{g_1} z_{g_2})/(z_{e_1} z_{e_2})$ is an *evolutionary parameter* that characterizes the state of the environment and its changes as the mass carriers dominate over electric ones, and in fact shows the distinction between material medium and vacuum. Hereinafter, we shall take it as being close to the unit in our case, while $l_k = 3.68 \times 10^{-13}$ m, and the vortical tube's mass is

$$m_k = \varepsilon_0 l_k = 1.19 \times 10^{-28} \text{ kg}. \tag{10}$$

In addition, if we express vortical tubes' masses in (9) — $z_g m_e$ as $\varepsilon_0 l$, then we shall get the following relation between the contour axes for unit charges with (7) in mind:

$$\frac{r_0}{l} = 2\pi \rho_e \gamma \times [c^2] = 17070 \approx a^2, \tag{11}$$

where ρ_e is the electron density equal to $m_e / r_e^3 = 4.071 \times 10^{13}$ kg/m^3; and a is the reverse fine structure constant equal to 137.036. Thus, the individual contour is most likely to have axes equal to the size of an electron r_e and Bohr atom $r_e a^2$.

3. Vortical tubes of a contour consist of a number of unidirectional parallel individual vortex threads spinning about the longitudinal axis of the contour with circumferential velocity V_{0i}. Their stability is ensured by the balance of *magnetic and inertial forces*, which give rise to the following formula:

$$V_{0i} = \left(\frac{z_{e_1} z_{e_2}}{z_g}\right)^{1/2} \left(\frac{r_e l}{2\pi}\right)^{1/2} \frac{1}{[\text{sec}]}. \tag{12}$$

Individual vortex filaments having length l and mass carriers in the number of $z_g = l/r_e$ are spinning about the longitudinal axis along an indefinite radius. In the case of unit charges, we have the following minimum circumferential velocity about the longitudinal axis:

$$V_{0i} = \frac{r_e}{(2\pi)^{1/2} \times [c]} = 1.124 \times 10^{-15} \text{ m/sec}. \tag{13}$$

The total number of contours (and the same of unit charges, respectively) may be as follows:

$$z = \frac{M}{m_k} = 1.56 \times 10^{19}. \tag{14}$$

The way these contours are packed in the volume of ball lightning is unclear. Possibly, a contour may be one-dimensional with the total length of $z \times l_k$. It can be expected that with transformation into the more energetically favourable structure the contour (folding repeatedly) forms a large number of loops or cells, which are enclosed in a spherical volume with a bright centre (nucleus). In both cases, with the elements being the most densely packed in the volume, the reduced minimum linear dimension of the outer spherical surface will be as follows:

$$l_{min} = z^{1/2} l_k = 0.00145 \text{ m}, \tag{15}$$

However, if we consider the ratio in (11) and take one of the axes of the Bohr radius instead of l_k for individual contours, then we can estimate the maximum size of such a sphere as $l_{max} \approx 0.00145 \times a \approx 0.2$ m.

Let us calculate the rest limit parameters of ball lightning — energy, charge, electric potential of streak lightning required for generation of ball lightning, and its ultimate density:

$$E_{lim} = Mc^2 = 1.67 \times 10^8 \text{ J}, \tag{16}$$

$$q_{lim} = z e_0 = 2.50 \text{ K}, \tag{17}$$

$$U_{lim} = \frac{E_{lim}}{q_{lim}} = 6.68 \times 10^7 \text{ V}. \tag{18}$$

The density will be calculated taking l_{min} as the sphere's diameter:

$$\rho_{lim} = \frac{M}{\frac{4}{3}\pi \left(\frac{l_{min}}{2}\right)^3} = 1.17 \text{ kg/m}^3, \tag{19}$$

which corresponds to the air density.

Contour branches with parallel unidirectional currents have to twist, so ball lightning contours are gradually opening losing the charge. Therefore, ball lightning has a sort of

an electrostatic tail behind. The maximum lifetime of ball lightning can be determined in a similar way as the neutron lifetime [2], i.e., as the *time constant* of the contour deformation (the ratio of the contour's characteristic dimension to the circumferential velocity):

$$\tau_{lim} = \frac{l_k}{V_{0i}} = 327 \text{ sec.} \tag{20}$$

When crossing the initial surface of our world, an open contour (vortical tube) actually forms an elementary charge (according to Wheeler). It can be assumed that the physical basis of ball lightning is formed by electrons. Their fermionic part is arranged into corresponding structures observable in the form of a fireball, while their bosonic parts converge in the centre of the ball going to the additional dimension (Y area) [2].

Let us also determine the capacity, electrostatic energy, and size required for ball lightning with q_{lim} charge. Paper [**?**] found a connection between new electrical units in a *Coulomb-free* form and SI-system units. It was shown that the mass of electrons 2.90×10^{-6} kg on the capacitor plates corresponded to one Farad. Velocity of 587 m/sec corresponded to 1 Volt, with the electrostatic capacity of the surface at which the charge begins to flow spontaneously into the external environment being $U_m = 511,000$ V.

Thus, ball lightning has the following capacity:

$$C = \frac{zm_e}{2.90\times10^{-6}} = 4.89\times10^{-6} \text{ F}, \tag{21}$$

with the same result in the SI system:

$$C = \frac{q_{lim}}{511,000} = 4.89\times10^{-6} \text{ F}, \tag{22}$$

and the maximum electrostatic energy of ball lightning being

$$E_m = \frac{1}{2} CU_m^2 = 6.39\times10^5 \text{ J}. \tag{23}$$

To have such a capacity, ball lightning must have a multilayer structure, e.g., the structure of a multilayer spherical capacitor. Paper [5] shows that the average distance between unit charges of a charged sphere with R radius is $\pi(Rr_e)^{1/2}$. Let us assume that the average linear dimension between the charges in the volume of ball lightning is the same. Then we can determine the size of ball lightning through the following equation:

$$\left(\frac{\frac{4}{3}\pi R^3}{z}\right)^{1/3} = \pi\left(Rr_e\right)^{1/2}, \tag{24}$$

therefore

$$R = \left(\frac{3\pi^2 z}{4}\right)^{2/3} r_e = 0.067 \text{ m}. \tag{25}$$

Let us determine the temperatures of the nucleus and the outer shell with the assumption that the radiation of ball lightning is the radiation of a blackbody. If the total energy is evenly lost over the lifetime of ball lightning, the average radiation power shall be as follows:

$$N = \frac{E_m}{\tau_{lim}} = 1950 \text{ W}, \tag{26}$$

then

$$T = \left(\frac{N}{\sigma S}\right)^{1/4}, \tag{27}$$

where σ is a Stefan-Boltzmann constant equal to 5.67×10^{-8} Wm^{-2} $(^\circ\text{K})^{-4}$; and S is the area of the spherical surface of ball lightning. Taking l_{min} as the nucleus diameter and l_{max} as the diameter of the outer shell, we calculate the respective areas S and determine their temperatures using formula (27) — 8.500°K and 724°K. External appearance of ball lightning, its behaviour, and results of its effect on the environment are extremely varied. Given its unpredictability, it is rarely possible to obtain objective instrumental data on ball lightning.

In his paper [6], Mikhail Dmitriev — a chemist having vast experience in working with low-temperature plasma — describes an encounter with ball lightning and an attempt to make a chemical analysis of ionized air behind it. Based on the analysis results, the author estimated the potential of ball lightning discharge at 300–400 kV. The temperature, degree of ionization, and concentration of charged particles in ball lightning was estimated at 1.14×10^{17} per cm^3 judging on its glow. It is easy to calculate that, in accordance with the proposed model and given such a concentration, the estimated diameter of ball lightning with z unit charges shall be 6.4 cm, which corresponds to its typical size. This means that the discharge potential and charge concentration of real ball lightning encountered by Dmitriev are consistent with the estimated model.

Since ball lightning does not consist of atoms and molecules, it does not interact with molecules of other media. This explains its ability to penetrate through obstacles and move against the wind, but actively respond to electric and magnetic fields at the same time.

Finally, it should be noted that people often associate ball lightning with a living being. Let us assume that life can be organized on another material basis. Then, indeed, given the number of unit elements ($z = 1.56\times10^{19}$) and complexity of their packing in the volume of ball lightning, it is appropriate to draw an analogy with a DNA strand, which is two meters long, packed in a microscopic cell nucleus, and contains information about the structure and behaviour of a living organism.

4 Conclusion

Thus, model ball lightning is a ball with its size ranging from 0.14 to 20 cm (its typical diameter is 13.4 cm), having density of no more than 1.17 kg per m^3, glow temperature of 724 to 8,500°K, and energy of 639 kJ concentrated in a small volume in the form of an electrostatic charge with 511 kV potential.

During the lifetime of ball lightning (up to 6.5 minutes), it is constantly losing the charge leaving an ionised trail behind. Ball lightning is able to penetrate obstacles.

In general, eyewitness accounts are in good agreement with the calculated characteristic parameters of the model object. Of course, some phenomena of real ball lightning fall outside the scope of the obtained characteristic parameters. At least this is due to the fact that its charge can be formed by not only electrons, but also by ions, and the evolutionary parameter ε may exceed the unit.

The ball lightning phenomenon and its complete internal organization can only be understood on the basis of an appropriate theory. However, from a phenomenological point of view, this model of ball lightning is in good agreement with the real object by its appearance and its basic aspects, and can serve as the basis for such a theory.

Submitted on April 28, 2016 / Accepted on May 20, 2016

References

1. Belyakov, A.V.: Charge of the electron, and the constants of radiation according to J.A.Wheeler's geometrodynamic model. *Progress in Physics*, 2010, vol. 6, issue 4, 90–94.
2. Belyakov A.V. Macro-analogies and gravitation in the micro-world: further elaboration of Wheeler's model of geometrodynamics. *Progress in Physics*, 2012, vol. 8, issue 2, 47–57.
3. Belyakov A.V. Evolution of stellar objects according to J. Wheeler's geometrodynamic concept. *Progress in Physics*, 2013, vol. 9, issue 1, 25–40.
4. Ignatov B.A. Ball lightning as a quasi-particle's child. In the book series: *Club of Fundamental Natural-Science Ideas. Hypotheses. Judgments. Opinions. Speculations*. Issue 1. Fenid, 1990.
5. Belyakov A.V. On the uniform dimension system. Is there the necessity for Coulomb? *Progress in Physics*, 2013, vol. 9, issue 3, 142–143.
6. Dmitriev M.T. Ball lightning nature. *Priroda*, June 1967, 98–106 (*in Russian*).

LETTERS TO PROGRESS IN PHYSICS

Dialogue Concerning the Two Chief World Views

Craig Alan Feinstein
2712 Willow Glen Drive, Baltimore, MD 21209, USA. E-mail: cafeinst@msn.com

In 1632, Galileo Galilei wrote a book called *Dialogue Concerning the Two Chief World Systems* which compared the new Copernican model of the universe with the old Ptolemaic model. His book took the form of a dialogue between three philosophers, Salviati, a proponent of the Copernican model, Simplicio, a proponent of the Ptolemaic model, and Sagredo, who was initially open-minded and neutral. In this paper, I am going to use Galileo's idea to present a dialogue between three modern philosophers, Mr. Spock, a proponent of the view that $\mathsf{P} \neq \mathsf{NP}$, Professor Simpson, a proponent of the view that $\mathsf{P} = \mathsf{NP}$, and Judge Wapner, who is initially open-minded and neutral.

Since 2006, I have published four proofs that $\mathsf{P} \neq \mathsf{NP}$ [5–8]. Yet at the present time, if one asks the average mathematician or computer scientist the status of the famous P versus NP problem, he or she will say that it is still open. In my opinion, the main reason for this is because most people, whether they realize it or not, believe in their hearts that $\mathsf{P} = \mathsf{NP}$, since this statement essentially means that problems which are easy to state and have solutions which are easy to verify must also be easy to solve. For instance, as a professional magician, I have observed that most laymen who are baffled by an illusion are usually convinced that the secret to the illusion either involves extraordinary dexterity or high technology, when in fact magicians are usually no more dexterous than the average layman and the secrets to illusions are almost always very simple and low-tech; as the famous designer of illusions, Jim Steinmeyer, said, "Magicians guard an empty safe" [13]. The thinking that extraordinary dexterity or high technology is involved in a magician's secret is, in my opinion, due to a subconscious belief that $\mathsf{P} = \mathsf{NP}$, that problems which are difficult to solve and easy to state, in this case "how did the magician do it?", must have complex solutions.

I have had many conversations in which I have tried to convince all types of people, from Usenet trolls to graduate students to professors to famous world-class mathematicians, that $\mathsf{P} \neq \mathsf{NP}$ with very little success; however, I predict that there will soon come a day when the mainstream mathematics and computer science community will consider people who believe that $\mathsf{P} = \mathsf{NP}$ to be in the same league as those who believe it is possible to trisect an angle with only a straightedge and compass (which has been proven to be impossible) [14].

I got the idea to write this paper after I learned of Galileo's book *Dialogue Concerning the Two Chief World Systems* [4], which presents a dialogue between three philosophers, Salviati, a proponent of the new Copernican model, Simplicio, a proponent of the old Ptolemaic model, and Sagredo, who was initially open-minded and neutral. The dialogue that follows is a dialogue between three modern philosophers, Mr. Spock, a proponent of the view that $\mathsf{P} \neq \mathsf{NP}$, Professor Simpson, a proponent of the view that $\mathsf{P} = \mathsf{NP}$, and Judge Wapner, who is initially open-minded and neutral. Professor Simpson, who is a fictitious anglicized straw man character like Simplicio, is a composite of many of the people whom I have had discussions with over the years about the P versus NP problem. He presents many challenges and questions, all of which have been raised before by real people, that Mr. Spock, the epitome of truth and logic, attempts to answer. And Judge Wapner, the epitome of open-mindedness and fairness, always listens to both sides of their arguments before drawing conclusions.

Spock: Yesterday we discussed the P versus NP problem [2, 3] and agreed that it is a problem of not only great philosophical importance, but also it has practical implications. We decided to look at a proof that $\mathsf{P} \neq \mathsf{NP}$ offered by Craig Alan Feinstein in a letter entitled "A more elegant argument that $\mathsf{P} \neq \mathsf{NP}$" [8]. The proof is surprisingly short and simple:

Proof: Consider the following problem: Let $\{s_1, \ldots, s_n\}$ be a set of n integers and t be another integer. Suppose we want to determine whether there exists a subset of $\{s_1, \ldots, s_n\}$ such that the sum of its elements equals t, where the sum of the elements of the empty set is considered to be zero. This famous problem is known as the SUBSET-SUM problem.

Let $k \in \{1, \ldots, n\}$. Then the SUBSET-SUM problem is equivalent to the problem of determining whether there exist sets $I^+ \subseteq \{1, \ldots, k\}$ and $I^- \subseteq \{k+1, \ldots, n\}$ such that

$$\sum_{i \in I^+} s_i = t - \sum_{i \in I^-} s_i.$$

There is nothing that can be done to make this equation simpler. Then since there are 2^k possible expressions on the left-hand side of this equation and 2^{n-k} possible expressions on the right-hand side of this equation, we can find a lower-bound for the worst-case running-time of an algorithm that solves the SUBSET-SUM problem by minimizing $2^k + 2^{n-k}$ subject to $k \in \{1, \ldots, n\}$.

When we do this, we find that $2^k+2^{n-k} = 2^{\lfloor n/2 \rfloor}+2^{n-\lfloor n/2 \rfloor} = \Theta(\sqrt{2^n})$ is the solution, so it is impossible to solve the SUBSET-SUM problem in $o(\sqrt{2^n})$ time; thus, because the Meet-in-the-Middle algorithm [10,11,15] achieves a running-time of $\Theta(\sqrt{2^n})$, we can conclude that $\Theta(\sqrt{2^n})$ is a tight lower-bound for the worst-case running-time of any deterministic and exact algorithm which solves SUBSET-SUM. And this conclusion implies that $\mathsf{P} \neq \mathsf{NP}$. □

To me, Feinstein's proof is not only logical but elegant too. Also, his conclusion is confirmed by history; just as Feinstein's theorem retrodicts, no deterministic and exact algorithm that solves SUBSET-SUM has ever been found to run faster than the Meet-in-the-Middle algorithm, which was discovered in 1974 [10, 15].

Simpson: But there is an obvious flaw in Feinstein's "proof": Feinstein's "proof" only considers a very specialized type of algorithm that works in the same way as the Meet-in-the-Middle algorithm, except that instead of sorting two sets of size $\Theta(\sqrt{2^n})$, it sorts one 2^k-size set and one 2^{n-k}-size set. Under these restrictions, I would agree that the Meet-in-the-Middle algorithm is the fastest deterministic and exact algorithm that solves SUBSET-SUM, but there are still many possible algorithms which could solve the SUBSET-SUM problem that the "proof" does not even consider.

Wapner: Professor Simpson, where in Feinstein's proof does he say that he is restricting the algorithms to the class of algorithms that you mention?

Simpson: He does not say so explicitly, but it is obviously implied, since there could be algorithms that get around his assertion that the minimum number of possible expressions on both sides is $\Theta(\sqrt{2^n})$.

Spock: How do you know that there could be such algorithms?

Simpson: I do not know, but the burden of proof is not on me; it is on Feinstein. And he never considers these types of algorithms.

Wapner: It is true that Feinstein never explicitly considers algorithms which work differently than the Meet-in-the-Middle algorithm, and the burden of proof is on Feinstein to show that these types of algorithms cannot run any faster than $\Theta(\sqrt{2^n})$ time.

Spock: Professor Simpson, is the burden of proof on Feinstein to consider in his proof algorithms which work by magic?

Simpson: No, only algorithms that are realistic.

Spock: Then why would you think that algorithms that get around the assertion that the minimum total number of possible expressions on both sides is $\Theta(\sqrt{2^n})$ are realistic?

Simpson: I do not know, but the burden of proof is not on me; it is on Feinstein.

Spock: Have you considered the fact that an algorithm which determines in $o(\sqrt{2^n})$-time whether two sets of size $\Theta(\sqrt{2^n})$ have a nonempty intersection *must* work by magic, unless there is a way to mathematically reduce the two sets into something simpler?

Wapner: Yes, I see your point; the minimum total number of possible expressions on each side of the SUBSET-SUM equation puts a natural restriction on the time that an algorithm must take to solve the SUBSET-SUM problem.

Simpson: But how do you know it is impossible to reduce the SUBSET-SUM problem into something simpler, so that the number of possible expressions on both sides is $o(\sqrt{2^n})$?

Spock: Simple algebra. Try to simplify the SUBSET-SUM equation above. You cannot do it. The best you can do is manipulate the equation to get $\Theta(\sqrt{2^n})$ expressions on each side.

Simpson: I'll agree that you cannot do it algebraically, but what about reducing the SUBSET-SUM problem to the 3-SAT problem in polynomial-time? This can be done since 3-SAT is NP-complete. If there is an algorithm that can solve 3-SAT in polynomial-time, then it would also be able to solve SUBSET-SUM in polynomial-time, contradicting Feinstein's lower-bound claim of $\Theta(\sqrt{2^n})$.

Spock: But this is magical thinking. If a problem is shown to be impossible to solve in polynomial time, then reducing the problem to another problem in polynomial-time will not change the fact that it is impossible to solve the first problem in polynomial time; it will only imply that the second problem cannot be solved in polynomial time.

Wapner: Spock is right about this. Do you have any other objections to Feinstein's argument?

Simpson: I have many objections. For instance, Feinstein's argument can be applied when the magnitudes of the integers in the set $\{s_1, \ldots, s_n\}$ and also t are assumed to be bounded by a polynomial to "prove" that it is impossible to solve this modified problem in polynomial-time. But it is well-known that one can solve this modified problem in polynomial-time.

Spock: But Feinstein's argument in fact *cannot* be applied in such a circumstance, because there would only be a polynomial number of possible values on each side of the equation,

even though the total number of possible expressions on each side is exponential. Feinstein's argument implicitly uses the fact that the total number of possible values on each side of the SUBSET-SUM equation is usually of the same order as the total number of possible expressions on each side, when there is no restriction on the magnitude of the integers in the set $\{s_1, \ldots, s_n\}$ and also t.

Simpson: Then here is a better objection: Suppose the set $\{s_1, \ldots, s_n\}$ and also t consist of vectors in $\mathbb{Z}_2^m$ for some positive integer m, instead of integers. Then one could use the same argument that Feinstein uses to "prove" that it is impossible to determine in polynomial-time whether this modified SUBSET-SUM equation has a solution, when in fact one can use Gaussian elimination to determine this information in polynomial-time.

Spock: Feinstein's argument would not apply to this situation precisely because one can reduce the equation

$$\sum_{i \in I^+} s_i = t - \sum_{i \in I^-} s_i.$$

to a simpler set of equations through Gaussian elimination. But when the set $\{s_1, \ldots, s_n\}$ and also t consist of integers, nothing can be done to make the above equation simpler, so Feinstein's argument is applicable.

Simpson: OK, then how would you answer this? Consider the Diophantine equation:

$$s_1 x_1 + \ldots + s_n x_n = t,$$

where x_i is an unknown integer, for $i = 1, \ldots, n$. One could use the same argument that Feinstein uses to "prove" that it is impossible to determine in polynomial-time whether this equation has a solution, when in fact one can use the Euclidean algorithm to determine this information in polynomial-time.

Spock: But again Feinstein's argument would not apply to this Diophantine equation, precisely because this Diophantine equation can be reduced via the Euclidean algorithm to the equation,

$$\gcd(s_1, \ldots, s_n) \cdot z = t,$$

where z is an unknown integer. And it is easy to determine in polynomial-time whether this equation has an integer solution by simply testing whether t is divisible by $\gcd(s_1, \ldots, s_n)$. No such reduction is possible with the SUBSET-SUM equation.

Simpson: The Euclidean algorithm is a clever trick that has been known since ancient times. But how do I know that another clever trick cannot be found to reduce the SUBSET-SUM equation to something simpler? Like for instance, if I take the greatest common denominator of any subset of $\{s_1, \ldots, s_n\}$ and it does not divide t, then I can automatically rule out many solutions to SUBSET-SUM, all at once.

Spock: But such a clever trick does not always work; what if the gcd *does* divide t? The P versus NP problem is a problem about the *worst-case* running-time of an algorithm, not whether there are clever tricks that can be used to speed up the running-time of an algorithm in some instances. Feinstein's proof only considers the *worst-case* running-time of algorithms which solve SUBSET-SUM.

Wapner: Also, it is simple high school algebra that it is impossible to make the SUBSET-SUM equation simpler than it is: Whenever one decreases the number of possible expressions on one side of the equation, the number of possible expressions on the other side increases. Mathematicians can be clever, but they cannot be clever enough to get around this fact.

Simpson: OK, but what about the fact that Feinstein never mentions in his proof the model of computation that he is considering? To be an valid proof, this has to be mentioned.

Spock: Feinstein's proof is valid in any model of computation that is realistic enough so that the computer cannot solve an equation with an exponential number of possible expressions in polynomial-time, unless it is possible to reduce the equation to something simpler.

Simpson: Or what about the fact that Feinstein never mentions in his paper the important results that one cannot prove that $\mathsf{P} \neq \mathsf{NP}$ through an argument that relativizes [1] or through a natural proof [12]?

Spock: Feinstein's proof does not relativize, because it implicitly assumes that the algorithms that it considers do not have access to oracles, and Feinstein's proof is not a natural proof, since it never even deals with the circuit complexity of boolean functions.

Simpson: What about the 2010 breakthrough by Howgrave-Graham and Joux [9] which gives a probabilistic algorithm that solves SUBSET-SUM in $o(\sqrt{2^n})$ time? I realize that the P versus NP problem is not about probabilistic algorithms, but what if their algorithm can be derandomized and solved in $o(\sqrt{2^n})$ time?

Spock: The algorithm by Howgrave-Graham and Joux does not in fact solve SUBSET-SUM, because it cannot determine for certain when there is no solution to a given instance of SUBSET-SUM; it can only output a solution to SUBSET-SUM in $o(\sqrt{2^n})$ time with high probability when a solution

exists. Furthermore, even if their algorithm can be derandomized, this does not guarantee that it will run in $o(\sqrt{2^n})$ time. And Feinstein has already proven that such a deterministic and exact algorithm is impossible.

Wapner: Are there any more objections to Feinstein's argument?

Simpson: I have no more specific objections. But the fact that the P versus NP problem has been universally acknowledged as a problem that is very difficult to solve and Feinstein's "proof" is so short and simple makes it almost certain that it is flawed. The fact that I could not give satisfactory responses to Spock's arguments does not mean that Feinstein is correct; Feinstein's proof has been out on the internet for a few years now, and still the math and computer science community as a whole does not accept it as valid. Hence, I believe that they are right and that Feinstein is wrong.

Wapner: Professor Simpson, isn't your reason for not believing Feinstein's proof the same reason Feinstein suggested for why most people do not believe his proof? Because most people believe in their hearts that P=NP, that problems which are difficult to solve and easy to state, in this case the P versus NP problem, cannot have short and simple solutions?

Spock: Indeed it is.

Wapner: And yes indeed, I am convinced that Feinstein's proof is valid and that $\mathsf{P} \neq \mathsf{NP}$.

Submitted on May 22, 2016 / Accepted on May 24, 2016

References

1. Baker T.P., Gill J., and Solovay R. Relativizations of the P ≠ NP Question. *SIAM Journal on Computing*, 1975, v. 4(4), 431–442.
2. Bovet P.B. and Crescenzi P. Introduction to the Theory of Complexity. Prentice Hall, 1994.
3. Cormen T.H., Leiserson C.E., and Rivest R.L. Introduction to Algorithms. McGraw-Hill, 1990.
4. Galilei G. Dialogue Concerning the Two Chief World Systems. University of California Press, 1953.
5. Feinstein C.A. Complexity science for simpletons. *Progress in Physics*, 2006, v. 2, issue 3, 35–42.
6. Feinstein C.A. An elegant argument that $P \neq NP$. *Progress in Physics*, 2011, v. 7, issue 2, 30–31.
7. Feinstein C.A. The computational complexity of the traveling Salesman Problem. *Global Journal of Computer Science and Technology*, 2011, v. 11, issue 23, 1–2.
8. Feinstein C.A. A more elegant argument that $P \neq NP$. *Progress in Physics*, 2012, v. 8, issue 1, L10.
9. Howgrave-Graham N., and Joux A. New generic algorithms for hard knapsacks. *IACR Cryptology ePrint Archive*, 2010, 189.
10. Horowitz E. and Sahni S. Computing partitions with applications to the Knapsack Problem. *Journal of the ACM*, 1974, v. 2l, no. 2, 277–292.
11. Menezes A., van Oorschot P., and Vanstone S. Handbook of Applied Cryptography. CRC Press, 1996.
12. Razborov A. and Rudich S. Natural proofs. *Journal of Computer and System Sciences*, 1997, v. 55, issue 1, 24–35.
13. Steinmeyer J. Hiding the Elephant: How Magicians Invented the Impossible and Learned to Disappear. Da Capo Press, 2004.
14. Angle Trisection. From MathWorld — A Wolfram Web Resource. http://mathworld.wolfram.com/AngleTrisection.html
15. Woeginger G.J. Exact algorithms for NP-hard problems. In: *Lecture Notes in Computer Science*, Springer-Verlag Heidelberg, 2003, v. 2570, 185–207.

Criteria for Aerial Locomotion in Exoplanetary Atmospheres: Revisiting the Habitable Zone for Flying Lifeforms

Robin James Spivey

Bangor University, Deiniol Road, Bangor, Gwynedd, Great Britain
E-mail: y.gofod@gmail.com

Liquid water is widely regarded as a hallmark of planetary habitability but, whilst its presence may be a prerequisite for life, aerial locomotion imposes additional constraints on the somewhat over-simplistic concept of a circumstellar habitable zone. Could animals of comparable physiology to birds be envisaged sustaining flight without environmental assistance on super-Earth planets of terrestrial density? A quantitative evaluation of flight athleticism in avian species provides the basis for extrapolation here. At constant atmospheric fraction, assuming a plentiful supply of combustible gas, the "aerial locomotion zone" would be restricted to planets $\lesssim$6.86 $M_\oplus$. However, due to the inevitable thermal impediments at higher altitudes, it is conceivable that the majority of the Earth's avian species could evolve sufficient athleticism for flight on temperate isoatmospheric planets of up to 15 $M_\oplus$, without adjustments in body mass.

1 Introduction

Birds, bats, insects and pterosaurs have independently surmounted the challenges of actively-powered flight [1], perhaps during hyperoxic episodes in the Earth's history [2]. Avian species span some four orders of magnitude in body mass [3] yet birds of all sizes undertake arduous seasonal migrations [4,5]. Flight is a complex and intrinsically dangerous activity especially in arboreal environments, over mountainous terrain, in regions where birds of prey are prevalent or during unfavourable weather [6]. Thus, there is a need for sophisticated neural control [7]. Exoplanet discoveries continue apace [8,9] and NASA's Kepler mission has already established that those of 1–2 Earth radii ("super-Earths") are remarkably abundant [10]. Neutrally buoyant aquatic animals are immune to changes in gravity and land animals can evolve sturdier bones or additional legs to cope with conditions on more massive planets. However, the feasibility of environmentally unassisted flight in stronger gravitational fields is clearly an intricate issue meriting more detailed scrutiny.

The Earth's oxygenated air provides birds not only with a breathable atmosphere but also a medium for generating propulsion and weight support during flight [11]. Consequently, gravity, atmospheric density and the chemical composition of an atmosphere influence the prospects for aerial locomotion. There is no evidence that the laws of physics vary either with time or location, so animals that are as anatomically and physiologically well-adapted to flight as any living here could have evolved elsewhere in the universe. This analysis therefore commences by evaluating the athleticism of Earth's avian species during environmentally unassisted horizontal flight. The limits of flight athleticism on Earth are then used as a basis for extrapolation to different planetary environments, leading to criteria that are likely to be satisfied if circumplanetary atmospheres are compatible with flight.

2 Flight power and athleticism

The following analysis concerns flying animals capable of supporting their own weight in still air conditions, building upon an established result from aerodynamic theory pertaining to hovering flight [12]. If a bird's wings have combined area A_{wing} and the air they sweep is on average accelerated to a downward velocity v_a then the volume of air being swept in unit time is $v_a A_{wing}$. In an atmosphere of density ρ, the mass of this parcel of air is $m_a = \rho v_a A_{wing}$ and so the rate of change of momentum in the air is $m_a v_a = \rho v_a^2 A_{wing}$. For a bird of body mass m_b, Newton's second law requires that this equals the bird's weight $m_b g$ which allows the downward velocity of the air to be obtained as $v_a = \sqrt{m_b g/\rho A_{wing}}$. The power required during hovering is the rate at which kinetic energy is imparted to the air

$$P_{hov} = \frac{m_a v_a^2}{2} = \frac{\rho A_{wing} v_a^3}{2} = \frac{1}{2}\sqrt{\frac{m_b^3 g^3}{\rho A_{wing}}}. \qquad (1)$$

Providing only a small fraction of the power relating to forward horizontal flight, P_f, is required to overcome the drag associated with forward motion, it can be argued that P_f and P_{hov} should scale almost identically. If, furthermore, avian anatomy scales isometrically then $A_{wing} \propto m_b^{2/3}$ and

$$P_f \propto \left(\frac{m_b^7 g^9}{\rho^3}\right)^{1/6}. \qquad (2)$$

For an individual animal this simplifies to $P_f \propto \sqrt{g^3/\rho}$, a term which concisely encapsulates environmental conditions. Thus, flight becomes more challenging on planets with stronger gravitational fields and reduced atmospheric densities [13]. On Earth, flying birds and flightless birds are delineated by the boundary $\sqrt{g^3/\rho}$ =27.7 $m^3\,s^{-3}\,kg^{-1/2}$. Departures from isometry are likely [14] and the allometrically

neutral relationship $A_{wing} \propto m_b^{2/3}$ is only marginally compatible with empirical data – actual wing measurements suggest $A_{wing} \propto m_b^{0.780\pm0.112}$ [15] . This implies the following modification involving an exponent $\alpha = 0.110 \pm 0.056$,

$$P_f \propto m_b^{1+\alpha} \sqrt{\frac{g^3}{\rho}}. \tag{3}$$

Since $\alpha > 0$, mass-specific flight power, P_f/m_b, generally increases with body mass [12]. A quantity χ is now introduced which is directly proportional to the mass-specific flight power needed to fly horizontally in still air. It is adopted as a proxy for flight athleticism and defined as

$$\chi = \left(\frac{m_b}{\tilde{m}}\right)^{\alpha} \sqrt{\frac{g^3}{\rho}} \tag{4}$$

where $\tilde{m}$ is a fiducial mass term used for normalisation and can be arbitrarily chosen. In particular, heavier animals capable of flight in hypodense air would score well on this measure. At a similar airspeed, aerodynamic drag is of less concern to large birds, in keeping with the earlier assumption that $P_f \propto P_{hov}$.

3 Aeronautical limits

Avian lungs utilise a cross-current airflow assisted by a complementary vasculature allowing for efficient gas exchange [16], advantageous during high altitude flight where the partial pressure of oxygen is reduced. At least three species appear to be capable of entirely self-powered flight 7000 m or more above sea level. An iconic example is the bar-headed goose, *Anser indicus*, whose seasonal migrations involve navigating the Himalaya [17] and the prominent obstacle of the Tibetan plateau. Having been satellite-tracked at 7290 m [18] they are more tolerant of hypoxia than brent geese, *Branta bernicla*, which have difficulty crossing the Greenland icecaps at altitudes of 2500 m [19]. Despite a body mass placing them in the 98$^{\text{th}}$ percentile of bird species [3], they have also been observed flying in formation at almost 8000 m by mountaineers climbing the Annapurna massif [20]. A number of cardiovascular, pulmonary, morphological and biochemical adaptation mechanisms could be responsible for this striking athleticism including high ventilation rates [21], relative immunity to respiratory alkalosis and haemoglobin of superior O_2 affinity, higher cardiac output [17] and tissue enhancements such as cardiac hypertrophy, greater capillary density and mitochondrial abundance [22].

Alpine choughs, *Pyrrhocorax graculus* also inhabit the Himalaya. Nesting as high as 6500 m [17], they have been known to follow climbers on Everest at altitudes approaching 8200 m – within the mountaineering "death zone". Small birds such as choughs readily take to the air but swans are much larger and typically require 15–20 wingbeats to become aloft when taking off from water, even though they can obtain some acceleration and weight support from webbed feet. On becoming airborne they continue to gain speed and gradually start to ascend, necessitating continued effort [23]. Thus, unlike smaller birds for which a short period of anaerobic exertion is adequate for take-off, swans must demonstrate aerobic athleticism at the commencement of each flight. This applies also to juveniles – cygnets only start to fly at 4–5 months of age. The athleticism demanded by take-off may confer upon swans an ability to sustain high altitude flight, even if they are not ecologically coerced to do so. Lowland species may be incapable of take-off in hypodense air but that does not preclude, *per se*, an ability to fly high – even though swans tend not to during migration [24]. In still air conditions, flying low in dense air facilitates flight – in accordance with (4). However, strong tailwinds capable of drastically curtailing migration times and total energy expenditure are sometimes available, especially at higher altitudes. During lengthier migratory flights, the additional costs of ascent and high altitude cruising can easily be fully recovered. In the cold and featureless seascape of the north Atlantic, which is neither conducive to the generation of strong thermals nor orographic updrafts, a flock of some 30 whooper swans, *Cygnus cygnus*, was detected in 1967 by radar then visually identified by a pilot to be flying at 8200 m with a ground-speed of 38 m s^{-1} towards the end of a ∼1000 km migration from Iceland to the UK [25].

The air density at 8200 m is 0.513 kg m^{-3}. Setting $\tilde{m} = m_w = 11$ kg, the mass of a whooper swan, and making allowances for variations in α, the maximum value of χ at which flight is possible at this altitude is 42.8 m^3 s^{-3} kg$^{-1/2}$ for whooper swans, 34.1–39.8 m^3 s^{-3} kg$^{-1/2}$ for bar-headed geese and 22.4–34.6 m^3 s^{-3} kg$^{-1/2}$ for Alpine choughs. Results for various species are presented in Fig. 1. Whooper swans appear to top the list for avian athleticism making them well-suited for astrobiological extrapolations. To compete, bar-headed geese, would need to be capable of flight at altitudes of 9.4–11.7 km, which seems unrealistically high [18].

4 Planetary environments

The radii, R, of terrestrial super-Earths are expected to scale with M, the planet mass, as $R \propto M^{0.274}$ [26]. Hence, surface gravity, g_s, should scale as $g_s \propto M^{0.452}$. The effective increase in sea level on a super-Earth planet with a similar water content to Earth can be estimated from the relationship plotted in Fig. 2. It is also relevant to mention that enhanced gravity tends to attenuate topographical features such as mountains and ridges. Super-Earth planets are variously taken to have a mass of 1–10 $M_\oplus$ or a radius of 1–2 $R_\oplus$ (1–12 $M_\oplus$) where the subscript $\oplus$ denotes the Earth. This analysis considers the slightly expanded range 1–15 $M_\oplus$ in order to encompass the largest planets capable of possessing hexagonally close-packed iron at their core [27, 28]. On Earth, a whooper swan can fly in air of density as low as $\rho_w = 0.513$ kg m^{-3}. Since athleticism is not an environmental variable, the min-

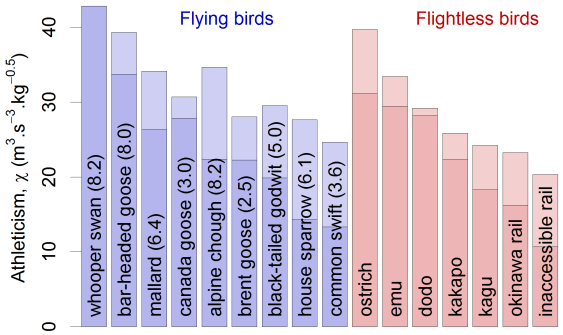

Fig. 1: Flight athleticism, χ, for various species. Estimated maximum altitudes are given in km (in parentheses) for unassisted flight. Selected results are also provided for flightless birds assuming (mostly with undue optimism) that they might be capable of flying in air slightly denser than that of sea level. Lightly shaded areas represent the uncertainty in the allometric scaling exponent, $0.054 < \alpha < 0.166$, using a fiducial mass $\tilde{m}$=11 kg.

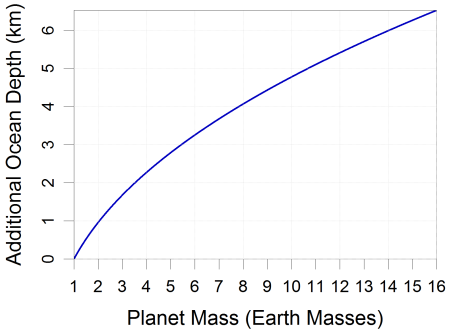

Fig. 2: Planets more massive than Earth but with an identical water fraction ($V_{H_2O} \propto M$) would have somewhat deeper oceans, the additional depth (in km) being at least 2.6 $[(M/M_\oplus)^{0.452} - 1]$ depending on topography. However, if planetary water is exclusively delivered from space via comets and asteroids whose spatial distribution varies little with galactic location, one would anticipate ocean depths to be largely independent of planet mass.

imum air density required by whooper swans on other planets is $\rho_{min} = \rho_w (M/M_\oplus)^{1.356}$. For a 15 Earth-mass planet, ρ_{min}=20.18 kg m^{-3}.

As with discussions of circumstellar habitable zones, some simplifying assumptions are helpful. Due to uncertainties such as cloud cover, humidity levels and fluctuations in atmospheric heating due to planetary rotation, an isothermal model atmosphere is adopted. In hydrostatic equilibrium the ideal gas law predicts that air density, ρ, is proportional to air pressure, $p = \beta\rho$. According to the International Standard Atmosphere, β has a value of about 82714 m^2 s^{-2} for a temperature of 15°C at sea level where ρ =1.225 kg m^{-3}. Gravity is taken to be insensitive to changes in altitude, z. By considering the weight of a thin horizontal layer of air,

$$\frac{dp}{dz} = \frac{dp}{d\rho} \times \frac{d\rho}{dz} = \beta \frac{d\rho}{dz} = -\rho g_s. \tag{5}$$

The air density at height h is obtained by integrating from $z = 0$ to $z = h$,

$$\beta (\ln \rho - \ln \rho_s) = -g_s h, \tag{6}$$

where ρ_s represents the air density at the surface. Thus, $\rho = \rho_s \exp(-g_s h/\beta)$ and the total mass contained by the atmosphere below height h is

$$M_h = 4\pi R^2 \int_0^h \rho(z)\, dz = \frac{4\pi\beta\rho_s R^2}{g_s} \left[1 - \exp\left(\frac{-g_s h}{\beta}\right)\right]. \tag{7}$$

M_h converges as $h \to \infty$ to yield the total mass of the entire atmosphere,

$$M_{atm} = \frac{4\pi\beta\rho_s R^2}{g_s}. \tag{8}$$

For an isothermal atmosphere, under the assumption of spherical symmetry, half the air mass lies below a scale height $\hat{h}$ given by

$$\hat{h} = \frac{\beta \ln 2}{g_s} = \frac{\beta \ln 2}{g_\oplus} \left(\frac{M_\oplus}{M}\right)^{0.452}. \tag{9}$$

This expression is entirely independent of ρ_s. Plots of surface gravity, planetary radius and atmospheric scale height against planetary mass are provided in Fig. 3.

5 Criteria for aerial locomotion

From (8) we have $\rho_s = g_s M_{atm}/4\pi\beta R^2$. Recalling that $g_s \propto M^{0.452}$ and $R \propto M^{0.274}$,

$$\rho_s = \frac{g_\oplus M_{atm} (M/M_\oplus)^{0.452}}{4\pi\beta R_\oplus^2 (M/M_\oplus)^{0.548}} \propto M^{0.904} \left(\frac{M_{atm}}{M}\right). \tag{10}$$

Since $g_s = g_\oplus (M/M_\oplus)^{0.452}$, the quantity $\sqrt{g_s^3/\rho_s}$, a factor previously found to be proportional to the power required by flight, can be expressed as follows

$$\begin{aligned} \frac{g_s^3}{\rho_s} &= \frac{4\pi\beta R_\oplus^2 (M/M_\oplus)^{0.096}}{g_\oplus M_{atm}} \left[g_\oplus (M/M_\oplus)^{0.452}\right]^3 \\ &= \frac{4\pi\beta R_\oplus^2 g_\oplus^2}{M_{atm}} \left(\frac{M}{M_\oplus}\right)^{1.452}, \end{aligned} \tag{11}$$

$$\sqrt{\frac{g_s^3}{\rho_s}} = \gamma \left(\frac{M}{M_\oplus}\right)^{0.226} \sqrt{\frac{M}{M_{atm}}} \tag{12}$$

where $\gamma = 2 g_\oplus R_\oplus \sqrt{\pi\beta/M_\oplus} = 0.026$ m^3 s^{-3} kg$^{-1/2}$. The maximum mass of an isoatmospheric planet (i.e. having a ratio M_{atm}/M identical to Earth's) that is compatible with flight for

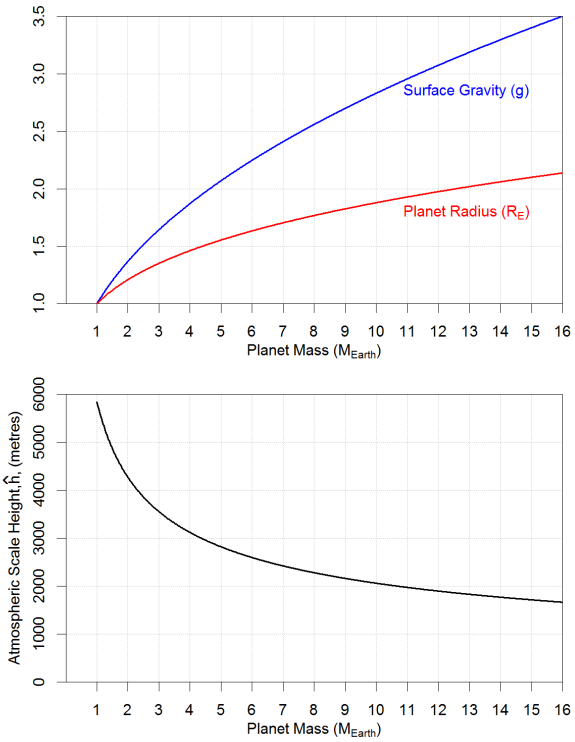

Fig. 3: Upper panel: planetary radius and surface gravity obey simple power law relationships according to planetary mass, $R \propto M^{0.274}$ and $g_s \propto M^{0.452}$ respectively. Lower panel: the scale height of the atmosphere, $\hat{h} = \beta \ln 2/g_s$, is independent of the surface air density and hence total mass of the atmosphere. It decreases for larger planets since a higher surface gravity is better able to confine the atmosphere close to the surface.

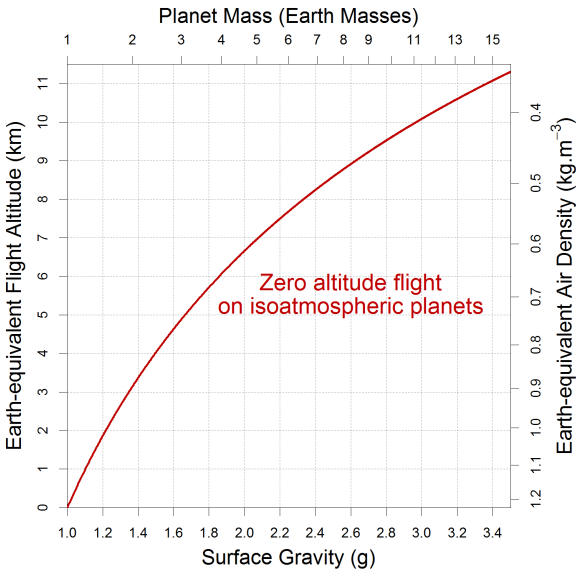

Fig. 4: For isoatmospheric planets the Earth-equivalent air density, ρ_{eq}, at the athleticism of zero-altitude flight, is inversely related to the surface gravity of a planet, $\rho_{eq} \propto 1/g_s$.

a whooper swan can be obtained by requiring that $\rho_s = \rho_{min}$. This implies that $\rho_w (M/M_\oplus)^{1.356} = \rho_\oplus (M/M_\oplus)^{0.904}$, and

$$M = M_\oplus \times \left(\frac{\rho_\oplus}{\rho_w}\right)^{1/0.452} \approx 6.86 M_\oplus. \qquad (13)$$

The surface gravity of this planet of 6.86 $M_\oplus$ would be $2.388g$. The maximum range and minimum power airspeeds of flying birds are expected to vary as $\rho^{-0.5}$ [12]. The surface air density of an isoatmospheric 6.86 Earth-mass planet would be $\sim 5.68\rho_\oplus$ so a typical airspeed of 21 m s^{-1} for a swan [24] might decline to 8.8 m s^{-1}, roughly the pace of an elite 400 m runner. In this same $2.385g$ environment, however, most people would struggle to walk at all and horses would be incapable of standing.

Since $P_f \propto m_b^{1+\alpha} \sqrt{g^3/\rho}$ and $\rho_s \propto M^{0.904}$, the flight power at zero altitude on isoatmospheric planets scales as $P_f \propto M^{0.226}$. Because $M/M_\oplus = (g_s/g_\oplus)^{1/0.452}$, it is apparent from (13) that $g_s/g_\oplus = \rho_\oplus/\rho_s$ for the limiting planet mass. Therefore, a particularly simple inverse relationship exists, $\rho_{eq} \propto 1/g_s$, allowing translation of the surface gravity of an isoatmospheric planet to the Earth-equivalent air density (and hence also equivalent maximum flight altitude via the International Standard Atmosphere). Results are presented in Fig. 4.

Might smaller birds be capable of flight on an isoatmospheric planet of 15 Earth masses? The surface air density would be $1.225\,(M/M_\oplus)^{0.904} = 14.17$ kg m^{-3}, lower than the minimum air density required by whooper swans for the same planet mass, $\rho_{min} = 20.18$ kg m^{-3}. Since $\chi \propto \rho_s^{-1/2}$, flight athleticism would have to be boosted by a factor of 1.1934. To achieve this, body mass could be reduced so that $m_b < m_w$ and flight would become feasible on a 15 Earth-mass planet if $m_b = m_w \times 1.1934^{-1/\alpha}$. Hence, flying animals of 0.42–3.8 kg or less (according to the value of α) may be capable of aerial locomotion on a 15 Earth-mass planet if they can match the flight athleticism of a whooper swan. Some ~88% of species have a body mass below 0.42 kg and ~99% have a body mass below 3.8 kg [3].

For an isoatmospheric 15-Earth mass planet one finds that $\sqrt{g_s^3/\rho_s} > 51.1$ m^3 s^{-3} kg$^{-1/2}$. On Earth this is equivalent to $\rho < 0.36$ kg m^{-3} or flight at altitudes $\gtrsim 11$ km. Even if smaller birds lack the athleticism of whooper swans, some may be able to fly in such rarefied air. The possibility could be investigated using a hypobaric wind tunnel operated at a comfortable flight temperature. Ruby-throated hummingbirds, which have a body mass of only 2–6 grams, can sustain hovering at densities down to 47% that of sea level air (0.576 kg m^{-3}) [29,30]. In forward flight, this species is likely to be capable of flying in yet more rarefied air. However,

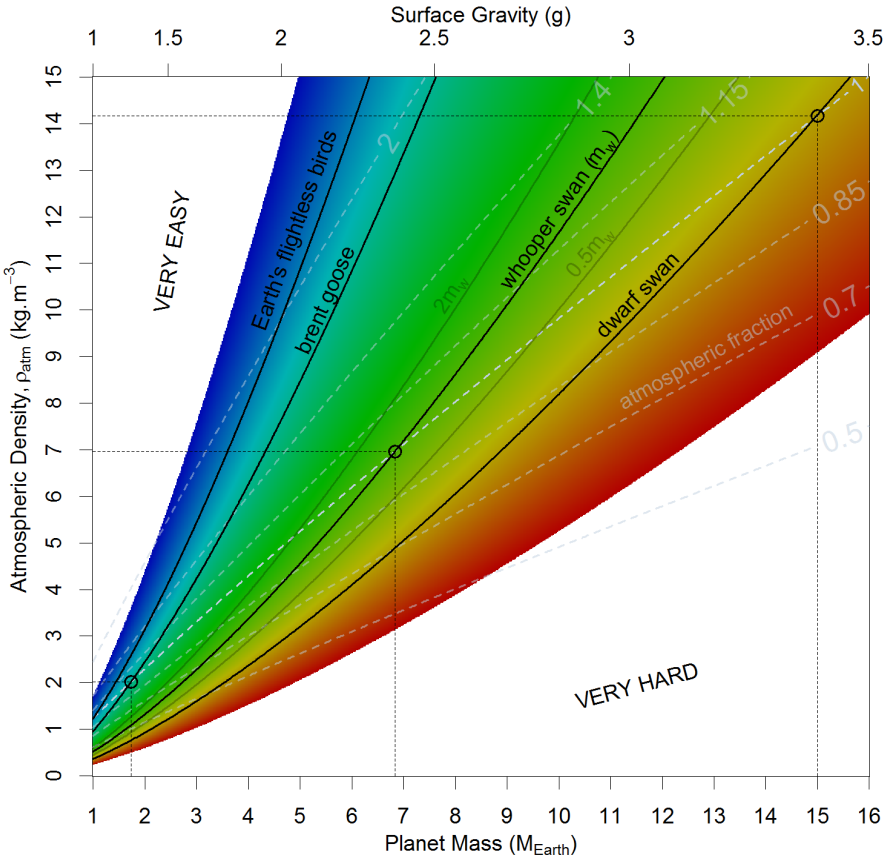

Fig. 5: Flight power (effort increases from blue to red) is a function of planet mass (or surface gravity) and atmospheric density. Conditions compatible with aerial locomotion lie upward of the solid contours. An 11 kg whooper swan appears capable of unassisted horizontal flight on isoatmospheric planets up to 6.86 $M_\oplus$. The influence of doubling or halving body mass relative to the whooper swan is shown for α=0.11. The trace marked *dwarf swan* corresponds to a hypothetical flying animal of the same flight athleticism as a whooper swan but of body mass 0.04–0.34m_w (corresponding to $0.054 < \alpha < 0.166$). Dashed contours represent atmospheric mass content relative to the Earth's fraction (862 parts per billion).

even then, due to its relatively small body mass, it is unlikely to challenge whooper swans for flight athleticism. The same argument applies also to flying insects.

Sufficient information has now been collected to describe circumstances compatible with environmentally unassisted circumplanetary flight in which buoyancy effects can be safely ignored. A planet would ideally occupy an orbit within the conventional circumstellar habitable zone [31] and, based upon the flight athleticism of whooper swans, the following criterion should also be satisfied:

$$\left(\frac{m_b}{m_w}\right)^{\alpha} \sqrt{\frac{g_s^3}{\rho_s}} \lesssim 42.8 \ \ \mathrm{m^3\,s^{-3}\,kg^{-1/2}}. \qquad (14)$$

By virtue of (12), an equivalent formulation involving only normalised mass terms is possible

$$\left(\frac{m_b}{m_w}\right)^{\alpha} \left(\frac{M}{M_\oplus}\right)^{0.226} \sqrt{\frac{M}{M_{atm}}} \lesssim 1646. \qquad (15)$$

Limitations in respiration or gas perfusion could potentially impinge upon the present analysis but oxygen delivery is not constrained in birds by the pulmonary system [23] and, in more inert atmospheres, flow-through breathing arrangements requiring little or no biomechanical effort can be imagined. Changes in atmospheric composition are likely over geological timescales [2]. Thus, it would ideally be useful to know whether an exoplanetary atmosphere has remained breathable and non-toxic for sufficient time to support the evolution of complex organisms.

Another factor which might well impact on these results is a change in atmospheric temperature, T_{atm}. The molar mass of the air, $M_{air} = 0.029\ \mathrm{kg\,mol^{-1}}$, the air temperature, T_{air}, and the universal gas constant, $R_{air} = 8,314\ \mathrm{N\,m\,mol^{-1}\,K^{-1}}$, obey the relationship $\beta = R_{air}T_{air}/M_{air}$. Since both γ and M_{atm} are linearly dependent on β, the value of $\sqrt{g_s^3/\rho_s}$ is proportional to $\sqrt{\beta}$. Since the value of β adopted here corresponds to an air temperature of 15°C, different atmospheric

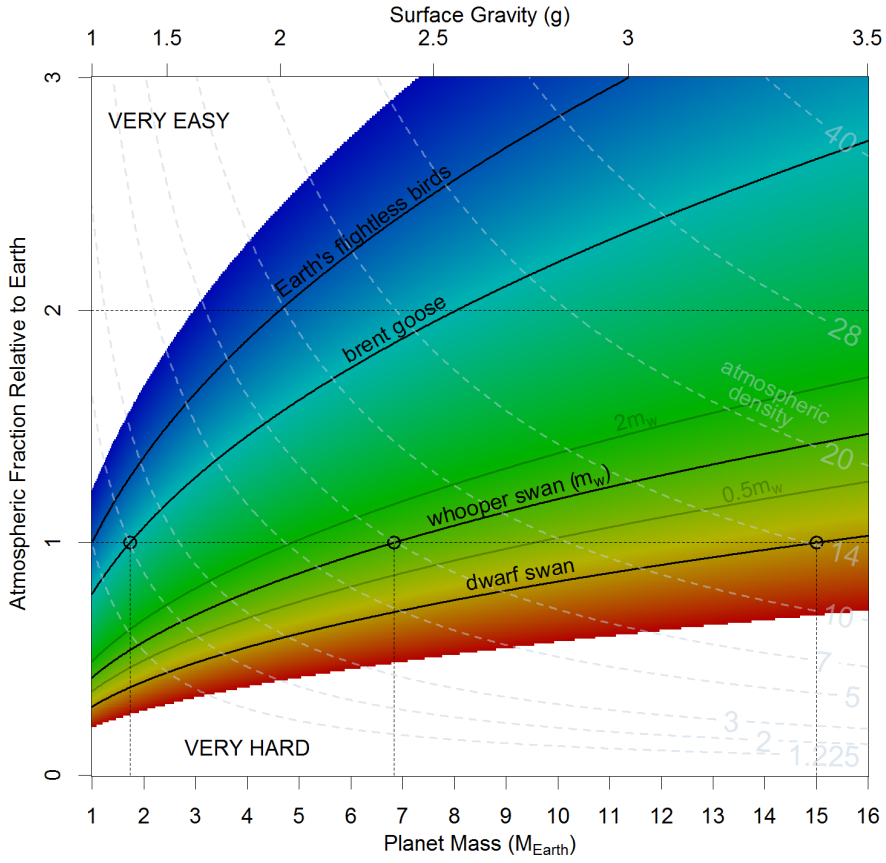

Fig. 6: Flight power according to planet mass and the atmospheric fraction relative to that of the Earth. Aerial locomotion is possible upwards of the solid contours. Dashed contours here represent the surface density of the atmosphere (kg m^{-3}), and correspond to isobars for the isothermal atmospheric model used here.

temperatures can be accommodated by applying a correction factor of $\sqrt{288.15/T_{atm}}$ to the right-hand sides of the inequalities (14) and (15).

The results of this analysis are presented graphically in Figs. 5 and 6. These limits are likely to be somewhat cautious since it is possible that, with determined effort, whooper swans may be capable of flying higher than the flock sighting at 8200 m. Although it has been conjectured that their initial ascent was aided by lee waves, such assistance would not have been present during the sea crossing from Iceland to Scotland [24]. Furthermore, this species regularly takes off in the dense air present at sea level which prohibits the evolution of larger wings that would tend to facilitate flight at extreme altitudes. Flying animals of extraterrestrial origin may not have been subjected to evolutionary pressures of this kind, particularly if their planets lack elevated land masses obstructing low altitude flight.

6 Discussion

Expressions (14) and (15) present criteria for aerial locomotion to be realistically possible in circumplanetary atmospheres. Comparisons of relative flight power under different environmental circumstances can utilise the expression $P_f \propto \sqrt{g_s^3/\rho_s}$. This predicts, for example, that flight in conditions resembling Saturn's moon Titan would be $\sim$ 23 times easier than at sea level on Earth. The wing-scaling exponent α has a small but positive value [15]. If this holds for a wide range of body masses then one can envisage animals flying in such conditions which are larger than any that have ever graced this planet. However, transport costs (or the energy/distance ratio), should approximately scale as $m_b^{0.7}$ during flight but only $m_b^{0.6}$ for running [1]. Above a certain body size, therefore, terrestrial locomotion would be energetically favoured to flight, though transit times might increase.

A primary finding is that, in the presence of a breathable atmosphere, winged animals of a body mass resembling the majority of the Earth's indigenous avian species could potentially evolve the ability to fly on isoatmospheric planets of at least 15 $M_\oplus$ ($g_s = 3.4g$). However, this work also highlights how even mildly reduced atmospheric fractions might potentially prohibit aerial locomotion. Novel techniques capable of remotely determining atmospheric composition, sur-

face atmospheric density and oceanic coverage could therefore be useful in augmenting future exoplanetary searches. Even worlds entirely covered in water could host flying animals. If in time the Earth were to become an ocean planet through continuing bombardment by comets and meteorites then seabirds could emulate penguins by mating, laying eggs and incubating them on floating icebergs.

That birds possess superb navigation skills has long been apparent but only recently have we appreciated that numerous species are adept problem-solvers [32] with an innate ability to fashion tools [33]. Eurasian Magpies (*Pica pica*) have demonstrated self-recognition when confronted with a mirror, a trait commonly associated with self-awareness [34]. Most birds are proficient hunters, potentially capable of stimulating the evolution of higher intelligence in land-based prey – such as our early mammalian ancestors. That cannot be said of insect-like creatures, which should in general cope more comfortably with higher gravitational fields due to the advantages of relatively small body masses and large area to volume ratios, facilitating respiration.

Flapping flight is a highly effective mode of locomotion for animals possessing sufficient athleticism. However, as aerial manoeuvres demand considerable coordination and spatiotemporal awareness, and body weight is critical, evolutionary pressures arise for efficient neurochemistry and neuroarchitecture. Volant organisms may well have played a pivotal role in shaping the Earth's natural history, enriching its biodiversity and accelerating the evolution of intelligent life. Avian species demonstrated considerable resilience in surviving the ecological catastrophe responsible for the extinction of most dinosaurs. In times of adversity, an ability to swiftly and efficiently relocate over planetary distances and flexibly forage on both land and sea may assist the propagation of flying animals over geological∼stellar timescales. Accurate determination of whether circumplanetary flight is possible should not be overlooked if future missions to extrasolar worlds are intent on maximising the chances of encountering complex lifeforms and, perhaps, even extraterrestrial civilisations of comparable sophistication to our own.

Submitted on May 13, 2016 / Accepted on May 17, 2016

References

1. Schmidt-Nielsen K. *Science*, 1972, v. 177 (4045), 222–228.
2. Dudley R. *Ann. Rev. Physiol.*, 2000, v. 62 (1), 135–155.
3. Codron D., Carbone C. & Clauss M. *PLoS One*, 2013, v. 8 (10), e77110.
4. Egevang C., Stenhouse I. J., Phillips R. A., Petersen A., Fox J. W. & Silk J. R. *Proceedings of the National Academy of Sciences*, 2010, v. 107 (5), 2078–2081.
5. Spivey R. J., Bishop C. M., Butler P. J. *et al. Science*, 2015, v. 347 (6219), 250–254.
6. Gatesy S. M. & Dial K. P. *Evolution*, 1996, v. 50 (1), 331–340.
7. Pearson R. G. The Avian Brain. Academic Press, 1972, ISBN 9780125480505.
8. Marcy G. W., Isaacson H., Howard A. W., Rowe J. F., Jenkins J. M., Bryson S. T., Christiansen J. *et al. Astrophys. J. Suppl. Ser.*, 2014, v. 210 (2), 20.
9. Cassan A., Kubas D., Beaulieu J. P., Dominik M., Horne K., Greenhill J., Pietrzyski G. *et al. Nature*, 2012, v. 481 (7380), 167–169.
10. Petigura E. A., Howard A. W. & Marcy G. W. *P.N.A.S.*, 2013, v. 110 (48), 19273–19278.
11. Tobalske B. W., Hedrick T. L., Dial K. P. & Biewener A. A. *Nature*, 2003, v. 421 (6921), 363–366.
12. Pennycuick C. J. Modelling the Flying Bird (Vol. 5). Elsevier, 2008, ISBN 9780123742995.
13. Altshuler D. L. & Dudley R. *Int. & Comp. Biol.*, 2006, v. 46 (1), 62–71.
14. Rayner J. *Mathematical Methods Applied Science*, 2001, v. 24 (17–18), 1485–1514.
15. Hedenstrom A. *J. Comp. Physiol. A*, 2008, v. 194 (7), 685–691.
16. Scott G. R. *J. Exp. Biol.*, 2011, v. 214 (15), 2455–2462.
17. Black C. R., Tenney S. M., *Respir. Physiol.*, 1980, v. 39, 217–239.
18. Hawkes L. A., Balachandran S., Batbayar et al. *Proc. Roy. Soc. B*, 2012, v. 280 (1750).
19. Gudmundsson G. A., Benvenuti S., Alerstam T., Papi F. & Lilliendahl K. *Proc. Roy. Soc. B*, 1995, v. 261 (1360), 73–79.
20. Blum A. Annapurna: A Woman's Place. Sierra Club Books, 1980, ISBN 9781578050222.
21. Tucker V. A. *Journal of Experimental Biology*, 1968, v. 48 (1), 55–66.
22. Fedde M. R., Faraci F. M., Kilgore Jr D. L., Cardinet III G. H. & Chatterjee A. Circulation, Respiration, and Metabolism. Springer Berlin Heidelberg, 1985, 149–163.
23. Tobalske B. W. & Dial K. P. *J. Exp. Biol.*, 2000, v. 203 (21), 3319–3332.
24. Pennycuick C. J., Einarsson O., Bradbury T. A. M. & Owen M. *J. Avian Biol.*, 1996, v. 27 (2), 118–134.
25. Liechti F. & Schaller E. *Naturwissenschaften*, 1999, v. 86 (11), 549–551.
26. Sotin C., Grasset O. & Mocquet A. *Icarus*, 2007, v. 191 (1), 337–351.
27. Spivey R. J. *J. Modern Physics*, 2013, v. 4, 20–47.
28. Spivey R. J. *Physics Essays*, 2015, v. 28 (2), 254–264.
29. Chai P. & Dudley R. *Nature*, 1995, v. 377 (6551), 722–725.
30. Chai P. & Dudley R. *J. Exp. Biol.*, 1995, v. 199 (10), 2285–2295.
31. Kasting J. F., Whitmire D. P. & Reynolds R. T. *Icarus*, 1993, v. 101 (1), 108–128.
32. Emery N. J. *Phil. Trans. Roy. Soc. B*, 2006, v. 361 (1465), 23–43.
33. Hunt G. R. *Nature*, 1996, v. 379 (6562), 249–251.
34. Prior H., Schwarz A. & Gunturkun O. *PLoS Biology*, 2008, v. 6 (8), e202.

The Structure of the Photon in Complex Vector Space

Kundeti Muralidhar

Physics Department, National Defence Academy, Khadakwasla, Pune-411023, India. E-mail: kundetimuralidhar@gmail.com

Considering the complex vector electromagnetic field, the energy of the photon is expressed as an even multivector consisting of a scalar kinetic energy part and a bivector rotational energy part. Since any even multivector can be expressed as a rotor representing internal rotations, the electromagnetic energy even multivector represents internal complex rotations. It has been shown that the spin angular momentum is the generator of rotations in the plane normal to the propagation direction and the orbital angular momentum is the generator of rotations in a plane normal to the spin plane. The internal structure of the photon may be visualized as a superposition of electromagnetic field flow or rotation in two normal orientations in complex vector space. The cause of such complex rotations is attributed to the presence of electromagnetic zeropoint field.

1 Introduction

Even after the photon inception into the field of physics over a century ago, the obscurity in understanding the photon structure persists. The concept of the photon, the energy quanta of electromagnetic radiation, was introduced by Planck in the blackbody radiation formula and Einstein in the explanation of the photoelectric effect. The photon is normally considered as a massless bundle of electromagnetic energy and the photon momentum is defined as the ratio between the energy of the photon and the velocity of light. It is well known from Maxwell's theory that electromagnetic radiation carries both energy and momentum [1]. The linear momentum density is given by the Poynting vector $\mathbf{E} \times \mathbf{B}$ and the angular momentum is the cross product of the Poynting vector with the position vector. Poynting suggested that circularly polarized light must contain angular momentum and showed it as the ratio between the free energy per unit volume and the angular frequency. In 1936, Beth [2] first measured the angular momentum of light from the inference that circularly polarized light should exert torque on a birefringent plate and that the ratio between angular momentum J and linear momentum P was found to be $\lambda/2\pi$, where λ is the wavelength of light. The measured angular momentum agreed in spin magnitude with that predicted by both wave mechanics and quantum mechanics. The Beth angular momentum is in general considered as the photon spin angular momentum.

The energy momentum tensor of the electromagnetic field $T^{\mu\nu}$ is not generally symmetric. By adding a divergence term $\partial_\mu U^{\mu\alpha\nu}$ to $T^{\mu\nu}$, one can construct a symmetric energy momentum tensor $\Theta^{\mu\nu}$ which is normally known as the Belinfante energy momentum tensor [3]. The tensor $U^{\mu\alpha\nu}$ is asymmetric in the last two indices. The symmetric energy momentum tensor satisfies the conservation law $\partial_\mu \Theta^{\mu\nu} = 0$. The advantage of the symmetric energy momentum tensor is that the angular momentum calculated from Θ^{k0} is a conserved quantity. Belinfante established the fact that the spin could be regarded as a circulating flow of energy and this idea was well explained by Ohanian [4]. In an infinite plane wave, the electric and magnetic field vectors are perpendicular to the propagation direction. In a finite transverse extent, the field lines are closed loops and represent circulating energy flow and imply the existence of angular momentum whose orientation is in the plane of circulation and it is the spin angular momentum. Further, as the electromagnetic waves propagate, the energy also flows along the direction of propagation. The translational energy flow implies the existence of additional orbital angular momentum. The magnetic field vector can be expressed as the curl of a vector potential $\mathbf{A}$ and the angular momentum density becomes $\mathbf{E} \times \mathbf{A}$. A close inspection shows that the total angular momentum has two components: one the spin angular momentum associated with the polarization and the other the orbital angular momentum associated with the spatial distribution [1]. The total angular momentum J can be split into a spin angular momentum S and an orbital angular momentum L [5]

$$J = \frac{1}{4\pi} \int \mathbf{E} \times \mathbf{A}\, d^3 r + \frac{1}{4\pi} \int \mathbf{E}^n (\mathbf{r} \times \nabla) \mathbf{A}^n\, d^3 r\,. \qquad (1)$$

The first term on the right is dependent on polarization and hence it is called spin angular momentum S and the second term is independent of polarization and depends on spatial distribution and identified with orbital angular momentum L. It has been argued that the photon angular momentum cannot be separated into a spin part and an orbital part in a gauge invariant way and the paradox was a subject for several papers and in standard textbooks for the past few decades [6].

In recent times the definitions of these angular momenta raised certain controversy. In all these definitions the angular momentum is defined as a vector product containing the position coordinate. The decomposition of total angular momentum of the photon into spin and orbital parts basically involves how we split the vector potential in a gauge invariant way and it has been studied by several authors and a detailed discussion is given in the review article by Leader and Lorce [7]. The absence of any rest frame for the photon suggests that the total angular momentum is observable but not separately

as spin angular momentum and orbital angular momentum. Though this separation is normally considered to be unphysical and not observable, Van Enk and Nienhuis [8] argued that both spin and orbital angular momenta are separately measurable quantities and gauge invariant. The gauge invariant spin and angular momentum parts are expressed as

$$J = \frac{1}{4\pi}\int \mathbf{E}\times\mathbf{A}\, d^3r + \frac{1}{4\pi}\int \mathbf{r}\times\mathbf{E}^n\nabla\mathbf{A}^n\, d^3r\,. \qquad (2)$$

In this expression $\mathbf{A} = \mathbf{A}_\perp$ and therefore both terms are gauge invariant. The canonical expression $\mathbf{E}^n\nabla\mathbf{A}^n$ gives pure mechanical momentum which is responsible for the orbital angular momentum of a photon. The azimuthal flow of electromagnetic field is given by $\mathbf{E}^n\nabla\mathbf{A}^n$ which is half of $\mathbf{E}\times\mathbf{B}$ and the other half is spin flow [9]. In an analogous way, in quantum chromodynamics, the gluon angular momentum can be decomposed into a spin part and an angular momentum part which plays an important role in understanding nucleon structure. Recently, Chen *et al.* [10] decomposed the gauge potential into pure and physical parts: $\mathbf{A} = \mathbf{A}_{pure} + \mathbf{A}_{phys}$, the pure part is related to gauge invariance and the physical part is related to physical degrees of freedom. In the decomposition by Wakamatsu *et al.* [11], the orbital angular momentum is defined similar to a classical expression $\mathbf{r}\times P_{kin}$, where $P_{kin} = -\frac{1}{4\pi}\int \mathbf{A}_{phys}\times\mathbf{E}\, d^3r$ and in this decomposition each term is gauge invariant and observable. Further studies by several authors revealed the fact that there could be infinitely many different ways to perform such decomposition in a gauge invariant way [7, 12]. In Beth's experiment, actually the spin angular momentum was measured. The measurement of orbital angular momentum has been performed in recent times. The amplitude of a Laguerre-Gaussian mode of light wave has an azimuthal angular dependence of $\exp(-il\phi)$, where l is the azimuthal mode index. The ratio between the angular momentum to the energy is $1/\omega$ or $L = l\,(\mathbf{E}\times\mathbf{B})/\omega$ and for Laguerre-Gaussian laser mode, it has been shown that the angular momentum is equal to $l\hbar$ and the total angular momentum of the whole light beam is $(l+\sigma_z)\hbar$, where σ_z is a unit vector along the direction of propagation [13]. The measurement of orbital angular momentum was reported by several authors [14–16].

Another important aspect of the photon is its internal zitterbewegung motion. It is well known from the first observation of Schrödinger [17] that a Dirac electron possesses zitterbewegung motion which is the oscillatory motion of the electron with very high frequency $\omega = 2mc^2/\hbar$ with internal velocity equal to the velocity of light. Such internal motion arises because of the classical electromagnetic fluctuating zeropoint field present throughout space [18]. The spin angular momentum of the electron is identified as the zeropoint angular momentum [18, 19]. On the basis of electron internal oscillations, classical models of electron were developed [20–22]. It is quite interesting that such zitterbewegung motion for the photon was derived from the relativistic Schrödinger like equation of the photon by Kobe [23]. It has been proved that the photon velocity contains parallel and perpendicular components with respect to the direction of propagation. The time dependent perpendicular component of velocity rotates about the direction of propagation with an angular frequency ω equal to the frequency of the electromagnetic wave. The finite special extension of internal rotation is equal to the reduced wavelength. The photon spin is then identified as the internal angular momentum due to zitterbewegung. Considering internal dynamical variables in the configuration space the zitterbewegung is attributed to the normal component of velocity vector oscillations about the particle centre [24]. In the quantum field theory, it has been shown that the zitterbewegung of a photon is attributed to the virtual transition process corresponding to the continuous creation and annihilation of virtual pairs of elementary excitations [25, 26]. Recently, Zhang [27] proposed that zitterbewegung of the photons may appear near the Dirac point in a two dimensional photonic crystal. In the case of an electron, the spin angular momentum is an intrinsic property. In the same way both spin and orbital momenta of the photon are intrinsic in nature [28, 29]. Thus one can anticipate that the photon is also having an internal spin structure described by the internal oscillations or rotations.

One of the most important applications of the photon angular momentum lies in the exploitation of the photon spin and angular momentum states for quantum computation and quantum information processing [30]. Superposition of polarization states can be used to construct qubits and transmit information. A standard approach to visualise the transformation of qubits is provided by the Poincaré sphere representation. Generally, any completely polarized state can be described as a linear superposition of spin states and corresponds to a point on the surface of a unit sphere. Analogous representation of orbital angular momentum states of the photon was introduced by Padgett and Courtial [31] and Agarwal [32]. Quantum entanglement of states is a consequence of quantum non-locality. The entanglement involving the spatial modes of electromagnetic field carrying orbital angular momentum was studied by Mair *et al.* [33] and Franke-Arnold *et al.* [34]. The phase dependence of angular momentum may provide multi-dimensional entangled states which are of considerable interest in the field of quantum information.

In vector algebra, the angular momentum is defined by a cross product of position and momentum vectors and identified as a vector normal to the plane containing the position and momentum vectors. However, the angular momentum is basically a planar quantity and better defined as a bivector in a plane [35]. Note that the cross product cannot be defined in a plane. In the case of the electron, the classical internal bivector spin was obtained from the multivector valued Lagrangian by Barut and Zhangi [20]. It has been shown that the particle executes internal complex rotations by absorbing zeropoint field and the angular momentum of these internal rotations

is identified as the bivector spin of the particle [36]. In an analogous way, the photon spin is a bivector quantity. Therefore, the photon spin may be visualised as a bivector product between an internal finite extension of the photon and an internal momentum. Similarly, one can visualise the orbital angular momentum of the photon as a bivector. Considering the electromagnetic field as a complex vector, it is possible to express the set of Maxwell's equations into a single form. The basics of complex vector algebra have been discussed in detail previously in references [37, 38].

Recently, the nature of the photon was discussed at length by several authors in the book edited by Roychoudhuri *et al.* [39]. The main views of understanding the nature or the structure of the photon are as follows. Einstein viewed the photon as a singular point which is surrounded by electromagnetic fields. In quantum electrodynamics, the photon is introduced as a unit of excitation associated with the quantised mode of the radiation field and it is associated with precise momentum, energy and polarization. In another view, the photon is interpreted as neither a quantum nor a wave but it can be a meson which produces off other hadronic matter and attains physical status. Photons are just fluctuations of random field or wave packets in the form of needles of radiation superimposed in the zeropoint field. However, understanding the photon structure still remains an open question.

The aim of this article is to explore the structure of the photon in complex vector space. To understand the structure of the photon, the electromagnetic field is expressed as a complex vector and the total energy momentum even multivector is developed in section 2. Section 3 deals with the internal angular momentum structure of the photon and conclusions are presented in section 4. Throughout this article, Lorentz–Heaviside units are used, i.e. $\epsilon_0 = \mu_0 = 1$ and energy terms are divided by 4π and conveniently we choose $c = 1$ [1]. However, for clarification sake in some places c is reintroduced.

2 Energy momentum multivector of the electromagnetic field

In the complex vector formalism, we express the electric field as a vector $\mathbf{E}$ and the magnetic field as a bivector $\mathbf{iB}$ and the electromagnetic field F is expressed as a complex vector [37, 38]

$$F = \frac{1}{2}(\mathbf{E} + \mathbf{iB}) . \tag{3}$$

Here, $\mathbf{i}$ is a pseudoscalar in geometric algebra of three dimensions [40], it commutes with all elements of the algebra and $\mathbf{i}^2 = -1$. A reversion operation changes the order of vectors and is indicated by an overbar

$$\bar{F} = \frac{1}{2}(\mathbf{E} - \mathbf{iB}) . \tag{4}$$

Now, the product $\bar{F}F$ is written as

$$\bar{F}F = \frac{1}{4}(E^2 + B^2) + \frac{1}{2}(\mathbf{E} \wedge \mathbf{iB}) . \tag{5}$$

Similarly, we find

$$F\bar{F} = \frac{1}{4}(E^2 + B^2) - \frac{1}{2}(\mathbf{E} \wedge \mathbf{iB}) . \tag{6}$$

The energy density of the electromagnetic field can be obtained from the scalar product

$$\bar{F} \cdot F = \frac{1}{2}(\bar{F}F + F\bar{F}) = \frac{1}{4}(E^2 + B^2) . \tag{7}$$

Further, the product $\bar{F} \wedge F$ gives a vector of the form

$$\mathbf{p} = -\frac{1}{c}\bar{F} \wedge F = -\frac{1}{2c}(\bar{F}F - F\bar{F}) = -\frac{1}{2c}(\mathbf{E} \wedge \mathbf{iB}) , \tag{8}$$

and the dual of $\mathbf{p}$ is expressed as

$$\mathbf{ip} = \frac{1}{2c}(\mathbf{E} \wedge \mathbf{B}) . \tag{9}$$

From the above expression, one can express the energy density of internal electromagnetic flux flow in the bivector plane normal to the propagation direction

$$\mathbf{ip}c = \frac{1}{2}(\mathbf{E} \wedge \mathbf{B}) . \tag{10}$$

This energy density of the photon can be identified as the rotational energy density. However, the energy density obtained in (7) represents the energy density of the photon as it propagates and it may be treated as the kinetic energy density of the photon. An even multivector is a sum of a vector and a bivector. The energy terms in (7) and (10) combine to give the total energy of the photon in even multivector form

$$\mathcal{E} = \frac{1}{4\pi}\int \frac{E^2 + B^2}{4} d^3r + \frac{1}{4\pi}\int \frac{\mathbf{E} \wedge \mathbf{B}}{2} d^3r = \mathcal{E}_{kin} + \mathcal{E}_{rot} . \tag{11}$$

The scalar part shows the flow of energy in the direction of propagation which can be identified as the kinetic part of energy $\mathcal{E}_{kin}$ and the bivector part can be identified as the rotational energy $\mathcal{E}_{rot}$ representing circulation of electromagnetic energy in a plane normal to the direction of propagation. In general, twice the kinetic energy is treated as the electromagnetic energy per unit volume and it is the energy of the photon. Since the energy of a photon is expressed as momentum times its velocity, we define kinetic momentum of a photon as $\mathbf{p}_k = \mathcal{E}_{kin}/\mathbf{v}$, where the velocity $\mathbf{v} = \mathbf{n}c$ and $\mathbf{n}$ is a unit vector along the direction of propagation. Introducing an internal velocity $\mathbf{u}$ satisfying the condition $\mathbf{u} \cdot \mathbf{v} = 0$ and $|\mathbf{u}| = c$, the internal momentum representing the rotational flux flow can be defined as $\mathbf{p}_r = \mathcal{E}_{rot}/\mathbf{u}$. From these definitions generalised photon velocity and momentum complex vectors can be constructed as

$$U = \mathbf{v} + \mathbf{iu} , \tag{12}$$

$$P = \mathbf{p}_k + \mathbf{i}\mathbf{p}_r \,. \tag{13}$$

A reversion operation on P gives $\bar{P} = \mathbf{p}_k - \mathbf{i}\mathbf{p}_r$. Since the magnitudes $|\mathbf{p}_k|$ and $|\mathbf{p}_r|$ are equal, we have $P^2 = \bar{P}^2 = 0$. Therefore, the complex vector P is a complex null vector which represents the lightlike nature of the photon. Similarly, the complex velocity vector is also a complex null vector. Now, the total energy of the photon is expressed as

$$\mathcal{E} = \mathbf{p}_k \cdot \mathbf{v} + \mathbf{p}_r \wedge \mathbf{u} \,. \tag{14}$$

The even multivector form given in the above equation can be compared to the symmetric energy momentum tensor $\Theta^{\mu\nu}$ with the identification of the scalar part with Θ^{00} and bivector part with Θ^{ij} . In three dimensions, the property of an even multivector is that it represents rotations in the bivector plane [35]. Then, the energy multivector can be expressed as a rotor with angular frequency ω

$$\mathcal{E} = \mathcal{E}_0 \, e^{\hat{J}\omega t} \,, \tag{15}$$

where $\hat{J}$ is a unit bivector in the plane normal to the propagation direction. This relation shows that the photon contains internal complex rotations and these rotations are analogous to the internal complex rotations or zitterbewegung of the electron. The cause of these internal rotations is attributed to the fluctuations of the zeropoint field [38]. In (15) the internal rotation represents the clockwise or right-handed rotation. A reversion operation on $\mathcal{E}$ gives

$$\bar{\mathcal{E}} = \mathcal{E}_0 \, e^{-\hat{J}\omega t} \,. \tag{16}$$

In this case, the internal rotation represents counterclockwise or left-handed rotation. The frequency of internal rotation is the rate per unit energy flux flow within the photon

$$\Omega = -\hat{J}\omega t = -\frac{1}{\mathcal{E}} \frac{d\mathcal{E}}{dt} \,. \tag{17}$$

Here, the frequency of internal rotation represents the counterclockwise direction. The internal complex rotations suggest that there exists an internal complex structure of the photon.

3 Internal structure of the photon

In general, the internal complex rotations represent the angular momentum of the photon. The angular momentum of a photon is defined as the ratio between the rotational energy of the photon and the frequency of internal rotation. Since the energy of the photon is a sum of kinetic and rotational energy components, we expect that the angular momentum of the photon contains two parts: one corresponding to the rotational flow of energy and the other to the translational flow of energy. According to the definition given in (2), the spin angular momentum bivector is in the orientation of the plane $\mathbf{A} \wedge \mathbf{E}$ which is a plane normal to the propagation direction. Let us consider a set of orthogonal unit vectors $\{\sigma_k; k = 1, 2, 3\}$ along x, y and z axes. If we choose the propagation direction along the z-axis, then the unit bivector along the spin orientation is $\mathbf{i}\sigma_3$. To understand the orientation of spin and orbital angular momenta, let us consider circularly polarized light waves propagating along the z-direction and the waves have finite extent in the $x-$ and y-directions. The propagating wave has cylindrical symmetry about the z-axis. The energy of the wave can be visualised as a sum of circulating energy flow in the x-y plane and a translational energy flow in the z-direction. In the case of circularly polarized light the vector potential $\mathbf{A}$ contains only two components

$$\mathbf{A} = \frac{\mathcal{E}_0}{\omega} \left[\sigma_1 cos(\mathbf{k}\cdot\mathbf{r} - \omega t) + \sigma_2 sin(\mathbf{k}\cdot\mathbf{r} - \omega t)\right] \,. \tag{18}$$

Here, $\mathbf{k}$ is the wave vector. The three vectors $\mathbf{E}$, $\mathbf{B}$ and $\mathbf{A}$ rotate in a plane normal to the propagation direction. Differentiation of (18) with respect to time gives the electric field vector

$$\mathbf{E} = \mathcal{E}_0 \left[-\sigma_1 sin(\mathbf{k}\cdot\mathbf{r} - \omega t) + \sigma_2 cos(\mathbf{k}\cdot\mathbf{r} - \omega t)\right] \,. \tag{19}$$

Then, the bivector product $\mathbf{A} \wedge \mathbf{E}$ becomes

$$\mathbf{A} \wedge \mathbf{E} = \mathbf{i}\sigma_3 \frac{\mathcal{E}_0^2}{\omega} \,, \tag{20}$$

where $\sigma_1\sigma_2 = \mathbf{i}\sigma_3$. The spin angular momentum of electromagnetic field or the photon is expressed as

$$S = \mathbf{i}\sigma_3 \frac{1}{4\pi} \int \frac{\mathcal{E}_0^2}{\omega} d^3r = \mathbf{i}\sigma_3 \hbar \,, \tag{21}$$

where the energy density of the electromagnetic wave is normalized so that the energy is one quantum. Normally, because of the fact $\sigma_3 \wedge \mathbf{k} = 0$, the z-component of angular momentum goes to zero but not the other components of the orbital angular momentum. From the second term on the right of (2), the angular momentum density is expressed as a sum of two terms

$$\mathbf{r} \wedge \mathbf{E}^n \nabla \mathbf{A}^n = \mathbf{r} \wedge \mathbf{E}^x \nabla \mathbf{A}^x + \mathbf{r} \wedge \mathbf{E}^y \nabla \mathbf{A}^y \,. \tag{22}$$

Substituting individual components of $\mathbf{E}$ and $\nabla\mathbf{A}$ in the above equation, we find the orbital angular momentum density

$$\mathbf{r} \wedge \mathbf{E}^n \nabla \mathbf{A}^n = \frac{\mathcal{E}_0^2}{\omega} \mathbf{r} \wedge \mathbf{k} \,. \tag{23}$$

Then the orbital angular momentum of the photon is expressed as

$$L = \mathbf{r} \wedge \mathbf{k} \frac{1}{4\pi} \int \frac{\mathcal{E}_0^2}{\omega} d^3r \,. \tag{24}$$

In the above equation, the vector $\mathbf{r}$ is restricted to the plane $\mathbf{i}\sigma_3$ and contains only x and y components. If the magnitude of $\mathbf{r}$ is equal to the reduced wavelength, then the product

$|\mathbf{r}||\mathbf{k}| = 1$ for circularly polarized light. The orbital angular momentum is now expressed as

$$L = \hat{\mathbf{r}} \wedge \sigma_3 \frac{1}{4\pi} \int \frac{\mathcal{E}_0^2}{\omega} d^3r = \mathbf{im}\hbar\,, \tag{25}$$

where the unit vector $\hat{\mathbf{r}}$, in an arbitrary direction, lies in the plane $\mathbf{i}\sigma_3$, the unit vector $\mathbf{m}$ is chosen normal to the orientation of $\hat{\mathbf{r}} \wedge \sigma_3$ and the integral term in (25) represents the ratio between the energy of the photon and the frequency. Thus the orientation of the orbital angular momentum is always normal to the orientation of spin angular momentum in a photon. In the case the photon is propagating in an arbitrary direction say $\mathbf{n}$ then from the above analysis, the spin angular momentum and orbital angular momentum are expressed as

$$S = \mathbf{in}\hbar\,, \tag{26}$$

$$L = \mathbf{im}\hbar\,. \tag{27}$$

The vectors $\mathbf{n}$ and $\mathbf{m}$ satisfy the condition $\mathbf{n} \cdot \mathbf{m} = 0$ and the vector $\mathbf{m}$ lies in the plane of the unit bivector $\mathbf{in}$. Since, the direction of unit vector $\mathbf{m}$ or the orientation of the plane $\mathbf{im}$ is arbitrary, the rotation of $\mathbf{m}$ is expressed by the relation $\mathbf{m}' = \bar{R}\mathbf{m}R$. Here, $R = e^{\mathbf{in}\phi/2}$ is a rotor and in this way the orbital angular momentum depends on the angle ϕ. The spin angular momentum describes the intrinsic angular momentum of a photon and commutes with the generator of translation $\mathbf{n}|\mathbf{k}|$. The spin angular momentum causes the complex vector field F to rotate in the $\mathbf{E} \wedge \mathbf{B}$ plane without changing the direction of propagation vector $\mathbf{k}$. The photon spin is the generator of rotations in the plane normal to the propagation direction. Whereas, the orbital angular momentum causes the plane having orientation defined by the bivector $\mathbf{r} \wedge \mathbf{k}$ to rotate without changing the direction of the vector $\mathbf{k}$ and the orientation of the plane $\mathbf{E} \wedge \mathbf{B}$. The orbital angular momentum does not commute with the generator of translation. The photon orbital angular momentum is the generator of rotations in a plane normal to the spin plane. Thus one can conclude that both the spin and orbital angular momenta of a photon are intrinsic. The intrinsic nature of orbital angular momentum was discussed by Berry [28]. Further, Allen and Padgett [29] argued that the spin and the orbital angular momenta are intrinsic in nature in the case when the transverse momentum is zero for the helical wave fronts. The spin and orbital angular momenta of the photon are fundamental quantities and produce complex rotations in space and such rotations are actually produced by the fluctuating zeropoint fields present throughout space [38, 41]. The internal complex rotations are not only limited to the rotations pertaining to the plane of spin angular momentum but also exists in the plane of orbital angular momentum. In the Laguerre-Gaussian modes of laser beams it has been shown explicitly in the quantum mechanical approach that the orbital angular momentum of light beams resembles the angular momentum of the harmonic oscillator [42].

4 Conclusions

The electromagnetic field per unit volume is represented by an energy momentum even multivector and expressed as a sum of scalar and bivector components, and we identify the scalar part as the kinetic part which shows the flow of energy in the direction of propagation and the bivector part as the rotational energy flow in the plane normal to the direction of propagation over a finite extent. The even multivector form of energy shows that there exist internal complex rotations of the electromagnetic field. The cause of these internal rotations is attributed to the fluctuations of the zeropoint field. In general, the internal complex rotations represent the angular momentum of the photon. The angular momentum of the photon is defined as the ratio between the rotational energy of the photon and the angular frequency of rotation. The spin angular momentum bivector represents a plane normal to the propagation direction. We find that the orientation of orbital angular momentum is always normal to the orientation of spin angular momentum in a photon. The photon spin is the generator of rotations in the plane normal to the propagation direction. The photon orbital angular momentum is the generator of rotations in a plane normal to the spin plane. Thus, one can conclude that both spin and orbital angular momenta of a photon are intrinsic in nature. The internal structure of the photon may be visualized as the superposition of electromagnetic field flow or rotation in two normal orientations in complex vector space. Because of the formal similarity between gluons and photons, the conclusions obtained here may be extended to the gluon structure.

Submitted on May 23, 2016 / Accepted on May 30, 2016

References

1. Jackson J. D. Classical Electrodynamics. Wiley Eastern Limited, New Delhi, 1978.
2. Beth R. A. Mechanical Detection and Measurement of the Angular Momentum of Light. *Phys. Rev.*, 1936, v. 50, 115.
3. Belinfante F. J. On the spin angular momentum of mesons. *Physica*, 1939, v. 6, 887–898.
4. Ohanian H. C. What is spin? *Am. J. Phys.*, 1986, v. 54, 500.
5. Cohen-Tannoudji C., Dupont-Roc J., Grynberg G. Photons and Atoms. John Wiley & Sons, New York, 1989.
6. Jauch J. M., Rohlich F. The Theory of Photons and Electrons. Springer-Verlag, Berlin, 1976.
7. Leader E., Lorce C. The angular momentum controversy: What's it all about and does it matter? *Physics Reports*, 2014, v. 541, 163–248.
8. Van Enk, Nienhuis G. Spin and Orbital Angular Momentum of Photons. *Euro. Phys. Lett.*, 1994, v. 25, 497.
9. Chen X. S., Chen G. B., Sun W. M., Wang F. Testing the correctness of the Poynting vector $\mathbf{E} \times \mathbf{B}$ as the momentum density of gauge fields. arXiv: hep-ph/0710.1427v1.
10. Chen X. S., Lu X. F., Sun W. M., Wang F., Goldman T. Spin and Orbital Angular Momentum in Gauge Theories: Nucleon Spin Structure and Multipole Radiation Revisited. *Phys. Rev. Lett.*, 2008, v. 100, 232002.
11. Wakamatsu M. Gauge-invariant decomposition of nucleon spin. *Phys. Rev. D.*, 2010, v. 81, 114010.

12. Sun W. M. Physical Decomposition of Photon Angular Momentum. arXiv: quant-ph/1407.2305v1.

13. Allen L., Beijersbergen M. W., Spreeuw R. J. C., Woerdman J. P. Orbital angular momentum of light and the transformation of Laguerre-Gaussian laser modes. *Phys. Rev. A*, 1992, v. 45, 8185.

14. Leach J., Padgett M. J., Barnett S. M., Franke-Arnold S., Courtial J. Measuring the orbital angular momentum of a single photon. *Phys. Rev. Lett.*, 2002, v. 88, 257901.

15. Allen L., Barnett S. M., Padgett M. J. Optical Angular Momentum. IOP Publishing, Bristol and Philadelphia, 2003.

16. Padgett M., Courtial J., Allen L. Light's Orbital Angular Momentum. *Physics Today*, May 2004, 35.

17. Barut A. O., Bracken A. J. Zitterbewegung and the internal geometry of electron. *Phys. Rev. D*, 1981, v. 23, 2454.

18. Hestenes D. Spin and uncertainty in the interpretation of quantum mechanics. *Am. J. Phys.*, 1979, v. 47, 399–415.

19. Hestenes D. Zitterbewegung in Quantum Mechanics. *Found. Phys.*, 2010, v. 40, 1–54.

20. Barut A. O., Zanghi A. J. Classical Model of the Dirac Electron. *Phys. Rev. Lett.*, 1984, v. 52, 2009–2012.

21. Hovarthy P. A. Mathisson's spinning electron: Noncommutative mechanics and exotic Galilean symmetry, 66 years ago. *Acta. Phys. Pol. B*, 2003, v. 34, 2611–2622.

22. Vaz J., Rodriguez W. A., Zitterbewegung and the electromagnetic field of the electron. *Phys. Lett. B*, 1993, v. 319, 243.

23. Kobe D. H. Zitterbewegung of a photon. *Phys. Lett. A*, 1999, v. 253, 7–11.

24. Unal N. A Simple Model of Classical Zitterbewegung: Photon Wave Function. *Found. Phys.*, 1997, v. 27, 731.

25. Wang Z. Y., Xiong C. D. Zitterbewegung by quantum field-theory considerations. *Phys. Rev. A*, 2008, v. 77, 045402.

26. Wang Z. Y., Xiong C. D., Qiu Q. Photon wave function and Zitterbewegung. *Phys. Rev, A*, 2009, v. 80, 032118.

27. Zhang X. Observing Zitterbewegung for Photon near the Dirac point of a Two-Dimensional Photonic Crystal. *Phys. Rev. Lett.*, 2008, v. 100, 113903.

28. Berry M. V. Paraxial beams of spinning light. In: Allen S., Barnett S. M., Padgett, M. J., eds. Optical Angular Momentum. IOP Publishing, Bristol, Philadelphia, 2003.

29. Allen L., Padgett M. J. The Poynting vector in Laguerre-Gaussian beams and the interpretation of their angular momentum density. *Opt. Commun.*, 2000, v. 184, 6771.

30. Calvo G. F., Picón A., Bagan E. Quantum field theory of photons with orbital angular momentum. *Phys. Rev. A*, 2006, v. 73, 01385.

31. Padgett M. J., Courtial J. Poincare-sphere equivalent for light beams containing orbital angular momentum. *Opt. Lett.*, 1999, v. 24, 430.

32. Agarwal G. S. SU(2) structure of the Poincaré sphere for light beams with orbital angular momentum. *J. Opt. Soc. Am. A*, 1999, v. 16, 2914–2916.

33. Mair A., Vaziri A., Weihs G., Zeilinger A. Entanglement of the orbital angular momentum states of photons. *Nature*, 2001, v. 412, 313–316.

34. Franke-Arnold S., Barnett S. M., Padgett M. J., Allen L. Two-photon entanglement of orbital angular momentum states. *Phys. Rev. A*, 2002, v. 65, 033823-1.

35. Hestenes D. Oersted Medal Lecture 2002: Reforming the Mathematical Language of Physics. *Am. J. Phys.*, v. 71, 104.

36. Muralidhar K. The Spin Bivector and Zeropoint Energy in Geometric Algebra. *Adv. Studies Theor. Phys.*, 2012, v. 6, 675–686.

37. Muralidhar K. Complex Vector Formalism of Harmonic Oscillator in Geometric Algebra: Particle Mass, Spin and Dynamics in Complex Vector Space. *Found. Phys.*, 2014, v. 44, 265–295.

38. Muralidhar K. Algebra of Complex vectors and Applications in electromagnetic Theory and Quantum Mechanics. *Mathematics*, 2015, v. 3, 781–842.

39. Roychoudhuri C., Kracklauer A. F., Creath K. The Nature of Light. What is a Photon? C. R. C. Press, Taylor and Francis Group, Boca Raton, FL, 2008.

40. Doran C., Lasenby A. Geometric Algebra for Physicists. Cambridge University Press, Cambridge, 2003.

41. Kobe D. H. A relativistic Schrödinger-like Equation for a Photon and Its second Quantization. *Found. Phys.*, 1999, v. 29, 1203–1231.

42. Nienhuis G., Allen L. Paraxial wave optics and harmonic oscillators. *Phys. Rev. A*, 1993, v. 48, 656–665.

Gravitational Waves from a Sinusoidially Varying Spherical Distribution of Mass

L. F. Obagboye[1], S. X. K. Howusu[1], E. N. Chifu[2] and E. J. N. Omaghali[3]

[1]Theoretical Physics Programme, National Mathematical Centre Abuja, Nigeria
E-mail: sxkhowusu@yahoo.com
[2]Physics Department, Federal University Dutse, Nigeria
E-mail: ebenechifu@yahoo.com
[3]Physics Department, University of Jos, Nigeria

A theory is developed for the study of spherical gravitational waves by constructing a Generalized Gravitational Field Equation from Newton's gravitational field equation. The Euclidean Laplacian ∇^2 is replaced with the Riemannian Laplacian ∇_R^2. A general gravitational field equation is obtained which resolves the incompleteness in Newton's gravitational field equation. The general gravitational field equation reduces to the pure Newtonian gravitational field equation in the limit of c^0 as required by the Principle of Equivalence of Physics. It also contains post Newton correction terms of orders of c^{-2} and all degrees of nonlinearity in the gravitational scalar potential and its derivatives. Considering a sinusoidally varying homogeneous spherical distribution of mass in the frame work of the obtained general gravitational field equation, gravitational waves are predicted with phase velocity equivalent to the speed of light in vacuo.

1 Introduction

According to General Relativity Theory, gravitational waves are oscillations of spacetime or small distortions of spacetime geometry, or ripples of spacetime curvature which propagate in the time through space as waves. Gravitational waves are produced mainly by extremely massive binary stellar objects, such as binary neutron stars or binary black holes. Though gravitational waves can be produced by all mass interactions, the amplitude of these waves is far too small to be detected. Normal solar systems produce gravitational waves when their planets orbit their primary, but again, these are incredibly tiny ripples. Even a binary black hole — which produces the most powerful gravitational waves we can imagine — requires measurements of distances of about 1/1000 of the diameter of a proton [1].

The search for gravitational waves has been the centre of current research in Astronomy and Cosmology. Higher precision and more sensitive detectors have been developed over the years. Experiments on gravitational waves started with Weber's experiments on gravitational antennae; in which he registered weak signals [2]. He concluded that some processes at the centre of the Galaxy were the origin of the detected signals. Other attempts were made in detecting gravitational waves such as [3-6]. The most recent experimental attempt by Abbott *et al.* in 2015 [7] claims that two detectors of the Laser Interferometer Gravitational-Wave Observatory simultaneously observed a transient gravitational-wave signal.

Much theoretical work has also been done to either proof or disproof the existence of gravitational waves. In a nutshell, theoretical studies of gravitational waves can be classified into three main groups [2]:

- Research targeted at giving an invariant definition for gravitational waves. These include Pirani [8], Bondi [9], and others.
- Searching for solutions to Einstein gravitational field equations by proceeding from physical considerations to describe gravitational radiations. These include studies by Einstein and Rosen [10], Petrov [11], Chifu and Taura [1] and others.
- Studying gravitational inertial waves, covariant with respect of transformations of spatial coordinates and also invariant with respect of transformations of time [12].

This research article falls in the second group. The so-called "Great Metric Tensor" [13-14] is used to deduce a general gravitational wave equation; which is later applied to a sinusoidally varying mass for a homogeneous spherical distribution of mass.

2 The general spherical gravitational field equation

Newton's gravitational field equation is given by

$$\nabla^2 f(\underline{r}, t) = 4\pi G \rho_0(\underline{r}, t) \tag{1}$$

where, ρ_0 is the density of proper mass in a distribution or system, ∇^2 is the pure Euclidean Laplacian, G is the universal gravitational constant and f is the gravitational scalar potential.

The incompleteness of equation (1) are as follows:

1. The density of proper mass (source of gravitational field) in equation (1) can vary with coordinate time and the Euclidean Laplacian cannot account for this possible variation.

2. Time variation of proper mass should result in the radiation of energy possibly in the form of gravitational waves or radiation that can propagate in space-time with or without gravitational field.
3. Newton's gravitational intensity vector $\underline{g}$ is given by

$$\underline{g} = -\underline{\nabla} f \tag{2}$$

where $\underline{\nabla}$ is the Euclidean gradient operator.

The Euclidean operator in equation (2) above has no variation with time and hence will not be sufficient for the complete description of gravitational intensity vector of time dependent gravitational fields.

From the foregoing it becomes necessary to seek a general gravitational field equation which will be sufficient for the description of all gravitational fields. Howusu in 2009 [13] proposed that a general gravitational field equation based on Riemannian coordinate geometry may be obtained by replacing the Euclidean Laplacian with Riemannian Laplacian to obtain

$$\nabla_R^2 f(\underline{r}, t) = 4\pi G \rho_0(\underline{r}, t) \tag{3}$$

where ∇_R^2 is the Riemannian Laplacian based on the great metric tensor for all possible gravitational fields. The gravitational intensity (acceleration due to gravity) for all possible gravitational fields can also be defined in terms of the Riemannian gradient operator ∇_R. The most general form of the Riemannian Laplacian is given as

$$\nabla_R^2 = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^\mu} \left(\sqrt{g} g^{\mu\nu} \frac{\partial}{\partial x^\nu} \right) \tag{4}$$

where $g^{\mu\nu}$ is the contravariant metric tensor. Thus, for any function $f(\underline{r}, t)$ we can write

$$\begin{aligned} \nabla_R^2 f(\underline{r}, t) = & \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} \left(\sqrt{g} g^{ij} \frac{\partial}{\partial x^j} \right) f(\underline{r}, t) \\ & + \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^0} \left(\sqrt{g} g^{00} \frac{\partial}{\partial x^0} \right) f(\underline{r}, t). \end{aligned} \tag{5}$$

Using Einstein's coordinates with $x^0 = ct$, equation (5) can be written explicitly as

$$\begin{aligned} \nabla_R^2 f(\underline{r}, t) = & \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} \left(\sqrt{g} g^{ij} \frac{\partial}{\partial x^j} \right) f(\underline{r}, t) \\ & + \frac{1}{c^2 \sqrt{g}} \frac{\partial}{\partial t} \left(\sqrt{g} g^{00} \frac{\partial}{\partial t} \right) f(\underline{r}, t). \end{aligned} \tag{6}$$

Hence equation (3) can be written more explicitly as

$$\begin{aligned} 4\pi G \rho_0(\underline{r}, t) = & \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} \left(\sqrt{g} g^{ij} \frac{\partial}{\partial x^j} \right) f(\underline{r}, t) \\ & + \frac{1}{c^2 \sqrt{g}} \frac{\partial}{\partial t} \left(\sqrt{g} g^{00} \frac{\partial}{\partial t} \right) f(\underline{r}, t). \end{aligned} \tag{7}$$

Equation (7) is the general field equation which resolves the incompleteness of Newton's gravitational field equation. Remarkably, the general gravitational field equation reduces to the pure Newton's gravitational field equation in the limit of c^0 (as required by the Principle of Equivalence of Physics). It may also be noted that the gravitational field equation contains post Newton correction terms of orders of c^{-2} and all degrees of nonlinearity in the gravitational scalar potential and its derivatives.

The Great Metric Tensor for all spherical gravitational fields in spherical polar coordinates (r, θ, ϕ, x^0) is given as [13-14]:

$$g_{11}(r, \theta, \phi, x^0) = \left(1 + \frac{2}{c^2} f(r, \theta, \phi, x^0) \right)^{-1}, \tag{8}$$

$$g_{22}(r, \theta, \phi, x^0) = r^2, \tag{9}$$

$$g_{33}(r, \theta, \phi, x^0) = r^2 \sin^2 \theta, \tag{10}$$

$$g_{00}(r, \theta, \phi, x^0) = -\left(1 + \frac{2}{c^2} f(r, \theta, \phi, x^0) \right) \tag{11}$$

where f is the gravitational scalar potential. From equation (8) to (11) it can be deduced that

$$\sqrt{g} = r^2 \sin \theta. \tag{12}$$

Equation (7) can thus be written as:

$$\begin{aligned} 4\pi G \rho_0(r, t) = & \frac{1}{r^2} \frac{\partial}{\partial r} \left[\left(1 + \frac{2}{c^2} f \right) r^2 \right] f \\ & + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) f \\ & + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} f \\ & - \frac{1}{c^2} \frac{\partial}{\partial t} \left[\left(1 + \frac{2}{c^2} f \right)^{-1} \frac{\partial}{\partial t} f \right]. \end{aligned} \tag{13}$$

Equation (13) is the general spherical gravitational field equation interms of the great metric tensor. The following important facts can be drawn from equation (13):

1. It contains the $\left(1 + \frac{2}{c^2} f\right)$ term which is not found in Newton's gravitational field equation. The consequence of this is that it predicts correction terms to the gravitational field of all massive spherical bodies.
2. The time component of this equation predicts the existence of gravitational waves with velocity which is equal to the speed of light in vacuo.

3 Special case: sinusoidally varying homogenous spherical distribution of mass

Now, consider a sinusoidally varying homogenous spherical distribution of mass. In this case, the mass varies in such a

way that f is independent of the polar angle θ and the azimuthal angle ϕ, [15] such that equation (13) reduces to

$$4\pi G\rho_0(r,t) = \frac{1}{r^2}\frac{\partial}{\partial r}\left[\left(1+\frac{2}{c^2}f\right)r^2\frac{\partial}{\partial r}\right]f - \frac{1}{c^2}\frac{\partial}{\partial t}\left[\left(1+\frac{2}{c^2}f\right)^{-1}\frac{\partial}{\partial t}f\right]. \quad (14)$$

Linearizing equation (14) we obtain:

$$f'' + \frac{2}{r}f' - \frac{1}{c^2}\ddot{f} = 4\pi G\rho_0. \quad (15)$$

Suppose we have a dipole antenna which consists of two spherical bodies where electrons are driven by an oscillator [1]; then the movement of the electric charges driven by the oscillator is equivalent to an exponential factor. We therefore modify equation (15) in such a way that the proper mass density varies sinusoidally within a homogeneous spherical mass distribution such that:

$$f'' + \frac{2}{r}f' - \frac{1}{c^2}\ddot{f} = 4\pi G\rho_e\, e^{i\omega t}. \quad (16)$$

In order to solve equation (16) we seek a solution such that

$$f(r,t) = R(r)\, e^{i\omega t} \quad (17)$$

where R is the radius of the spherical mass distribution. Equation (15) will thus become

$$R''(r) + \frac{2}{r}R'(r) + \frac{1}{c^2}\omega^2 R(r) = 4\pi G\rho_e. \quad (18)$$

Let

$$R(r) = \frac{1}{r}F(r),$$

then

$$R' = -\frac{1}{r^2}F(r),$$

and

$$R''(r) = \frac{1}{r}F''(r) - \frac{2}{r^2}F'(r) + \frac{2}{r^3}F(r).$$

It therefore follows that equation (18) becomes

$$\frac{1}{r}F''(r) + \frac{\omega^2}{c^2 r}F(r) = 4\pi G\rho_e. \quad (19)$$

Hence, the interior field equation for this distribution of mass is given as

$$\frac{1}{r}F''(r) + \frac{\omega^2}{c^2 r}F(r) = 4\pi G\rho_e\,;\ r < R \quad (20)$$

and the corresponding exterior field equation as:

$$\frac{1}{r}F''(r) + \frac{\omega^2}{c^2 r}F(r) = 0\,;\ r > R. \quad (21)$$

Equation (21) is a simple harmonic function which can have three solutions viz:

$$F(r) = Be^{ikr}, \quad (22)$$

$$F(r) = D\cos(kr), \quad (23)$$

and

$$F(r) = E\sin(kr). \quad (24)$$

Taking the first and second derivatives of equation (22) we have

$$F'(r) = ikB\, e^{ikr}$$

and

$$F''(r) = -k^2 B\, e^{ikr},$$

which can be substituted into (21) to yield

$$-k^2 B\, e^{ikr} + \frac{\omega^2}{c^2}B\, e^{ikr} = 0, \quad (25)$$

hence

$$k = \pm\frac{\omega}{c}. \quad (26)$$

We thus state the complimentary solution as

$$F_c^-(r) = E\sin\left(\frac{\omega}{c}r\right);\ r > R \quad (27)$$

$$F_c^+(r) = D\cos\left(\frac{\omega}{c}r\right);\ r < R. \quad (28)$$

The particular solution for the interior field equation is given by

$$\frac{1}{r}F''(r) + \frac{\omega^2}{c^2 r}F(r) = 4\pi G\rho_e; r < R. \quad (29)$$

Let $F(r) = Ar$, then $F'(r) = A$ and $F''(r) = 0$ and equation(29) yields

$$A = \frac{4\pi G c^2\rho_e}{\omega^2}, \quad (30)$$

and hence

$$F_p^-(r) = \frac{4\pi G c^2\rho_e}{\omega^2}r. \quad (31)$$

Equation (31) is thus the particular solution for the exterior field equation. The general solution for the exterior field is then given as

$$R^+(r) = \frac{D}{r}\cos\left(\frac{\omega}{c}r\right)\frac{4\pi G c^2\rho_e}{\omega^2}. \quad (32)$$

Equation (17) can thus be fully expressed as

$$f^+(r,t) = \frac{D}{r}\cos\left(\frac{\omega}{c}r\right)\cos(\omega t) + \frac{iD}{r}\cos\left(\frac{\omega}{c}r\right)\sin(\omega t) \quad (33)$$

with independent solutions

$$f^+(r,t) = \frac{D}{r}\cos\left(\frac{\omega}{c}r\right)\cos(\omega t) \quad (34)$$

and

$$f^{+}(r,t) = \frac{D}{r}\cos\left(\frac{\omega}{c}r\right)\sin(\omega t). \tag{35}$$

The two solutions (34) and (35) can be combined to yield

$$f^{+}(r,t) = \frac{1}{2}\frac{D}{r}\left[\cos\left(\omega\left(\frac{r}{c}+t\right)\right)+\cos\left(\omega\left(\frac{r}{c}-t\right)\right)\right]. \tag{36}$$

From equation (36) it is clear that the phase of the wave equation ϕ is given by

$$\phi = \frac{\omega r}{c} \pm \omega t, \tag{37}$$

hence

$$\frac{dr}{dt} = c. \tag{38}$$

4 Concluding remarks

In this paper we have shown [equation (36)] that in the limit of linear terms, the general gravitational field equation predicts gravitational waves with phase velocity which is equal to the speed of light in empty space. These waves will not vary with any angle, hence they will move along radial lines from inside the sphere outwards(radial waves). A sinusoidally varying mass thus radiates spherical gravitational waves. The obtained results gives similarlar predictions as in [1, 16] in the limit c^{-2} though in the limit c^{0} [16] predicts gravitational waves with imaginary phase.

Submitted on May 27, 2016 / Accepted on June 2, 2016

References

1. Chifu E.N. and Taura L.S. *Progress in Physics*, 2013, v. 9, issue 3, 7.
2. Borissova L. *Progress in Physics*, 2005, v. 1, issue 2, 30.
3. Braginsky V.B., Manukin A.B., Popov E.J., Rudenko V.N. *Phys. Rev. Ser. A*, 1973, v. 45, 271.
4. Douglas D.N., Gram R.Q., Tyson J.A., Lee R.W. *Phys. Rev. Lett.*, 1975, v. 35, 480.
5. Levine J.L., Garwin R.L. *Phys. Rev. Lett.*, 1973, v. 31, 173.
6. Tyson J.A., Macleannan C.G., Lanzorotti L.J. *Phys. Rev. Letts*, 1973, v. 30, 1006.
7. Abbott B.P. et al. *Phys. Rev. Lett*, 2015, v. 116, 061102.
8. Bondi H., Pirani F., Robinson J. *J. Proc. Roy. Soc. A*, 1959, v. 251, 519.
9. Bondi H. *Nature*, 1957, v. 179, 1072.
10. Rosen N. *Res. Council Israel*, 1954, v. 4, 328.
11. Petrov A.Z. Einstein Spaces. Pergamon, london, 1969.
12. Borissova L. *The Abraham Zelmanov Journal*, 2010, v. 3, 26.
13. Howusu S.X.K. The Metric Tensors for Gravitational Fields Physics and the Mathematical Principles of Riemannian Theoretical Physics. Jos University Press, 2009.
14. Howusu S.X.K. Riemannian Revolutions in Physics and Mathematics. Jos University Press, 2013.
15. Chifu E.N. and Howusu S.X.K. *Physics Essays*, 2009, v. 22(1), 73.
16. Chifu E.N. and Howusu S.X.K. *Progress in Physics*, 2009, v. 5, issue 3, 45.

Gravitational Shielding as Viewed in the Planck Vacuum Theory

William C. Daywitt

National Institute for Standards and Technology (retired), Boulder, Colorado. E-mail: wcdaywitt@me.com

This paper argues that gravitational shielding does not exist, because gravitational waves travel within the vacuum state rather than free space.

1 Introduction

The concept of gravitational shielding has been around for a long time and it would incorrectly assert, for example, that when the earth lines up between the moon and the sun, the moon-sun gravitational attraction is reduced. The fact that this shielding (by the earth in this case) does not occur is one of the great mysteries in the history of physics. The theory of the Planck vacuum (PV) state, however, offers an easy explanation for the absence of such shielding.

As a counter example to the gravitational force, consider the free space Coulomb force (e^2/r^2) between two charges e separated by the distance r. If a shield of any type whatsoever is placed between the charges, the resulting force is changed dramatically. Indeed, if a large enough grounded screen were inserted between the charges, the force would vanish entirely.

2 Newton Force

Now consider the gravitational force between two free space masses m in the center-of-mass (CoM) coordinate frame defined by $m\mathbf{r}_1 + m\mathbf{r}_2 = 0$ (with $-\mathbf{r}_1 = \mathbf{r}_2 = \mathbf{r}$):

$$F_{gr}(r) = -\frac{m^2G}{(2r)^2} = -\frac{m^2c^4}{4r^2(c^4/G)} = -\frac{(mc^2/2r)^2}{(c^4/G)} \tag{1}$$

$$= -\frac{(mc^2/2r)^2}{(m_*c^2/r_*)} = -n_{2r}^2\frac{m_*c^2}{r_*} \tag{2}$$

where the n-ratio

$$n_{2r} = \frac{mc^2/2r}{m_*c^2/r_*} < 1 \tag{3}$$

is the normalized force either mass m exerts on the PV at the position of the opposite mass, where the masses are centered at $\pm\mathbf{r}$ from the origin of the CoM coordinates. The normalization force m_*c^2/r_* is the maximum force the PV can sustain before breaking down. This force also normalizes the Einstein field equation [1, eqn.15].

The three ratios in (1) are the force equation expressed in terms of Newton's secondary constant G, an experimental constant that makes (1) agree with the experimental data. As such, however, G hides a significant amount of physics. The substitution $c^4/G = m_*c^2/r_*$ [1, eqn.5] replaces G by a combination of primary (fundamental) constants that lead to (2) and the following nonrelativistic explanation of the gravitational force.

The gravitational field $g(r)$ of either mass can be defined in the usual manner and yields

$$g(r) = \frac{F_{gr}(r)}{m} = -\frac{c^2n_{2r}}{2r} \tag{4}$$

which is again centered at the radii $\pm r$ from the CoM origin, where r is the coordinate radius common to both free space and its underlying PV state.

From (2) and (4) it is easy to carry these calculations a step further. Newton's second law applied to either mass gives the acceleration

$$\ddot{r} = \frac{d\dot{r}}{dt} = \dot{r}\frac{d\dot{r}}{dr} = -\frac{c^2n_{2r}}{2r} = -\frac{c^2\cdot mc^2}{4\cdot m_*c^2/r_*}\frac{1}{r^2} \tag{5}$$

or

$$\dot{r}d\dot{r} = -\frac{c^2\cdot mc^2}{4\cdot m_*c^2/r_*}\frac{dr}{r^2} \tag{6}$$

of the masses. Integrating both sides of (6) from r_0 to r leads to

$$\frac{\dot{r}^2 - \dot{r}_0^2}{2} = \frac{c^2n_{2r0}}{2}\left(\frac{r_0}{r} - 1\right) \tag{7}$$

where $\dot{r}_0$ is the velocity of either mass at $r = r_0$, and $r \le r_0$. Without changing the final conclusions, it is convenient to set $\dot{r}_0 = 0$ — i.e., to assume that the masses are released from rest at $\pm r_0$. Then (7) yields their relative velocities toward the origin

$$\frac{\dot{r}}{c} = -\left[n_{2r0}\left(\frac{r_0}{r} - 1\right)\right]^{1/2} \tag{8}$$

where

$$n_{2r0} = \frac{mc^2/2r_0}{m_*c^2/r_*} \tag{9}$$

and $[\cdots]$ in (8) is the normalized force either mass exerts on the PV at the position of the other mass.

3 Conclusions

Three important observations are evident from the previous calculations: equations (2), (4), and (8) are all expressed in terms of PV parameters, implying that the vacuum state mediates the dynamics of the gravitational force between free space masses. A corollary to this conclusion is that gravitational waves, the carrier of the gravitational force, do not propagate in free space — they propagate within the degenerate PV state. Thus free-space gravitational shielding does not change the gravitational force between free space masses.

Finally, the fact that the PV is a degenerate state implies that the Planck particles making up the PV quasi-continuum cannot execute macroscopic (as opposed to microscopic) motions. Thus the gravitational waves that propagate through the PV state must be percussion-like waves, similar to the waves traveling on the surface of a kettle drum.

Submitted on May 26, 2016 / Accepted on June 5, 2016

References

1. Daywitt W.C. The Trouble with the Equations of Modern Fundamental Physics. *American Journal of Modern Physics. Special Issue: Physics Without Higgs and Without Supersymmetry*, 2016, v. 5, no. 1-1, 22. See also www.planckvacuum.com.

The Journal on Advanced Studies in Theoretical and Experimental Physics, including Related Themes from Mathematics

PROGRESS IN PHYSICS

A quarterly issue scientific journal, registered with the Library of Congress (DC, USA). This journal is peer reviewed and included in the abstracting and indexing coverage of: Mathematical Reviews and MathSciNet (AMS, USA), DOAJ of Lund University (Sweden), Scientific Commons of the University of St. Gallen (Switzerland), Open-J-Gate (India), Referativnyi Zhurnal VINITI (Russia), etc.

Electronic version of this journal:
http://www.ptep-online.com

Postal Address

Department of Mathematics and Science,
University of New Mexico,
705 Gurley Ave., Gallup, NM 87301, USA

ISSN: 1555-5534 (print)
ISSN: 1555-5615 (online)

Standard Address Number: 297-5092
Printed in the United States of America

October 2016 **Vol. 12, Issue 4**

CONTENTS

Consiglio J. Take Fifteen Minutes to Compute the Fine Structure Constant (*Letters to Progress in Physics*) 305

Bilbao L. Does the Velocity of Light Depend on the Source Movement? 307

Marquet P. Vacuum Background Field in General Relativity 313

Marquet P. On the Physical Nature of the de Broglie Wave 318

Daywitt W. C. Antigravity and Vacuum Propulsion in the Planck Vacuum Theory 323

Ogiba F. On the Quantum-Relativistic Behavior of Moving Particles 325

Faizuddin A. Type III Spacetime with Closed Timelike Curves 329

Seaver J. R. Experimental and Theoretical Test of Cahill's Detection of Absolute Velocity in Gas-mode Interferometer Experiments 332

Heymann Y. Dark Matter, the Correction to Newton's Law in a Disk 347

Zhang T. X. Principle of Spacetime Black Hole Equivalence 353

Seaver J. R. Attempt to Replicate Cahill's Quantum Gravity Experiment to Measure Absolute Velocity 362

Marquet P. Some Insights on the Nature of the Vacuum Background Field in General Relativity (*Letters to Progress in Physics*) 366

Heymann Y. Conservation Laws and Energy Budget in a Static Universe 368

Information for Authors

Progress in Physics has been created for rapid publications on advanced studies in theoretical and experimental physics, including related themes from mathematics and astronomy. All submitted papers should be professional, in good English, containing a brief review of a problem and obtained results.

All submissions should be designed in LaTeX format using *Progress in Physics* template. This template can be downloaded from *Progress in Physics* home page http://www.ptep-online.com

Preliminary, authors may submit papers in PDF format. If the paper is accepted, authors can manage LaTeXtyping. Do not send MS Word documents, please: we do not use this software, so unable to read this file format. Incorrectly formatted papers (i.e. not LaTeXwith the template) will not be accepted for publication. Those authors who are unable to prepare their submissions in LaTeXformat can apply to a third-party payable service for LaTeX typing. Our personnel work voluntarily. Authors must assist by conforming to this policy, to make the publication process as easy and fast as possible.

Abstract and the necessary information about author(s) should be included into the papers. To submit a paper, mail the file(s) to the Editor-in-Chief.

All submitted papers should be as brief as possible. Short articles are preferable. Large papers can also be considered. Letters related to the publications in the journal or to the events among the science community can be applied to the section *Letters to Progress in Physics*.

All that has been accepted for the online issue of *Progress in Physics* is printed in the paper version of the journal. To order printed issues, contact the Editors.

Authors retain their rights to use their papers published in *Progress in Physics* as a whole or any part of it in any other publications and in any way they see fit. This copyright agreement shall remain valid even if the authors transfer copyright of their published papers to another party.

Electronic copies of all papers published in *Progress in Physics* are available for free download, copying, and re-distribution, according to the copyright agreement printed on the titlepage of each issue of the journal. This copyright agreement follows the *Budapest Open Initiative* and the *Creative Commons Attribution-Noncommercial-No Derivative Works 2.5 License* declaring that electronic copies of such books and journals should always be accessed for reading, download, and copying for any person, and free of charge.

Consideration and review process does not require any payment from the side of the submitters. Nevertheless the authors of accepted papers are requested to pay the page charges. *Progress in Physics* is a non-profit/academic journal: money collected from the authors cover the cost of printing and distribution of the annual volumes of the journal along the major academic/university libraries of the world. (Look for the current author fee in the online version of *Progress in Physics*.)

LETTERS TO PROGRESS IN PHYSICS

Take Fifteen Minutes to Compute the Fine Structure Constant

Jacques Consiglio

52 Chemin de Labarthe, 31600 Labastidette, France
E-mail: Jacques.Consiglio@gmail.com

This note complements the calculus of the fine structure constant provided in [2] in agreement with the theory of mass/resonances developed therein. It shows that the value of α can be predicted from geometry using a) the assumption of integral resonances, b) de Broglie's thesis, and c) the Wheeler-Feynman absorber theory and its time-symmetry; hence independently of precision measurement.

1 Introduction

Using Quantum Electro-Dynamics (QED), precise measurement of the electron magnetic moment anomaly enables to compute the value of the fine structure constant.

In this note, we show that the resulting value of α pulls us back almost to square one, namely Bohr's model and de Broglie's thesis, since the assumption of integral resonances used in [2] and its use of the Wheeler-Feynman absorber theory [5], [6] give the same result, straight from geometry.

2 The calculus

In order to complete the calculus, we shall need two assumptions used sequentially:

- All elementary particles are integral-number based resonances of physical currents. We uses the verb "to be" in its full sense: there is nothing else to deal with.
- The Wheeler-Feynman absorber theory [5], [6], is close to the right picture. The universe expands in a 4^{th} spatial dimension and we live at some sort of boundary or membrane that expands spherically. Up and Down-time currents exist making particles.

Now according to de Broglie [1] the phase coherence of the wave gives the Bohr orbits. Second, consider the first orbit and imagine the figure, a helicoid, in x, y, t. Considering a system of unit where the Bohr radius is 1 in x, y, and its Compton frequency is 1 on the time axis, the helix length is:

$$L_h^2 = 137^2 + (2\pi)^2.$$

According to the assumptions, this expression is the effect of a resonance, but α is the coupling of the electron with the field; therefore it is the amplitude *and the geometry* from which L_h develops. Since $\alpha < 1$, we necessarily have:

$$\alpha \leftarrow L_h^{-1}.$$

But the electron makes one turn when the helix makes two turns. With respect to the electron "being" a resonance, its rotational path length must be reduced by half and we get a resonance length:

$$L_r^2 \approx 137^2 + \pi^2.$$

Now we need to take into account the wavelength h/p as part of the electron resonance. According to de Broglie, its phase velocity is $V = c^2/v$, with v the electron velocity; here distances are inverted and velocity dependent. Its length around the proton is then $1/274$ (the electron phase repeats every 274 Compton periods). But when the wave makes one turn the electron progresses; therefore the resonance makes 275 turns when the electron resonance makes a full turn. The wave misses 1 turn over 275, which gives a negative term:

$$L_r^2 \approx 137^2 + \pi^2 - \frac{1}{275}.$$

Here the negative term is not squared. The explanation is a little less trivial than the rest of the calculus. Denoting a_n the radius of the n^{th} Bohr orbit and λ_{dn} the associated de Broglie wavelength, we have:

$$a_n = n^2\, a_0; \lambda_{dn} = n\, \lambda_{d0}\,.$$

Those quantities are physical. The round trip of the wave is $n\,\lambda_{dn} = n^2\,\lambda_{d0}$ and corresponds to quantized angular momentum; at the opposite, the same trip includes $137\,n^2$ Compton periods. Therefore a different treatment is needed for 137 and $1/275$. The former is squared in (1) and associated to n^2; then since the latter is associated to n, it cannot be squared; otherwise this expression would be orbit dependent in n. This is the physical aspect, it means that on any Bohr orbit we can use a system of units in the space dimensions where $n^2\,a_0 = 1$, and the de Broglie wavelength and its angle (its phase velocity) defines the unit for $V > c$. We end-up with a system of units which is entirely defined by de Broglie's geometry, where all quantities are defined by h or $\hbar$, and the electron mass.

Let us now use the second assumption. The field is time-symmetrical for an observer which is fixed in time (this is also the perspective of QED). Time symmetry implies that the electron is composite of up and down-time currents: Up-time $= -e/2$, down-time $= +e/2$. Those currents are cen-

tered like the electron resonance (on the helix) and manifest an electric charge which contribution (sign) depends on their own sign and time-orientation. Their interaction gives $(-e/2)(+e/2) = -e^2/4$, which compares to $-e^2$, the interaction electron-proton.

We must apply the same reasoning to the wave; by symmetry it is also composed of two currents of opposite directions, but of identical charges, centered on the electron. Then we just add $1/4$ as follows:

$$L_r^2 = 137^2 + \pi^2 - \frac{1}{275}\left(1 + \frac{1}{4}\right). \quad (1)$$

Last we compute the inverse of this length to get α:

$$(1) \rightarrow L_r^{-1} = 7.29\,735\,256\,656\,433\,e^{-3}. \quad (2)$$

Compare with CODATA 2014:

$$\alpha = 7.29\,735\,256\,64\,(12)\,e^{-3}. \quad (3)$$

The difference is on the last digit and $1/7^{th}$ the uncertainty.

You can stop your chronometer.

3 Conclusions

The fine structure constant was computed from de Broglie's geometry under the following assumptions:

- The electron "is" an integral resonance,
- The existence of symmetrical currents, where we see the signature of a resonant system,
- Asymmetry in currents between space and time, which is implicit in the reasoning.

This result completes the calculus provided in [2] where a logical origin of 137 is uncovered.

Interestingly, it was possible to predict this value of α about 70 years ago pushing Wheeler-Feynman's absorber theory to its natural consequences in terms of time-symmetry, since $\alpha \approx 1/137$ was known.

By the way, it also requires to use de Broglie's geometry in its full extent; not only the wavelength $\lambda_d = h/p$, but also the phase velocity $V = c^2/v > c$ for which no experimental verification exists. We showed that this velocity is consistent with the current best estimate of α.

Last but not least, the coefficient $1/4$ in (1) addresses the wave compositeness; an aspect of importance, or rather a possibility meaning the incompleteness of wave mechanics, quantum mechanics and field theory.

Submitted on July 5, 2016 / Accepted on July 7, 2016

References

1. De Broglie L. Recherches sur la théorie des quanta. *Annales de Physique* — 10e sérié — Tome III — Janvier-Février 1925.
2. Consiglio J. On quantization and the resonance paths. *Progress in Physics*, 2016, v. 12(3), 259–275.
3. Consiglio J. On the absorber in gravitation. *Progress in Physics*, 2016, v. 12(1), 20–25.
4. Cramer J. The transactional interpretation of quantum mechanics. *Rev. Mod. Phys.*, 1986, v. 58(3).
5. Wheeler J.A., Feynman R.P. Interaction with the absorber as the mechanism of radiation. *Reviews of Modern Physics*, 1945, v. 17(2–3), 157–161.
6. Wheeler J.A., Feynman R.P. Classical electrodynamics in terms of direct interparticle action. *Reviews of Modern Physics*, 1949, v. 21(3), 425–433.

Does the Velocity of Light Depend on the Source Movement?

Luis Bilbao
Universidad de Buenos Aires. Consejo Nacional de lnvestigaciones Científicas y Técnicas.
Instituto de Física del Plasma (INFIP). Facultad de Ciencias Exactas y Naturales. Buenos Aires, Argentina.
E-mail: bilbao@df.uba.ar

Data from spacecrafts tracking exhibit many anomalies that suggest the dependence of the speed of electromagnetic radiation with the motion of its source. This dependence is different from that predicted from emission theories that long ago have been demonstrated to be wrong. By relating the velocity of light and the corresponding Doppler effect with the velocity of the source at the time of detection, instead of the time of emission, it is possible to explain quantitatively and qualitatively the spacecraft anomalies. Also, a formulation of electromagnetism compatible with this conception is possible (and also compatible with the known electromagnetic phenomena). Under this theory the influence of the velocity of the source in the speed of light is somewhat subtle in many practical situations and probably went unnoticed (i.e. below the detection limit) in other measurements.

1 Introduction

In these lines I intend to show that there exists consistent evidence pointing to the need of revision and further study of what seem at present a settled issue, namely the independence of the speed of electromagnetic radiation on the motion of its source.

The main point in the evidence is the range disagreement during the Earth flyby of the spacecraft NEAR in 1998. Its range was measured near the point of closest approach using two radar stations, Millstone and Altair, of the Space Surveillance Network, and compared with the trajectory obtained from the Deep Space Network [1]. As for the range, the two measurements should match within a meter-level accuracy (the resolution is 5 m for Millstone and 25 m for Altair), but actual data showed a difference that varies linearly with time (with different slopes for the two radar stations) up to a maximum difference of about 1 km, i.e. more than 100 times larger than the accuracy of the equipment used (see figure 10 of [1]). Further, when NEAR crossed the orbits of Global Positioning System (GPS) satellites, orbital radius 26,600 km, the measured range difference was 650 m, that is, a time difference of 2 μs. Is it reasonable that any standard GPS receiver performs better than the Deep Space Network or the Space Surveillance Network?

There has not been a complete explanation for the range discrepancy. It is very difficult to find any physical reason that may produce this anomaly, for any physical disturbance of the path of the spacecraft should manifest equally in the Deep Space Network and the Space Surveillance Network data. Guruprasad [2] proposed an explanation that points to a time lag in the Deep Space Network signals proportional to the range, but the model is, at best, within 10% of the measured data (i.e. larger than the instrumental error) and, more important, it fails to explain an important feature, that is, the different slope for the two radars. If we assume that systems are working properly, then the measured range difference (time lag) could be due to different propagation time of the employed signals.

Additional points in the evidence come from anomalies related to the tracking of spacecrafts, present in both Doppler and ranging data. The Pioneer anomaly [3] and the flyby anomaly [4] refer to small residuals of the differences between measured and modeled Doppler frequencies of the radio signals emitted by the spacecrafts. Although these residuals are very small (less than 1 Hz on GHz signals) the problem is that they follow a non-random pattern, indicating failures of the model. According to the temporal variation of those residuals the Pioneer anomaly exhibits a main term, an annual term, a diurnal term and a term that appears during planetary encounters. It should be clarified that a few years ago an explanation of the Pioneer anomaly was published [5]. However, it is a very specific solution that applies only to the main term of the Pioneer spacecraft anomaly, but left unresolved many other anomalies, including those of the spaceships Cassini, Ulysses and Galileo; the annual term; the diurnal term; the increases of the anomaly during planetary encounters; the flyby anomaly; and the possible link between all them (it is hard to think that there are so many different causes for the mentioned anomalies). For all this, I believe that the issue can not be closed as it stands.

2 Range disagreement

As a matter of fact, the range difference between the Space Surveillance Network and the Deep Space Network, δR, is perfectly fitted with

$$\delta R(t) = -\frac{\mathbf{R}(t)\cdot\mathbf{v}(t)}{c}\,, \tag{1}$$

where $\mathbf{R}(t)$ is a vector range pointing from the spacecraft to the radar, $\mathbf{v}(t)$ the spacecraft velocity relative to the radar,

and c the speed of light. Figure 1 shows this fit and its comparison with measured data. The orbital and measured data were taken from [1]. Although the exact location of the radar stations are unknown to the author (approximate values are: Millstone 42.6° N 71.43° W, and Altair 9.18° N 167.42° E), the fit is statistically significant for both radar stations ($p < 10^{-3}$) including the first outliers points. It reproduces the (almost) linear dependence with time during the measured interval, and the two different slopes for Millstone and Altair stations due to their different locations.

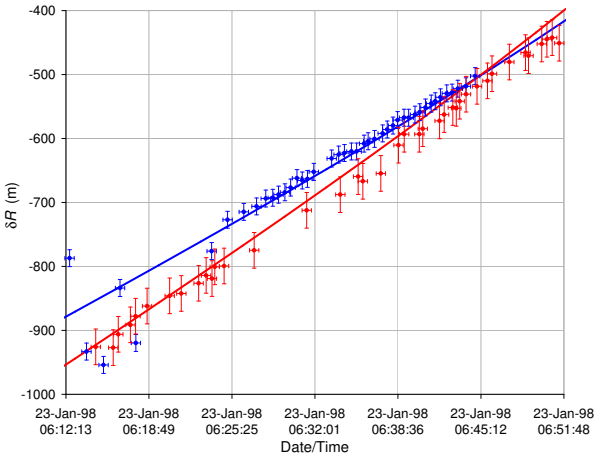

Fig. 1: Range disagreement between the Space Surveillance Network and the Deep Space Network, for 1998 NEAR flyby (Millstone blue points, upper trace, and Altair red points, lower trace). Also the fit (1) is plotted (full lines, Millstone in blue and Altair in red). For Millstone, the error bars refer to the uncertainties in the extraction of the data from figure 10 of [1], rather than to its tracking error (5 m), while for Altair, the accuracy is 25 m.

Since range measurements are based on time-of-flight techniques, the validity of (1) means that the electromagnetic waves (microwave) of the Deep Space Network and the Space Surveillance Network travel at different speeds. Specifically, in the radar frame of reference, if the Space Surveillance Network waves travel at c, then the Deep Space Network waves travel at c plus the projection of the spacecraft velocity in the direction of the beam, in sharp contrast with the Second Postulate of the Special Relativity Theory.

In view of the above result one may ask what is established, at present, about the relation of the speed of electromagnetic radiation (light for short) to the motion of the source. In order to elaborate this point the following questions are of relevance:

1. Are there *simultaneous* measurements of the speed of light from different moving macroscopic sources (not moving images) with different velocities?;
2. Since ballistic (emission) theories are ruled out (see, for example, DeSitter [6,7], Brecher [8] and Alväger et al [9]), how else could the speed of light depend on the source movement?;
3. How is it possible that there is a first order difference in v/c in spacecraft range measurements, while at the same time there are many experiments on time dilation that are consistent with Special Relativity Theory to second order in v/c (see, for example, [10])?;
4. If the velocity of light depend on the velocity of the source, why has this not been observed in other phenomena in the past?

In answer to the previous questions, so far as the author is aware, there is no known experimental work that simultaneously measures the speed of light from two different sources (not images), or that simultaneously measures the speed of light and that of its source. For example, in the work by Alväger et al, [9] the speed of light is measured at a later time (≈ 200 ns) than the emission time, and there is no measurement of the speed of the source at the time of the *detection* of the light.

Note that measurements involving moving images produce different results from those produced by mobile sources. For example, under Special Relativity Theory, a moving source is affected by time dilation while a moving image is not. Therefore, to ensure the independence of the speed of light from its source movement, it is essential to have two sources with different movements.

Although controversial and beyond the scope of the this note, time dilatation phenomena may be of different physical origin from first order terms, as it may be inferred from the work of Schrödinger [11]. Thus, measurements of time dilatation phenomena in accordance with Special Relativity Theory, does not necessarily imply the independence of the speed of light with the movement of the source.

The experiments mentioned above [6–9] only rule out ballistic theories in which radiation maintains the speed of the source at the time of *emission*, but do not rule out other ideas, like Faraday's 1846 [12].

3 Faraday's ray vibrations

In order to remove the ether, Faraday introduced the concept of vibrating rays [12], in which an electric charge is conceived as a center of force with attached "rays" that extend to infinity. The rays move with their center, but without rotating. According to this view, the phenomenon of electromagnetic radiation corresponds to the vibration of these "rays", that propagates at speed c relative to the rays (and the center). That is, the radiation remains linked to the source even after emitted. Today we could describe the interaction as a kind of entanglement between the charge and the photon. A framework for the electromagnetic phenomena according to Faraday's ideas was developed. It was called "Vibrating Rays Theory" [13] in reference to Faraday's "vibrating rays".

Under Faraday's idea, the velocity of radiation at a given epoch will be equal to c plus the velocity of the source at the *same* epoch, in contrast with ballistic theories in which

the emitted light retains the speed of the source at the *emission* epoch. In this sense the radiation is always linked to the charge at every time after the emission. Consequently, the measured Doppler Effect corresponds to the speed of the source at the time of *reception*, as well.

Further, a difference between active and passive reflection is expected, since the latter is still related to the original source according to Vibrating Rays Theory. The Deep Space Network works with the so called active reflection (the spacecraft re-emits in real time a signal in phase with the received signal from Earth), while the Space Surveillance Network works with passive radar reflection. In consequence, the down-link signal from the approaching spacecraft will propagate faster that the reflected one. Using the available orbital data [1] we found that, under Vibrating Rays Theory, the theoretical time-of-flight difference between active and passive reflection gives exactly the same range disagreement as (1), see Part 6 of [13].

4 Pioneer anomaly

The Pioneer anomaly refers to the fact that the received Doppler frequency differs from the modeled one by a blue shift that varies almost linearly with time, and whose derivative is

$$\frac{d(\Delta f)}{dt} \approx -6 \times 10^{-9}\ \text{Hz/s}, \tag{2}$$

where Δf is the frequency difference between the measured and the modeled values.

In the case of a source with variable speed, the main difference in Doppler (to first order) between Vibrating Rays Theory and Special Relativity Theory, is that Special Relativity Theory relates to the speed of the source at the time of *emission*, while Vibrating Rays Theory relates to the speed of the source at the time of *reception*. Precisely, this difference seems to be present in the spacecraft anomalies.

If Vibrating Rays Theory is valid, it automatically invalidates all calculations and data analysis of spacecraft tracking which are based on Special Relativity Theory. So, it is not easy to make a direct comparison between the expected results from Special Relativity Theory and Vibrating Rays Theory. However, to see whether or not the main features predicted by Vibrating Rays Theory are present in the measurements, we can evaluate the residual by simulating a measured Doppler signal assuming that light propagates in accordance to Vibrating Rays Theory but analyzed according to Special Relativity Theory.

Calling t_2 the emission time of the downlink signal from the spacecraft toward Earth and t_3 the reception time at Earth, the first order difference of the Doppler shift between Vibrating Rays Theory and Special Relativity Theory is (see [13] Part 4)

$$\Delta f = f_{VRT} - f_{SRT} \approx f_0 \hat{\mathbf{r}} \cdot \frac{\mathbf{v}_2 - \mathbf{v}_3}{c}, \tag{3}$$

where $\mathbf{v}_2$ and $\mathbf{v}_3$ represent the velocities of the spacecraft at the corresponding epoch, $\hat{\mathbf{r}}$ is the unit vector from the spaceship to the antenna, and f_0 the proper frequency of the signal. That is, the velocity used in the Special Relativity Theory formula is that at the time of *emission* while according to Vibrating Rays Theory is that corresponding at the time of *reception*.

Since the spacecraft slows down as it moves away, then $\hat{\mathbf{r}} \cdot (\mathbf{v}_2 - \mathbf{v}_3) > 0$, therefore the difference corresponds to a small blue shift mounted over the large red shift, as it has been observed in the Pioneer anomaly. It should be noted that this difference appears because of the active reflection produced by the on-board transmitter. In case of a passive reflection (for example, by means of a mirror) the above difference vanishes.

4.1 Main term

An estimate of the order of magnitude of 3 is obtained by using that the variation of the velocity of the spacecraft between the time of emission and reception is approximately

$$\mathbf{v}_2 - \mathbf{v}_3 \approx \mathbf{a}\,(t_2 - t_3), \tag{4}$$

where $\mathbf{a}$ is a mean acceleration during the down-link interval. An estimate for the duration of the down-link is simply

$$t_3 - t_2 \approx \frac{r}{c}, \tag{5}$$

where r is a mean position of the spaceship between t_2 and t_3, therefore

$$\Delta f \approx -f_0 \frac{\mathbf{r} \cdot \mathbf{a}}{c^2}.$$

Since

$$\mathbf{a} = -\frac{GM}{r^2}\hat{\mathbf{r}},$$

where G is the gravitational constant, and M the mass of the Sun, then, the time derivative becomes

$$\frac{d\,(\Delta f)}{dt} \approx f_0 \frac{\mathbf{v} \cdot \mathbf{a}}{c^2}. \tag{6}$$

If the difference (6) is interpreted as an anomalous acceleration we get

$$a_a \approx \frac{v}{c}\, a, \tag{7}$$

that is, the so-called anomalous acceleration is v/c times the actual acceleration of the spacecraft.

Using data from HORIZONS Web-Interface [14] for the spacecraft ephemeris, some characteristic value for a_a can be obtained. Consider the anomalous acceleration detected at the shortest distance of the Cassini spacecraft during solar conjunction in June, 2002. The spacecraft was at a distance of 7.42 AU moving at a speed of 5.76 km/s. The anomalous acceleration given by (7) is $a_a \approx 2 \times 10^{-9}$ m/s^2 of the same order of the measured one ($\approx 2.7 \times 10^{-9}$ m/s^2). Also, the closest distance at which the Pioneer anomaly has been detected was

about 20 AU. the anomalous acceleration predicted by (7) at that distance is $a_a \approx 7.3 \times 10^{-10}$ m/s^2 of the same order as the measured one.

The "anomaly" given by (7) decreases in time in a way that has not been observed. Note, however, that according to Markwardt [15] the expected frequency at the receiver includes an additional Doppler effect caused by small effective path length changes, given by

$$\Delta f_{path} = -\frac{2f_0}{c}\frac{dl}{dt}, \tag{8}$$

where dl/dt is the rate of change of effective photon trajectory path length along the line of sight. This is a first order effect that can partially hide the difference between Special Relativity Theory and Vibrating Rays Theory. Therefore, a more careful analysis should take into account the additional contribution of (8) in (7).

Further, other first order effects may appear, for example, by a slight rotation of the orbital plane. Due to spacecraft maneuvers or random perturbations the orbital parameters are obtained by periodically fitting the measurements with theoretical orbits. Therefore there is no straightforward way to weight the importance of these fittings in (7). In other words, data acquisition and analysis may hide part of the Vibrating Rays Theory signature.

4.2 Annual term

Apart from the residual referred to in the preceding paragraph there is also an annual term. According to Anderson et al [16] the problem is due to modeling errors of the parameters that determine the spacecraft orientation with respect to the reference system. Anyway, Levy et al [17] claim that errors such as errors in the Earth ephemeris, the orientation of the Earth spin axis or the stations coordinates are strongly constrained by other observational methods and it seems difficult to modify them sufficiently to explain the periodic anomaly.

The advantage of studying the annual term over the main term, is that the former is less sensitive to the first order correction mentioned above, and, for the case of Pioneer, also to the thermal propulsion correction [5]. Clearly, the Earth orbital position does not modify those terms.

As before, the annual term is explained by the difference between the velocity of the spacecraft at the time of emission and that at the moment of detection, which depends on whether the spaceship is in opposition or in conjunction relative to the Sun. When the spacecraft is in conjunction, light takes longer to get back to Earth than in opposition. The time difference between emission and reception will be increased by the time the light takes in crossing the Earth orbit. Specifically, taking into account the delay due to the position of Earth in its orbit, in opposition equation (5) should be written as

$$t_3 - t_2 \approx \frac{r + R_{orb}}{c}, \tag{9}$$

while in conjunction it would be

$$t_3 - t_2 \approx \frac{r - R_{orb}}{c}, \tag{10}$$

where R_{orb} is the mean orbital radius of Earth.

Therefore, an estimate of the magnitude of the amplitude of the annual term is

$$\Delta f \approx f_0 \frac{aR_{orb}}{c^2}. \tag{11}$$

For the case of Pioneer 10 at 40 AU we get

$$\Delta f \approx 14 \text{ mHz}, \tag{12}$$

and at 69 AU

$$\Delta f \approx 4.8 \text{ mHz}, \tag{13}$$

in good agreement with the observed values.

Using data from HORIZONS Web-Interface [14] a more complete analysis of the time variation of Δf has be performed. The residual (that is, simulated Doppler using Vibrating Rays Theory but interpreted under Special Relativity Theory) during 12 years time span is plotted in figure 2. Also the dumped sine best fit of the 50 days average measured by Turyshev et al [18] is plotted showing an excellent agreement between measurements and Vibrating Rays Theory prediction. The negative peaks (i.e., maximum anomalous acceleration) occur during conjunction when the Earth is further apart from the spacecraft, and positive peaks during opposition. Also, the amplitude is larger at the beginning of the plotted interval and decreases with time, as it was observed [4, 18].

5 Flyby anomaly

Like the Pioneer anomaly, the Earth flyby anomaly can be associated to a modeling problem, in the sense that relativistic Doppler includes terms that are absent in the measured signals. The empirical equation of the flyby anomaly is given by Anderson et al [4], which, notably, can be derived using Vibrating Rays Theory, as is done in Part 6 of [13].

Consider the case of NEAR tracked by 3 antennas located in USA, Spain, and Australia (a full description of the tracking system is found in a series of monographs of the Jet Propulsion Laboratory [19]). The receiving antenna was chosen as that having a minimum angle between the spacecraft and the local zenith.

Using available orbital data, a simulated Doppler signal has been calculated using Vibrating Rays Theory. Thus, the simulated residual is obtained by subtracting the theoretical Special Relativity Theory Doppler, from the Vibrating Rays Theory calculation. We observed, however, that the term that contains the velocity of the antennas, that is

$$d = \frac{\gamma_{u_3}}{\gamma_{u_1}} \frac{1 - \hat{\mathbf{r}}_{23} \cdot \mathbf{u}_3/c}{1 - \hat{\mathbf{r}}_{12} \cdot \mathbf{u}_1/c}, \tag{14}$$

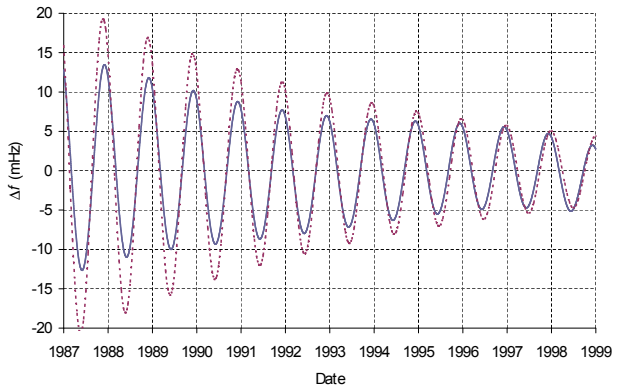

Fig. 2: Annual variation of the frequency difference between Vibrating Rays Theory and Special Relativity Theory (full line) and anomalous dumped sine best fit of the 50 days average measured by Turyshev et al [18] (dashed line), for Pioneer 10 from January 1987 to January 1999.

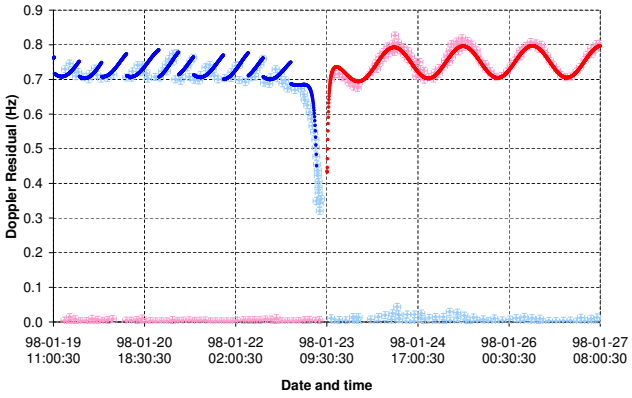

Fig. 3: Fitting the pre- (right half, in red) and post-encounter (left half, in blue) X-band Doppler data residual, for the NEAR flyby under an ideal hyperbolic orbit. Solid lines simulated according to Vibrating Rays Theory. Crosses, actual data extracted from reference [4].

is not enough to completely remove the first order (in u/c) Earth signature (**u** is the velocity of the antenna, 1 refers to the emission epoch and 3 to the reception epoch, as in [13] Part 4).

This is so because the velocity of the antennas is not uniform and the evaluation of the emission time is different for Vibrating Rays Theory and Special Relativity Theory. Then, a small first order term remains. Anyway, since orbital parameters are obtained by periodically fitting the measurements to theoretical orbits, thus a similar procedure is needed for Vibrating Rays Theory. Curiously, by doing so, the first order term is removed. The only difference between orbits adjusted by Special Relativity Theory and Vibrating Rays Theory is a slight rotation of the orbit plane, as mentioned above. Note that in the case of range disagreement (discussed above) two different orbital adjustment would be needed by the Deep Space Network and the Space Surveillance Network due to the different propagation speed. In consequence, it will be impossible to fit a simultaneous measurement, as it seems to happen with the range disagreement.

The final result shows that each antenna produces a sinusoidal residual with a phase shift at the moment of maximum approach. Therefore, if we fit the data with the pre-encounter sinusoid a post-encounter residual remains and vice versa.

In figure 3 are simultaneously plotted the result of fitting the residual by pre-encounter data (right half in red, corresponding to figure 2a of [4]) and by post-encounter data (left half in blue, corresponding to figure 2b of [4]).

Note that the simulated plots are remarkably similar to the reported ones, including the amplitude and phase (i.e., minima and maxima) of the corresponding antenna. The fitting of post-encounter data (blue) can be improved by appropriately setting the exact switching times of the antennas (which are unknown to the author). The flyby Doppler residual exhibits a clean signature of the Vibrating Rays Theory.

6 Conclusions

In this work I have presented observational evidence favoring a dependence of the speed of light on that of the source, in the manner implied in Faraday's ideas of "vibrating rays".

It is remarkable and very suggestive that, as derived from Faraday's thoughts, simply by relating the velocity of light and the corresponding Doppler effect with the velocity of the source at the time of detection, is enough to quantitatively and qualitatively explain a variety of spacecraft anomalies.

Also, it is worth mentioning that a formulation of electromagnetism compatible with Faraday's conception is possible, as shown in [13] Part 8, which is also compatible with the known electromagnetic phenomena. The most remarkable fact of this new formalism is the simultaneous presence of instantaneous (static terms) and delayed (radiative terms) interactions (i.e., local and nonlocal phenomena in the same interaction).

Finally, under Vibrating Rays Theory the manifestation of the movement of the source in the speed of light is more subtle than the naive $c + kv$ hypothesis (k is a constant, $0 \leq k \leq 1$) usually used to test their dependence [8]. Thus, it is also of fundamental importance the fact that, from the experimental point of view, it is very difficult to detect differences between Vibrating Rays Theory and Special Relativity Theory, as discussed in [13], which is also manifest in the smallness of the measured anomalies, and in the non clear manifestation of the effect in usual experiments and observations. For example, it produces a negligible effect on satellite positioning systems, see Part 7 of [13].

I am aware of how counterintuitive these conceptions are to the modern scientist, but also believe that, given the above evidence, a conscientious experimental research is needed to settle the question of the dependence of the speed of light on that of its source as predicted by Vibrating Rays Theory, and

that has been observed during the 1998 NEAR flyby. As a closure, I recall Fox's words regarding the possibility of conducting an experiment on the propagation of light relative to the motion of the source: "*Nevertheless if one balances the overwhelming odds against such an experiment yielding anything new against the overwhelming importance of the point to be tested, he may conclude that the experiment should be performed*" [20].

Acknowledgements

I am thankful to Fernando Minotti who read this paper and improved the manuscript significantly, although he may not agree with all of the interpretations provided in this paper.

Submitted on July 4, 2016 / Accepted on July 8, 2016

References

1. Antreasian P.G., Guinn J. R. Investigations into the unexpected Delta-V increases during the Earth gravity assists of Galileo and NEAR. AIAA Paper No. 98-4287 presented at the AIAA/AAS Astrodynamics Specialist Conference and Exhibit (Boston, August 10-12, 1998).
2. Guruprasad V. Observational evidence for travelling wave modes bearing distance proportional shifts. *E*PL, 2015, v. 110, 54001.
3. Anderson J.D., Laing P.A., Lau E.L., Liu A.S., Nieto M.M., Turyshev S.G. Indication, from Pioneer 10/11, Galileo, and Ulysses Data, of an Apparent Anomalous, Weak, Long-Range Acceleration. *Phys. Rev. Lett.*, 1998, v. 81, 2858–2861.
4. Anderson J.D., Campbell J.K., Ekelund J.E., Ellis J., Jordan J.F. Anomalous Orbital-Energy Changes Observed during Spacecraft Flybys of Earth. *Phys. Rev. Lett.*, 2008, v. 100, 091102.
5. Turyshev S.G., Toth V.T., Kinsella G., Lee S.-C., Lok S.M., Ellis J. Support for the Thermal Origin of the Pioneer Anomaly. *Phys. Rev. Lett*, 2012, v. 108, 241101.
6. de Sitter W. Ein astronomischer Beweis für die Konstanz der Lichgeshwindigkeit. *Z. Phys.*, 1913, v. 14, 429.
7. de Sitter W. Über die Genauigkeit, innerhalb welcher die Unabhängigkeit der Lichtgeschwindigkeit von der Bewegung der Quelle behauptet werden kann. *Z. Phys.*, 1913, v. 14, 1267.
8. Brecher K. Is the Speed of Light Independent of the Velocity of the Source? *Phys. Rev. Lett.*, 1977, v. 39, 1051.
9. Alväger T., Farley F.J.M., Kjellman J., Wallin I. Test of the second postulate of special relativity in the GeV region. *Phys. Lett.*, 1964, v. 12, 260–262.
10. Botermann B., Bing D., Geppert C., et al. Test of Time Dilation Using Stored Li+ Ions as Clocks at Relativistic Speed. *Phys. Rev. Lett.*, 2014, v. 113, 120405.
11. Schrödinger E. Die erfüllbarkeit der relativitätsforderung in der klassischen mechanik. *Ann. der Physik*, 1925, v. 77, 325–336.
12. Faraday M. Thoughts on ray-vibrations. *Phil. Mag. Series 3*, 1846, v. 28, 345.
13. Bilbao L., Bernal L., Minotti F. Vibrating Rays Theory. arXiv: 1407.5001.
14. http://ssd.jpl.nasa.gov
15. Markwardt C.B. Independent Confirmation of the Pioneer 10 Anomalous Acceleration. arXiv: gr-qc/0208046.
16. Anderson J.D., Laing P.A., Lau E.L., Liu A.S., Nieto M.M., Turyshev S.G. Study of the anomalous acceleration of Pioneer 10 and 11. *Phys. Rev. D*, 2002, v. 65, 082004.
17. Levy A., Christophe B., Reynaud S., Courty J-M., Brio P., Mtris G. Pioneer 10 data analysis: investigation on periodic anomalies. Journées scientifiques de la SF2A, Paris, France, 2008, 133–136 (hal-00417743).
18. Turyshev S.G., Anderson J.D., Laing P.A., Lau E.L., Liu A.S., Nieto M.M. The Apparent Anomalous, Weak, Long-Range Acceleration of Pioneer 10 and 11. arXiv: gr-qc/9903024.
19. DESCANSOTeam, Jet Propulsion Laboratory, California Institute of Technology. http://descanso.jpl.nasa.gov/Monograph/mono.cfm (accessed July 2014).
20. Fox J.G. Experimental Evidence for the Second Postulate of Special Relativity. *Am. J. Phys.*, 1962, v. 30, 297.

Editorial Comment

This paper plays an importance in the understanding of the physical observable velocity of light that differs from the world-invariant in the General Theory of Relativity.

Defining physical observable quantities in the General Theory of Relativity is not a trivial problem. This is because we are looking at objects in a four-dimensional space-time, and we have to determine which components of these four-dimensional tensor quantities are physically observable. A complete mathematical theory for calculating physically observable quantities in the four-dimensional space (space-time) of General Relativity was introduced in 1944 by Abraham Zelmanov, and is known as the *theory of chronometric invariants*[*]. Landau and Lifshitz in §84 of their *The Classical Theory of Fields* also introduced physically observable time and observable three-dimensional intervals similar to Zelmanov. But they limited themselves only to this particular case, while only Zelmanov arrived at the versatile mathematical theory. A compendium of Zelmanov's theory of physical observable quantities can also be found in the books[†].

In short, physically observable are the projections of four-dimensional quantities onto the time line and the three-dimensional spatial section of the observer, which can be non-uniform, deformed, curved and rotating. These projections are calculated through the special projecting operators which take all the aforementioned factors into account. In particular, the physical observable velocity of light differs from the world-invariant, and is depended on the gravitational potential and the rotation velocity of the observer's space. In ultimate physical conditions, as is shown in Chapter 5 of *Particles Here and Beyond the Mirror*[†], the observable velocity of light can even become zero, that is verified by the frozen light experiment (Lene Hau, 2001).

Even more. In a physical space (space-time metric) wherein is a shift at one of the spatial directions (that means a spatial anisotropy), the observable velocity of light is depended on the signal source's velocity at this preferred direction. We drafted such a space-time metric in the last decade.

Einstein's postulates have now only a historical meaning. Once Einstein moved his theory on the mathematical basis of Riemannian geometry, he found that all the postulates are the manifestations of geometry of Riemannian spaces. It is as well true about the world-invariant of the velocity of light. In a space, which is free of gravitation, is uniform, non-deformed, and non-rotating, the physical observable velocity of light coincides with the world-invariant. However in a real physical space it does not.

For this reason the experimental compendium and the analysis presented in Bilbao's paper will maybe give a new fresh stream in search for the further theoretical predictions of the General Theory of Relativity.

Dmitri Rabounski, Editor-in-Chief
Larissa Borissova, Assoc. Editor

[*]Zelmanov A. Chronometric invariants. American Research Press, Rehoboth (NM), 2006. Zelmanov A. Chronometric invariants and accompanying frames of reference in the General Theory of Relativity. *Soviet Physics Doklady*, 1956, v. 1, 227–230.

[†]Borissova L. and Rabounski D. Fields, Vacuum, and the Mirror Universe. 2nd ed., Svenska fysikarkivet, Stockholm, 2009. Rabounski D. and Borissova L. Particles Here and Beyond the Mirror. 3rd ed., American Research Press, Rehoboth (NM), 2012.

Vacuum Background Field in General Relativity

Patrick Marquet

18 avenue du Président Wilson, 62100 Calais, France
E-mail: patrick.marquet6@wanadoo.fr

We assume here a slightly varying cosmological term which readily induces a permanent background field filling the physical vacuum. A precise form of the variable cosmological term is introduced containing an infinitesimal Killing vector which accounts for the space-time variation of this term. As a result the term can be added to the Einstein Lagrangian without affecting the varied action δS. As a result, we showed in an earlier publications that the permanent background field filling the vaccum is excited in the vicinity of matter which precisely corresponds to its gravitational field classically described by a pseudo-tensor. With this preparation, the global energy-momentum tensor of matter and gravity field is no longer a pseudo-tensor and is formally conserved like the Einstein tensor. In the excited state, this antisymmetric tensor can be conveniently symmetrized by applying the Belinfante procedure which automatically self excludes far from matter since the background field tensor is naturally symmetric.

Introduction

The substance of this study is inspired by the following considerations. In the framework of the Theory of General Relativity (GR), the Einstein tensor exhibits a *conceptually* conserved property, while any corresponding stress-energy tensor does not, which leaves the theory with a major inconsistency. When pure matter is the source, a so-called "pseudo-tensor" describing its gravitational field is introduced so that the four-momentum of both matter and its gravity field is conserved [1]. Unfortunately in this approach, the gravitational field maybe transformed away at any point and by essence, its pseudo-tensor cannot appear in the Einstein's field equations, as it should be.

We will tackle the problem in another way : Restricting our study to neutral massive flow, we proceed as follows. We introduce a space-time variable term that supersedes the so-called cosmological term Λg_{ab} in the Einstein's field equations [2]. Under this latter assumption, we formally show that the gravity field of a massive source is no longer described by a vanishing *pseudo-tensor*, but it is represented by a *true* tensor which can explicitly appear with the bare matter tensor together with another specific field, on the right hand side of the Einstein's field equations. Inspection also shows that this *global* stress-energy tensor now complies with the intrinsic conservation property of the Einstein tensor as it should be. As a result, the physical vacuum is here filled with a *homogeneous vacuum background field* which is always present in the so-called Einstein's "source free" equations and whose tensor exhibits a conserved property. Our theory leads to admit that matter causes the surrounding background field to produce its gravitational field which decreases asymptotically to the level of this vacuum field. Naturally, since we will deal with energy-momentum canonical field tensors which are not symmetric, the total angular momentum of the isolated system is not conserved. In this case, it is always possible to apply the symmetrizing procedure to these tensors according to J. Belinfante [3]. In the absence of matter, the inferred Belinfante tensor reduces to the symmetric background field tensor as it should be.

Notations

Space-time Latin indices run from $a = b$: 0, 1, 2, 3, while spatial Greek indices run from $\alpha = \beta$: 1, 2, 3. The space-time signature is -2. In the present text, $\varkappa$ is Einstein's constant $4\pi G/c^4$, where G is Newton's gravitational constant.

1 The field equations in General Relativity

1.1 The problem of the conserved gravity tensor

The General Theory of Relativity requires a 4-dimensional pseudo-Riemannian manifold. A Riemannian manifold is characterized by the line element $ds^2 = g_{ab}\, dx^a dx^b$. It is well known that by varying the action $\mathcal{S} = \mathfrak{L}_E d^4x$ with respect to the metric tensor g_{ab} with the Lagrangian density given by

$$\mathfrak{L}_E = \sqrt{-g}\, g_{ab}\left[\left\{{e \atop ab}\right\}\left\{{d \atop de}\right\} - \left\{{d \atop ae}\right\}\left\{{e \atop bd}\right\}\right], \tag{1.1}$$

$$g = \det \| g_{ab} \|. \tag{1.2}$$

Also one infers the symmetric Einstein tensor

$$G_{ab} = R_{ab} - \frac{1}{2}\, g_{ab}\, R\,, \tag{1.3}$$

where

$$R_{bc} = \partial_a \left\{{a \atop bc}\right\} - \partial_c \left\{{a \atop ba}\right\} - \left\{{d \atop bc}\right\}\left\{{a \atop da}\right\} - \left\{{d \atop ba}\right\}\left\{{a \atop dc}\right\} \tag{1.4}$$

is the Ricci tensor with its contraction R, the curvature scalar (the $\left\{{e \atop ab}\right\}$ denote the Christoffel symbols of the second kind). The 10 source free field equations are

$$G_{ab} = 0\,. \tag{1.5}$$

The second rank Einstein tensor G_{ab} is symmetric and is only function of the metric tensor components g_{ab} and their first and second order derivatives. The relations

$$\nabla_a G^a_b = 0 \tag{1.6}$$

are the conservation identities provided that the tensor G_{ab} has the form [4]

$$G_{ab} = k\left[R_{ab} - \frac{1}{2}\, g_{ab}\,(R + 2\Lambda)\right], \tag{1.7}$$

where k is a constant, which is here assumed to be 1, while Λ is usually named the *cosmological constant*.

Einstein's field equations for a source free field are

$$G_{ab} = R_{ab} - \frac{1}{2}\, g_{ab}\, R - \Lambda g_{ab} = 0\,. \tag{1.8}$$

In the case where the field source is present, the field equations become

$$G_{ab} = R_{ab} - \frac{1}{2}\, g_{ab}\, R - \Lambda g_{ab} = \varkappa\, T_{ab}\,, \tag{1.8 bis}$$

where T_{ab} is the energy-momentum tensor of the source.

However, unlike the Einstein tensor G_{ab} which is *conceptually conserved*, the conditions

$$\nabla_a T^a_b = 0 \tag{1.9}$$

are never satisfied in a general coordinates system [5]. Therefore, the Einstein tensor G_{ab} which *intrinsically* obeys a conservation condition inferred from the Bianchi's identities, is generally related with a tensor T_{ab} which obviously *fails to satisfy the same requirement*.

Hence, we are faced here with a major inconsistency in GR which can be removed in the case of a neutral massive source upon a small constraint.

1.2 The tensor density representation

We first set

$$\mathfrak{g}^{ab} = \sqrt{-g}\, g^{ab} \tag{1.10}$$

thus the Einstein tensor density is

$$\mathfrak{G}^{ab} = \sqrt{-g}\, G^{ab}\,, \tag{1.10 bis}$$

$$\mathfrak{G}^c_a = \sqrt{-g}\, G^c_a\,, \tag{1.10 ter}$$

$$\mathfrak{R}^{ab} = \sqrt{-g}\, R^{ab}\,. \tag{1.11}$$

In the density notations, the field equations with a source (1.8) will read

$$\mathfrak{G}^{ab} = \mathfrak{R}^{ab} - \frac{1}{2}\, g^{ab}\, \mathfrak{R} - \sqrt{-g}\, g^{ab}\, \Lambda = \varkappa\, \mathfrak{T}^{ab}\,, \tag{1.12}$$

where $\mathfrak{T}^{ab} = \sqrt{-g}\, T^{ab}$.

2 The new approach on gravity

2.1 The canonical gravity pseudo-tensor

Let us consider the energy momentum tensor for neutral matter density ρ

$$T_{ab} = \rho\, c^2 u_a u_b \tag{2.1}$$

as the right hand side of the standard field equations

$$G_{ab} = R_{ab} - \frac{1}{2}\, g_{ab}\, R = \varkappa\, T_{ab}\,. \tag{2.2}$$

The conservation condition for this tensor are written

$$\nabla_a T^a_b = \sqrt{-g}\, \partial_a T^a_b - \frac{1}{2}\, T_{ac} \partial_b\, g_{ac} = 0 \tag{2.3}$$

with the tensor density

$$\mathfrak{T}^a_b = \sqrt{-g}\, T^a_b\,. \tag{2.4}$$

However, across a given hypersurface dS_b, the integral

$$P^a = \frac{1}{c} \int T^{ab} \sqrt{-g}\, dS_b \tag{2.5}$$

is conserved only if [6]

$$\partial_a\, \mathfrak{T}^a_b = 0\,. \tag{2.6}$$

This problem can be cured only if the metric admits a Killing vector field [7]. If this is not so, we write (2.3) for the *bare* matter tensor density

$$\partial_a (\mathfrak{T}^a_b)_{\text{matter}} = \frac{1}{2}\, (\mathfrak{T}^{cd})_{\text{matter}}\, \partial_b\, g_{cd}\,. \tag{2.7}$$

Inspection then shows that

$$R_{il}\, d\mathfrak{g}^{il} = \sqrt{-g}\left[-R^{ie} + \frac{1}{2}\, g^{ie} R\right] dg_{ie} = \\ = -\varkappa\, (\mathfrak{T}^{ie})_{\text{matter}}\, dg_{ie}\,. \tag{2.8}$$

Taking now into account the Lagrangian formulation for R_{il} which is

$$R_{il} = \frac{d\mathfrak{Q}_E}{\mathfrak{g}^{il}} = \partial_k \left[\frac{\partial \mathfrak{Q}_E}{\partial(\partial_k\, \mathfrak{g}^{il})}\right] - \frac{d\mathfrak{Q}_E}{\partial\, \mathfrak{g}^{il}}\,, \tag{2.9}$$

we obtain

$$-\varkappa\, (\mathfrak{T}^{il})_{\text{matter}}\, dg_{il} = \left\{\partial_k \left[\frac{\partial \mathfrak{Q}_E}{\partial\, (\partial_k\, \mathfrak{g}^{il})}\right] - \frac{\partial \mathfrak{Q}_E}{\partial \mathfrak{g}^{il}}\right\} d\mathfrak{g}^{il} = \\ = \partial_k \left[\frac{\partial \mathfrak{Q}_E\, d\mathfrak{g}^{il}}{\partial\, (\partial_k\, \mathfrak{g}^{il})}\right] - d\mathfrak{Q}_E$$

or

$$-\varkappa\, (\mathfrak{T}^{il})_{\text{matter}}\, \partial_m\, g_{il} = \partial_k \left[\frac{\partial \mathfrak{Q}_E\, \partial_m(\partial \mathfrak{g}^{il})}{\partial(\partial_k\, \mathfrak{g}^{il})} - \delta^k_m\, \mathfrak{Q}_E\right] = \\ = 2\varkappa\, \partial_k (\mathfrak{t}^k_m)_{\text{field}}\,, \tag{2.10}$$

where $(\mathfrak{t}^k_m)_{\text{field}}$ denotes the field tensor density extracted from

$$2\varkappa\,(\mathfrak{t}^k_m)_{\text{field}} = \frac{\partial \mathfrak{Q}_E\,\partial_m(\partial \mathfrak{g}^{il})}{\partial(\partial_k\,\mathfrak{g}^{il})} - \delta^k_m\,\mathfrak{Q}_E \tag{2.11}$$

so that we have the explicit canonical form

$$(\mathfrak{t}^k_m)_{\text{field}} = \frac{1}{2\varkappa}\left\{\frac{\partial \mathfrak{Q}_E\,\partial_m(\partial \mathfrak{g}^{il})}{\partial(\partial_k\,\mathfrak{g}^{il})} - \delta^k_m\,\mathfrak{Q}_E\right\} \tag{2.12}$$

where

$$\partial_k(\mathfrak{T}^k_i)_{\text{matter}} = \frac{1}{2}\,(\mathfrak{T}^{ek})_{\text{matter}}\,\partial_k\,g_{ei} = -\partial_k(\mathfrak{t}^k_i)_{\text{field}}$$

that is, the required conservation relation is

$$\partial_k\left[(\mathfrak{T}^k_i)_{\text{matter}} + (\mathfrak{t}^k_i)_{\text{field}}\right] = 0\,. \tag{2.13}$$

Looking back of the deduction, (2.12) defines the *canonical gravity pseudo-tensor density of matter*

$$(\mathfrak{t}^k_m)_{\text{pseudogravity}} = \frac{1}{2\varkappa}\left\{\frac{\partial \mathfrak{Q}_E\,\partial_m(\partial \mathfrak{g}^{il})}{\partial(\partial_k\,\mathfrak{g}^{il})} - \delta^k_m\,\mathfrak{Q}_E\right\}. \tag{2.14}$$

Expressed with the explicit form of the Lagrangian density $\mathfrak{Q}_E$ (1.1), (2.14) can be written in the form

$$(\mathfrak{t}^k_m)_{\text{pseudogravity}} = \\ = \frac{1}{2\varkappa}\left(\left\{{k \atop il}\right\}\partial_m\,\mathfrak{g}^{il} - \left\{{i \atop il}\right\}\partial_m\,\mathfrak{g}^{lk} - \delta^k_m\,\mathfrak{Q}_E\right). \tag{2.15}$$

This is the *mixed Einstein-Dirac pseudo-tensor density* [8] which is not symmetric on k and m, and therefore is not suitable for basing a definition of angular momentum on.

Thus, our aim is to look for:

- A *true* tensor;
- A *symmetric* tensor.

2.2 The new canonical tensor

In the density notations, the field equations with a massive source (1.8 bis) can be re-written as

$$\mathfrak{G}^{ab} = \mathfrak{R}^{ab} - \frac{1}{2}\,\mathfrak{g}^{ab}\,\mathfrak{R} - \mathfrak{g}^{ab}\,\zeta = \varkappa\,(\mathfrak{T}^{ab})_{\text{matter}}\,, \tag{2.16}$$

where in place of the constant cosmological term $\Lambda\sqrt{-g}$, we have introduced a *scalar density* denoted as

$$\zeta = \Xi\sqrt{-g}\,. \tag{2.17}$$

Unlike Λ, the scalar Ξ is slightly space-time variable and can be regarded as a Lagrangian characterizing a *specific vacuum background field*.

We will choose the variation of Ξ as follows

$$\Xi = \nabla_a\,\kappa^a, \tag{2.17 bis}$$

where κ^a is a Killing vector. Hence

$$\zeta = \sqrt{-g}\,\nabla_a\,\kappa^a. \tag{2.17 ter}$$

We will first write the field equations with a massive source together with its gravity tensor density

$$\mathfrak{G}^{ab} = \mathfrak{R}^{ab} - \frac{1}{2}\,g^{ab}\,\mathfrak{R} = \varkappa\left[(\mathfrak{T}^{ab})_{\text{matter}} + (\mathfrak{t}^{ab})_{\text{gravity}}\right] \tag{2.18}$$

where $(\mathfrak{t}^{ab})_{\text{gravity}}$ is related to ζ as

$$\mathfrak{G}^{ab} = \mathfrak{R}^{ab} - \frac{1}{2}\,g^{ab}\,\mathfrak{R} = \varkappa\left[(\mathfrak{T}^{ab})_{\text{matter}} + \frac{g^{ab}\zeta}{2\varkappa}\right]. \tag{2.19}$$

Re-instating the term ζ accordingly, the gravitational field tensor density now reads

$$(\mathfrak{t}^k_m)_{\text{gravity}} = \frac{1}{2\varkappa}\left\{\frac{\partial \mathfrak{Q}_E\,\partial_m(\partial \mathfrak{g}^{il})}{\partial(\partial_k\,\mathfrak{g}^{il})} - \delta^k_m\,(\mathfrak{Q}_E - \zeta)\right\}. \tag{2.20}$$

A first inspection shows that ζ represents the Lagrangian density of the background field, therefore the modified field equations (2.19) should be derived from an Einstein Lagrangian density different from $\mathfrak{Q}_E$ (1.1) and which includes ζ.

By choosing the form (2.17 ter), we check that

$$\zeta = \sqrt{-g}\,\nabla_a\,\kappa^a = \partial_a\left(\sqrt{-g}\,\kappa^a\right).$$

Now, if we write the new action as

$$\mathcal{S}_M = \int \mathfrak{Q}_M\,d^4x = \int \mathfrak{Q}_E\,d^4x + \int \partial_a\left(\sqrt{-g}\,\kappa^a\right)d^4x$$

due to Gauss' theorem we see that the last integral can be transformed in an integral extended to an hyperfurface which does not contribute in the variation of $\mathcal{S}_M$ and

$$\delta\int \mathfrak{Q}_M\,d^4x = \delta\int \mathfrak{Q}_E\,d^4x\,.$$

Therefore, it is legitimate to maintain $(\mathfrak{t}^k_m)_{\text{gravity}}$ as per (2.20).

The presence of the scalar density ζ characterizing the background field is here of central importance, as it means that $(\mathfrak{t}^k_m)_{\text{gravity}}$ can never be zero in contrast to the classical theory where the gravitational field is only described by an awkward *pseudo-tensor*.

The quantity $(\mathfrak{t}^k_m)_{\text{gravity}}$ constitutes thus a *true* tensor density describing the gravity field attached to the neighbouring matter.

It is then easy to show that we have the conserved quantity

$$\partial_a\left[(\mathfrak{T}^b_a)_{\text{matter}} + (\mathfrak{t}^b_a)_{\text{gravity}}\right] = 0\,. \tag{2.21}$$

In this picture and examining (2.20), we clearly see that the gravitational field of matter appears as an *excited state* of the homogeneous background energy field which permanently fills the physical vacuum.

Far from its matter source, the field sharply decreases down to the level of the background field described by the tensor density $(\mathfrak{t}^{ab})_{\text{background field}}$. Therefore the "source free" field equations should always retain a non-zero right hand side according to

$$\mathfrak{G}^{ab} = \mathfrak{R}^{ab} - \frac{1}{2}\, g^{ab}\, \mathfrak{R} = \varkappa\, (\mathfrak{t}^{ab})_{\text{background field}} \qquad (2.22)$$

which are the equivalent of (1.8)

$$\mathfrak{G}^{ab} = \mathfrak{R}^{ab} - \frac{1}{2}\, g^{ab}\, \mathfrak{R} = \varkappa\, \frac{g^{ab}\,\zeta}{2\varkappa}\,. \qquad (2.23)$$

In this case, the conservation law applied to the right hand side of the tensor density field equations is straightforward

$$\partial_a(\mathfrak{t}^b_a)_{\text{background field}} = \partial_a\left(\frac{\zeta}{2\varkappa}\,\delta^b_a\right) = 0\,. \qquad (2.24)$$

2.3 Symmetrization of the gravity tensor

Let us consider the new gravity tensor expressed with the explicit form of the Lagrangian density $\mathfrak{L}_E$ (1.1):

$$(\mathfrak{t}^k_m)_{\text{gravity}} = \\ = \frac{1}{2\varkappa}\left[\left\{{k \atop il}\right\}\partial_m\mathfrak{g}^{il} - \left\{{i \atop il}\right\}\partial_m\mathfrak{g}^{lk} - \delta^k_m\,(\mathfrak{L}_E - \zeta)\right]. \qquad (2.25)$$

Like we mentioned, this tensor includes the Einstein-Dirac pseudo-tensor which is not symmetric. We can however follow the *Belinfante procedure* used to symmetrize the canonical tensor $(\Theta^k_m)_{\text{gravity}}$ that extracted from $(\mathfrak{t}^k_m)_{\text{gravity}} = \sqrt{-g}\,(\Theta^k_m)_{\text{gravity}}$.

The *total angular momentum* is known to be the sum

$$M^{cba} = x^b(\Theta^{ca})_{\text{gravity}} - x^a(\Theta^{cb})_{\text{gravity}} + S^{cab}, \qquad (2.26)$$

where S^{cab} is the contribution of the *intrinsic angular momentum*. By definition,

$$S^{cab} = -S^{cba}.$$

Local conservation of the total angular momentum, i.e. $\nabla_c M^{cab} = 0$, requires that

$$\nabla_c S^{cab} = (\Theta^{ab})_{\text{gravity}} - (\Theta^{ba})_{\text{gravity}}\,. \qquad (2.27)$$

We now add a tensor Υ^{bca} which is antisymmetric with respect to the first two indices b, c:

$$(\mathfrak{t}^{ca})_{\text{gravity}} = (\Theta^{ca})_{\text{gravity}} + \nabla_b\,\Upsilon^{bca}, \qquad (2.28)$$

where

$$\Upsilon^{cba} = \frac{1}{2}\left(S^{cba} + S^{bab} - S^{acb}\right). \qquad (2.29)$$

The $(\mathfrak{t}^{ab})_{\text{gravity}}$ should be identified to the *Belinfante-Rosenfeld tensor* [9] which is found to be symmetric.

In addition, the antisymmetry of Υ^{cba} guarantees that the conservation law remains unchanged

$$\nabla_a(\Theta^a_b)_{\text{gravity}} = \nabla_a(\mathfrak{t}^a_b)_{\text{gravity}} = 0\,. \qquad (2.30)$$

Staying far distant from matter (unexcited state), we have

$$(\Theta^{ab})_{\text{gravity}} \longrightarrow (\mathfrak{t}^{ab})_{\text{background field}}\,, \qquad \Upsilon^{cba} = 0\,.$$

By essence, $(\mathfrak{t}^{ab})_{\text{background field}}$ is thus symmetric.

Conclusions and outlook

Like we mentioned in an earlier publication, from the beginning of General Relativity, the cosmological constant Λ has played an unsavory rôle Einstein included this constant in his theory, because he wanted to have a cosmological model of the Universe which he wrongly thought static. Shortly after the works published by De Sitter and Lemaître, he decided to reject it.

But to-day, despite its smallness, a term like Λ seems to be badly needed to explain some astronomical observations, all related with the basic dynamical expanding model (Robertson-Walker et al.), even though its occurrence was never clearly explained.

In the classical General Relativity, the space-time is either filled with ponderomotive energy or devoid of source, which is accepted as a physical vacuum. However, numerous experiments predict that quantum vacuum is not "empty" but permanently subjected to virtual particles exchanges of energy.

Heisenberg's Uncertainty Principle, which allows for this process to take a place, has not been used in our demonstration, but it certainly plays a role in the variable property of the cosmological background field which our study relied on.

To sum up all that above, we have eventually reached the following important results:

- The gravitational energy can be represented by a true tensor;
- Its nonlocalizability doesnot hold anymore;
- The existence of a vacuum field is inferred from GR, which confirms the quantum predictions.

This last conclusion is noteworthy since our theory shows that General Relativity and Quantum Physics have convergent results.

Submitted on July 8, 2016 / Accepted on July 14, 2016

References

1. Landau L. and Lifshitz E. The Classical Theory of Fields. Addison-Wesley, Reading (Massachusetts), 1962, p. 402 (French translation).
2. Marquet P. The gravitational field: A new approach. *Progress in Physics*, 2013, v. 9, issue 3, 62–66.
3. Belinfante J. *Physica*, 1939, v. 6, 887.
4. Cartan E. La Géométrie des Espaces de Riemann. Re-print de 1925, Gauthier-Villars, Paris, 1946.

5. Straumann N. General Relativity and Relativistic Astrophysics. Springer-Verlag, Berlin, 1984, p. 159.
6. Tonnelat M.A. Les Théories Unitaires de l'Electromagnétisme et de la Gravitation. Gauthier-Villars, Paris, 1959, p. 18.
7. Hawking S.W., Ellis G.F.R. (1987). The Large Scale Structure of Space time. Cambridge University Press, 1987, p. 62.
8. Dirac P.A.M. General Theory of Relativity. Princeton University Press, 2nd edition, 1975, p. 61.
9. Rosenfeld L. Sur le tenseur d'Impulsion-Energie. Palais des Académies Roy. de Belgique, Serie: "Mémoires de Classes de Sciences", tome 18, Bruxelles, 1940.

On the Physical Nature of the de Broglie Wave

Patrick Marquet

18 avenue du Président Wilson, 62100 Calais, France
E-mail: patrick.marquet6@wanadoo.fr

Here is revisited de Broglie's Wave Mechanics Theory of Double Solution wherein a particle endowed with a variable proper mass is required to propagate within a hidden medium in order to describe a physical scalar wave carrying its own associated mass. Since the experiment that detected the wave applied to electrons, we extend the de Broglie's theory to the Dirac spinor, so that we can outline the physical reality of this fermion field.

Introduction

Some hundred years ago, was established the famous relation $E = h\nu$ later verified for the photon. On this basis, in 1924, Louis de Broglie extended the wave dualism to all massive particles. The predicted original wave function associated with a given particle was soon detected in 1927 by Davisson and Germer in their famous experience on electrons diffraction by a nickel crystal lattice [1]. The wave producing physical effects, was an overwhelming evidence of its true existence.

Nevertheless, since the *Brussels Solvay Symposium* was held in 1927, official physics interpretation prevailed which considered quantum mechanics on the pure statistical grounds and then leading to accept the notion of *non-real* wave functions.

Although it is unquestionable that use of a probabilistic wave and its generalization did lead to accurate prediction and fruitful theories, de Broglie could never believe that observable physical phenomena follow from abstract mathematical wave functions. In his opinion, the wave function had to remain an objective physical entity which is intimately related with its mass, rather than the subjective probabilistic representation currently adopted in modern quantum physics. Since the real wave was detected by means of electrons scattering, we will here formally show that there is a strict identity between its phase and the one of its associated wave which therefore physically carries the particle. To make this identity possible, the electron proper mass must be variable according to the Planck-Laue relation [2]. Within this frame, de Broglie's theory inferred a so-called "guidance formulae" which forces the electron to be always in motion. However, because of the stationary property of energy levels inside an atom, a static electron is not compatible with its dynamic guided state. de Broglie then postulated the existence of a *hidden medium* which permanently exchanges energy and momentum with the electron causing it to oscillate and then avoiding a motionless location.

When I first met Louis de Broglie in summer 1966, this issue was debated with a great deal of speculation. Today, another explanation can be pushed forward.

Notations

Space-time Latin indices run from $a = b$: 0, 1, 2, 3, while spatial Greek indices run from $\alpha = \beta$: 1, 2, 3. The space-time signature is -2.

1 Spinor field-electron duality

1.1 The origins of the Double Solution Theory

1.1.1 Basics of the wave mechanics

From standard optics, we first recall the definition of the classical wave with a frequency v

$$\psi = a(n)\exp\left[i(\nu t - \boldsymbol{k}\cdot\boldsymbol{r})\right] \tag{1.1}$$

which propagates along the direction of the unit vector $\boldsymbol{n}$. (Here $\boldsymbol{k}$ is the 3-wave vector, $\boldsymbol{k}\cdot\boldsymbol{r} = \phi$ is the wave spatial phase, n is the refractive index of the medium.)

Formula (1.1) is a solution of the classical propagation equation

$$\Delta\psi = \frac{1}{w^2}\frac{\partial^2\psi}{c^2\partial t^2}\,, \tag{1.2}$$

where w is the wave phase velocity of the wave moving in a dispersive medium whose refractive index is $n(\nu)$ generally depending on the coordinates, and which is defined by

$$\frac{1}{w} = \frac{n(\nu)}{c}\,. \tag{1.3}$$

This medium is assumed to be homogeneous and only depends on the frequency ν. The (constant) phase ϕ of the wave is progressing along the given direction with a separation given by a distance $\lambda = w/\nu$, called wavelength.

Consider now the superposition of a group of stationary (monochromatic) waves having each a very close frequency along the x-axis

$$\psi = \int_{\nu_0-\Delta\nu}^{\nu_0+\Delta\nu} a(n)\exp\left[i(\nu t - \phi(\nu))\right]. \tag{1.4}$$

Such a group of waves moves with a constant velocity called group velocity v_g according to the Rayleigh's formula

$$\frac{1}{v_g} = \frac{d(\nu/v)}{d\nu} = \frac{1}{\nu_0}\frac{\partial n\nu}{\partial\nu}\,. \tag{1.5}$$

The wave mechanics eventually shows that the group velocity v_g of waves associated with a particle of rest mass m_0, coincides with the velocity of this particle whose momentum along the x-axis (in vacuum) is given by the famous de Broglie's relation [3]

$$p_x = m_0 v_x = \frac{h}{\lambda}\,. \tag{1.6}$$

We clearly note that there is an obvious *first* physical link between the particle and its associated wave which will be further substantiated.

1.1.2 Double nature of the wave function

Like we mentioned above, de Broglie was firmly convinced that the wave associated with a massive particle should be a *real observable quantity*, therefore, he introduced a *true plane wave* of the usual form

$$\psi = a(x^\alpha) \exp\left[\frac{i}{h}\,\phi(x^\alpha)\right], \tag{1.7}$$

which is connected to a probabilistic Ω-wave by the relation

$$\Omega = f\,\psi\,, \tag{1.8}$$

where f is a constant normalizing factor.

The original wave mechanics is thus complemented with the Double Solution Theory [4], for Ω and ψ are two solutions of the same propagation equation. The Ω-wave (normed in the usual quantum machanical formalism), has the nature of a *subjective* probability representation formulated by means of the objective ψ-wave.

Defining ψ^* as the complex conjugate of ψ, it is well known that $\psi^2 dV = \psi\,\psi^* dV$ gives the absolute value of finding the particle in the volume element dV so that the normalization condition is adapted with f as

$$\int_V \Omega\,\Omega^* dV = 1\,. \tag{1.8 bis}$$

This guarantees that the particle is present in the arbitrary volume V.

The Ω and ψ have the same phase ϕ, but the constant f ought to be much larger than 1. Indeed, the current theory which only uses the Ω-function assumes this quantity to be spread out over the whole wave, i.e. spread out over a related physical quantity b (e.g. energy of the particle) according to

$$\int_V \Omega\,\Omega^* dV = b\,. \tag{1.8 ter}$$

In the double solution theory however, b should be concentrated in a very small region occupied by the particle and the integral of $a^2 b\,dV$ taken over the ψ-wave in the volume V is much smaller than b, which eventually leads to $|b| \gg 1$.

2 Extension to the spinor

2.1 The real spinor wave

2.1.1 The Dirac operators and Dirac equation (reminder)

In order to write the Schrödinger equation under a relativistic form, P. A. M. Dirac has defined a specific four-components wave function Ψ^A called *spinor* [5] which must necessarily apply to any spin-1/2 particles thus in our case, the electron. (Capital Latin spinorial indices are: $A = B = 1, 2, 3, 4$.)

To this effect, he introduced a system of (4×4) non local trace free matrices $\gamma_a = (\gamma_a{}^A{}_B)$. (In the classical theory, it is customary to omit the spinorial indices.)

The matrices γ a can display the standard following components [6]:

$$\gamma_0 = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad \gamma_1 = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix},$$
$$\gamma_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad \gamma_3 = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \tag{2.1}$$

in order to satisfy the fundamental relation

$$\gamma_a \gamma_b + \gamma_b \gamma_a = -2\,\eta_{ab}\,\mathbf{I}\,, \tag{2.2}$$

where η_{ab} is the Minkowskian tensor, and $\mathbf{I}$ is the unit matrix.

Formula $W = \gamma^a \partial_a$ is known as the Dirac operator where the Planck constant h is absorbed in the ∂_a.

For a free massive spin 1/2-field, the Dirac equation is eventually written as

$$(W - m_0 c)\,\Psi = 0\,, \tag{2.3}$$

where the proper mass m_0 is attributed to the associated spin 1/2-electron.

2.1.2 The normed spinor density

Since we are here considering a spin 1/2-fermion particle we must look for a wave which is a *real* spinor Ψ that physically carries the electron. From the classical Dirac theory, it is well known that the probability density of the electron's presence. is the time component of the (real) Dirac current vector density [7]

$$(J^a)_{\mathrm{D}} = i\,(^\sharp\Psi \gamma^a \Psi)\,, \tag{2.4}$$

where $^\sharp\Psi$ is the Dirac adjoint spinor $\Psi^+\gamma^0$, and Ψ^+ is the (complex) conjugate transpose of Ψ. So, this density of the electron reads

$$(J^0)_{\mathrm{D}} = i\,(^\sharp\Psi \gamma^0 \Psi) \tag{2.4 bis}$$

which is easily shown to be always definite and positive. Without the loss of generality, we could express Ψ under the form of a plane wave spinor [8] as

$$\Psi = \varpi(x^a) \exp\left[\frac{i}{h}\,\phi(x^a)\right],$$

where the wave spinor amplitude ϖ and the phase ϕ are *real local* functions. The Dirac spinor amplitudes ϖ could then be tuned so as to possess the orthogonality and completeness properties that guarantee that the plane waves Ψ have the adequate *normalization* to delta functions [9]. However, only a single spinor Ψ can be considered as a physical wave function, whereas we are left with 4-components Ψ^A. Then, at first glance, one might be tempted to consider the simple combination

$$\mathbf{\Psi} = \Psi^1 + \Psi^2 + \Psi^3 + \Psi^4.$$

Unfortunately, the Ψ-components are defined with respect to a spinorial frame $S(V_4)$ distinct from the structural Minkowski space, which renders those physically irrelevant. Instead, we will follow another extremely simple way: since ρ is here a real value, we have always the freedom to define a scalar wave function Φ such that

$$\Phi\,\Phi^* = \rho\,. \tag{2.5}$$

Moreover, we assume that this wave function has the same form as ψ (1.7)

$$\Phi = \omega(x^a)\exp\left[\frac{i}{h}\,\phi(x^a)\right]. \tag{2.6}$$

We state that Φ is the true wave function of the electron which was actually detected in the Davisson and Germer experiment upon a given set of gamma matrices γ^a, simply because it is derived from a real quantity which is itself inferred from the 1/2-spinor definition (2.4 bis) as it should.

Thus, we apply the same hypothesis conjectured by de Broglie (1.8 bis), and we are now able to write the normed expression as

$$\int_V \Xi\,\Xi^*\,dV = 1\,, \tag{2.7}$$

where

$$\Xi = g\,\Phi \tag{2.7 bis}$$

is the subjective wave function and g is a normalizing factor which satisfies (2.7).

In all the following text, Φ will be denoted as the "spinor wave".

2.1.3 Internal frequency of the electron

From (2.6), the energy and momentum of the electron located at x^a are

$$E = \partial_t\phi\,, \tag{2.8}$$

$$\boldsymbol{P} = P_a = -\mathrm{grad}\,\phi\,. \tag{2.9}$$

In order to outline the physical nature of the Φ-spinor wave, we start from the following consideration: in the framework of the Special Theory of Relativity, the frequency of a plane monochromatic wave is transformed as

$$\nu = \frac{\nu_0}{\sqrt{1-\mathbf{v}^2/c^2}}\,, \qquad \mathbf{v} = \mathrm{v}_a\,, \tag{2.10}$$

whereas the clock's frequency ν_c is transformed according to

$$\nu_c = \nu_0\sqrt{1-\mathbf{v}^2/c^2}\,. \tag{2.11}$$

If an electron is assumed to contain a rest energy $m_0c^2 = h\nu_0$, it is likened to a small clock of frequency ν_0, so that when moving with velocity $\mathbf{v}$, its frequency ν_c differs from that of the wave which is here noted ν.

In this concept, our main task will consist of showing that the electron is permanently in phase with its associated spinor wave, thus justifying the true nature of Φ that *physically* carries the electron

2.2 The physical nature of the spinor-electron duality

2.2.1 The Planck-Laue relation

We now postulate that the electron possesses a variable proper mass m_0' from which an important useful equation will be inferred.

Let us first write the Lagrange function for an observer who sees the electron of variable proper mass m_0' moving at the 3-velocity $\mathbf{v}$

$$L = -m_0'c^2\sqrt{1-\mathbf{v}^2/c^2} \tag{2.12}$$

so that the least action principle applied to this Lagrangian be still expressed by

$$\delta\int_{t_0}^{t_1} L dt = \delta\int_{t_0}^{t_1}\left(-m_0'c^2\sqrt{1-\mathbf{v}^2/c^2}\right)dt = 0\,. \tag{2.13}$$

From this principle are inferred the equations of motion

$$\frac{d}{dt}\frac{\partial L}{\partial\dot{x}_a} = \frac{\partial L}{\partial x_a} \tag{2.14}$$

with $\dot{x}_a = dx_a/dt$. It leads to

$$\frac{d\boldsymbol{P}'}{dt} = -c^2\sqrt{1-\mathbf{v}^2/c^2}\ \ \mathrm{grad}\,m_0' \tag{2.15}$$

(since m_0' is now variable).

Hence, by differentiating the well know relativistic relation

$$\frac{E'^2}{c^2} = \boldsymbol{P}'^2 + m_0'^2c^2 \tag{2.16}$$

we obtain

$$\frac{dE'}{dt} = c^2\sqrt{1-\mathbf{v}^2/c^2}\,\frac{\partial m_0'}{\partial t}\,. \tag{2.17}$$

Combining (2.15) and (2.17) readily gives

$$\frac{dE'}{dt} - \frac{\mathbf{v}\,d\boldsymbol{P}'}{dt} = c^2\sqrt{1-\mathbf{v}^2/c^2}\,\frac{dm_0'}{dt} \tag{2.18}$$

where

$$\frac{dm_0'}{dt} = \frac{\partial m_0'}{\partial t} + \mathrm{grad}\,m_0'$$

is the variation of the mass in the course of its motion.

On the other hand, we have

$$\frac{d\,(\boldsymbol{P}'\cdot\mathbf{v})}{dt} = \frac{\mathbf{v}\,d\boldsymbol{P}'}{dt} + \frac{m'_0c^2}{\sqrt{1-\mathbf{v}^2/c^2}}\,\frac{\mathbf{v}}{c}\,\frac{d\frac{\mathbf{v}}{c}}{dt} = \\ = \frac{\mathbf{v}\,d\boldsymbol{P}'}{dt} - m'_0c^2\,\frac{d}{dt}\,\sqrt{1-\mathbf{v}^2/c^2} \qquad (2.19)$$

or

$$\frac{d}{dt}\left(m'_0\sqrt{1-\mathbf{v}^2/c^2}\right) = \\ = c^2\,\sqrt{1-\mathbf{v}^2/c^2}\;\frac{dm'_0}{dt} + m'_0c^2\,\frac{d}{dt}\,\sqrt{1-\mathbf{v}^2/c^2}\,.$$

Hence, (2.18) can be written as

$$\frac{d}{dt}\left(E' - \mathbf{v}\cdot\boldsymbol{P}' - m'_0c^2\,\sqrt{1-\mathbf{v}^2/c^2}\right) = 0 \qquad (2.20)$$

which is satisfied when the electron is at rest (that is $\mathbf{v} = 0$, $E'_0 = m'_0c^2$).

Therefore, we must have always

$$E' = \frac{m'_0c^2}{\sqrt{1-\mathbf{v}^2/c^2}} = m'_0c^2\,\sqrt{1-\mathbf{v}^2/c^2} \;+\; \frac{m'_0\,\mathbf{v}^2}{\sqrt{1-\mathbf{v}^2/c^2}}\,. \qquad (2.21)$$

This is known as the Planck-Laue formula which plays a central rôle in our present theory.

2.2.2 Phase identity of the electron and its spinor wave

Let us first recall the relativistic form of the Doppler formula

$$\nu_0 = \nu\,\frac{1-\mathbf{v}/w}{\sqrt{1-\mathbf{v}^2/c^2}}\,, \qquad (2.22)$$

where ν_0 is the wave's frequency in the frame attached to the electron, ν and w are respectively the frequency and phase velocity of the spinor wave in a reference frame where this electron has a velocity $\mathbf{v}$.

With this formula, and taking the classical Planck relation $E = h\nu$ into account, we find

$$E = E_0\,\frac{1-\mathbf{v}^2/c^2}{1-\mathbf{v}/w}\,. \qquad (2.23)$$

However, inspection shows that the usual equation

$$E = \frac{E_0}{\sqrt{1-\mathbf{v}^2/c^2}} \qquad (2.24)$$

holds only if

$$1-\mathbf{v}/w = 1-\mathbf{v}^2/c^2 \qquad (2.25)$$

that implies

$$w\mathbf{v} = c^2. \qquad (2.26)$$

This latter relation is satisfied provided we set up

$$E' = \frac{m'_0c^2}{\sqrt{1-\mathbf{v}^2/c^2}}\,, \qquad (2.27)$$

$$\boldsymbol{P}' = \frac{m'_0\mathbf{v}}{\sqrt{1-\mathbf{v}^2/c^2}}\,. \qquad (2.28)$$

A variable proper mass is then required to insure that the electron as it moves, remains constantly in phase with that of the associated spinor wave. To see this, let us first multiply the Planck-Laue equation by dt

$$\left(\frac{m'_0c^2}{\sqrt{1-\mathbf{v}^2/c^2}} - \frac{m'_0\,\mathbf{v}^2}{\sqrt{1-\mathbf{v}^2/c^2}}\right)dt = m'_0c^2\,\sqrt{1-\mathbf{v}^2/c^2}\;dt. \qquad (2.29)$$

If $\boldsymbol{n}$ is the unit vector normal to the phase surface, we then consider that the electron whose internal frequency is $\nu_0 = m'_0c^2/h$ has travelled a distance $d\boldsymbol{n}$ during a time interval dt, so that its internal phase ϕ_i has been changed by

$$d\phi_i = h\nu_0\,\sqrt{1-\mathbf{v}^2/c^2}\;dt = m'_0c^2\,\sqrt{1-\mathbf{v}^2/c^2}\;dt\,. \qquad (2.30)$$

At the same time, the corresponding spinor wave phase variation is

$$d\phi = \partial_t\phi\,dt + \partial_{\boldsymbol{n}}\phi\,d\boldsymbol{n} = (\partial_t\phi + \mathbf{v}\,\mathrm{grad}\,\phi)\,dt$$

and, by analogy with the classical formulae (2.8) and (2.9), one can write

$$\boldsymbol{P}' = -\,\mathrm{grad}\,\phi = \frac{m'_0\mathbf{v}}{\sqrt{1-\mathbf{v}^2/c^2}}\,,$$

$$E' = \partial_t\phi = \frac{m'_0c^2}{\sqrt{1-\mathbf{v}^2/c^2}}\,,$$

so we find

$$d\phi = \left(\frac{m'_0c^2}{\sqrt{1-\mathbf{v}^2/c^2}} - \frac{m'_0\,\mathbf{v}^2}{\sqrt{1-\mathbf{v}^2/c^2}}\right)dt. \qquad (2.31)$$

Hence, from (2.29) we obtain the fundamental result which states that the internal phase of the electron is identical to that of its associated spinor wave

$$d\phi = d\phi_i\,. \qquad (2.32)$$

With (1.6), there is an obvious *second physical link* between the electron and the spinor wave Φ which clearly *carries* the lepton.

This is what we wanted to show.

Conclusions and outlook

Within the above theory, the electron is guided by its spinor wave which means that it is always in motion. In this case, the electron doesnot apparently comply with atomic quantum stationary states for which the electron is required to have zero velocity. De Broglie et al. [10] thus postulated a vacuum *hidden thermostat* whereby the electron is permanently exchanging energy and momenta. According to the authors this sub-quantum medium would cause the electron to fluctuate in a *Brownian-like* manner so as to exhibit a static situation only at the atomic level. In this way, the wavy-electron would be allowed to undergo perpetual infinitesimal propagation. Our opinion however differs from this hypothesis which we believe, would mark the limitation of the Double Solution theory. Preferably, we suggest that each energy level of an atom be characterized by a stationary limited spinor *wave packet* carrying a dynamical electron: the mean energy of the pair *wavepacket-moving electron* would then represent the quantized energy level of the atom.

"Squeezing" stepwise the wave packet (i.e. increasing the frequency) would mean jumping to a higher energy level and vice versa, which actually could reflect the excited/desexcited states of the atom. This process tends to validate the spectroscopic sharpness of the atomic rays as it is observed. All in all, the exposed theory seems to cope with an electron whose *physical* wave interacts with a *physical* diffraction device, and yet satisfies the established relativistic features of Dirac's theory.

Submitted on July 14, 2016 / Accepted on July 16, 2016

References

1. Davisson C.J., Germer L.H. Reflection of electrons by a crystal of nickel. *Proc. National Acad. of Sciences of the USA*, 1 April 1928, v. 14(4), 317–322.
2. de Broglie L. Thermodynamique Relativiste et Mécanique Ondulatoire. *Ann. Institut Henri Poincaré*, 1968, v. IX, no. 2, 89–108.
3. de Broglie L. Doctorate Thesis (1924). 2nd edition, Masson, Paris, 1963.
4. de Broglie L. Etude du mouvement des particules dans un milieu réfringent. *Ann. Inst. Henri Poincaré*, 1973, v. XVIII, no. 2, 89–98.
5. Dirac P.A.M. The Principles of Quantum Mechanics. PUF, Paris, 1931.
6. Lichnérowicz A. Champ de Dirac, champ du neutrino et transformations C, P, T sur un espace courbe. *Ann. Institut Henri Poincaré Sec. A*, 1964, v. 3, 233–290.
7. Moret-Bailly F. Le Champ Neutrinique en Relativité Générale. *Ann. Institut Henri Poincaré Sec. A*, 1966, v. 4, 301–355.
8. Berestetski V., Lifshitz E., Pitayevski L. Électrodynamique quantique. (Series: Physique théorique, tome 4), 2ème édition, traduit du Russe par V. Kolimeev, Editions de la Paix, Moscou, 1989.
9. Greiner W., Reinhardt J. Field Quantization. Springer (Berlin), 1993, p. 124.
10. Vigier J.P., Bohm D. Model of the causal interpretation of quantum theory in terms of a fluid with irregular fluctuations. *Physical Review*, 1954, v. 96, no. 1, 208–216.

Antigravity and Vacuum Propulsion in the Planck Vacuum Theory

William C. Daywitt

National Institute for Standards and Technology (retired), Boulder, Colorado, USA
E-mail: wcdaywitt@me.com

This paper explores the ideas of antigravity and vacuum propulsion from a fundamental-physics point of view, making use of the Planck vacuum (PV) model of the vacuum state.

1 Introduction

It is shown in a previous paper [1] that free-space gravitational shielding is ineffective, because gravitational waves, the carrier of the gravitational force, propagate within the PV state rather than free space. This result suggests that, as the gravitational waves are interior to the vacuum state, they may be affected by perturbations to that state. The following calculations are focused on that assumption in an attempt to determine if antigravity and vacuum propulsion are viable concepts.

Section 2 examines Newton's gravitational force between the earth (or any other large object), and a much smaller mass, from the viewpoint of the PV theory. The structure of that force is revealed in equations (2) and (3) in terms of n-ratios, which are normalized mass/PV coupling forces between the free space masses and the invisible vacuum state.

2 Newton's gravity

Newton's gravitational force F_{gr} between the two spherical masses $m \ll M$ separated by a distance $r\,(= a + h + A)$ can be expressed as

$$-F_{gr}(r) = \frac{mMG}{r^2} = \frac{(mc^2/r)(Mc^2/r)}{c^4/G} \tag{1}$$

$$= \frac{(mc^2/r)(Mc^2/r)}{m_*c^2/r_*} = n_r(m)n_r(M)\frac{m_*c^2}{r_*} \tag{2}$$

$$= \frac{aA}{r^2}\, n_a(m)\, n_A(M)\, \frac{m_*c^2}{r_*} = \frac{mc^2}{r}\,\frac{An_A(M)}{r} \tag{3}$$

using $G = e_*^2/m_*^2$ and $r_*m_*c^2 = e_*^2$ [2], where a and A are the radii of the masses m and M respectively, and h is the shortest distance between their surfaces. The mass m_* and Compton radius r_* belong to the separate Planck particles making up the degenerate PV state. The n-ratios in (2) and (3) are defined as

$$n_r(m) = \frac{mc^2/r}{m_*c^2/r_*}, \qquad n_r(M) = \frac{Mc^2/r}{m_*c^2/r_*}, \tag{4}$$

and

$$n_a(m) = \frac{mc^2/a}{m_*c^2/r_*}, \qquad n_A(M) = \frac{Mc^2/A}{m_*c^2/r_*}. \tag{5}$$

The coupling forces and the n-ratios are all less that one.

The force $m_*c^2/r_*\,(= c^4/G)$ is the maximum coupling force sustainable by the PV state [3, Fig.1]. Of particular interest to the present paper is $n_a(m)$, which is the normalized coupling force the mass m exerts on the PV at the surface of m. It is noted that the force m_*c^2/r_* normalizing the coupling forces also normalizes the Einstein field equation, and that the n-ratios are at the core of the metrics associated with the Schwarzschild equation [2] [4].

The mass/PV coupling forces in the numerators of (4) and (5) represent gravity-like forces the various free-space masses exert on the PV state. For example

$$\frac{mc^2}{r} = \frac{mc^2G}{r\cdot G} = \frac{mc^2G}{r\cdot e_*^2/m_*^2} = \frac{mm_*^2c^2G}{r\cdot r_*m_*c^2} = \frac{mm_*G}{rr_*} \tag{6}$$

is the force the mass m exerts on the Planck particles within the PV that are at a radius r from the center of m. The other coupling forces in (4) and (5) are similarly interpreted — e.g., $Mc^2/A = Mm_*G/Ar_*$.

Newton's dynamical equations start from his second law of motion ($m\ddot{r} = F_{gr}$) and the final expression in (3). With $dr = dh$ and $\ddot{r} = \ddot{h}$:

$$m\ddot{r} = m\ddot{h} = -\frac{mc^2}{r}\,\frac{An_A(M)}{r} \tag{7}$$

or

$$\ddot{r} = \ddot{h} = -\frac{c^2}{r}\,\frac{An_A(M)}{r} \tag{8}$$

for the acceleration of m toward M. Equation (8) is easily integrated over r via

$$\ddot{r} = \frac{d\dot{r}}{dt} = \dot{r}\,\frac{d\dot{r}}{dr} = \frac{d(\dot{r}^2/2)}{dr} = -\frac{Ac^2}{r^2}\, n_A(M) \tag{9}$$

from $r_0\,(\geqslant r + a)$ to r, and yields

$$\dot{r}^2 - \dot{r}_0^2 = 2c^2\left(\frac{r_0}{r} - 1\right)n_{r_0}(M) = 2c^2\left[n_r(M) - n_{r_0}(M)\right] \tag{10}$$

or

$$\left(\dot{r}^2 - \dot{r}_0^2\right)^{1/2} = -\left\{2c^2\left[n_r(M) - n_{r_0}(M)\right]\right\}^{1/2} \tag{11}$$

which is an equation involving only n-ratios and implying that the gravity dynamic takes place within the vacuum state.

3 PV state

The PV state [2] is assumed to be a degenerate state of negative energy Planck particles $(-e_*, m_*)$. Its degenerate nature implies that the Planck particle eigenstates within the vacuum are fully occupied. Thus the Planck particles are not free to exhibit macroscopic motion. The vacuum is bathed, however, in microscopic zero-point Planck-particle agitation.

Due to this degeneracy, when the PV is perturbed it exhibits percussion-like response waves, much like the waves on the surface of a kettle drum. For example, the free space electron core $(-e_*, m_e)$ perturbs the vacuum with the two-term coupling force "$e_*^2/r^2 - m_e c^2/r$", leading to the Dirac-equation response [5], where that response does not involve macroscopic Planck particle motion.

The previous section outlines the PV response to coupling forces of the form mc^2/r — thus the PV response is of the nature of gravitational percussion waves traveling within the PV between the positions of the free space masses m and M. It is this type of wave motion that is envisioned in the discussion to follow; i.e., wave motion that does not involve macroscopic Planck particle motion.

4 Summary, conclusions, and comments

From a survey of equations (1)–(11), it is clear: that the final expression in (3) is the springboard for the Newtonian dynamics, equations (7)–(11); and that none of the expressions in (2) and (3) show any sign of a direct free-space gravitational force acting between m and M — the force is channeled through the vacuum state. The second conclusion implies that there can be no free-space gravitational shielding [1].

The first expression in (3),

$$F_{gr}(r) = -\frac{aA}{r^2}\, n_a(m)\, n_A(M)\, \frac{m_* c^2}{r_*} \qquad (12)$$

suggests that, if the coupling force mc^2/a in $n_a(m)$ could be masked or eliminated, then $F_{gr} = 0$ and m would experience no gravitational attraction toward the mass M; so the mass m would be effectively weightless. The vanishing of the n-ratio $n_a(m)$ thus leads to a simple explanation for antigravity, once the physical mechanism for nullifying $n_a(m)$ is specified.

It is difficult to find experimental data in the open literature that addresses the preceding theoretical calculations. The one source germane to the present work the author could find is contained in the e-book entitled "What Goes Up..." [6], which is a novel that claims to discuss real experimental data. The principle interest here is the composite electrical coil that is at the heart of a craft that is claimed to exhibit antigravity and vacuum propulsion.

The doughnut shaped coil consists of two current loops each of which supports a separate a.c.-d.c. signal, where the two a.c. signals in the two loops are set at different frequencies. This heuristic description is sketchy due to unavailable details in the coil design. The book claims that the magnetic fields (or the magnetic flux) produced by the coil are the source of antigravity and vacuum propulsion (though the book doesn't use the term "vacuum propulsion"). The a.c. field destroys the gravity force F_{gr}; and the (±) d.c. field causes the craft to move up or down at a high rate of speed. The second paragraph of the present section and the a.c. currents in the coil thus account for antigravity. (The a.c. and d.c. stand for "alternating current" and "direct current" respectively.)

Although much theoretical knowledge concerning the PV state exists [2], there is much still to be learned. The antigravity conundrum was readily resolved with the force equation (12). Even then, details of how the a.c. flux from the coil nullifies the effect of $n_a(m)$ is not fully understood. Concerning vacuum propulsion, things are even worse. For closure sake, then, it will just be stated (assumed) that the ± d.c. flux interacting with the charges $(-e_*)$ of the separate Planck particles within the PV result in the rapid systematic movement of the coil, and hence the space-craft. Reflecting upon the intricacy of the electron spinor field caused by the electron/PV interaction [5], the idea of vacuum propulsion doesn't seem so strange.

Submitted on July 14, 2016 / Accepted on July 17, 2016

References

1. Daywitt W.C. Gravitational shielding as viewed in the Planck Vacuum Theory. *Progress in Physics*, 2016, v. 12, issue 3, 301.
2. Daywitt W.C. The trouble with the equations of modern fundamental physics. *American Journal of Modern Physics*, *Special Issue*: *Physics Without Higgs and Without Supersymmetry*, 2016, v. 5, no. 1-1, 22. See also http://www.planckvacuum.com
3. Daywitt W.C. Limits to the validity of the Einstein field equations and General Relativity from the viewpoint of the negative-energy Planck vacuum state. *Progress in Physics*, 2009, v. 5, issue 3, 27.
4. Daywitt W.C. The Crothers metrics and the black hole metric as viewed from the Planck vacuum perspective. *Galilean Electrodynamics*, 2014, Sept./Oct. 2014, 82.
5. Daywitt W.C. Understanding the Dirac equation and the electron-vacuum system. *Progress in Physics*, 2013, v. 9, issue 4, 78.
6. Stover J.C. What Goes Up... Publish Green, Minneapolis (MN), 2015.

On the Quantum-Relativistic Behavior of Moving Particles

Fernando Ogiba
E-mail: fogiba@gmail.com

The *zitterbewegung* of massless elementary electrical charges consists of two distinct vacuum induced fluctuations. The first, random loops (spin) at the light speed (co-moving frame) [1], is attributed to absorptions and emissions of zero-point radiation at the Compton's rate (stochastic electrodynamics). It will be shown that the second (de Broglie) emerges because such radiation, just passing but tangled for a while (rest mass), doesn't submit to the ordinary motion of bodies; its light speed is ensured by truncations and restoration of the translational motion (inertia). Synchronized with absorption-emission, kinetic energy becomes vibrational energy (x-ray), and vice versa. The implied works are due to back and forth self-stresses (contractions) triggered by imminent violations of the light speed limit (loops at the light speed plus ordinary motion) implicit in the improper de Broglie phase velocity. Time spent to preserve the normal motility of the tangled radiation is observed only in the fixed frame (time dilation).

1 Introduction

Due to permanent interactions with the Planck's vacuum [2–4], massless elementary electrical charges (MEEC) are induced to move along quantum-relativistic paths at the speed of the interacting radiation, independently of the observed ordinary motion of particles, as implicit in the approach originating the concept of *zitterbewegung* [5]. In such approach, it was considered a particle (an electron) of rest mass m_0, which, therefore, must be attributed (respecting the peculiarities of the interaction) to the mass-equivalent of the zero-point energy absorbed (incident momentum) and emitted (reaction momentum) by MEEC. It means that MEEC, on average, retain zero-point radiation; a boson giving the rest mass.

In the particular case of free particles, the argued paths are continual random "jumps" (diffusion of probability) among trajectories belonging to the ensemble dictated by the Dirac equation [6]. Theoretical results indicate that such trajectories are curvilinear, over which particles are found at the light speed, which agrees with experimental facts. Indeed, if they are seen as random loops of electrical current in the co-moving frame (a charge e moving at the light speed c over a spherical shell of average radius r_c), then we find that the corresponding magnetic moment,

$$\mu_z = IA = \frac{ec}{2\pi r_c}\pi r_c^2 = \frac{ecr_c}{2}\,, \tag{1}$$

matches the observed magnetic moment of spin-1/2 particles,

$$\mu_z \approx \frac{e\hbar}{2m_0}\,, \tag{2}$$

if $2\pi r_c = \lambda_c$, where $\lambda_c = h/m_0c$ is the Compton's wavelength.

Alternatively, if an electron can be found over circles at the light speed (co-moving frame), then its momentum components should fluctuate like $p' = m_0\dot{q}' = m_0c\cos(\omega' t' + \phi_{q'})$, where $\phi_{q'}$ are random phases. It implies the coordinates

$$q' = \frac{c}{\omega_c}\sin(\omega' t' + \phi_{q'})\,, \tag{3}$$

where c/ω' is the radius of the loops of current (fluctuations with spherical shape). Inserting the corresponding variances (averaging over random phases),

$$\Delta p'^2 = \frac{1}{2}(m_0c)^2, \quad \Delta q'^2 = \frac{1}{2}\frac{c^2}{\omega'^2}\,, \tag{4}$$

into the minimum uncertainty relation, $\Delta p'\Delta q' = \hbar/2$, yields

$$\omega' = \frac{m_0c^2}{\hbar}\,, \quad r_c = \frac{c}{\omega'} = \frac{\lambda_c}{2\pi}\,, \tag{5}$$

that is, the Compton's angular frequency ($\omega' = \omega_c$).

Considering the center of mass of the fluctuations (vibrations) at the origin of the co-moving frame ($\mathbf{x}' = 0$), it implies that a free particle moving in the x-direction of the fixed frame will be seen as a material wave of wave number $\mathbf{k} = (k, 0, 0)$. Phase invariance, considering special relativity, i.e.

$$\omega' t' - \mathbf{k}'\cdot\mathbf{x}' = \omega t - \mathbf{k}\cdot\mathbf{x}, \tag{6}$$

implies

$$t' = \frac{\omega}{\omega_c}\left(t - \frac{x}{v_p}\right), \quad v_p = \frac{\omega}{k}\,, \tag{7}$$

where v_p is the phase velocity. Comparing the Eq. (7) with the Lorentz transformation

$$t' = \gamma\left(t - \frac{v}{c^2}\,x\right) \tag{8}$$

one gets the parameters of the material wave [7]:

$$\omega = \gamma\frac{m_0c^2}{\hbar}\,, \quad k = \gamma\frac{m_0v}{\hbar}\,, \quad v_p = \frac{\omega}{k} = \frac{c^2}{v}\,, \tag{9}$$

from which we can see that v_p is a violation of the natural speed of electromagnetic waves. This fact makes v_p meaningless in the context of the special relativity, which is reinforced

by the existence of a group velocity (transport of matter) coinciding with the particle velocity, i.e.

$$v_g = \frac{\partial \omega}{\partial k} = \frac{\partial E}{\partial p} = v\,. \tag{10}$$

Technically, the concept of group velocity requires that the resultant material wave be a superposition of waves of different frequencies, which agrees with the successful concept of wave packet [8]. However, a wave packet implies a set of phase velocities. As the phase velocity of the resultant material wave is a violation of the natural speed of radiation, then we should expect that the phase velocities of the constituent waves also are speed violations (at least mostly).

Here, is it wise keep in mind that such speed violations, being in full agreement with the concepts expressed by equations (6), (8) and (10), cannot be meaningless. In effect, notice that an evolution at the phase velocity ($x = v_p t$) implies that time "stops" in the co-moving frame ($t' = 0$). Emphasizing, in this particular situation, time is computed only in the fixed frame. Remarkably, despite of being an improper evolution, it agrees with the ultimate meaning of time dilation.

Until now, we have seen that single frequency (ω_c) fluctuations of spherical shape ($r_c = c/\omega_c$) become multi-frequency fluctuations in the fixed frame [9], which manifest as a wave packet (material wave). The emergence of multiple angular frequencies implies that the translational motion cause a break of the spherical shape of the fluctuations, given that for each emerging angular frequency there must correspond a different radius ($\omega_i r_i = c$, where c is invariant). Coincidentally, this agrees with length contraction, i.e., according to the theory of special relativity, in the fixed frame the fluctuations must present an ellipsoidal shape.

The above argumentation implies that the phase velocity v_p is a statistical quantity; given that all frequencies implied in the wave packet do not exist simultaneously but in the elapsed time of an ordinary measurement (much greater than $2\pi/\omega_c$).

Physically, contractions of the vacuum induced fluctuations requires back and forth forces, whose resultant, at least on average, must be zero. Moreover, these forces — defined only in the fixed frame — do not have the same nature of the electromagnetic forces (from the Planck's vacuum) responsible by the fluctuations in the co-moving frame.

The search for forces triggered by the translational motion must begin noting that the speed violation v_p is dominant in the Lorentz transformations (LT), i.e.

$$x' = \gamma\,(x - vt)\,, \quad t' = \gamma\left(t - \frac{x}{v_p}\right), \quad \gamma = \left(1 - \frac{v}{v_p}\right)^{-\frac{1}{2}}, \tag{11}$$

which suggests that LT — to account for the light speed limit in both reference frames — just consider imminent velocity violations when the linear translational motion takes place. In other words, the emerging vibrations, whose *statistical superposition* gives v_p, must relate to a mechanism ensuring the speed limit of the zero-point radiation (ZPR) tangled for a moment by MEEC (co-moving frame), given that the relative velocity is lower than c.

Let us analyze, heuristically, the complete motion. From the equations (1) to (5) and the presence of the Planck's vacuum, it is implicit that in the co-moving frame MEEC are found over circular trajectories at the speed of the "impregnating" zero-point radiation (ZPR). Therefore, it be expected the occurrence of all sort of violations of the light speed limit when the ordinary translational motion is added. However, resulting velocities for MEEC — imbued with the properties of radiation — either greater or smaller than c are forbidden by the well-known Maxwell's relation $\mu\varepsilon c^2 = 1$. So, it is plausible to think that the vibrations implied in the wave packet — related to radii contraction of the fluctuations — arise to avoid any possible speed violation of the tangled ZPR, which would result from the simple combination of random orbits at light speed with the observed motion of matter.

In the next sections, based on well-known physical facts, it will be presented some evidences that the periodical motion induced by the Planck's vacuum combined with the ordinary motion of particles implies the appearance of periodical back and forth self-stresses, which are imposed by the *normal motility* of the *tangled radiation*. Here, it must be emphasized the following: First, *tangled radiation* is ZPR continually imprisoned during an infinitesimal time (less than $2\pi/\omega_c$) by MEEC. Second, *normal motility* relates to evolutions of free radiation; assumed to be extensible to the *tangled radiation*, given the massless nature of the "host".

2 The need for periodical longitudinal self-stresses

The energy carried by the material wave is the vibrational energy, $E = \hbar\omega$, which must be the energy of the particle, $E = \gamma m_0 c^2$. Therefore,

$$\omega = \frac{m_0 c^2}{\hbar} + \frac{(\gamma - 1) m_0 c^2}{\hbar} = \omega_c + \omega_T, \tag{12}$$

where the Compton's frequency (ω_c) expresses the rate at which zero-point energy is going in and out of the MEEC (on average remaining as rest energy), and ω_T accounts for all vibrations implied in the wave packet; likewise that v_p represents all corresponding phase velocities (one at a time).

Given the statistical nature of the wave packet (in the sense of the quantum superposition), it implies that particles can present, at a given time, only kinetic energy, or only vibrational energy, or a mix of them; all these possibilities occurring, in accordance with energy conservation, at a very high rate (synchronized with ω_c).

Coincidentally, for $v \ll c$, ω_T is the *maximum* frequency emitted by electrons in a x-ray apparatus (Duane-Hunt formula, $\hbar\omega_{max} = eV = m_0 v^2/2$), which does not contradict the fact that electrons can collide presenting frequencies different from ω_{max}. In effect, these other frequencies can be built into

the well-known wavelength spread of x-ray data; the complementary energy (kinetic) simply warm the target.

The above facts suggest that kinetic energy becomes vibrational energy, and vice-versa, but the sum of them, at any time, is $(\gamma - 1)m_0c^2$ or $\hbar\omega_T$, as required by energy conservation. Inexorably, such changes of the kinetic energy imply positive and negative works on the particle. Nonetheless, if one takes into account that the MEEC-ZPR *electromagnetic* interaction is completely resolved, in the sense that it yields well-defined rest energy (mass), spin and Compton's parameters, then there must be another reason for the emergence of vibrations triggered by the translational motion. Only remains to appeal to the dynamics allowed by the tangled ZPR, which, in view of the above, only can be attributed to periodical back and forth self-stresses, whose sole purpose is to ensure its light speed limit; an imposition of the hindmost nature of radiation.

Here, it should be pointed up that these self-stresses — ensuring the normal motility of the tangled radiation — cannot be interpreted in the same sense of Poincaré stresses [10], which were postulated in order to guarantee the stability of the Abraham-Lorentz model for the electron. In effect, the semi-classical electron stability should be understood as an electromagnetic pressure balance involving the Planck's vacuum, as proposed by Casimir [11].

3 The Zitterbewegung and self-stresses

A formal account for the two kind of vacuum induced fluctuations, as exposed elsewhere, can be seen in the quantum-relativistic approach of the *zitterbewegung* [12], although not working the properties of the Planck's vacuum of explicit way; that is, using the recipes of the stochastic electrodynamics. This is possible because Lorentz transformations as well as quantum equations takes into account non-localized statistical features of the wave packet (the ultimate product of the matter-vacuum interaction). Hence, the following results, despite of evidencing the co-moving loops of electrical current (spin), should be interpreted statistically [13].

Inserting the Dirac Hamiltonian, $H = c\alpha_j p_j + \beta m_0 c^2$, into the Heisenberg picture of quantum mechanics and considering that the matrices α_j and β commute with momentum (p_j) and position (x_j) operators, one gets

$$\frac{dp_j}{dt} = \frac{i}{\hbar}\left[H, p_j\right] = 0\,, \quad \frac{dx_j}{dt} = \frac{i}{\hbar}\left[H, x_j\right] = c\alpha_j\,, \qquad (13)$$

where the first implies that H and p_j commute (constants of the motion), and the second, in the full sense of the operation

$$c\alpha_j\psi = \pm c\psi\,, \qquad (14)$$

where ψ represents a four-component spinor, means that "a measurement of a component of the velocity of a free electron is certain to lead to the result $\pm c$" [5, p. 262], which is not the ordinary velocity of free particles, but that of the tangled ZPR.

The result (14) means that — on average, everywhere, in all directions and with equal probability — electrons go forth and back at the light speed; an expected behavior, considering the main properties of the interacting ZPR (homogeneity, isotropy, randomness and Lorentz invariant spectral density).

Whenever the electron is on a permitted Dirac trajectory, despite of being temporarily, it must obey the parameters of such trajectory. As the trajectories are curvilinear, then there are accelerations. In fact, they are given by $\ddot{x}_j = (i/\hbar)[H, \dot{x}_j]$, where $\dot{x}_j = c\alpha_j$, which corresponds to the equations

$$\ddot{x}_j = \frac{2i\left(H\dot{x}_j - c^2 p_j\right)}{\hbar}\,, \quad \ddot{x}_j = \frac{2i\left(c^2 p_j - \dot{x}_j H\right)}{\hbar}\,, \qquad (15)$$

since $Hc\alpha_j + c\alpha_j H = 2cp_j$. Integrating, yields respectively

$$\dot{x}_j = c^2 p_j H^{-1} + \eta_j e^{i2Ht/\hbar}; \quad \dot{x}_j = c^2 p_j H^{-1} + \eta'_j e^{-i2Ht/\hbar}; \qquad (16)$$

where the operators η and η' (constants of integration) must take into account that these *components* must match, periodically, the tangential velocity ($c\alpha_j$), as implicit in Eq. (14), which implies that $\eta = \eta' = c\alpha_j - c^2 p_j H^{-1}$. Moreover, on average the velocity must be the observed one ($c^2 p_j H^{-1}$). Therefore, the velocity operator becomes

$$\dot{x}_j = c^2 p_j H^{-1} + \left(c\alpha_j - c^2 p_j H^{-1}\right)\cos\left(2Ht/\hbar\right)\,, \qquad (17)$$

from which, considering the same above conditions, one gets the position operator

$$x_j(t) = c^2 p_j H^{-1} t + \frac{\left(\hbar c\alpha_j H^{-1} - \hbar c^2 p_j H^{-2}\right)}{2}\sin\left(\frac{2H}{\hbar}t\right). \qquad (18)$$

Notice, for $p_j = 0$ the operators (17) and (18) violate the minimum uncertainty relation ($m_0 \Delta\dot{x}_j \Delta x_j = \hbar/2$) by a factor 2 (the eigenvalues of α_j are unitary and $H \to m_0c^2$). This happens because the Dirac Hamiltonian takes into account matter and antimatter, whose energy gap is $2H$ [14, p. 949]. For only one kind of particle (e.g. free electrons in the two slit experiment, where is not verified the presence of positrons), it suffices to ignore the factor 2 in the equations (15).

Regardless of the comment made in the previous paragraph, the statistical components of the velocity of an electron (moving at the speed of light), as expressed by Eq. (17), show that — in order to maintain the speed imposed by the tangled radiation — the translational velocity $c^2 p_j H^{-1}$ is periodically subtracted and added, depending on the sign of c. Indeed, apart intermediary values, for forward evolutions of the local motion ($+c$), the translational motion is completely subtracted, and for backward evolutions ($-c$), it is completely restored, as can be seen from the allowed values of the cosine and the Eq. (14). Clearly, synchronized with absorptions and emission of zero-point energy (rest energy), the kinetic energy changes at the Compton's rate (considering only one kind of particle). As truncations and restorations of the translational motion behaves as vibrations, then kinetic energy is

being transformed into vibrational energy, and vice versa. These positive and negative works, necessarily, imply back and forth forces (zero, on average). However, as there are no external forces — other than those yielding the well-defined evolutions in the co-moving frame — then such works must be assigned to periodical longitudinal self-stresses (PLSS), which are imposed by the very motility of radiation, as inferred in the preceding paragraph.

From the position operators (18) — statistical coordinates defined in the fixed frame — we can verify the following: First, they do not explicit a set of vibrations composing the wave packet, but the motion of the resulting material wave, whose statistical frequency is $\omega = H/\hbar$. Second, for $p_j = 0$, these coordinates agree with the proposed equations (3); evolutions with spherical shape in the co-moving frame. Third, in the fixed frame ($p_j \neq 0$), the amplitude of the vibration (enclosed difference of operators) suffers a contraction in the direction of the motion; evolutions with ellipsoidal shape.

4 Final remarks

Fluctuations with spherical shape (co-moving frame) becoming fluctuations with ellipsoidal shape (fixed frame) explains the emergence of all vibrations implied in the wave packet, but in the sense that a motion with constant tangential velocity (light speed) over an ellipsoid implies an infinite number of angular frequencies. The wave packet is a statistical concept; it simply expresses the fact that during the time of an ordinary measurement the particle can be found at any position on the ellipsoidal surface; each one corresponding to a given angular frequency (particle states). This is the fundamental feature of quantum superposition.

As "self-impulses", in principle, cannot be observed in the co-moving frame, then the corresponding time intervals also not. From another point of view, the strength of self-stresses depends of the relative velocity, but an observer in the co-moving frame cannot decide about the constant velocity of such frame (principle of relativity); therefore, also cannot decide about self-stresses (and its duration). This is implicit in the LT, as can be seen inserting the improper evolution $x = (c^2/v)t$ (triggering a given self-stress), which gives $t' = 0$. In short, the time spent to preserve the "integrity" of the tangled ZPR is computed only in the fixed frame, which is in full agreement with the cumulative time dilation.

The vibrational energy corresponding to self-stresses only are emitted as radiation under non-uniform decelerations (as in a x-ray apparatus). Contrasting with thermal excitations (external forces), PLSS only imply restrictions to the mobility of vacuum induced fluctuations (without external forces); so, radiationless.

The corresponding back and forth strains (restrictions to the translational motion) explain the non-cumulative length contraction.

Newton's inertia relates to the de Broglie periodicity [15]; that is, the periodicity of the wave packet, whose corresponding vibrations come from PLSS; "opposing forces".

To finalize, truncations of the ordinary motion followed by complete restoration of the kinetic energy, as implicit in the Eq. (17), is in full agreement with the observed energy conservation (first Newton's law).

5 Conclusion

In the light of the foregoing, the quantum relativistic behavior of particles emerge because the ZPR, continually entrapped by MEEC during the time of an absorption-emission of zero-point energy, does not submit to the ordinary motion of bodies. From another point of view, the quantum of the Higgs field (Higgs boson) does not move with the observed velocities of the corresponding particle; its light speed is ensured by conservative periodical truncations and restorations of the ordinary motion, whose momentum dependent strength (amplitude of emerging vibrations) explain why inertia (mass) increases with the particle velocity.

Submitted on July 26, 2016 / Accepted on July 28, 2016

References

1. Huang K. On the Zitterbewegung of the Dirac Electron. *American Journal of Physics*, 1952, v. 20, no. 8, 479–484.
2. Boyer T. Random Electrodynamics: The theory of classical electrodynamics with classical electromagnetic zero-point radiation. *Physical. Review D*, 1975, v. 11, no. 4, 790–808.
3. Milonni P.W. The Quantum Vacuum: An Introduction to Quantum Electrodynamics. Academic Press, Inc, 1994.
4. Ogiba F. Planck's Radiation Law: Thermal excitations of vacuum induced Fluctuations, *Progress in Physics*, 2015, v. 11, no. 2, 145–148.
5. Dirac P. A. M. The Principles of Quantum Mechanics, 4ed. Oxford University Press, 1999.
6. Ogiba F. Addendum to Phenomenological Derivation of the Schrödinger Equation. *Progress in Physics*, 2014, v. 10, no. 2, 108–110.
7. de Broglie L. Ondes et quanta. *Comptes rendus des séances de l'Académie des Sciences*, 1923, v. 177, 507–510.
8. Burkhardt C.E., Leventhal J.J. Foundations of Quantum Physics. Springer, 2008.
9. Dávid G., Cserti J. General theory of Zitterbewegung. *Physical Review B*, 2010, v. 81, 121417-1–121417-4.
10. Poincaré H. Sur la dynamique de l'électron. *Rendiconti del Circolo Matematico di Palermo*, 1906, v. 21, no. 1, 129–175.
11. Puthoff H.E. Casimir Vacuum Energy and the Semiclassical Electron. *International Journal of Theoretical Physics*, 2007, v. 46, no. 12, 3005–3008.
12. Sakurai J.J. Advanced Quantum Mechanics. Addison-Wesley, 1967.
13. Pauli W. The Connection Between Spin and Statistics. *Physical Review*, 1940, v. 58, no. 8, 716–722.
14. Messiah A. Quantum Mechanics. Volume II. North Holland Publishing Company, 1962.
15. Wignall J.W.G. De Broglie waves and the nature of mass. *Foundations of Physics*, 1985, v. 15, no. 2, 207–227.

Type III Spacetime with Closed Timelike Curves

Ahmed Faizuddin
Hindustani Kendriya Vidyalaya, Dinseh Ojha Road, Bhangagarh, Guwahati-05, Assam, India
E-mail: faizuddinahmed15@gmail.com

We present a symmetric spacetime, admitting closed timelike curves (CTCs) which appear after a certain instant of time, *i.e.*, a time-machine spacetime. These closed timelike curves evolve from an initial spacelike hypersurface on the planes $z = constant$ in a causally well-behaved manner. The spacetime discussed here is free from curvature singularities and a 4D generalization of the Misner space in curved spacetime. The matter field is of pure radiation with cosmological constant.

1 Introduction

One of the most intriguing aspects of Einstein's theory of gravitation is that solutions of field equations admit closed timelike curves (CTC). Presence of CTC in a spacetime leads to time-travel which violates the causality condition. The first one being Gödel's spacetime [1] which admits closed timelike curves (CTC) everywhere and an eternal time-machine spacetime. There are a considerable number of spacetimes in literature that admitting closed timelike curves have been constructed. A small sample would be [1–21]. One way of classifying such causality violating spacetimes would be to categorize the metrics as either eternal time-machine in which CTC always exist (in this class would be [1, 2]), or as time-machine spacetimes in which CTC appear after a certain instant of time. In the latter category would be the ones discussed in [18–20]. Many of the models, however, suffer from one or more severe drawbacks. For instance, in some of these solutions, for example [13, 14, 20], the weak energy condition (WEC) is violated indicating unrealistic matter-energy content and some other solutions have singularities.

Among the time-machine spacetimes, we mention two: the first being Ori's compact core [17] which is represented by a vacuum metric locally isometric to pp waves and second, which is more relevant to the present work, the Misner space [22] in 2D. This is essentially a two dimensional metric (hence flat) with peculiar identifications. The Misner space is interesting in the context of CTC as it is a prime example of a spacetime where CTC evolve from causally well-behaved initial conditions.

The metric for the Misner space [22]

$$ds^2_{Misn} = -2\,dt\,dx - t\,dx^2 \tag{1}$$

where $-\infty < t < \infty$ but the co-ordinate x is periodic. The metric (1) is regular everwhere as $\det g = -1$ including at $t = 0$. The curves $t = t_0$, where t_0 is a constant, are closed since x is periodic. The curves $t < 0$ are spacelike, but $t > 0$ are timelike and the null curves $t = t_0 = 0$ form the chronology horizon. The second type of curves, namely, $t = t_0 > 0$ are closed timelike curves (CTC). This metric has been the subject of intense study and quite recently, Levanony and Ori [23], have studied the motion of extended bodies in the 2D Misner space and its flat 4D generalizations. A non-flat 4D spacetime, satisfying all the energy conditions, but with causality violating properties of the Misner space, primarily that CTC evolve smoothly from an initially causally well-behaved stage, would be physically more acceptable as a time-machine spacetime.

In this paper, we shall attempt to show that causality violating curves appear in non-vacuum spacetime with comparatively simple structure. In section 2, we analyze the spacetime; in section 3, the matter distribution and energy condition; in section 4, the spacetime is classified and its kinematical properties discussed; and concluding in section 5.

2 Analysis of the spacetime

Consider the following metric

$$\begin{aligned} ds^2 = {} & 4\,r^2\,dr^2 + e^{2\alpha r^2}\left(dz^2 - t\,d\phi^2 - 2\,dt\,d\phi\right) + \\ & + 4\beta\,z\,r\,e^{-\alpha r^2}\,dr\,d\phi \end{aligned} \tag{2}$$

where ϕ coordinate is assumed periodic $0 \leqslant \phi \leqslant \phi_0$, where α is an integer and $\beta > 0$ is a real number. We have used co-ordinates $x^1 = r$, $x^2 = \phi$, $x^3 = z$ and $x^4 = t$. The ranges of the other co-ordinates are $t, z \in (-\infty, \infty)$ and $0 \leqslant r < \infty$. The metric has signature $(+,+,+,-)$ and the determinant of the corresponding metric tensor $g_{\mu\nu}$, $\det g = -4\,r^2\,e^{6\alpha r^2}$. The non-zero components of the Einstein tensor are

$$\begin{aligned} G^{\mu}_{\mu} &= 3\,\alpha^2, \\ G^{t}_{\phi} &= -\frac{1}{2}\,e^{-6\alpha r^2}\,\beta^2\,. \end{aligned} \tag{3}$$

Consider an azimuthal curve γ defined by $r = r_0$, $z = z_0$ and $t = t_0$, where r_0, z_0, t_0 are constants, then we have from the metric (2)

$$ds^2 = -\,t\,e^{2\alpha r^2}\,d\phi^2. \tag{4}$$

These curves are null for $t = 0$, spacelike throughout for $t = t_0 < 0$, but become timelike for $t = t_0 > 0$, which indicates the presence of closed timelike curves (CTC). Hence CTC form at a definite instant of time satisfy $t = t_0 > 0$.

It is crucial to have analysis that the above CTC evolve from a spacelike $t = constant$ hypersurface (and thus t is a

time coordinate) [17]. This can be ascertained by calculating the norm of the vector $\nabla_\mu t$ (or by determining the sign of the component g^{tt} in the inverse metric tensor $g^{\mu\nu}$ [17]). We find from (2) that

$$g^{tt} = t\, e^{-2\,\alpha\, r^2} + \beta^2\, z^2\, e^{-6\,\alpha\, r^2}\,. \tag{5}$$

A hypersurface $t = constant$ is spacelike provided $g^{tt} < 0$ for $t = t_0 < 0$, but becomes timelike provided $g^{tt} > 0$ for $t = t_0 > 0$. Here we choose the z-planes defined by $z = z_0$, (z_0, a constant equal to zero) such that the above condition is satisfied. Thus the spacelike $t = constant < 0$ hypersurface can be chosen as initial conditions over which the initial may be specified. There is a Cauchy horizon for $t = t_0 = 0$ called Chronology horizon which separates the causal and non-causal parts of the spacetime. Hence the spacetime evolves from a partial Cauchy hypersurface (initial spacelike hypersurface) in a causally well-behaved manner, up to a moment, *i.e.*, a null hypersurface $t = 0$ and CTC form at a definite instant of time on $z = constant$ plane.

Consider the Killing vector $\eta = \partial_\phi$ for metric (2) which has the normal form

$$\eta^\mu = (0, 1, 0, 0)\,. \tag{6}$$

Its co-vector is

$$\eta_\mu = \left(2\,\beta\, z\, r\, e^{-\alpha\, r^2},\ -t\, e^{2\,\alpha\, r^2},\ 0,\ -e^{2\,\alpha\, r^2}\right). \tag{7}$$

The (6) satisfies the Killing equation $\eta_{\mu\,;\,\nu} + \eta_{\nu\,;\,\mu} = 0$. For cyclicly symmetric metric, the norm $\eta_\mu\,\eta^\mu$ of the Killing vector is spacelike, closed orbits [24–28]. We note that

$$\eta^\mu\,\eta_\mu = -t\, e^{2\,\alpha\, r^2} \tag{8}$$

which is spacelike for $t < 0$, closed orbits (ϕ co-ordinate being periodic).

An important note is that the Riemann tensor $R_{\mu\nu\rho\sigma}$ can be expressed in terms of the metric tensor $g_{\mu\nu}$ as

$$R_{\mu\nu\rho\sigma} = k\left(g_{\mu\rho}\, g_{\nu\sigma} - g_{\mu\sigma}\, g_{\nu\rho}\right) \tag{9}$$

where $k = -\alpha^2$ for the spacetime (2).

Another important note is that if we take $\beta = 0$, then the spacetime represented by (2) is maximally symmetric vacuum spacetime and locally isometric anti-de Sitter space in four-dimension. One can easily show by a number of transformations the standard form of locally isometric AdS_4 metric [29]

$$ds^2 = \frac{3}{(-\Lambda)\, x^2}\left(-dt^2 + dx^2 + d\phi^2 + dz^2\right) \tag{10}$$

where one of the co-ordinate ϕ being periodic.

3 Matter distribution of the spacetime and energy condition

Einstein's field equations taking into account the cosmological constant

$$G^{\mu\nu} + \Lambda\, g^{\mu\nu} = T^{\mu\nu}\,, \quad \mu, \nu = 1, 2, 3, 4\,. \tag{11}$$

Consider the energy-momentum tensor of pure radiation field [30]

$$T^{\mu\nu} = \rho\, n^\mu\, n^\nu \tag{12}$$

where n^μ is the null vector defined by

$$n^\mu = (0, 0, 0, 1)\,. \tag{13}$$

The non-zero component of the energy-momentum tensor

$$T^t_\phi = -\rho\, e^{2\,\alpha\, r^2}\,. \tag{14}$$

Equating field equations (11) using (3) and (14), we get

$$\begin{aligned} &\Lambda = -3\,\alpha^2, \\ &\rho = \frac{1}{2}\,\beta^2\, e^{-8\,\alpha\, r^2}, \quad 0 \leqslant r < \infty\,. \end{aligned} \tag{15}$$

The energy-density of pure radiation or null dust decreases exponentially with r and vanish at $r \to \pm\infty$. The matter field pure radiation satisfy the energy condition and the energy density ρ is always positive.

4 Classification and kinematical properties of the spacetime

For classification of the spacetime (2), we can construct the following set of null tetrads $(k, l, m, \bar{m})$ as

$$k_\mu = (0, 1, 0, 0)\,, \tag{16}$$

$$l_\mu = \left(-2\,\beta\, z\, r\, e^{-\alpha\, r^2},\ \frac{t}{2}\, e^{2\,\alpha\, r^2},\ 0,\ e^{2\,\alpha\, r^2}\right), \tag{17}$$

$$m_\mu = \frac{1}{\sqrt{2}}\,(2\, r, 0, i\, e^{\alpha\, r}, 0)\,, \tag{18}$$

$$\bar{m}_\mu = \frac{1}{\sqrt{2}}\left(2\, r, 0, -i\, e^{\alpha\, r^2}, 0\right), \tag{19}$$

where $i = \sqrt{-1}$. The set of null tetrads above are such that the metric tensor for the line element (2) can be expressed as

$$g_{\mu\nu} = -k_\mu\, l_\nu - l_\mu\, k_\nu + m_\mu\, \bar{m}_\nu + \bar{m}_\mu\, m_\nu\,. \tag{20}$$

The vectors (16)–(19) are null vectors and are orthogonal except for $k_\mu\, l^\mu = -1$ and $m_\mu\, \bar{m}^\mu = 1$. Using this null tetrad above, we have calculated the five Weyl scalars

$$\begin{aligned} &\Psi_3 = -\frac{i\,\alpha\,\beta\, e^{-2\,\alpha\, r^2}}{2\,\sqrt{2}}\,, \\ &\Psi_4 = -\frac{1}{4}\,\beta\, e^{-2\,\alpha\, r^2}\left(i + 2\,\alpha\, z\, e^{\alpha\, r^2}\right) \end{aligned} \tag{21}$$

are non-vanishing, while $\Psi_0 = \Psi_1 = \Psi_3 = 0$. The spacetime represented by (2) is of type III in the Petrov classification scheme. Note that the non-zero Weyl scalars Ψ_3 and Ψ_4 are finite at $r \to 0$ and vanish as $r \to \pm\infty$ indicating asymptotic flatness of the spacetime (2). The metric (2) is free

from curvature singularities. The curvature invariant known as Kretchsmann scalar is given by

$$R^{\mu\nu\rho\sigma} R_{\mu\nu\rho\sigma} = 24\,\alpha^4 \tag{22}$$

and the curvature scalar

$$R = -12\,\alpha^2 \tag{23}$$

are constant being non-zero.

Using the null tetrad (16) we have calculated the *Optical* scalars [30] the *expansion*, the *twist* and the *shear* and they are

$$\begin{aligned} \Theta &= \frac{1}{2}\, k^{\mu}_{;\mu} = 0\,, \\ \omega^2 &= \frac{1}{2}\, k_{[\mu\,;\,\nu]}\, k^{\mu\,;\,\nu} = 0\,, \\ \sigma\,\bar{\sigma} &= \frac{1}{2}\, k_{(\mu\,;\,\nu)}\, k^{\mu\,;\,\nu} - \Theta^2 = 0 \end{aligned} \tag{24}$$

and the null vector (16) satisfy the geodesics equation

$$k_{\mu\,;\,\nu}\, k^{\nu} = 0\,. \tag{25}$$

Thus the spacetime represented by (2) is non-diverging, has shear-free null geodesics congruence. One can easily show that for constant r and z, the metric (2) reduces to conformal Misner space in 2D

$$ds^2_{confo} = \Omega\, ds^2_{Misn} \tag{26}$$

where $\Omega = e^{2\,\alpha\, r^2}$ is a constant.

5 Conclusion

Our primary motivation in this paper is to write down a metric for a spacetime that incorporates the Misner space and its causality violating properties and to classify it. The solution presented here is non-vacuum, cyclicly symmetric metric (2) and serves as a model of time-machine spacetime in the sense that CTC appear at a definite instant of time on the z-plane. Most of the CTC spacetimes violate one or more energy conditions or unrealistic matter source and are unphysical. The model discussed here is free from all these problems and matter distribution is of pure radiation field with negative cosmological constant satisfying the energy condition.

Submitted on July 25, 2016 / Accepted on August 9, 2016

References

1. Godel K. An example of a new type of cosmological solutions of Einsteins field equations of gravitation. *Rev. Mod. Phys.*, 1949, v. 21 (3), 447–450.
2. van Stockum W. J. The gravitational field of a distribution of particles rotating about an axis of symmetry. *Proc. R. Soc. Edin.*, 1937, v. 57, 135–154.
3. Tipler F. J. Rotating Cylinders and the Possibility of Global Causality Violation. *Phys. Rev. D*, 1974, v. 9 (8), 2203–2206.
4. Gott J. R. Closed timelike curves produced by pairs of moving cosmic strings: Exact solutions. *Phys. Rev. Lett.*, 1991, v. 66 (9), 1126–1129.
5. Bonnor W. B. An exact, asymptotically flat, vacuum solution of Einstein's equations with closed timelike curves. *Class. Quantum Grav.*, 2002, v. 19, 5951–5957.
6. Bonnor W. B. Closed timelike curves in general relativity. *Int. J. Mod. Phys. D*, 2003, v. 12, 1705–1708.
7. Bonnor W. B., Steadman B. R. Exact solutions of the Einstein Maxwell equations with closed timelike curves. *Gen. Rel. Grav.*, 2005, v. 37 (11), 1833–1844.
8. Kerr R. P. Gravitational Field of a Spinning Mass as an Example of Algebraically Special Metrics. *Phys. Rev. Lett.*, 1963, v. 11 (5), 237–238.
9. Carter B. Global structure of the Kerr family of gravitational fields. *Phys. Rev. D*, 1968, v. 174, 1559–1571.
10. Gürses M., Karasu A., Sarioğlu Ö. Godel-type metrics in various dimensions. *Class. Quantum Grav.*, 2005, v. 22 (9), 1527–1543.
11. Alcubierre M. The warp drive: hyper-fast travel within general relativity. *Class. Quantum Grav.*, 1994, v. 11, L73–L77.
12. Lobo F. S. N., Crawford P. Time, closed timelike curves and causality. The Nature of Time : Geometry, Physics and Perception. NATO Sci. Ser. II, 2003, v. 95, 289–296.
13. Morris M. S., Thorne K. S., Yurtsever U. Time Machines and the Weak Energy Conditions. *Phys. Rev. Lett.* , 1988, v. 61 (13), 1446–1449.
14. Morris M. S., Thorne K. S. Wormholes in space-time and their use for interstellar travel: A tool for teaching general relativity. *Amer. J. Phys.*, 1988, v. 56, 395–412.
15. Krasnikov S. V. Hyper-fast interstellar travel in general relativity. *Phys. Rev. D*, 1998, v. 57 (8), 4760–4766.
16. Everett A. E., Roman T. A. A superluminal subway: the Krasnikov tube. *Phys. Rev. D*, 1997, v. 56, 2100–2108. arXiv: gr-qc/9702049.
17. Ori A. A new time-machine model with compact vacuum core. *Phys. Rev. Lett.* , 2005, v. 95, 021101–021107.
18. Ori A. Formation of closed timelike curves in a composite vacuum/dust asymptotically-flat spacetime. *Phys. Rev. D*, 2007, v. 76, 044002–044015.
19. Ori A. Improved time-machine model. *Phys. Rev. D*, 1996, v. 54 (8), 4858–4861.
20. Ori A. Causality violation and the weak energy condition. *Phys. Rev. D*, 1994, v. 49 (8), 3990–3997.
21. Sarma D., Patgiri M., Ahmed F. U. Pure radiation metric with stable closed timelike curves. *Gen. Rel. Grav.*, 2014, v. 46, 1633–1641.
22. Misner C. W. Taub-NUT Space as a Counterexample to Almost Anything. In Ehlers J., ed. Relativity Theory and Astrophysics 1. Relativity and Cosmology, Am. Math. Soc., Providence, RI, 1967, 160–169.
23. Levanony D., Ori A. Extended time-travelling object in Misner space. *Phys. Rev. D*, 2011, v. 83, 044043. arXiv: gr-qc/1102.0907.
24. Mars M., Senovilla J. M. M. Axial symmetry and conformal Killing vectors. *Class. Quantum Grav.*, 1993, v. 10, 1633–1647.
25. Carter B. The commutation property of a stationary, axisymmetric system. *Commun. Math. Phys.*, 1970, v. 17, 233–238.
26. Branes A. Some restrictions on the symmetry groups of axially symmetric spacetimes. *Class. Quantum Grav.*, 2001, v. 18, 5511–5520.
27. Bičák J., Schmidt B. D. Isometries compatible with gravitational radiation. *J. Math. Phys.*, 1984, v. 25 (3), 600–606.
28. Wald R. M. General Relativity. Cambridge University Press, Cambridge, 1984.
29. Zofka M., Bičák J. Cylindrical spacetimes with $\Lambda \neq 0$ and their sources. *Class. Quantum Grav.*, 2008, v. 25 (1), 015011. arXiv: gr-qc/0712.2144
30. Stephani H., Kramer D., MacCallum M., Hoenselaers C., Herlt E. Exact Solutions to Einstein's Field Equations, Cambridge Univ. Press, 2003.

Experimental and Theoretical Test of Cahill's Detection of Absolute Velocity in Gas-mode Interferometer Experiments

Jay R. Seaver

Energy Matters Foundation, PO BOX 2588, Longmont, CO 80502, USA
E-mail: jay@energy-matters.org

Several papers by Cahill, et al. assert that Michelson-Morley type experiments performed in gas have small but non-null results which, when properly analyzed, show that the absolute speed of the earth was detected. Here we show that Cahill made a fundamental error in his assumptions and that the mathematical analysis upon which he based his conclusions is invalid. We also include a report on an experiment that verifies these mathematical conclusions. The experiment uses water instead of air as the wave medium. The much larger index of refraction of water (1.33 vs. 1.00029) greatly amplifies the effect Cahill predicts and makes the null result of the new experiment dramatically apparent. This confirms both theoretically and experimentally that absolute velocity was not and cannot be detected in Michelson-Morley type experiments regardless of the refractive medium in which they are performed.

1 Introduction

I was intrigued by several papers by Cahill [1–4] that purport to re-evaluate the original Michelson-Morley (MM) and other "gas-mode" interferometer experiments and prove that they actually measured the absolute speed of the earth through space. Cahill shows in these papers that the index of refraction of air caused results that although small were not completely null. He asserts that the absolute velocity of the earth was measured and that absolute space was detected — but was it?

I set out to test Cahill's assertions by designing an experiment capable of getting a larger non-null result. This experiment uses water as the medium through which the light propagates so that the "incomplete cancellation of the geometrical effects" (according to Cahill) would be greatly amplified by the much larger index of refraction. This allows easy detection of the interference-fringe shifts in a low-cost Michelson-type interferometer.

The experiment had a resolution that was more than 10^3 times greater than the effect Cahill's equations predicted. The results of the experiment were unequivocally null. Based on the null results, I set out to reexamine Cahill's assumptions and mathematical derivations. It was through this reexamination that I derived the correct equations and proved that the so-called "cancellation of the geometrical effects" is complete and the results of any MM type experiment must be null whether done in vacuum or in a refractive medium. We show that both the herein derived equations and the results of the present experiment are in complete agreement that absolute space cannot be detected with these types of experiments.

Our derivations (and Cahill's) are based on classical physics. By "classical physics" we mean merely that the equations of the special theory of relativity (SRT) will not be used to transform values between inertial reference frames. All measurements in the derivations are made in the rest frame (or what Cahill calls the "quantum foam" frame) where light-speed is constant and isotropic. But in SRT, light-speed is constant and isotropic in *all* frames. Therefore our derivations will be in complete compliance with the formalism of SRT, while at the same time satisfying Cahill and his followers that they are also valid in Cahill's absolute frame.

The value measured in the experiment is the shift, measured in wavelengths, of the interference pattern of two light beams. Because this measurement is a scalar value, independent of the actual length of the wavelength, it is invariant in all reference frames. This is what allows us to do the entire analysis from the rest frame but make the actual measurement in the laboratory frame — they must agree.

2 Correcting Cahill's derivations

We will use Cahill's equations as derived in [1] for this analysis.

Cahill begins his analysis by making the following (incorrect) assumption regarding the speed of light in the refractive medium of air: "If the gas is moving with respect to the quantum foam, as in an interferometer attached to the earth, then the speed of light relative to the quantum foam is still $V = c/n$ up to corrections due to the Fresnel drag. *But this dragging is a very small effect and is not required in the present analysis*". [emphasis added]) He is correct that Fresnel drag is a very small effect, but as will soon be evident, it is not small compared to the effect he is trying to measure and it cannot be ignored.

The laboratory frame is assumed to have an arbitrary velocity v with respect to the rest frame. We also make the following two assumptions which Cahill made in his analysis and which are entirely consistent with SRT: 1) clocks slow down with velocity and 2) lengths contract with velocity. The

factor by which they slow down is defined as

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

For convenience, we also make the following definition:

$$\beta = \frac{v}{c} \quad \Rightarrow \quad \gamma = \frac{1}{\sqrt{1 - \beta^2}}. \quad (2)$$

If both arms of the interferometer are of rest-length L and one is aligned parallel to the velocity of the laboratory and the other is aligned at right angles to this velocity, then the length of the orthogonal arm in the rest frame is still L, but the length of the parallel arm experiences a contraction if measured in the rest frame,

$$L_{\|} = L\sqrt{1 - \beta^2} = \frac{L}{\gamma}. \quad (3)$$

Cahill defines n to be the index of refraction of the gas and uses the same value n in both frames. This seems perfectly reasonable, since n is a scalar and therefore invariant. But just because it has the same value in both frames does not mean that it affects the path of the waves in both frames the same way. This will be demonstrated by observing *from within the rest frame* how observers *within the moving frame* measure and define n. It then becomes apparent that in the rest frame the velocity of light in a moving refractive medium is not simply c/n plus the traditional drag term.

Before observing how n is measured, we must first understand how clocks are synchronized using Einstein's method. We will do this by observing from the rest frame as clocks are synchronized in the laboratory frame. Let there be clocks at each end of the arm aligned parallel to the velocity which we designate as clock A and clock B. According to Eq. (3) this distance between the two clocks is L/γ in the rest frame. The procedure for synchronizing the two clocks in the moving frame is as follows:

1. A light wave leaves clock A at time 0 on clock A in the moving frame and also at time 0 in the rest frame.
2. The light beam propagates towards clock B at velocity c in both frames. In the rest frame clock B is moving at velocity v in the same direction as the light beam.
3. The light arrives at B at time t_1 in the rest frame.
4. The total distance the light travels in the rest frame on the outbound path is $c\,t_1$. This can be separated into two distances: 1) the length of the contracted arm L/γ and the distance clock B moved during the time t_1 which is $v\,t_1$. Solving for t_1, we get

$$t_1 = \frac{L}{\gamma\,(c - v)}. \quad (4)$$

5. The light reflects from a mirror at B and returns to A at time t_2 in the rest frame. Since the clock at A was moving towards the light during this leg, the distance that the light traveled before reaching A was $L/\gamma - v\,(t_2 - t_1)$. Using the same logic as above, the time $t_2 - t_1$ to make the return trip as measured in the rest frame is

$$t_2 - t_1 = \frac{L}{\gamma\,(c + v)}. \quad (5)$$

6. Solving for t_2, the total time to make the round trip as measured in the rest frame is

$$t_2 = \frac{L/\gamma}{c + v} + \frac{L/\gamma}{c - v} = \frac{2\,L/\gamma}{c\,(1 - v^2/c^2)} = \frac{2\,L}{c}\gamma. \quad (6)$$

7. The clocks in the moving frame run slower by a factor of γ than the clocks in the rest frame. Therefore, the time on clock A when the light returns is

$$t_A = \frac{t_2}{\gamma} = \frac{2\,L}{c}. \quad (7)$$

8. Using Einstein's method of synchronization, clock B is defined to be synchronized to clock A if at the moment of reflection the time on clock B is set to $t_A/2$.

$$t_B = \frac{L}{c}. \quad (8)$$

As expected, the observers in the laboratory frame measure the speed of light to be c in both directions. But notice that at the moment of reflection of the light from clock B, the time is t_1 in the rest frame and t_B on clock B in the moving frame. But what is the time on clock A at that moment? Since clock A was defined to be 0 at time 0 in the rest frame, and since clock A runs slower by a factor of γ than clocks in the rest frame, the time on clock A must be t_1/γ. But that means that to an observer in the rest frame, there is a bias between clocks A and B,

$$t_{bias} = t_B - \frac{t_1}{\gamma} = \frac{L}{c} - \frac{L}{\gamma^2\,(c - v)} = -\frac{v\,L}{c^2}. \quad (9)$$

Please note that this is in complete agreement with SRT. Position-dependent clock biases are the source of relative simultaneity in SRT. Events are defined to be simultaneous in the moving frame when the clocks at the sites of the two events read the same value. But because of the permanent bias between the clocks (when observed from the rest frame), those same two events are never simultaneous within the rest frame. From this exercise we see that there is nothing mysterious or magical about relative simultaneity — it is simply a byproduct of defining the one-way time of flight of a light wave to be 1/2 of the two-way time of flight.

The bias in Eq. (9) is the same position-dependent bias that occurs in the transformation of time between frames using the Lorentz transformation of SRT. But we have determined its value not by performing this transformation but by simply observing from the rest frame as clocks were synchronized in the moving frame. We have used nothing more than

this definition and classical physics to derive the same bias between the clocks as defined in SRT.

Now that we understand how clocks in the moving frame appear to observers in the rest frame, we are ready to see how the index of refraction, when measured in the laboratory, appears to observers in the rest frame. To measure the index of refraction in the laboratory, a light beam is sent from clock A at time 0 through a refractive material and arrives at clock B at time t_{Bn}, where the n in the subscript indicates time through the refractive material. This is the time of flight of the light beam as measured in the laboratory. The index of refraction is then defined as

$$n = \frac{c\, t_{Bn}}{L}. \tag{10}$$

This corresponds to a velocity of light in the refractive medium of c/n as measured in the laboratory. Let us now look at that same velocity as measured in the rest frame. Because of the bias on clock B, although the time on clock A is 0 when the light is emitted, the observer in the rest frame sees the light wave leave clock A when clock B reads $-v\,L/c^2$. The elapsed time on clock B for the time of flight is therefore

$$\Delta t_{Bn} = t_{Bn} + \frac{v\,L}{c^2}. \tag{11}$$

Using Eq. (10) to substitute for t_{Bn}, and remembering that clocks in the moving frame run slower by a factor of γ, the elapsed time in the rest frame for the time of flight is

$$\Delta t_0 = \frac{L\,(c\,n + v)\,\gamma}{c^2}. \tag{12}$$

We defined the direction from A to B to be the same direction as the velocity of the moving frame. Since lengths contract with velocity, the total distance the light propagated during this time, as measured in the rest frame, is

$$\Delta d_0 = \frac{L}{\gamma} + v\,\Delta t_0 = \frac{L}{\gamma} + \frac{v\,L\,(c\,n + v)\,\gamma}{c^2}. \tag{13}$$

The velocity of the light beam in the refractive material as measured in the rest frame is this distance divided by the propagation time, which simplifies to

$$c_{n0+} = \frac{\Delta d_0}{\Delta t_0} = \frac{c\,(c + n\,v)}{c\,n + v}. \tag{14}$$

Notice that this can be put in the following form:

$$c_{n0+} = \frac{c/n + v}{1 + \dfrac{(c/n)\,v}{c^2}}. \tag{15}$$

In this form it is very obvious that we have derived the velocity addition formula of SRT where the two velocities are c/n and v. This shows that there is nothing mysterious about the velocity addition formula of SRT. It is easily derived using classical physics if one acknowledges that clocks and lengths change with velocity. The only mystery is what causes velocity-dependent lengths and clock-rates in the first place. But that is a topic for a separate paper.

We can also write this equation in a different form,

$$c_{n0+} = \frac{c}{n} + \left(\frac{n^2 - 1}{n^2}\right)\left(\frac{n}{n + \beta}\right)v. \tag{16}$$

In this form, we can clearly see that the Fresnel drag coefficient is simply a consequence of the velocity addition formula. They are not separate phenomena. Prior to Lorentz and Einstein, it was thought that the Fresnel drag term consisted only of the $\dfrac{n^2 - 1}{n^2}v$ term. The $\dfrac{n}{n + \beta}$ term is so close to 1 that except for extremely high velocities it was unobservable.

What we have shown in this derivation is that the Fresnel drag term is automatically included in our derivation once we acknowledge that lengths and times change with velocity. In fact, Fresnel drag is proof that lengths and times really do change with velocity.

When the light is sent in the opposite direction through the refractive medium, the sign of the laboratory's velocity v in equation (14) is inverted resulting in a reverse speed of

$$c_{n0-} = \frac{c\,(c - n\,v)}{c\,n - v}. \tag{17}$$

Summing the times of propagation for these out and back velocities, we can calculate the total time for a round trip on the parallel arm in the rest frame if the light is passing through a moving refractive medium with an index of refraction n:

$$\Delta t_{\|0} = \frac{L}{\gamma\,(c_{n0+} - v)} + \frac{L}{\gamma\,(c_{n0-} + v)} = 2\frac{L}{c/n}\gamma. \tag{18}$$

Not surprisingly, this is the same value we would have calculated if we had simply used the Lorentz transforms of SRT to transform the time on clock A into the rest frame for a round trip of length $2\,L$ at velocity c/n. Be we have derived it using nothing but classical physics and the two assumptions regarding length contraction and the slowing of clocks with velocity.

We will now look at the time for the round trip on the orthogonal arm. In the laboratory frame, n has the same value in all directions.* Therefore, as measured in the laboratory frame, the round-trip time in the orthogonal direction is

$$\Delta t_{\perp} = \frac{2\,L}{c/n} = \frac{2\,L\,n}{c}. \tag{19}$$

With the arm oriented orthogonal to the velocity, the light-propagation times for the outbound and return trips are equal in the rest frame so there is no bias between clocks A and B. Since clocks in the moving frame run slower when observed

*This is proven in Section 4.2 where the velocity of light in a moving refractive medium is derived for any arbitrary direction.

from the stationary frame, this same time in the stationary frame is simply the elapsed time in the rest frame multiplied by γ,

$$\Delta t_{\perp 0} = 2\frac{L}{c/n}\gamma. \tag{20}$$

We see that this is exactly the same as the time for the parallel path given in equation (18) so the MM experiment is doomed to give null results regardless of the index of refraction of the medium.

3 Comparing to Cahill's results

We now compare these results to Cahill's results (we use subscript C for Cahill's times), which come from his equations (7) and (10) in [1]:

$$\left.\begin{aligned} \Delta t_{\parallel C} &= \frac{2L}{\gamma\frac{c}{n}\left(1-\frac{v^2}{c^2}n^2\right)} = \frac{2L\gamma}{c/n}\left(\frac{1}{\gamma^2\left(1-n^2\beta^2\right)}\right) \\ \Delta t_{\perp C} &= \frac{2L}{\sqrt{\frac{c^2}{n^2}-v^2}} = \frac{2L\gamma}{(c/n)}\left(\frac{1}{\gamma\sqrt{1-n^2\beta^2}}\right) \end{aligned}\right\}. \tag{21}$$

The right-most terms in parenthesis are the error factors Cahill introduced by ignoring the "drag" effect. Without these terms, the times are identical. Notice that both of these error terms are very close to 1. In fact for a velocity of 360 km/sec and $n = 1.00029$ (which are the approximate values Cahill used in his paper), the two terms in parenthesis are $(1 + 8 \times 10^{-10})$ and $(1 + 4 \times 10^{-10})$, respectively. It is easy to see why Cahill thought they could be ignored and simply set to 1.

The difference between equations (21) is Cahill's measured time difference between the parallel and orthogonal orientations. It can be shown that for $v/c = \beta << 1$ this difference can be approximated by

$$\Delta t_{\parallel C} - \Delta t_{\perp C} = \frac{Ln}{\gamma c}\left(\frac{\left(n^2-1\right)\beta^2}{1-n^2\beta^2}\right). \tag{22}$$

In the original MM experiment, $L = 11$ and $n \approx 1.00029$. The absolute velocity that Cahill calculated was on the order of 360 km/sec, which results in $\beta \approx 0.0012$. Substituting these into equation (22) results in a measured time difference of

$$\Delta t_{\parallel C} - \Delta t_{\perp C} \approx 3.1 \times 10^{-17}. \tag{23}$$

This confirms Cahill's estimate of a difference on the order of 10^{-17} sec. The wavelength of light used in the original experiment was approximately 600 nm which for a velocity of c has a temporal period of about 2.0×10^{-15} sec. Since there is one spatial period (wavelength) for each temporal period, the fringe shift in wavelengths is the total time delay of Eq. (23) divided by the temporal period of the light wave:

$$\Delta\lambda = \frac{3.1 \times 10^{-17}}{2.0 \times 10^{-15}}\lambda \approx 0.016\lambda. \tag{24}$$

This represents a predicted fringe shift of about 1.6% of a wavelength in the original experiment. It is this value that Cahill used to predict the non-null results.

We conclude that Cahill made a fatal mistake when he assumed he could ignore the Fresnel drag effects. It is precisely the ignoring of Fresnel drag that creates the 1.6% difference in phase. Quoting Cahill, "Of course experimental evidence is the final arbiter in this conflict of theories." In that spirit, we will present the design and results of an experiment that proves that an index of refraction greater than 1 does not give non-null results in Michelson-interferometer experiments as Cahill asserts.

Cahill's analysis of the raw data from the original MM experiment shows a non-null result which is sidereal in nature and which agrees, according to Cahill, with his above calculations. It is beyond the scope of this paper to address the source of the non-null, sidereal effect found in the raw data. But one paper that has addressed this issue shows that the very large drift in the experiment combined with an improper statistical analysis is entirely responsible for the apparent non-null result [5].

4 Design of the new experiment

The analysis of the experiment to test Cahill's results is again done as if we are an observer in a rest frame where light speed is isotropic. Since we are constrained to make all of the actual measurements in the moving frame of our laboratory, we define the results of the experiment in terms of an invariant scalar value that will have the same value in all frames. This is done by measuring the shift of an interference pattern in units of wavelengths. This is a scalar value that must be the same in all frames and allows us to make measurements in the moving frame that are in full agreement with those same measurements made in the hypothetical rest frame.

As mentioned above, the non-null result that Cahill predicted is less than 2% of a wavelength. This is much too small to be measured in an inexpensive, home-built interferometer. To increase the sensitivity of the experiment, the index of refraction was increased from 1.00029 of air to 1.33 of water. Of course, the experiment cannot be done completely submerged in water, so a refractive block containing water was introduced into one of the paths.

Figure 1 shows the physical layout of the experiment. A laser emits a beam that is split into two separate beams. One beam travels exclusively through air on its path to the detector. The other beam travels the same distance, but part of this path passes through a block of refractive material of length L that slows the wave down. When it exits the refractive block (RB), it then continues at the normal speed of light until it is recombined with its sister beam at the detector. Distilled water with an index of refraction of 1.33 is used for the refractive block. Unfortunately, using a refractive block is not the same as performing the entire experiment while im-

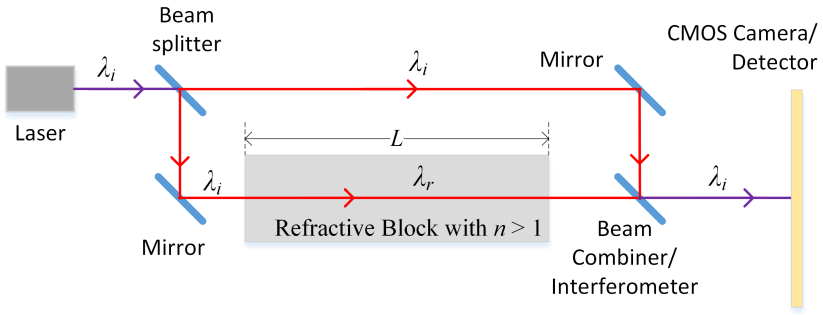

Fig. 1: Layout of experiment.

mersed in a high-refractive medium. Specifically, it complicates the mathematics by introducing refraction in the beam as it passes through the boundary between the air and the water. But the complication is worth it because it allows a large enough fringe shift (according to Cahill's equations) that even our inexpensive interferometer is sensitive enough to measure it.

The wavelength that is emitted from the laser after taking into account the velocity-dependent slowing of the clocks and the Doppler shift, is designated as the incident wavelength λ_i. At the first beam splitter, one wave goes straight along an unrefracted path. The other beam gets reflected downward (in the figure) before reflecting off a second mirror that puts it on a trajectory that is parallel to the first beam but which passes through the refractive block. The two beams recombine at the beam combiner and propagate together to the camera detector. The phase and frequency shifts due to the reflection of the mirrors and beam splitters are exactly the same for the two paths and exactly cancel one another so they can be ignored. The wavelength leaving the laser and arriving at the phase detector is also the same for both paths.

It is the phase relationship between the two beams at the detector that we are interested in. Since the entire path is identical for both beams except for the length L of the RB, we only need to calculate the phase shift that occurs through the RB and compare it to the phase change that occurs over this same distance in the other path to account for the entire phase shift at the detector. All other effects will be identical on both paths and cannot alter the phase difference caused by the delay through the RB. By rotating the experiment 90 degrees we can measure the phase shift in each direction. Any difference between the two directions is a measure of absolute velocity through space — which Cahill predicts will be non-zero.

4.1 Velocity and the path of the beam

In this analysis, we are only going to look at the two cases where the velocity of the laboratory is orthogonal to the beam and parallel to the beam, respectively. We will be discussing multiple angles in this analysis. To keep these angles straight, the following definitions will be used:

1. The symbol φ will be used for the angle between the velocity vector of the refractive medium (laboratory) and the light wave path within the medium. It will have no subscript in the moving frame and a 0 subscript in the rest frame.
2. The symbol θ_i will be used for the incident angle of the wave path at the surface of the refractive block. It is defined as the angle between the light wave path and the normal to the refractive surface, which is the standard definition from geometric optics. It will have a subscript 0 when measured in the rest frame and no additional subscript in the moving frame.
3. The symbol θ_r will be used for the refracted angle of the wave path within the refractive block. It is defined as the angle between the light wave path in the RB and the line that is normal to the refractive surface, which is again the standard definition from geometric optics. It will have a subscript 0 when measured in the rest frame and no additional subscript in the moving frame.
4. In the case where the velocity is parallel to the line that is normal to the refractive surface, the θ angles will have an additional $\|$ symbol in the subscript. If the velocity is orthogonal to the normal a $\perp$ symbol will be used. Since the φ angles are by definition between the light path and the velocity, no subscript is necessary to indicate velocity direction.

Figure 2 shows a laser diode with a highly divergent beam that is collimated using an aperture. In actual lasers, a collimating lens is used instead of an aperture because a lens can capture most of the light. Obviously the aperture loses all of the light that doesn't pass through it. But for our purposes the math and visualization is easier with the aperture and the principle is the same. The view in Figure 2 is for a laser that is stationary with respect to the observer.

Figure 3 shows what happens to the path of the beam if the laser is moving up (orthogonal to beam) in this figure at velocity v. The laser and aperture position are shown at time t for an emission that occurred at time 0. Notice that during the time that a wave front in the beam travels a distance ct (in vacuum), the aperture and laser move a distance vt. This

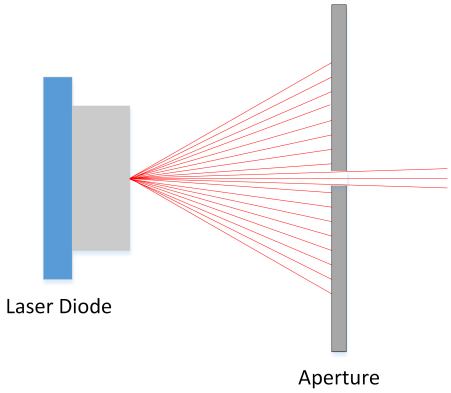

Fig. 2: Laser collimation using an aperture.

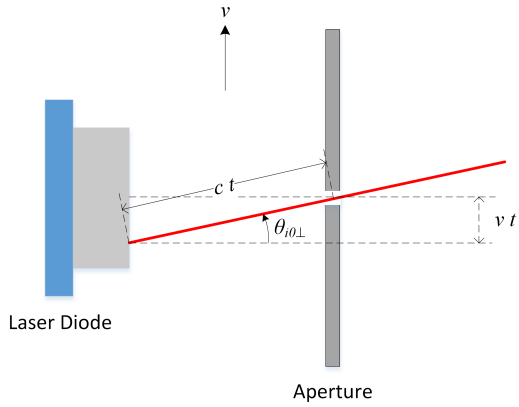

Fig. 3: Path of a single wavelet/photon for orthogonal direction.

means that only waves that left the laser at an angle of $\theta_{i0\perp}$ (in the rest frame) make it through the aperture — hence we have only shown one path in the figure. This angle assures that the orthogonal component of the velocity of the wave is exactly v and the parallel velocity of the wave is $\sqrt{c^2 - v^2}$.

Since every wave leaving the laser that makes it through the aperture follows a similar path, the resulting beam, which is made up of all of these individual wavelets, appears to remain perfectly aligned with the laser and with the aperture. The solid red line in Figure 3 shows the path of an individual wavelet from its emission at the laser surface to its exit from the aperture. Although the wavelet moves at an angle , the beam one would see at any instant in time is the collection of all of the wavelets that have left the laser. A "snapshot" of the positions of several of these wavelets, each on its own unique path, is shown in Figure 4 . Notice that the three wavelets that have been propagating for times t_1, t_2 and t_3 each remain perfectly aligned with each other and with the center of the laser because the aperature assures that their velocity component in the orthogonal direction is exactly v. Any wavelets with different orthogonal velocities are blocked by the aperture.

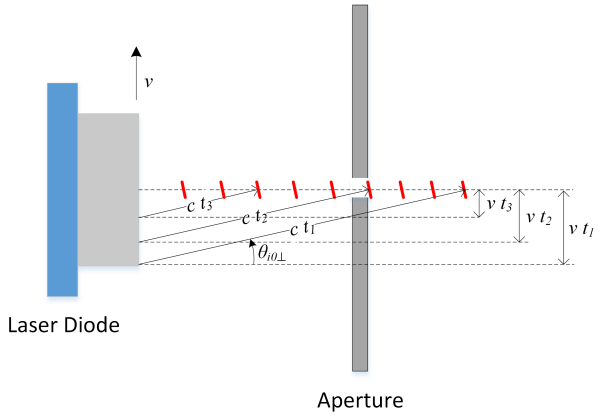

Fig. 4: Snapshot of laser beam for orthogonal direction.

We see that Mother Nature has conspired with light so that an observer in any frame sees a straight, horizontal beam going from the center of the laser through the center of the aperture and arriving at a distant target still centered — just as it appears when the system is stationary. This assures that the path of the composite beam relative to the laboratory is independent of the velocity of the laboratory even though the individual wavelets are moving at a velocity-dependent angle.

Since the index of refraction of air is so close to 1 and since the effect of the index of refraction of the refractive block is so much larger, we are going to simplify the math by treating the air as if the index of refraction were exactly 1. From Figure 3, we can see that the sine and cosine of $\theta_{i0\perp}$ are given by

$$\left.\begin{aligned} \sin\theta_{i0\perp} &= \frac{v}{c} = \beta \\ \cos\theta_{i0\perp} &= \sqrt{1-\sin^2\theta_{i0\perp}} = \sqrt{1-\beta^2} = \frac{1}{\gamma} \end{aligned}\right\}. \qquad (25)$$

4.2 Velocity of light in a moving medium at arbitrary angle

In the orthogonal direction, we can see that the wavelets enter the refractive block at an angle. This means that the wavelet angle will be refracted upon passing through the surface of the RB. The angle of refraction of a moving block cannot be determined by Snell's law alone – it is much more complicated.

Before calculating exactly how a beam refracts in a moving medium, we will first derive the general term for the velocity of light in a moving medium where the angle between the wavelet path and the velocity of the medium is an arbitrary angle between 0 and π.

In the rest frame of the medium, the geometry is as shown in Figure 5. The path AB is that of a laser beam propagating a distance L in a medium with an index of refraction of n. The source A and destination B are on opposite ends of an arm

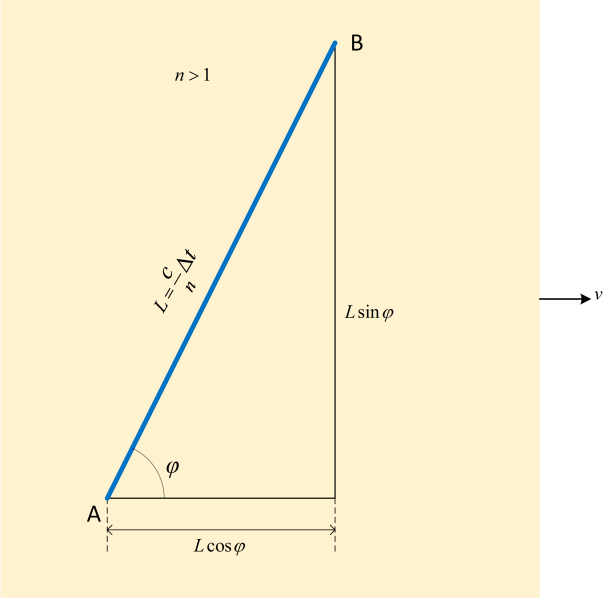

Fig. 5: Beam path at arbitrary angle in rest frame of medium.

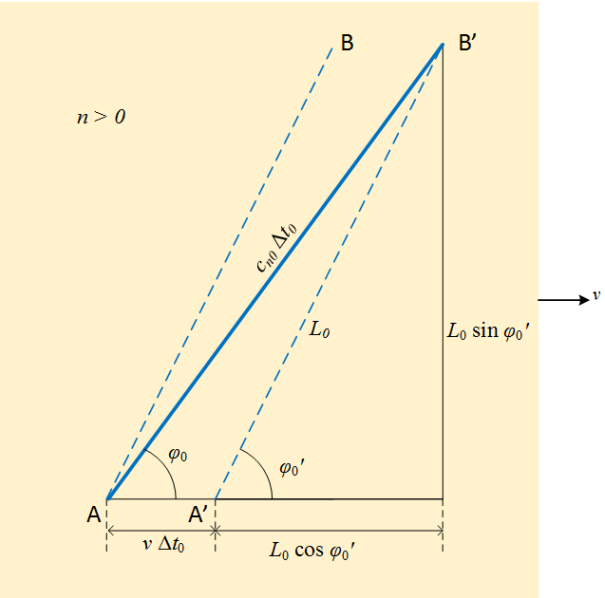

Fig. 6: Wavelet path at arbitrary angle in absolute frame.

of the experiment. The arm and the medium are both moving in the direction shown at velocity v with respect to the rest frame. The values given in the figure are for measurements made by an observer within the moving frame. In this frame, the time for a wavelet in the beam to travel from the source to the destination is by definition of the index of refraction n,

$$\Delta t = \frac{L\,n}{c}. \tag{26}$$

Figure 6 shows the path of the same wavelet if the beam is observed from the rest frame. The dotted lines show the instantaneous positions of the ensemble of wavelets that make up the beam at two different times. And the angle φ_0' is the angle that the visible composite beam makes with the velocity. The bold solid line shows the path that an individual wavelet takes.

In the moving frame, the clocks at A and B are assumed to have been synchronized using Einstein's method. As we derived earlier, synchronizing the clocks in the moving frame will create a bias between the clocks when observed from the rest frame:

$$t_{0bias} = \frac{v\,L\,\cos\varphi}{c^2}. \tag{27}$$

Taking this into account and also accounting for the fact that clocks run slower in the moving frame, the time for a wavelet to propagate from A to B′ in the rest frame is

$$\Delta t_0 = (\Delta t - t_{0bias})\,\gamma = \left(\frac{L\,n}{c} + \frac{v\,L\,\cos\varphi}{c^2}\right)\gamma. \tag{28}$$

Length contraction in the direction of the velocity causes the angle φ in the moving frame to increase to φ_0' in the rest frame (beam and wavelet path are the same within the moving frame). The length of the arm L will decrease in the rest frame to L_0:

$$L_0 = L\sqrt{\frac{\cos^2\varphi}{\gamma^2} + \sin^2\varphi}. \tag{29}$$

Since lengths do not contract in directions orthogonal to the velocity,

$$L\,\sin\varphi = L_0\,\sin\varphi_0'. \tag{30}$$

From the right triangle with hypotenuse AB′ in Figure 6, we get the following relationship for angle φ_0:

$$\sin\varphi_0 = \frac{L\,\sin\varphi}{c_{n0}\,\Delta t_0}. \tag{31}$$

Pythagorean's Theorem requires that

$$(c_{n0}\Delta t_0)^2 = \left(v\,\Delta t_0 + L_0\sin\varphi_0'\right)^2 + \left(L_0\sin\varphi_0'\right)^2. \tag{32}$$

Using equations (30), (31) and (32) we can solve for c_{n0} and $\sin\varphi_0$:

$$c_{n0} = \sqrt{\frac{L_0^2}{\Delta t_0^2} + v^2 \pm \frac{2\sqrt{\left(L_0^2 - L^2\sin^2\varphi\right)v^2}}{\Delta t_0}}, \tag{33}$$

$$\sin\varphi_0 = \frac{L\sin\varphi}{\sqrt{\frac{L_0^2}{\Delta t_0^2} + v^2 \pm \frac{2\sqrt{\left(L_0^2 - L^2\sin^2\varphi\right)v^2}}{\Delta t_0}}}. \tag{34}$$

Substituting equations (28) and (29) into these equations, results in solutions involving only the angle φ in the rest frame of the medium:

$$c_{n0} = c\sqrt{1 - \frac{n^2 - 1}{\gamma^2 (n + \beta\cos\varphi)^2}}\,, \tag{35}$$

$$\sin\varphi_0 = \frac{\sin\varphi}{\sqrt{\gamma^2 (n + \beta\cos\varphi)^2 - n^2 + 1}}\,. \tag{36}$$

From which we can also calculate the cosine:

$$\cos\varphi_0 = \sqrt{1 - \frac{\sin^2\varphi}{\gamma^2 (n + \beta\cos\varphi)^2 - n^2 + 1}}\,. \tag{37}$$

Equation (35) is the speed of an individual wavelet as measured in the rest frame when the medium (i.e. laboratory) is moving at velocity $v = c\beta$. It demonstrates that there is not a unique index of refraction $n_0 = c/c_{n0}$ for a moving medium. The speed of light through the medium is a function of both the velocity of the medium and the angle which the beam makes with that velocity. Cahill ignored the "drag" component and assumed the velocity in the moving medium was the same c/n as in the stationary medium. This is what introduced his error.

The angle φ is the angle as measured in the moving frame between the velocity of the frame and the direction of the light waves. Equations (36) and (37) describe the angle φ_0 at which the light waves are moving in the rest frame in terms of φ in the moving frame.

The velocity of the light wavelets can be separated into two components, one parallel to the laboratory velocity and one orthogonal to the laboratory velocity. In Figure 6, these two components are

$$\left.\begin{aligned} c_{n0\perp} &= c_{n0}\sin\varphi_0, \\ c_{n0\parallel} &= c_{n0}\cos\varphi_0 \end{aligned}\right\}. \tag{38}$$

Substituting equations (35), (36) and (37) into these equations gives us the expressions for the parallel and orthogonal components of wavelet velocity in the rest frame:

$$\left.\begin{aligned} c_{n0\perp} &= \frac{c}{n}\left(\frac{\sin\varphi}{\gamma\left(1 + \frac{\beta}{n}\cos\varphi\right)}\right) \\ c_{n0\parallel} &= \frac{c}{n}\left(\frac{\cos\varphi + \beta n}{1 + \frac{\beta}{n}\cos\varphi}\right) \end{aligned}\right\}. \tag{39}$$

With the parallel and orthogonal components of the velocity, we know everything about the velocity and direction of the wavelets within the moving medium. We are now ready to investigate how this affects the refraction of a beam that is entering a moving medium as observed from the rest frame.

4.3 Refraction of light entering a moving refractive medium

For our analysis of refraction, we will refer to Figure 7 where we have added the incident and refracted angles. This figure again shows a moving refractive medium with index of refraction n as measured in the moving frame. A laser source is attached to and moving along with the refractive medium. Both the medium and the laser are moving at velocity v in the rest frame in the direction shown, which is orthogonal to the line which is normal to the refractive surface. The line normal to the refractive surface will be referred to as the normal line. The medium is shown at two different positions separated in time. The laser source is shown at three different times.

The dotted lines leaving the laser again show the location of the ensemble of wavelets that make up the composite visible laser beam at these times. This is the apparent path of the laser beam. The bold line shows the path that is actually taken by an individual wavelet or photon within the beam in propagating from the source to A and then through the medium to B′.

It is readily apparent that the relationship between the angles is

$$\theta_{r0\perp} = \frac{\pi}{2} - \varphi_0. \tag{40}$$

This can be expressed as

$$\left.\begin{aligned} \sin\varphi_0 &= \cos\theta_{r0\perp} \\ \cos\varphi_0 &= \sin\theta_{r0\perp} \end{aligned}\right\}. \tag{41}$$

These angles as measured in the moving frame will have a similar relationship:

$$\left.\begin{aligned} \sin\varphi &= \cos\theta_{r\perp} \\ \cos\varphi &= \sin\theta_{r\perp} \end{aligned}\right\}. \tag{42}$$

From Snell's law, the incident and refracted angles in the moving frame (i.e. in the rest frame of the RB) are related by

$$\sin\theta_{i\perp} = n\,\sin\theta_{r\perp}. \tag{43}$$

Substituting this into equations (42) results in

$$\left.\begin{aligned} \sin\varphi &= \sqrt{1 - \frac{\sin^2\theta_{i\perp}}{n^2}} \\ \cos\varphi &= \frac{\sin\theta_{i\perp}}{n} \end{aligned}\right\}. \tag{44}$$

Substituting these into equations (39) gives us the parallel and orthogonal components of the wave velocity in the rest frame as a function of the incident angle in the moving frame:

$$c_{n0\parallel} = \frac{c}{n}\left(\frac{\sin\theta_{i\perp} + \beta n^2}{n + \frac{\beta}{n}\sin\theta_{i\perp}}\right) \tag{45}$$

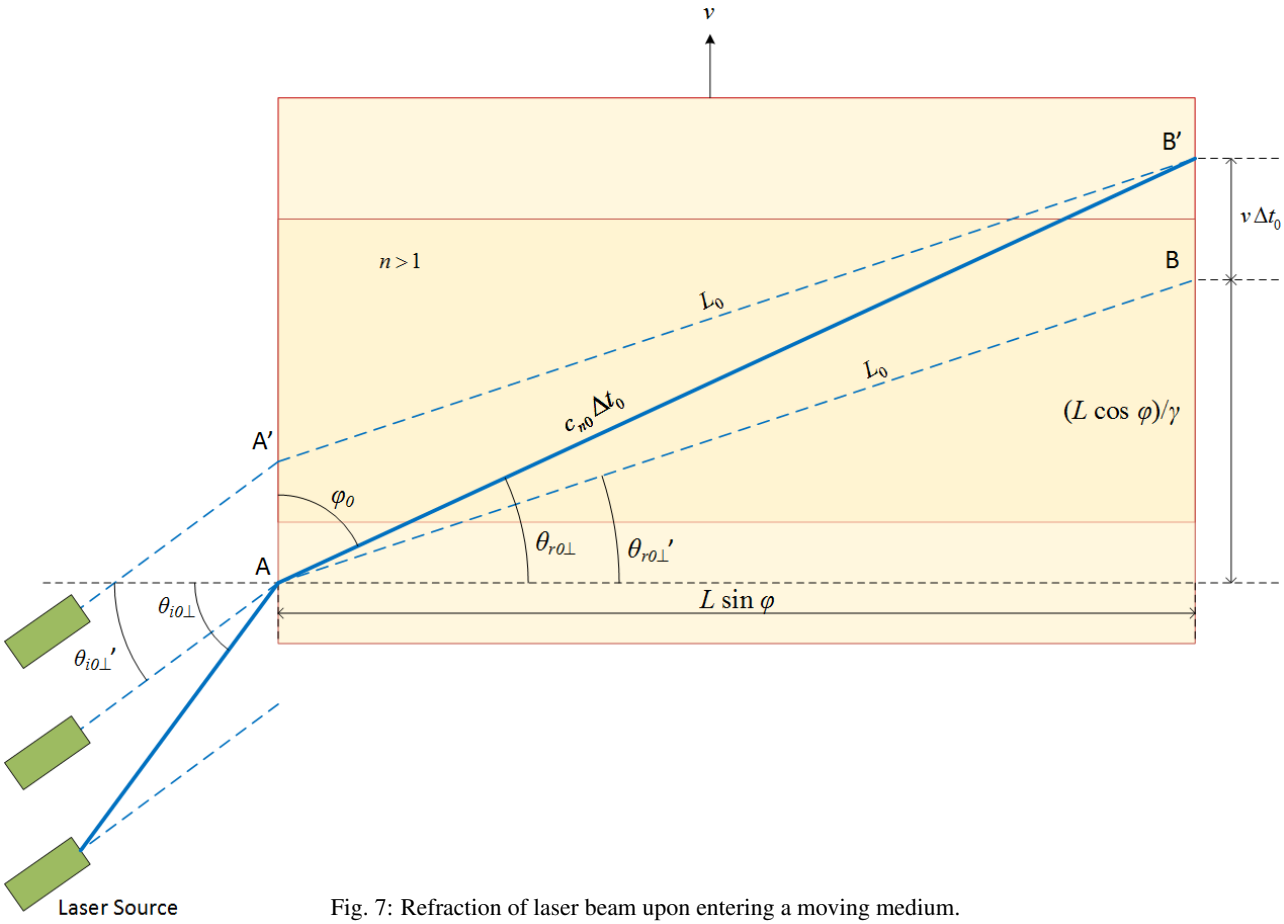

Fig. 7: Refraction of laser beam upon entering a moving medium.

$$c_{n0\perp} = \frac{c}{n}\left[\frac{n\sqrt{1-\dfrac{\sin^2\theta_{i\perp}}{n^2}}}{\gamma\left(n+\dfrac{\beta}{n}\sin\theta_{i\perp}\right)}\right]. \tag{46}$$

A similar set of equations can be obtained when the velocity is parallel to the normal line. Although the parallel case is not shown in the figure, it is easy to see that in this case the refracted angle is equal to φ_0 so that

$$\left.\begin{aligned}\sin\varphi = \sin\theta_{r\parallel} &= \frac{\sin\theta_{i\parallel}}{n}\\ \cos\varphi = \cos\theta_{r\parallel} &= \sqrt{1-\frac{\sin^2\theta_{i\parallel}}{n^2}}\end{aligned}\right\}. \tag{47}$$

Substituting these into equation (39) gives the orthogonal and parallel components of the wave velocity for the parallel orientation:

$$c_{n0\parallel} = \left(\frac{c}{n}\right)\frac{\left(n\beta+\sqrt{1-\dfrac{\sin^2\theta_{i\parallel}}{n^2}}\right)}{1+\dfrac{\beta}{n}\sqrt{1-\dfrac{\sin^2\theta_{i\parallel}}{n^2}}}, \tag{48}$$

$$c_{n0\perp} = \left(\frac{c}{n}\right)\frac{\sin\theta_{i\parallel}}{\gamma\left(n+\beta\sqrt{1-\dfrac{\sin^2\theta_{i\parallel}}{n^2}}\right)}. \tag{49}$$

We now have a complete description of how an incident wave is refracted when it enters a moving refractive medium. Equations (45) and (46) govern the refraction if the medium is moving orthogonal to the normal line. Equations (48) and (49) govern when the medium is moving parallel to the normal line.

4.4 Refraction with $\theta_i = 0$ and velocity orthogonal to beam

With the general equations derived, we are now ready to analyze the specific situation of this experiment. The incident angle, as measured in the moving frame (i.e. rest frame of the medium) is zero whether the direction is orthogonal or parallel:

$$\theta_{i\parallel} = \theta_{i\perp} = 0. \tag{50}$$

Substituting these into the expressions for the wave velocity components when the velocity is orthogonal to the normal

line (Equations (45) and (46)), we get

$$c_{n0\parallel} = \frac{c}{n}\left(\frac{\beta n^2}{n}\right) = v, \tag{51}$$

$$c_{n0\perp} = \frac{c}{n}\left[\frac{n\sqrt{1}}{\gamma(n)}\right] = \frac{c}{n\gamma}. \tag{52}$$

And for the case when the velocity is parallel to the normal line (Equations (48) and (49)), we have

$$c_{n0\parallel} = \left(\frac{c}{n}\right)\frac{(n\beta+1)}{1+\frac{\beta}{n}} = \frac{c+cn\beta}{n+\beta} = \frac{c(c+nv)}{cn+v}, \tag{53}$$

$$c_{n0\perp} = 0. \tag{54}$$

We already derived this expression for the parallel case in equation (14). We repeat it here to show that equations (48) and (49) are consistent with the earlier derivation.

The angle of propagation in the parallel case is 0, but in the orthogonal case it is defined by

$$\sin\varphi_0 = \frac{c_{n0\parallel}}{\sqrt{{c_{n0\parallel}}^2 + {c_{n0\perp}}^2}} = \frac{v}{\sqrt{v^2 + \dfrac{c^2}{n\gamma^2}}}. \tag{55}$$

Equations (51) and (52) are quite remarkable. Equation (51) shows that the velocity component of the wave velocity that is parallel to the medium velocity v is always exactly equal to v. It is completely independent of the index of refraction n. This is what guarantees that the path that a wave takes through the medium will not change relative to the medium no matter how fast the source and medium are moving or no matter what the index of refraction is. This is why observers in the medium cannot detect any change in the trajectory of the waves when their velocity changes.

Equation (52) gives the component of wave speed that is orthogonal to the velocity of the medium. This is the term that guarantees that the time measured for the wave to pass through the medium is always measured to be c/n in the moving frame (the rest frame of the medium). For example, if the laboratory is at rest the velocity of a wave is c/n, and the time to pass through a block that is of length L is Ln/c. If the laboratory is then accelerated to a velocity of v in the orthogonal direction, the clocks in that frame slow so that the time Ln/c becomes $Ln\gamma/c$. But from equation (52) we see that the orthongonal component of the wave speed slows down by the same factor of γ so that the time measured in the laboratory to traverse length L remains at Ln/c.

5 Calculating time delays and phase shifts

Knowing the incident wavelengths, velocities and directions, we can calculate the change in phase shift that occurs with velocity. The only place that the phase can be different between the two paths is when the beam is passing through the refractive block. The distance that the unrefracted beam races ahead of the refracted wave while the refracted wave is slowed down by the RB is proportional to the phase difference between the two paths.

We will begin by analyzing the parallel direction where the velocity of the medium and velocity of the light beam are aligned.

5.1 Time delay and phase shift with light beam parallel to velocity

In this case, the laboratory is moving at velocity v with the light beam parallel to the velocity. The length of the RB will contract to

$$L_{0\parallel} = \frac{L}{\gamma}. \tag{56}$$

The velocity of the light within the refractive material, with respect to the rest frame, is given by equations (53) and (54):

$$\left.\begin{aligned} c_{n0\parallel} &= \frac{c}{n}\left(\frac{1+\beta n}{1+\dfrac{\beta}{n}}\right) = \frac{c(c+nv)}{cn+v} \\ c_{n0\perp} &= 0 \end{aligned}\right\}. \tag{57}$$

The refractive block itself is moving at velocity v, so the effective velocity of the light with respect to the RB is

$$c_{n0\parallel e} = c_{n0\parallel} - v = \frac{c(c+nv)}{cn+v} - v = \frac{c}{\gamma^2}\left(\frac{1}{n+\beta}\right). \tag{58}$$

At this relative velocity, the total time it takes a wave to propagate through the RB is

$$\Delta t_{0\parallel} = \frac{L_{0\parallel}}{c_{n0\parallel e}} = \frac{L}{\gamma\dfrac{c}{\gamma^2}\left(\dfrac{1}{n+\beta}\right)} = \frac{L(n+\beta)\gamma}{c}. \tag{59}$$

The total distance a wavelet propagates in the parallel direction while inside the RB is measured in the rest frame to be

$$\Delta x_{0\parallel} = \frac{L}{\gamma} + v\,\Delta t_{0\parallel} = \frac{L}{\gamma} + L\beta(n+\beta)\gamma = L\gamma(1+n\beta). \tag{60}$$

The total distance the unrefracted beam propagates in this same time is

$$\Delta x_{0u\parallel} = c\,\Delta t_{0\parallel} = L(n+\beta)\gamma. \tag{61}$$

The difference between these two distances for the refracted and unrefracted paths is the spatial phase shift that occurred between the two waves as a result of the path through the RB:

$$\Delta x_{\parallel} = \Delta x_{0\parallel} - \Delta x_{0u\parallel} = L\gamma(n-1)(1-\beta). \tag{62}$$

Dividing this difference by the wavelength of the incident wave gives the phase shift in wavelengths in the parallel orientation:

$$k_{\parallel} = \frac{L}{\lambda_{i0\parallel}}\,\gamma\,(n-1)\,(1-\beta)\,. \tag{63}$$

This is a scalar value. Like the number of marbles in a bowl, it is the same for all observers in all frames. It represents the phase difference between the refracted path and the unrefracted path in the parallel direction as measured in wavelengths.

Since it is an invariant, we should be able to verify that it is the same value as measured in the moving frame. The phase shift in that frame that would be expected is

$$k_{\parallel} = \left(c - \frac{c}{n}\right)\frac{\Delta t}{\lambda_i} = \left(c - \frac{c}{n}\right)\left(\frac{L\,n}{\lambda_i\,c}\right) = \frac{L}{\lambda_i}\,(n-1)\,. \tag{64}$$

To show that equations (63) and (64) are, in fact, the same scalar value, we note that the frequency of the laser source will be reduced in the rest frame and there will also be a Doppler shift of the wavelength in that frame. Thus, the wavelength of the incident wave in the rest frame is

$$\lambda_{i0\parallel} = \lambda_i\,\gamma\,(1-\beta) = \lambda_i\sqrt{\frac{(1-\beta)^2}{(1+\beta)\,(1-\beta)}} = \lambda_i\sqrt{\frac{1-\beta}{1+\beta}}. \tag{65}$$

Substituting this into equation (63) gives the total phase shift in wavelengths between the two paths in the parallel orientation:

$$k_{\parallel} = \frac{L}{\lambda_i}\sqrt{\frac{1+\beta}{1-\beta}}\,\gamma\,(n-1)\,(1-\beta) = \frac{L}{\lambda_i}\,(n-1)\,. \tag{66}$$

This is, of course, the same scalar value measured in the moving frame in equation (64). The interesting thing about this number is that it is completely independent of the velocity of the medium. That is just another way of saying that no matter what the velocity of the frame, all observers will always measure exactly the same phase shift.

Notice that k is a very large number since L is measured in meters and the wavelength is measured in hundreds of nanometers. This number is not measurable by the interferometer. It is only able to measure *differences* in phase. Fortunately it is the difference between the orthogonal and parallel phase shifts that we are interested in. We will now repeat the above procedure to determine the phase shift for the orthogonal direction.

5.2 Time delay and phase with the light beam orthogonal to velocity

When the light beam is orthogonal to the velocity of the laboratory, no contraction occurs and the length of the RB remains at its rest length of L. Since the individual wavelets are moving through the RB at an angle, the time that it takes for an individual wavelet to travel through the block is determined by the component of its velocity that is parallel to the normal line.

This is obtained from equation (52):

$$c_{n0\,\perp} = \frac{c}{n\,\gamma}. \tag{67}$$

Of course, it propagates a distance L in this direction at this speed. Since the velocity of the laboratory is orthogonal to the RB, this is also the velocity of a wave relative to the RB. The total time for a wave to propagate through the RB is

$$\Delta t_{0\perp} = \frac{L}{c_{n0\perp}} = \frac{L\,n\,\gamma}{c}. \tag{68}$$

During this same time, the unrefracted beam is propagating at speed c but not exactly orthogonal. It's velocity in the orthogonal direction is also given by equation (52), but with $n = 1$, since it is moving through vacuum:

$$c_{n0u\perp} = \frac{c}{\gamma}. \tag{69}$$

The distance that the unrefracted beam travels in this time is

$$\Delta x_{0u\perp} = \frac{c}{\gamma}\,\Delta t_{0\perp} = \frac{c}{\gamma}\,\frac{L\,n\,\gamma}{c} = L\,n. \tag{70}$$

The difference between the two distances is

$$\Delta x_{\perp} = \Delta x_{0u\perp} - \Delta x_{0\perp} = n\,L - L = L\,(n-1)\,. \tag{71}$$

We divide this by the wavelength in the orthogonal direction to get the total phase shift:

$$k_{\perp} = \frac{L}{\lambda_{i0\perp}}\,(n-1)\,. \tag{72}$$

For calculating $\lambda_{i0\perp}$ we must again account for the longer wavelength due to the slowing of the frequency source. While there is no Doppler shift orthogonal to a moving source, we must consider the change in wavelength due to the angle at which it is propagating in the rest frame. So

$$\lambda_{i\,0\perp} = \lambda_i\,\gamma\,\cos\theta_{i\,0\perp}. \tag{73}$$

Since this wavelength is measured in vacuum while the wave is moving at velocity c, from equation (69), we see that

$$\cos\theta_{i\,0\perp} = \frac{c/\gamma}{c} = \frac{1}{\gamma}. \tag{74}$$

Thus, $\lambda_{0_{\perp}} = \lambda_i$ and the total phase shift in wavelengths from equation (72) becomes

$$k_{\perp} = \frac{L}{\lambda_i}\,(n-1)\,. \tag{75}$$

Comparing this to the phase shift for the parallel case in equation (66), we see that they are identical. We have now proven mathematically that regardless of whether or not the experiment is performed in vacuum or in a refractive medium there is no difference in phase between the two orientations — it will always be a null experiment.

6 Numerical values of Cahill's predictions

Cahill, on the other hand, predicted that there will be a measurable phase difference. Cahill predicts in his equation (11) in [1] that the time difference between the two paths will approximate to his equation (12). Again using a "C" in the subscripts to indicate Cahill's predictions, his time difference is

$$\Delta t_C \approx L n \left(n^2 - 1\right)\left(\frac{\beta^2}{c}\right). \tag{76}$$

But this is for a two-way experiment. Our experiment is a one-way measurement. Cahill's one-way time in the parallel direction through the refractive block is derived from his equations (1) and (2):

$$\Delta t_{0\parallel C} = \frac{L n}{c}\left(\frac{1}{\gamma\,(1 - n\beta)}\right). \tag{77}$$

His time in the orthogonal direction is given in his equation (8) in [1]:

$$\Delta t_{0\perp C} = \frac{L n}{c}\left(\frac{1}{\sqrt{1 - n^2\,\beta^2}}\right). \tag{78}$$

Both of these times are as measured in the rest reference frame and represent the total time between a wavelet entering the refractive block until it exits.

According to Cahill, the speed of light in the refractive material is approximately c/n in both cases and Fresnel drag is insignificant. In the orthogonal case, from his Figure 1 (b) this requires that the direction of the wave is actually at an angle that satisfies the equations

$$\left.\begin{aligned} \sin\theta_{r0\perp C} &= \frac{v\,n}{c} \\ \cos\theta_{r0\perp C} &= \sqrt{1 - \left(\frac{v\,n}{c}\right)^2} = \sqrt{1 - n^2\,\beta^2} \end{aligned}\right\}. \tag{79}$$

Therefore the distances traveled in the parallel and orthogonal directions during the times of equations (77) and (78) are respectively

$$\left.\begin{aligned} \Delta x_{0\parallel C} &= \frac{c}{n}\Delta t_{0\parallel} = \frac{L}{\gamma\,(1 - n\beta)} \\ \Delta x_{0\perp C} &= \frac{c}{n}\,\Delta t_{0\perp}\cos\theta_{r0\perp} \\ &= L\left(\frac{1}{\sqrt{1 - n^2\beta^2}}\right)\sqrt{1 - n^2\,\beta^2} = L \end{aligned}\right\}. \tag{80}$$

On the other hand, the distances traveled by the light in the unrefracted paths in these times are

$$\left.\begin{aligned} \Delta x_{0u\parallel C} &= c\,\Delta t_{0\parallel} = \frac{L n}{\gamma\,(1 - n\beta)} \\ \Delta x_{0u\perp C} &= c\,\Delta t_{0\perp}\cos\theta_{i0\perp} = \frac{L n}{\gamma}\left(\frac{1}{\sqrt{1 - n^2\,\beta^2}}\right) \end{aligned}\right\}. \tag{81}$$

For each direction, respectively, the differences between the refracted and unrefracted lengths are

$$\left.\begin{aligned} \Delta x_{\parallel C} &= \Delta x_{0u\parallel C} - \Delta x_{0\parallel C} = \frac{L}{\gamma\,(1 - n\beta)}\,(n - 1) \\ \Delta x_{\perp C} &= \Delta x_{0u\perp C} - \Delta x_{0\perp C} = L\left(\frac{n}{\gamma\,\sqrt{1 - n^2\beta^2}} - 1\right) \end{aligned}\right\}. \tag{82}$$

Using equation (65) for the parallel incident wavelength (orthogonal incident wavelength is unchanged), we can convert these distances to wavelengths:

$$\left.\begin{aligned} k_{\parallel C} &= \frac{\Delta x_{\parallel C}}{\lambda_{i0\,\parallel C}} = \frac{L}{\gamma\,\lambda_i\,(1 - n\beta)}\,(n - 1)\,\sqrt{\frac{1 + \beta}{1 - \beta}} \\ k_{\perp C} &= \frac{\Delta x_{\perp C}}{\lambda_{i0\,\perp C}} = \frac{L}{\lambda_i}\left(\frac{n}{\gamma\,\sqrt{1 - n^2\beta^2}} - 1\right) \end{aligned}\right\}. \tag{83}$$

The total phase shift predicted by Cahill's equations is the difference between these two values, which simplifies to

$$\Delta k_C = \frac{L}{\lambda_i\,(1 - n\beta)}\left(n - \beta - \frac{n}{\gamma}\sqrt{\frac{1 - n\beta}{1 + n\beta}}\right). \tag{84}$$

In this experiment

$$L = 1\text{ m} \qquad n = 1.33 \qquad \lambda_i = 650\text{ nm}. \tag{85}$$

Cahill claims that the original MM experiment measured a velocity of about 360 km/sec. Thus,

$$v = 3.6 \times 10^5 \qquad \Rightarrow \qquad \beta = 0.0012. \tag{86}$$

Substituting all these values into equation (84) gives us the phase shift that Cahill predicts for this experiment:

$$\Delta k_C = 1421\text{ wavelengths}. \tag{87}$$

This is an enormous phase difference which would easily be detected by this experiment if it existed.

7 Results of experiment

The present experiment is capable of measuring phase differences with a resolution of about 0.1 wavelengths. The phase shift was measured between a north-south orientation and an east-west orientation each hour for 12 hours. Had there been any significant velocity difference in any direction, one or more of these measurements would have been able to detect it.

The peak phase difference (after averaging) was measured to be 0.1 wavelengths at 10 a.m. This is within the error tolerance of the experiment and is therefore not statistically different from zero. After averaging the 10 measurements at each time, the measured phase shifts in wavelengths are graphed in Figure 8.

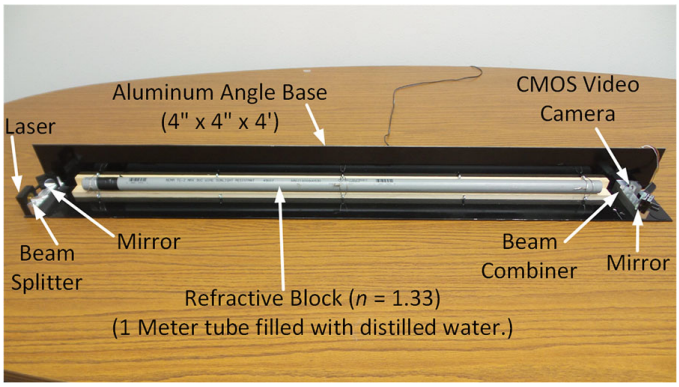

Fig. 9: Overview of interferometer system.

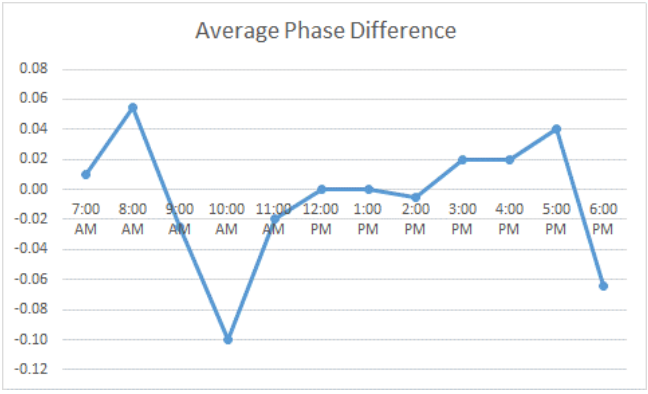

Fig. 8: Measured phase shift.

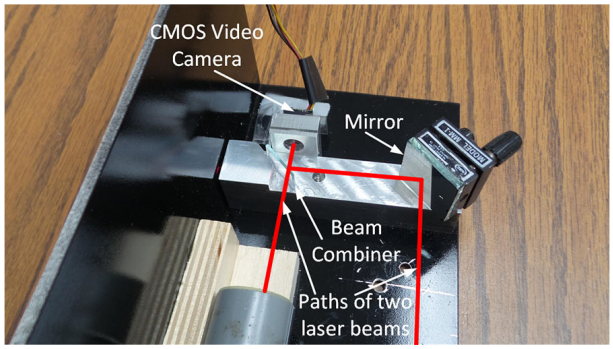

Fig. 10: Closeup of camera/detector end.

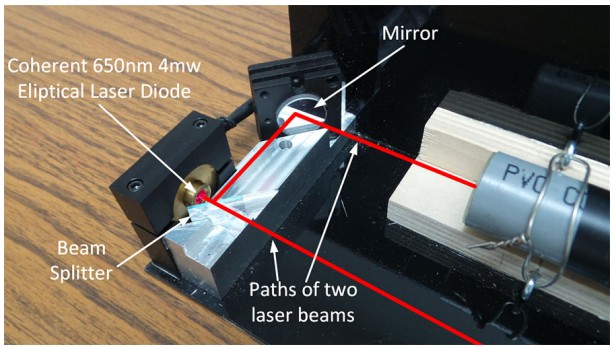

Fig. 11: Closeup of laser diode end.

The results of this experiment are the "final arbiter" and clearly rule in favor of the derivation in this paper and against Cahill's derivation. The measured phase shifts are 4 orders of magnitude less than those predicted by Cahill and they are within the measurement tolerances of the null prediction of this paper. We can conclude that the mathematical derivations in this paper are correct and that it is impossible to detect the absolute velocity of the earth using MM type experiments regardless of the index of refraction of the medium used.

8 Description and procedures of experiment

Figures 9, 10 and 11 show actual annotated photographs of the interferometer system used in the experiment. It is arranged according to the layout shown in Figure 1. Not shown in these pictures are two polarizers — one at the output of the laser and one at the input to the camera. These were rotated relative to one another to attenuate the light to just the right brightness so that the camera image was optimized for visualization of the fringe pattern. Without them the image was too bright and the camera's CMOS detector bloomed to an all-white image.

8.1 Measurement considerations

The fringe shifts are measured by displaying the output of the camera on a computer monitor. Figure 12 shows the camera output plus two drafting triangles that were placed on the monitor as references to assist in measuring fringe movement.

The entire system is mounted on a 4-foot (1.22 m) aluminum base that is painted black. The thermal expansion coefficient of aluminum causes it to expand about 29 μm per degree C. That is 45 wavelengths of light per degree C or about one half wavelength for each hundredth of a degree C.

Fig. 12: Fringe pattern output from camera.

An even larger sensitivity occurs due to the fluctuations in barometric pressure which change the index of refraction of the air. Because of this extreme sensitivity to temperature and pressure, there is a constant drifting of the fringe patterns that must be taken out of the measurement.

To minimize the thermal drift, the following mitigating techniques were employed:

1. The entire interferometer was placed inside a cardboard tubular shipping container and sealed on both ends.
2. The system was allowed to warm up and reach a stable temperature prior to making any measurements.
3. The measurements were taken inside a room with no outside walls or windows.
4. The heating and air conditioning system was turned off so that only slow, convection heating from outside the building could affect the temperature inside the room.
5. A 4 foot wooden dowel was used to rotate the system so that human body temperature was kept away from the system.
6. The system was rotated very slowly (about 30 seconds for a 90 degree rotation) to minimize the cooling and pressure effects of the air flow.

By doing all of these things, the drift was reduced to significantly less than 1 fringe per minute (probably mostly due to barometric pressure drift), which was easy to remove from the measurements.

Mechanical disturbances were minimized by placing the system on pillows and attaching it to a rotatable platform with a bungee cord pressing it into the pillows. The platform is made from an aluminum trailer hitch-mounted cargo carrier with the hitch attachment removed. The platform was mounted to the base of a rotating office chair (after removing the seat) so that it could be rotated very smoothly and with little effort. The pillows prevented any residual vibrations of the platform from propagating to the interferometer. The result is that almost no vibrations affected the fringes so they were very easy to follow as they drifted slowly across the screen.

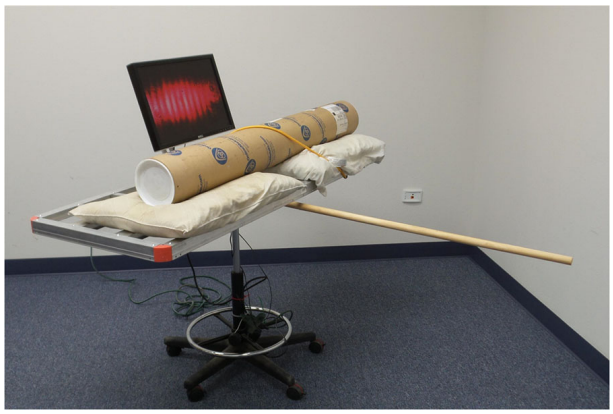

Fig. 13: Complete system with vibrational and thermal mitigation.

Figure 13 shows the system after employing these temperature and vibration mitigating techniques. The interferometer is sealed inside the tubular cardboard shipping container with the camera output coming through a small hole in the back of the container into the monitor.

8.2 Measurement procedure

To improve accuracy and resolution, 10 measurements were made at 1 hour intervals for 12 hours – which corresponds to 10 measurements every 15 degrees of earth's rotation for 180 degrees total rotation. The measurements were performed in Longmont, Colorado between 7 am and 6 pm on September 22 and 23, 2015. The following procedure was used:

1. Turn on the system and let it warm up for 2 hours.
2. At the top of each hour, position the system in a north-south orientation.
3. Place the edge of a triangle in the middle of the fringe nearest to the center of the screen.
4. Very slowly rotate the system clockwise 90 degrees until it reaches an east-west orientation. (about 30 seconds)
5. Estimate the movement of the fringe to the nearest 0.1 wavelength – including any drift that occurred. Record this as phase 1.
6. Reposition the edge of the triangle in the middle of the center fringe.
7. Very slowly rotate the system counterclockwise to return to the north-south orientation.
8. Estimate the movement of the fringe to the nearest 0.1 wavelength – including any drift that occurred. Record this as phase 2.
9. Repeat steps 2 to 8 until 10 pairs of phase 1 and phase 2 measurements have been recorded.
10. Wait until the top of the next hour and repeat steps 2 to 9 until data for 12 hours have been recorded.

After all data were recorded, the phase shift of each measurement was calculated as

$$PhaseShift = \frac{1}{2}\,(Phase1 - Phase2)\,. \tag{88}$$

This removes any drift from the measurement because it will be constant in both phases.* For example, suppose *Phase1* includes a real shift of k and a drift of d. Then when returning, *Phase2* will measure a real shift of $-k$ and the same drift d. The phase shift recorded will be

$$PhaseShiftR = \frac{(k+d)-(-k+d)}{2} = k. \tag{89}$$

This was done for each of the 10 measurements at each hour. The 10 measurements for each hour were averaged. This improves the resolution of the final answer and averages out drift errors due to each "slow" rotation not being exactly the same amount of time. These results are tabulated in Table 1 and were graphed earlier in Figure 8.

Time	Average Phase Shift
7:00	0.01
8:00	0.06
9:00	−0.03
10:00	−0.10
11:00	−0.10
12:00	0.00
13:00	0.00
14:00	−0.01
15:00	0.02
16:00	0.02
17:00	0.04
18:00	−0.06

Table 1: Measured phase shifts.

9 Conclusions

We have now shown both mathematically and experimentally that Michelson-Morely-type interferometer experiments cannot detect the absolute speed of the earth through space regardless of the medium through which the light is propagating. This experiment and the accompanying mathematical analysis show that the conspiracy between Mother Nature and light is complete. They have conspired to make it impossible to detect our absolute speed using light signals.

Acknowledgements

I am sincerely grateful to Doug Klingbeil for his insight and suggestions regarding the optical components. He also designed and machined the aluminum blocks upon which the optical components are mounted. Special thanks also goes to my daughter Lindsea Seaver for her assistance in precisely assembling the apparatus.

Submitted on August 16, 2016 / Accepted on September 4, 2016

References

1. Cahill R.T., Kitto K. Michelson-Morley Experiments Revisited and the Cosmic Background Radiation Preferred Frame. *Apeiron*, April 2003, v. 10(2), 104–117.
2. Cahill R.T. The Michelson and Morley 1887 Experiment and the Discovery of Absolute Motion. *Progress in Physics*, 2005, v. 1(3), 25–29.
3. Cahill R.T. A New Light-Speed Anisotropy Experiment: Absolute Motion and Gravitational Waves Detected. *Progress in Physics*, 2006, v. 2(4), 73–92.
4. Cahill R.T. Absolute Motion and Gravitational Effects. *Apeiron*, 2004, v. 11(1), 53–111.
5. Roberts T.J. An Explanation of Dayton Miller's Anomalous "Ether Drift" Result. arXiv: physics/0608238.

*The rate of change of the temperature and the pressure are assumed to be nearly constant over the 1 minute required to take each measurement.

Dark Matter, the Correction to Newton's Law in a Disk

Yuri Heymann

3 rue Chandieu, 1202 Geneva, Switzerland. E-mail: y.heymann@yahoo.com

The dark matter problem in the context of spiral galaxies refers to the discrepancy between the galactic mass estimated from luminosity measurements of galaxies with a given mass-to-luminosity ratio and the galactic mass measured from the rotational speed of stars using the Newton's law. Newton's law fails when applied to a star in a spiral galaxy. The problem stems from the fact that Newton's law is applicable to masses represented as points by their barycenter. As spiral galaxies have shapes similar to a disk, we shall correct Newton's law accordingly. We found that the Newton's force exerted by the interior mass of a disk on an adjacent mass shall be multiplied by the coefficient η_{disk} estimated to be 7.44 ± 0.83 at a 99% confidence level. The corrective coefficient for the gravitational force exerted by a homogeneous sphere at it's surface is 1.00 ± 0.01 at a 99% confidence level, meaning that Newton's law is not modified for a spherical geometry. This result was proven a long time ago by Newton in the shell theorem.

1 Introduction

Dark matter is an hypothetical type of matter, which refers to the missing mass of galaxies, obtained from the difference between the mass measured from the rotational speed of stars using the Newton's law and the visual mass. The visual mass is estimated based on luminosity measurements of galaxies with a given mass-to-luminosity ratio.

The problem of galaxy rotational curves was discovered by Vera Rubin in the 1970s [1–3], with the assistance of the instrument maker Kent Ford. In Figure 1, we show the rotational velocity curve of stars versus the expected rotational velocity curve from visible mass as a function of the radius of a typical spiral galaxy. According to [4], the estimated dark matter to visible matter ratio in the universe is about 5.5.

It has been hypothesized that dark matter is made of invisible particles which do not interact with electromagnetic radiations. The hunt for the dark matter particle has already begun. The Xenon dark matter experiment [5] is taking place in a former gold mine nearly a mile underground in South Dakota. The idea is to find hypothetical dark matter particles underneath the earth to avoid particule interference from the surface.

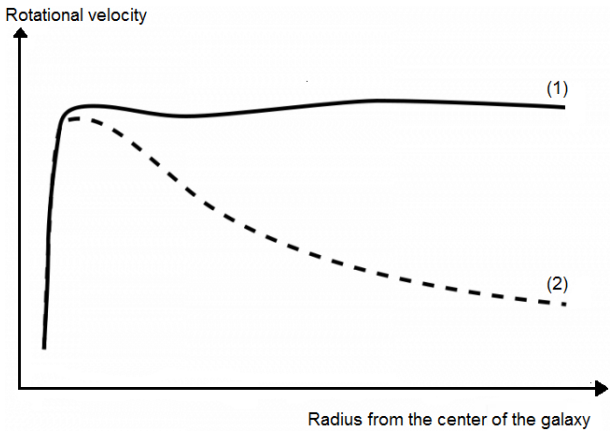

Fig. 1: The problem of galaxy rotational curves, where (1) is the actual rotational velocity curve of stars; and (2) the expected rotational velocity curve from the visible disk.

Other experiments seek dark matter in space. In 2011, NASA lauched the AMS (Alpha Magnetic Spectrometer) experiment, a particle detector mounted on the ISS (International Space Station) aimed at measuring antimatter in cosmic rays and search for evidence of dark matter. In December 2015, the Chinese Academy of Sciences lauched the DAMPE (Dark Matter Particle Explorer), a satellite hosting a poweful space telescope for cosmic ray detection and investigating particles in space and hypothetical dark matter.

An investigation of the amount of planetary-mass dark matter detected via gravitational microlensing concluded that these objects only represent a small portion of the total dark matter halo [6]. The study of the distribution of dark matter in galaxies led to the development of two models of the dark matter halo. These models are known as the dark matter halo profile of Navarro, Frenk and White [7], and the Burkert dark matter halo profile [8, 9].

Dark matter is a hot topic in particle physics, and has led to the development of various theories. According to [10], the favoured candidates for dark matter are axions, supersymmetric particles, and to some extent massive neutrinos. The Majorana fermion has also been proposed as a candidate for dark matter [11, 12]. Other candidates for dark matter would be dark pions, a set of pseudo-Goldstone bosons [13]. Many alternatives have been proposed including modified Newtonian gravity. Mordehai Milgrom proposed the MOND theory, according to which Newton's law is modified for large distances [14, 15]. Moffat proposed a modified gravity theory based on the action principle using field theory [16, 17]. James Feng and Charles Gallo proposed to model galaxy ro-

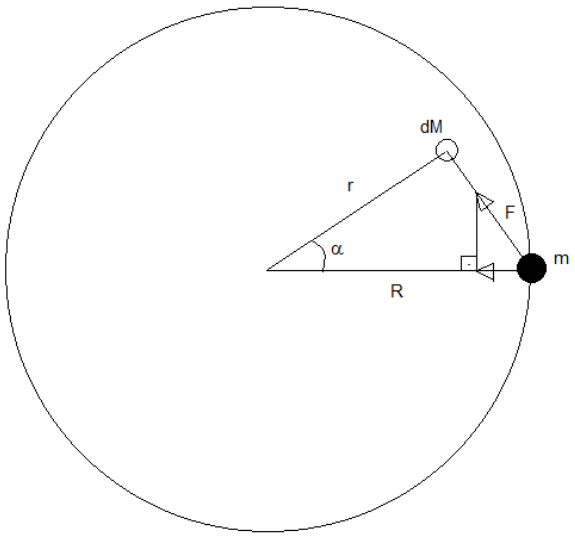

Fig. 2: Force exerted by an infinitesimal mass dM of the disk on a mass m located at the edge of the disk using polar coordinates. The radius of the disk is R. Let the mass dM be at a distance r from the center of the disk. Let α be the angle between the two axis passing by the center of the disk in the direction of the two masses dM and m.

tational curves by applying Newtonian dynamics to a rotating thin disk [18, 19]. Their approach is similar to the route we undertake in the current work, although the latter was done independently.

According to Pavel Kroupa, the dark matter crisis is a major problem for cosmology [20]. In addition, he states that the hypothesis that exotic dark matter exists must be rejected [21]. In the present study we find that dark matter is mainly a problem of geometry because Newton's law is applicable to masses which can be approximated by a point in space. Below, we compute the corrective coefficient to Newton's law in a disk and in a sphere.

2 Calculation of the gravitational force in a disk

The Newton's law states that the gravitational force between two bodies is expressed as follows:

$$F_{\text{Newton}} = \frac{G\,M\,m}{R^2}\,, \tag{1}$$

where G is the gravitational constant, M and m the respective masses of the two bodies in interaction, and R the distance between the barycenters of the two masses.

The shape of spiral galaxies allows us to use the gravitational force computed for a disk. Let us assume a homogeneous disk of surface density ρ_s, and radius R. A mass m is located at the edge of this disk at a distance R from the center of the disk.

In Figure 2, we represent the force exerted by an infinitesimal mass dM of the disk on the mass m using polar coordinates. Because of the symmetry of the disk with respect to

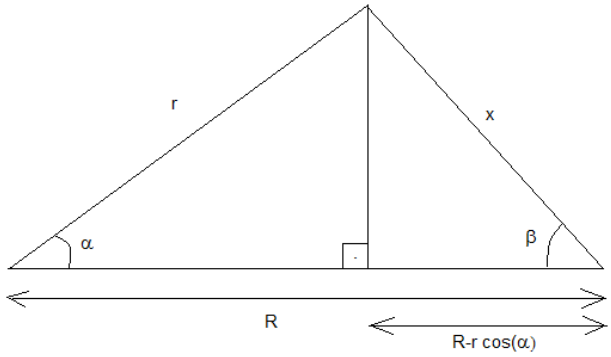

Fig. 3: Triangle to compute the projection of the force exerted by the infinitesimal mass dM on mass m on the axis passing by the center of the disk to the mass m

the axis passing between its center and the mass m, we need to compute the projection of the force exerted by the infinitesimal mass dM on this axis. For this purpose we apply basic trigonometric rules (see figure 3). For convenience, we consider the polar coordinates (r, α) to describe the position of dM, where r is the radial distance, and α the angle between the mass dM and an arbitrary direction as viewed from the center of the disk.

Let us say x is the distance between the mass dM and m. From trigonometry we calculate x as follows:

$$x^2 = r^2 \sin^2\alpha + (R - r\cos\alpha)^2\,. \tag{2}$$

Hence, we get:

$$x^2 = r^2 + R^2 - 2Rr\cos\alpha\,. \tag{3}$$

Let β be the angle between the center of the disk and the mass dM as viewed from the mass m. The angle β is calculated as follows:

$$\cos\beta = \frac{R - r\cos\alpha}{x}\,. \tag{4}$$

By Newton's law, the infinitesimal force exerted by dM on m projected on the axis passing through the center of the disk and the mass m is as follows:

$$dF = \frac{G\,m\,dM}{x^2}\cos\beta\,. \tag{5}$$

Combining (4) and (5), we get:

$$dF = \frac{G\,m\,dM}{x^3}\,(R - r\cos\alpha)\,. \tag{6}$$

Combining (3) and (6), we get:

$$dF = \frac{G\,m\,dM\;(R - r\cos\alpha)}{(r^2 + R^2 - 2Rr\cos\alpha)^{\frac{3}{2}}}\,. \tag{7}$$

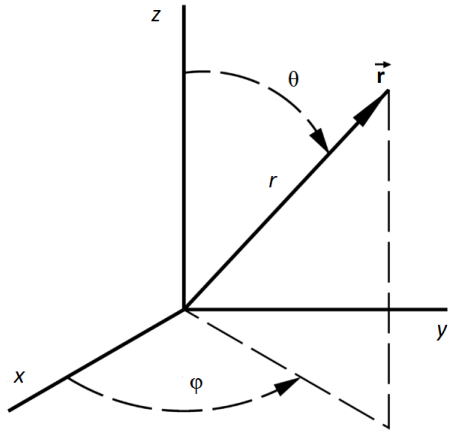

Fig. 4: Spherical coordinate system, where r is the radial distance, θ the polar angle, and φ the azimutal angle.

Because we are using polar coordinates, the surface element dA is as follows:

$$dA = r\,dr\,d\alpha\,. \tag{8}$$

To obtain the infinitesimal mass dM, we multiply the infinitesimal surface dA by the surface density ρ_s; hence, we get:

$$dM = \rho_s\, r\,dr\,d\alpha\,. \tag{9}$$

Therefore, the infinitesimal force dF is as follows:

$$dF = \frac{\rho_s\, G\, m\left(Rr - r^2\cos\alpha\right)}{\left(r^2 + R^2 - 2Rr\cos\alpha\right)^{\frac{3}{2}}}\,dr\,d\alpha\,. \tag{10}$$

Because the total mass of the disk is $M = \rho_s\,\pi\,R^2$, we get:

$$dF = \frac{G\,M\,m}{\pi\,R^2}\,\frac{\left(Rr - r^2\cos\alpha\right)}{\left(r^2 + R^2 - 2Rr\cos\alpha\right)^{\frac{3}{2}}}\,dr\,d\alpha\,. \tag{11}$$

The total force F exerted by the disk on the mass m is obtained by the following integral:

$$F = \frac{G\,M\,m}{\pi\,R^2}\int_{r=0}^{R}\int_{\alpha=0}^{2\pi}\frac{\left(Rr - r^2\cos\alpha\right)}{\left(r^2 + R^2 - 2Rr\cos\alpha\right)^{\frac{3}{2}}}\,dr\,d\alpha\,. \tag{12}$$

We rearrange the terms in the integral to obtain:

$$F = \frac{G\,M\,m}{\pi\,R^2}\times \times\int_{r=0}^{R}\int_{\alpha=0}^{2\pi}\frac{R^2\left(\frac{r}{R} - \left(\frac{r}{R}\right)^2\cos\alpha\right)}{R^3\left(\left(\frac{r}{R}\right)^2 + 1 - 2\left(\frac{r}{R}\right)\cos\alpha\right)^{\frac{3}{2}}}\,dr\,d\alpha\,. \tag{13}$$

Hence:

$$F = \frac{G\,M\,m}{\pi\,R^3}\times \times\int_{r=0}^{R}\int_{\alpha=0}^{2\pi}\frac{\left(\frac{r}{R} - \left(\frac{r}{R}\right)^2\cos\alpha\right)}{\left(\left(\frac{r}{R}\right)^2 + 1 - 2\left(\frac{r}{R}\right)\cos\alpha\right)^{\frac{3}{2}}}\,dr\,d\alpha\,. \tag{14}$$

We apply the change of variable $u = \frac{r}{R}$, hence $dr = R\,du$. Therefore, we get:

$$F = \frac{G\,M\,m}{\pi\,R^2}\int_{u=0}^{1}\int_{\alpha=0}^{2\pi}\frac{\left(u - u^2\cos\alpha\right)}{\left(u^2 + 1 - 2u\cos\alpha\right)^{\frac{3}{2}}}\,du\,d\alpha\,. \tag{15}$$

From (15), we see that in a disk, Newton's force $F_{\text{Newton}} = \frac{G\,M\,m}{R^2}$ needs to be multiplied by the following coefficient:

$$\eta_{disk} = \frac{1}{\pi}\int_{u=0}^{1}\int_{\alpha=0}^{2\pi}\frac{\left(u - u^2\cos\alpha\right)}{\left(u^2 + 1 - 2u\cos\alpha\right)^{\frac{3}{2}}}\,du\,d\alpha\,. \tag{16}$$

3 Calculation of the gravitational force in a sphere

Let us consider a homogeneous sphere of radius R and average mass density ρ. We consider an infinitesimal mass dM of the sphere represented by its spherical coordinates (r, θ, φ), where r is the radial distance, θ the polar angle, and φ the azimuthal angle (see Figure 4). Let the volume of the sphere be defined by the following boundaries: $r \in [0, R]$, $\theta \in [0, \pi]$, and $\varphi \in [0, 2\pi]$. We assume that a mass m is located at the surface of this sphere on the x-axis.

In Cartesian coordinates we have $x = r\sin\theta\,cos\varphi$, $y = r\sin\theta\,sin\varphi$ and $z = r\cos\theta$. Hence, the distance x between the mass dM and m is as follows:

$$x = \sqrt{(R - r\sin\theta\,\cos\varphi)^2 + r^2\sin^2\theta\,\sin^2\varphi + r^2\cos^2\theta}. \tag{17}$$

Let β be the angle as viewed from the mass m between the direction of the center of the sphere and the mass dM. Hence, we get:

$$\cos\beta = \frac{R - r\sin\theta\cos\varphi}{x}\,. \tag{18}$$

The volume element in spherical coordinates is as follows:

$$dV = r^2\sin\theta\,d\theta\,d\varphi\,dr\,. \tag{19}$$

Therefore, the infinitesimal force exerted by dM on m projected in the axis passing through m and the center of the sphere is as follows:

$$dF = \frac{G\,m\,\rho\,r^2\sin\theta\,\cos\beta}{x^2}\,d\theta\,d\varphi\,dr = \frac{G\,m\,\rho\,r^2\sin\theta\,(R - r\sin\theta\cos\varphi)}{x^3}\,d\theta\,d\varphi\,dr\,. \tag{20}$$

Let $M = \rho\frac{4}{3}\pi R^3$ be the total mass of the sphere, hence:

$$F = G\,m\,M\frac{3}{4\pi R^3}\int_{r=0}^{R}\int_{\theta=0}^{\pi}\int_{\varphi=0}^{2\pi}\frac{r^2\sin\theta\,(R - r\sin\theta\,\cos\varphi)}{\left(R^2 + r^2\sin^2\theta\,\cos^2\varphi - 2Rr\sin\theta\,\cos\varphi + r^2\sin^2\theta\,\sin^2\varphi + r^2\cos^2\theta\right)^{\frac{3}{2}}}\,d\theta\,d\varphi\,dr\,. \tag{21}$$

We rearrange the terms in the integral to obtain a function of ratios of r/R, and apply the substition $u = \frac{r}{R}$; hence, we get:

$$F = \frac{G\,m\,M}{R^2}\frac{3}{4\pi}\int_{u=0}^{1}\int_{\theta=0}^{\pi}\int_{\varphi=0}^{2\pi}\frac{u^2\sin\theta\,(1 - u\sin\theta\cos\varphi)}{\left(1 + u^2\sin^2\theta\,\cos^2\varphi - 2u\sin\theta\,\cos\varphi + u^2\sin^2\theta\,\sin^2\varphi + u^2\cos^2\theta\right)^{\frac{3}{2}}}\,d\theta\,d\varphi\,dr\,. \tag{22}$$

Therefore, the corrective coefficient to Newton's law in a sphere is as follows:

$$\eta_{sphere} = \frac{3}{4\pi}\int_{u=0}^{1}\int_{\theta=0}^{\pi}\int_{\varphi=0}^{2\pi}\frac{u^2\sin\theta\,(1 - u\sin\theta\,\cos\varphi)}{\left(1 + u^2\sin^2\theta\,\cos^2\varphi - 2u\sin\theta\,\cos\varphi + u^2\sin^2\theta\,\sin^2\varphi + u^2\cos^2\theta\right)^{\frac{3}{2}}}\,d\theta\,d\varphi\,dr\,. \tag{23}$$

4 Numerical evaluation of the gravitational corrective coefficients

Because the integrals in (16) and (23) do not have a known closed-form solution, we need to evaluate them numerically. Monte Carlo simulation is an appropriate method for computing multidimentional integrals. Using Monte Carlo simulation we can compute both an estimate of the integral and its standard deviation.

4.1 Numerical evaluation of the double integral over the disk

Let us consider the integration of a function $f(r,\alpha)$ over a disk of radius R in polar coordinates, where r is the radius and α an angle from a reference direction. The integral to evaluate is expressed as follows:

$$\int_0^{2\pi}\int_0^{R} f(r,\alpha)\,r\,dr\,d\alpha\,. \tag{24}$$

We shall apply the following change of variables:

$$\alpha = 2\pi u_1\,, \tag{25}$$

and

$$r = R\sqrt{u_2}\,, \tag{26}$$

where u_1 and u_2 are two independent random variables of uniform distribution over [0, 1]. This change of variables gives a uniform distribution on the disk of radius R.

Let N be the number of times we generate the random set (u_1, u_2). Hence, the integral of $f(r,\alpha)$ over the disk converges towards the following estimate for N large:

$$I = \pi R^2\frac{\sum_1^N f_i}{N}\,, \tag{27}$$

where f_i is the function $f(r,\alpha)$ evaluated for each draw of the random set (u_1, u_2) with the change of variables (25) and (26).

Because the variance of a random variable X is given by $Var(X) = E[X^2] - (E[X])^2$ and the variance of the sample mean is $Var(\bar{X}) = \frac{Var(X)}{N}$, the variance of the estimate is computed as follows:

$$\mathrm{Var}(I) = \frac{\pi^2R^4\frac{\sum_1^N f_i^2}{N} - \left(\pi R^2\frac{\sum_1^N f_i}{N}\right)^2}{N}\,. \tag{28}$$

The standard deviation of the estimate of η_{disk} is equal to the square root of the variance of the estimate of the double integral on the disk divided by π. To evaluate the integral in (16), we used the Mersenne Twister pseudo-random number generator [22] with N= 1.2×10^{10}. We obtained $\eta_{disk} = 7.44$ with standard deviation of 0.320.

4.2 Numerical evaluation of the triple integral over the sphere

As for the disk, let us use Monte Carlo simulation to evaluate the triple integral of $f(r,\theta,\varphi)$ over the sphere of radius R in the spherical coordinate system. The integral to evaluate is expressed as follows:

$$\int_0^{R}\int_0^{\pi}\int_0^{2\pi} f(r,\theta,\varphi)\,r^2\,\sin\theta\,d\varphi\,d\theta dr\,. \tag{29}$$

For this pupose we generate a set of three independent random variables (u_1, u_2, u_3), each with a uniform distribution over the interval [0, 1]. We apply the following change of variables, which gives a uniform distribution over the sphere:

$$\theta = 2\arcsin\left(\sqrt{u_1}\right)\,, \tag{30}$$

and

$$\varphi = 2\pi u_2 \,, \tag{31}$$

and

$$r = R\, u_3^{\frac{1}{3}} \,. \tag{32}$$

Let N be the number of time we generate the random set (u_1, u_2, u_3). Hence, the triple integral over the sphere converges towards the following estimate for N large:

$$I = \frac{4\pi R^3}{3} \frac{\sum_1^N f_i}{N} \,, \tag{33}$$

where f_i is the function $f(r, \theta, \varphi)$ evaluated for each draw of the random set (u_1, u_2, u_3) using the change of variables (30), (31) and (32).

The variance of the estimate is computed as follows:

$$\mathrm{Var}(I) = \frac{\left(\frac{4\pi R^3}{3}\right)^2 \frac{\sum_1^N f_i^2}{N} - \frac{4\pi R^3}{3}\left(\frac{\sum_1^N f_i}{N}\right)^2}{N} \,. \tag{34}$$

The standard deviation of the estimate of η_{sphere} is equal to the square root of the variance of the estimate of the triple integral on the sphere multiplied by $\frac{3}{4\pi}$. To evaluate the integral in (23), we used the Mersenne Twister pseudo-random number generator with N= $1{\times}10^8$. We obtained $\eta_{sphere} = 1.00$ with standard deviation of $3.85{\times}10^{-3}$.

5 Interpretation

In the present study, we have solved the dark matter puzzle in the context of spiral galaxies by considering the geometry of massive bodies. Dark matter is a hypothetical mass introduced to fill the discrepancy between galaxy mass as measured from the rotational speed of stars and visible mass. Isaac Newton proved the shell theorem [23], which applies to objects of spherical geometry. The shell theorem states that:

1. A spherical body affects external objects gravitationally as though all of its mass were concentrated in a point at its barycenter.
2. For a spherical body, no net gravitational force is exerted by the external shell on any object inside the sphere, regardless of the position.

Because spiral galaxies have shapes which can be approximated by a disk, the distribution of matter will directly affect the perceived gravitational force for a mass rotating on such a disk, and the shell theorem does not apply. By considering an interior mass distributed in space according to an idealised homogeneous disk, we found that Newton's law is corrected by a multiplicative coefficient. This coefficient is estimated to be about 7.44 based on our calculations above of the dark matter to visible mass ratio of 5.5. This coefficient can be interpreted as if the mass of the disk was excentered towards the object perceiving it. In our calculations, we only considered the interior mass of the disk for radii below the position of the object. For an object located on the disk, the outer mass of the disk for radii above of the position of the object may also exert a gravitational force of opposite direction on the object, mitigating the gravitational force exerted by the interior of the disk. This effect which was not quantified should create the asymptotic behavior for galaxy rotational curves when moving far away from the galaxy's central bulge.

Furthermore, for a spiral galaxy, the mass density may increase as we move closer to the center of the disk, causing a departure from the idealised homogeneous disk. In addition, the closer we move towards the central supermassive black hole, which is spherical, the more the interior mass tends towards a sphere and the gravitational corrective coefficient converges towards unity. The shift in the gravitational corrective coefficient at different radii on the galactic disk ought to explain the observed shape of galaxy rotational curves.

Let us illustrate the impact of the gravitational coefficient we found on the mass of the Milky Way. The centripetal force of a star in orbit is expressed as $F_c = \frac{mv^2}{R}$, where m is the mass of the star, v the tangential velocity and R the radius to the center of the galaxy. Hence, the interior mass of the galaxy for a given star is expressed as $M = \frac{Rv^2}{\eta G}$, where $v = wR$ with w the angular velocity, η the gravitational coefficient, and G the gravitational constant. The apparent mass of the Milky Way was estimated to be around $6.82{\times}10^{11}\ M_\odot$ [24]. Let us approximate the Milky Way by a homogeneous disk; therefore, the gravitational coefficient at the periphery of the disk is about $\eta = 7.44$. This leads to an intrinsic mass of the Milky Way of $9.17{\times}10^{10} M_\odot$.

6 Conclusion

To address the discrepancy between galaxy mass estimated from the rotational velocity of stars and visual mass estimated from luminosity measurements, the existence of dark matter was hypothesized. A number of approaches have taken to hunt for both the dark matter particle and modified gravity. For instance, Milgrom proposed that Newton's law should be modified for large distances. Dark matter remains an unresolved problem challenging cosmology and particle physics.

In the present study, we propose a geometrical approach as Newton's law applies to masses that can be approximated by a point in space corresponding to their barycenter. As spiral galaxies have shapes close to a disk, we derived the corrective coefficient to Newton's law in an idealised disk of homogeneous mass distribution. We found that the Newton's law in a homogeneous disk shall be multiplied by the coefficient η_{disk} estimated to be 7.44 ± 0.83 at a 99% confidence level, which fills the dark matter gap in galaxy haloes. We conclude that dark matter in spiral galaxies is a problem of geometry, and that Newton's law needs to be corrected to account for the geometry of the mass. For a spherical geometry, we found that the corrective gravitational coefficient η_{sphere} is

1.00 ± 0.01 at a 99% confidence level.

This means that the Newton's law is not modified for spherical geometry, which was proven a long time ago by Newton.

Submitted on September 5, 2016 / Accepted on September 7, 2016

References

1. Rubin V.C., Ford W.K. Jr. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions *The Astrophysical Journal*, 1970, v. 159, 379–403.
2. Rubin V.C., Thonnard N., Ford W.K. Jr. Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/ *The Astrophysical Journal*, 1980, v. 238, 471–487.
3. Rubin V.C., Burstein D., Ford W.K. Jr., and Thonnard N. Rotation velocities of 16 SA galaxies and a comparison of Sa, Sb, and SC rotation properties *The Astrophysical Journal*, 1985, v. 289, 81–104.
4. Planck collaboration. Planck 2013 results. I. Overview of products and scientific results. *Astronomy and Astrophysics*, 2014, v. 571, Table 10.
5. Aprile E., Alfonsi M., Arisaka K., Arneodo F., Balan C., Baudis L., Behrens A., Beltrame P., Bokeloh K., Brown E., Bruno G., Budnik R., Cardoso J.M.R., Chen W.-T., Choi B., Cline D.B., Contreras H., Cussonneau J.P., Decowski M.P., Duchovni E., Fattori S., Ferella A.D., Fulgione W., Gao F., Garbini M., Giboni K.-L., Goetzke L.W., Grignon C., Gross E., Hampel W., Kish A., Lamblin J., Landsman H., Lang R.F., Le Calloch M., Levy C., Lim K.E., Lin Q., Lindemann S., Lindner M., Lopes J.A.M., Lung K., Marrodán Undagoitia T., Massoli F.V., Mei Y., Melgarejo Fernandez A.J., Meng Y., Molinario A., Nativ E., Ni K., Oberlack U., Orrigo S.E.A., Pantic E., Persiani R., Plante G., Priel N., Rizzo A., Rosendahl S., dos Santos J.M.F., Sartorelli G., Schreiner J., Schumann M., Scotto Lavina L., Scovell P.R., Selvi M., Shagin P., Simgen H., Teymourian A., Thers D., Vitells O., Wang H., Weber M., and Weinheimer C. Analysis of the XENON100 Dark Matter Search Data *Astroparticle Physics*, 2014, v. 54, 11–24.
6. EROS Collaboration and MACHO Collaboration. EROS and MACHO Combined Limits on Planetary Mass Dark Matter in the Galactic Halo. arXiv: astro-ph/9803082.
7. Navarro J.F., Frenk C.S., and White S.D.M. A universal density profile from hierarchical clustering. *The Astrophysical Journal*, 1997, v. 490, 493–508.
8. Salucci P., Walter F., and Borriello A. ΛCDM and the distribution of dark matter in galaxies: A constant-density halo around DDO 47. *Astronomy and Astrophysics*, 2003, v. 409, 53–56.
9. Burkert A. The Structure of Dark Matter Halos in Dwarf Galaxies. *The Astrophysical Journal Letters*, 1995, v. 447, L25.
10. Bergstrom L. Non-Baryonic Dark Matter arXiv: astro-ph/9712179.
11. Ho C.M. and Scherrer R.J. Anapole Dark Matter. arXiv: 1211.0503.
12. Jacques T., Katz A., Morgante E., Racco D., Rameez M., and Riotto A. Complementarity of DM Searches in a Consistent Simplified Model: the Case of Z. arXiv: 1605.06513.
13. Bhattacharya S., Melic B., and Wudka J. Dark Matter Pions. *Acta Physica Polonica Series B*, 2013, v. 44, no. 11, 2359–2366.
14. Milgrom M. A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis *Astrophysical Journal* , 1983, v. 270, 365–370.
15. Milgrom M. A modification of the Newtonian dynamics — Implications for galaxies. *Astrophysical Journal* , 1983, v. 270, 371–389.
16. Brownstein J.R. and Moffat J.W. Galaxy Rotation Curves Without Non-Baryonic Dark Matter. *The Astrophysical Journal*, 2006, v. 636, 721–741.
17. Moffat J.W. Nonsymmetric Gravitational Theory. *Journal of Mathematical Physics*, 1995, v. 36, 3722.
18. Feng J.Q. and Gallo C.F. Modeling the Newtonian dynamics for rotation curve analysis of thin-disk galaxies. *Research in Astronomy and Astrophysics*, 2011, v. 11, 1429–1448.
19. Feng J.Q. and Gallo C.F. Mass distribution in rotating thin-disk galaxies according to Newtonian dynamics. arXiv: 1212.5317.
20. Kroupa P. The Dark Matter Crisis: Falsification of the Current Standard Model of Cosmology. *Publications of the Astronomical Society of Australia*, 2012, v. 29, 395–433.
21. Kroupa P. Lessons from the Local Group (and beyond) on dark matter. arXiv: 1409.6302.
22. Matsumoto M., and Nishimura T. Mersenne Twister: a 623 dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation — Special issue on uniform random number generation*, 1998, v. 8, 3–30.
23. Newton I. Philosophiae Naturalis Principia Mathematica. London, 1687, Theorem XXXI.
24. Eadie G. M. and Harris W. E. Bayesian Mass Estimates of the Milky Way II: The Dark and Light Sides of Parameter assumptions. arXiv: 1608.04757.

Principle of Spacetime Black Hole Equivalence

T. X. Zhang

Department of Physics, Alabama A & M University, Normal, Alabama 35762. E-mail: tianxi.zhang@aamu.edu

A new principle of spacetime black hole equivalence (SBHEP) is proposed. In addition to Einstein's general relativity and the cosmological principle, the SBHEP principle provides the third base for the black hole universe model that was recently developed by the author in attempt to model the universe, explain existing observations, and overcome cosmic problems and difficulties without relying on a set of hypothetical entities. A black hole universe does not have the horizon and flatness problems so that an inflation epoch is not required. Its origin from starlike and supermassive black holes removes the initial big bang singularity and magnetic monopole problems. A black hole or spacetime is static or in equilibrium when it does not accrete or merge with others, otherwise it becomes dynamic, expands, and emits. Gamma ray bursts, X-ray flares from galactic centers, and quasars can be self-consistently explained as emissions of dynamic starlike, massive, and supermassive black holes. Cosmic microwave background radiations are blackbody radiations of the black hole universe, an ideal blackbody. A black hole universe can accelerate if it accretes matter in an increasing rate, so that an explanation of the supernova measurements does not need dark energy.

1 Introduction

Cosmology is the study of the origin and development of the universe. The currently accepted standard big bang model of the universe (BBU) stands on two bases, which are (1) Einstein's general relativity (GR) that describes the effect of matter on spacetime and (2) the cosmological principle (CP) of spacetime isotropy and homogeneity that generates the Friedmann-Lemaître-Robertson-Walker (FLRW) metric of spacetime [1–4]. The Einstein field equation given in GR along with the FLRW metric of spacetime derived from CP produces the Friedmann equation (FE) that governs the development and dynamics of the universe. Although the big bang theory has made incredible successes in explaining the universe, there still exists innumerable problems and difficulties. Solutions of these problems and difficulties severely rely on an increasing number of hypothetical entities (HEs) such as dark matter, dark energy, inflation, big bang, and so on [5]. Therefore, the BBU consists of GR, CP, and innumerable HEs, *i.e.* BBU = {GR, CP, HE, HE, HE,....} (see the blue part of Fig. 1). Although it has only two bases (GR and CP), the BBU severely relies on an increasing number of HEs that have not yet been and may never be tested or falsifiable.

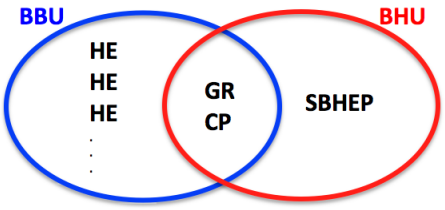

Fig. 1: The comparison of fundamentals between BBU and BHU (see Section 1 for details).

Describing the universe without relying on a set of HEs to explain observations and overcome cosmic problems and difficulties is essential to developing a physical cosmology. Recently, the author has developed a new physical cosmology called black hole universe (BHU) [6–7]. Instead of making many those HEs as the BBU did, the BHU proposes a new principle to the cosmology – the Principle of Spacetime Black Hole Equivalence (SBHEP) – in an attempt to explain all the existing observations of the universe and overcome all the existing problems and difficulties [8–12]. Standing on the three bases (GR, CP, and SBHEP), the new cosmological theory – BHU = {GR, CP, SBHEP} (see the red part of Fig. 1) – can fully explain the universe in various aspects as well as to conquer all the cosmic problems according to the well-developed physics without making any other HEs and including any other unsolved difficulties. GR and CP are common to both BBU and BHU. The BBU stands on two legs unstably so that needs many crutches for support, while the BHU stands on three legs stably without needing any other props. In the BHU, a single SBHEP removes all of innumerable HEs made in the BBU. This paper describes how this new black hole universe model explains the universe and conquers the cosmic difficulties with the principle of spacetime black hole equivalence.

2 Equivalence between spacetime and black hole

The effect of matter on spacetime can be obtained by solving Einstein's field equation provided in GR [13],

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{1}$$

where the subscripts μ and ν are the four-dimensional (4D) spacetime coordinate indices running through 0–3. $G_{\mu\nu}$ is

Einstein's curvature tensor, G is the gravitational constant, c is the speed of light in free space, and $T_{\mu\nu}$ is the 4D energy-momentum tensor. Adding a term of $\Lambda g_{\mu\nu}$ to the left hand side of (1), Einstein developed a static cosmology [14] and de Sitter developed a dynamic cosmology [15]. Here Λ is the cosmological constant and $g_{\mu\nu}$ is the metric of spacetime,

According to the cosmological principle, the universe, if it is viewed on a scale that is sufficient large, is homogeneous and isotropic. This principle implies that there is no special location and direction in the universe. The properties of the universe are the same for all observers in the universe. More strongly, physical laws are all universal. If a physical law is applicable to the Earth, then it can be applied to everywhere. The isotropic and homogeneous spacetime can be described by the FLRW metric [1–4],

$$\begin{aligned} ds^2 &= -g_{\mu\nu}dx^\mu dx^\nu \\ &= c^2dt^2 - R^2(t)\left[\frac{dr^2}{1-kr^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right], \end{aligned} \tag{2}$$

where $R(t)$ is the radius of curvature of the space and k is the curvature constant of the space.

Substituting the FLRW metric of spacetime into the Einstein field equation of general relativity, one can have the Friedmann equation [16],

$$H^2(t) \equiv \frac{\dot{R}^2(t)}{R^2(t)} = \frac{8\pi G\rho(t)}{3} - \frac{kc^2}{R^2(t)}, \tag{3}$$

where $H(t)$ is the Hubble parameter or the expansion rate of the universe and ρ is the density of matter. The dot sign refers to the derivative of the quantity with respect to time, $\dot{R}(t) \equiv dR(t)/dt$. Including the cosmological constant, (3) has a term of $\Lambda/3$ on the right hand side.

According to the Schwarzschild field solution of (1) [17], the metric of spacetime surrounding a spherical body with mass M appears to be singular at the Schwarzschild radius $r_g = 2GM/c^2$, which divides the space into two disconnected patches. This indicates that a region of space, where matter accumulates up to a critical level such that the mass-radius (M-R) ratio reaches up to $M/R = c^2/(2G) \simeq 6.67 \times 10^{26}$ kg/m, forms a black hole and constructs its own spacetime, which is singular in space and non-causal in time to the outside. Therefore, it is reasonable to suggest or postulate that a black hole once formed constructs its own spacetime and a spacetime encloses its own unique black hole [6–7]. In other words, spacetime and black hole are equivalent. This postulate of the equivalence between spacetime and black hole plays a fundamental role in the modeling of the universe; therefore, we raise it as a new principle of the cosmology [18]. Without matter, a physical spacetime cannot be formed; without a spacetime, matter cannot become into existence. As a moral idea or belief, we cannot prove its correctness mathematically, but the truth for the principle of spacetime black hole equivalence can be justified and validated through explaining various observations of the universe, such as CMB and supernova measurements, *etc.*, and overcoming cosmic problems, such as dark energy and inflation problems, *etc.*, in accordance with the black hole model of the universe that is developed on the basis of this principle. In the following sections, we will demonstrate how the black hole model of the universe developed on the basis of this new principle to explain the observations of the universe and overcome the cosmic problems and difficulties.

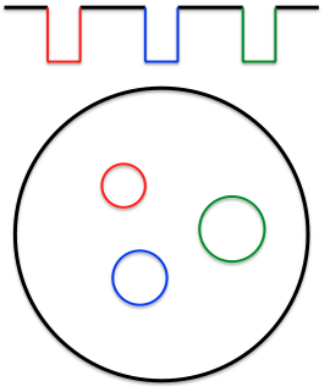

Fig. 2: The hierarchically layered structure of black hole universe. Inside our spacetime or black hole universe (the region represented or circled by the solid black lines), there are a number of subspacetimes (the regions represented or circled by the solid color lines), which are the observed star-like, massive, and supermassive black holes.

3 Black hole universe

From this principle of spacetime black hole equivalence, we understand that our universe, because it is constructed in a 4D spacetime, is or wraps a black hole, which is extremely supermassive and fully expanded. Its big radius and enormous mass can be determined in terms of the measurement of matter density of the universe as detailed below in the subsection 3.1. The inside observed star-like, massive, and supermassive black holes can be considered as subspacetimes (or called child universes) of our black hole universe (see Fig. 2). This hierarchically layered structure of spacetimes and subspacetimes genuinely overcomes the horizon problem, which was identified to exist in the big bang model of the universe primarily by Charles Misner in 1960s [19–20] and solved by Alan Guth in 1980s with the hypothesis of cosmic inflation [21] according to a field that does not correspond to any physical field. Therefore, in the black hole model of the universe, there does not exist the horizon problem at all.

3.1 Mass-radius relation and spacetime equilibrium

The boundary of a spacetime or black hole is determined, according to the Schwarzschild solution, by

$$\frac{2GM}{c^2R} = 1, \tag{4}$$

which is also the relation of the effective mass and radius of the universe according to Mach's principle [22–24] as well as

the relation of the observable mass and radius of the universe according to observations. The mass and radius of a spacetime or black hole satisfy this Mach-Schwarzschild M-R relation. The space curvature constant of a closed spacetime or black hole is positive, *i.e.*

$$k = 1. \tag{5}$$

It is noted here that the big bang model suggests that the spacetime of the universe is flat (*i.e.* $k = 0$).

The cosmological principle expresses the matter inside a spacetime or black hole to be uniformly (*i.e.* isotropically and homogeneously) distributed in a scale which is sufficiently large (*i.e.* comparable rather than too small) in comparison with respect to the size of the spacetime. Then, the density of matter in a spacetime or black hole is given by

$$\rho \equiv \frac{M}{V} = \frac{3c^2}{8\pi GR^2} = \frac{3c^6}{32\pi G^3M^2}, \tag{6}$$

which is inversely proportional to the square of radius or the square of mass. This square-law density expression (ρR^2 = constant or ρM^2 = constant) naturally removes the flatness problem of the universe, which was first pointed out by Robert Dicke in the BBU [25–26].

Therefore, the flatness (or fine-tuning) problem does not exist in the black hole model of the universe. Furthermore, by measuring the density, we can determine both the radius and mass of the universe. For the density of the present universe to be about the critical density $\rho_0 \sim \rho_c \equiv 3H_0^2/(8\pi G)$, we have the mass and radius of the present universe to be $M_0 \sim 8.8\times10^{52}$ kg (about a half hundred sextillions of solar masses) and $R_0 \sim 1.32 \times 10^{26}$ m (about forty-three hundred Mpc or one Hubble length). Here, according to measurements [27–30], the Hubble constant is chosen as $H_0 = 70$ km/s/Mpc. Therefore, the present universe is an extremely supermassive and fully expanded black hole with extremely low density and weak gravity. The gravitational field at its surface is $g_0 = GM_0/R_0^2 \sim 3 \times 10^{-10}$ m/s^2 and thus a 100-kg object at the surface or inside only weighs 3×10^{-8} N or less.

The big bang universe is an isolated system and the total energy or mass (though unknown) is a constant, so that the density is inversely proportional to the cube of radius (*i.e.* ρR^3 = constant). Fig. 3 plots the density of a black hole as a function of its radius in the unit of kilometers (the solid line) or a function of the mass in the unit of 0.337 solar masses (the same line). The dashed line plots the density of the big bang universe as a function of its radius with mass equal to M_0 (for a bigger mass, the line is shifted to a larger radius). The dotted line marks the density of the present universe (ρ_0) and its intersection with the solid line shows the mass (M_0), density (ρ_0), and radius (R_0) of the present universe. Three circles along the solid line represent a star-like black hole with three solar masses, a supermassive black hole with three billion solar masses, and the present black hole universe with mass M_0. The black hole universe is not an isolated system because its mass increases as it expands. The density decreases inversely proportional to the square of the radius (or the mass) of the black hole universe. Considering that matter can enter but cannot exit a black hole, we can suggest that the black hole universe is a semi-open system surrounded by outer space and matter. The entire space is infinite, existed forever, and isolated. It contains everything without outside and edge. Inside the entire space, any universe has outside space and matter and thus cannot be isolated.

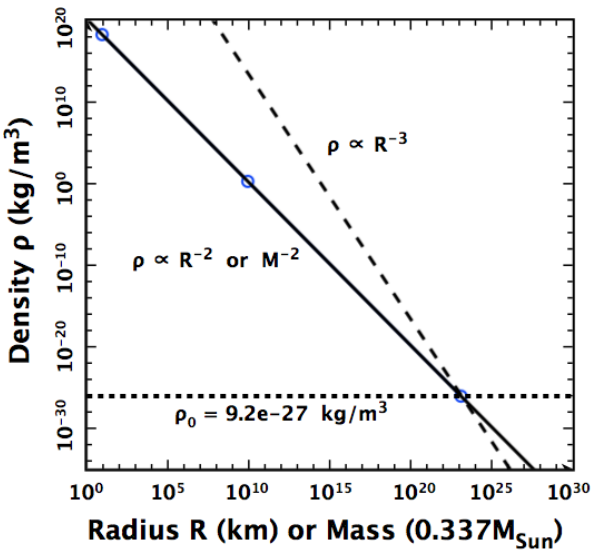

Fig. 3: The density of the black hole universe versus its mass and radius (the solid line). The dotted line refers to $\rho = \rho_0$, so that the intersection of the solid and dotted lines represents the density, radius, and mass of the present universe. The dashed line plots the density of the big bang universe, if it has mass M_0, as a function of the radius.

In the black hole universe model, we have that the effective and observable radii are the actual radius of the universe at all time, so that the black hole universe is always all observable and Mach's principle holds forever. In the big bang theory, the ratio between the effective radius and the radius of the universe increases as the universe expands and will reach the unity at a point, which is the present time if the universe has mass M_0. Before the point, the effective radius is less than the radius of the universe. While, after the point, the effective radius will be greater than the radius of the universe, at which Mach's principle does not hold, so that other physical laws neither hold.

According to GR and the stellar physics, a star with 20 or more solar masses, at the end of its life, will form a black hole after a supernova explosion [31]. Therefore, the black hole model of the universe does not need a big bang. The universe can be considered to originate from a star-like black hole (child universe) with several solar masses, which grows through a supermassive black hole with billion solar masses to the present universe with hundred-sextillion solar masses

by accreting ambient matter and merging with other black holes. Which one was the first or initial black hole that the universe has grown up from is not critical or important to the present universe because the mass of the original one only takes one part of a sextillion in the present universe. This origin of the universe from black holes not only overcomes the fine-tuning problem but also conquer the difficulty of banging the universe out from nothing that violates the law of conservation of energy. This resolves the big bang singularity problem. In addition, if the universe originated from a star-like black hole, it would not be hot enough to create a magnetic monopole in any time period, thus solving the magnetic monopole problem. The recent discovery of gravitational waves confirms the existence and merger of black holes [32] and thus support this black hole model of the universe.

Substituting the Mach-Schwarzschild M-R relation (4) (or density (6)) and positive space curvature constant (5) into the Friedmann equation (3), we have $\dot{R} = 0$ and $\dot{M} = 0$ – a zero rate of change in radius or mass. This indicates that a spacetime or black hole is static [7,33] when it neither accretes matter from the outside nor merges with other black holes. In the static state, the spacetime reaches equilibrium because the positive curvature balances the gravity entirely. A spacetime with the curvature radius R can hold the matter with mass equal to $c^2R/(2G)$ in equilibrium. Hawking [34] theorized the surface radiation of a static black hole with the quantum effect. For a star-like or more massive black hole, the Hawking radiation is negligibly weak and ultra-cold, which leads to the entropy of a static black hole to be 20 orders higher than its massive parent star. Including the cosmological constant, (3) determines Λ in the static state as $\Lambda = 3H^2$, which is $\sim 1.55 \times 10^{-35}$ s^{-2}.

3.2 Expansion and acceleration of spacetime

When a spacetime or black hole accretes its ambient matter or merges with other black holes, it becomes dynamic and expands. The rate of expansion or Hubble parameter is given by

$$H = \frac{\dot{R}}{R} = \frac{\dot{M}}{M}, \tag{7}$$

and the deceleration parameter is given by

$$q = -\frac{R\ddot{R}}{\dot{R}^2} = -\frac{M\ddot{M}}{\dot{M}^2}. \tag{8}$$

Here, the double dot symbol refers to the second order derivative of the parameter with respect to time. A spacetime or black hole expands if it gains matter, *i.e.* $\dot{M} > 0$, and accelerates if it gains matter in an increasing rate, *i.e.* $\ddot{M} > 0$. The expansion of spacetime is physical and outward without violating Einstein's light-speed upper limit and conservation of energy. A spacetime or black hole grows its space as it accretes matter by taking the outside space rather than by stretching the space of itself geometrically.

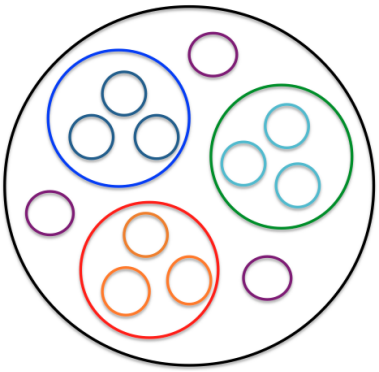

Fig. 4: A simple sketch of the innermost three layers of the entire space that is structured hierarchically. The black circle represents the mother universe. Our black hole universe is coded as red, in which three child universes (*i.e.* star-like or supermassive black holes) were drawn. Parallel to our universe, there are sister universes. Here two adult sister universes (blue and green circles) and three little sister universes (brown circles). The adult sister universes have also their child or baby universes, but the little sister universes are too young to have their babies.

For a spacetime or black hole including our black hole universe to expand, it must have an outside, where matter is available for accretion. The black hole model of the universe suggests that the entire space is structured with layers, hierarchically and family-like. Fig. 4 sketches the innermost three layers of the black hole universe including the mother universe (black circle), our universe itself (red circle), and child or baby universes (*i.e.* star-like or supermassive black holes). We have only drawn three child universes (yellow circles). We have also drawn two adult sister universes (blue and green circles) and three little sister universes (brown circles), which are universes parallel to our black hole universe. The adult sister universes have also their child universes. There should have a number of child universes and may also have many sister universes. A child universe grows by accreting material from its outside or by merging with other child universes. This universe grows up by accreting material from the mother universe or by merging with sister universes. The mother universe will also grow up if it has outside; otherwise, it is static. If the whole space is finite, then the number of layers is finite. Otherwise, it has infinite layers and the outermost layer corresponds to the limit of zero Kelvin for the absolute temperature, zero for the density, and infinite for the radius and mass.

From the data of type Ia supernova measurements, Daly et al. obtained the deceleration parameter of the present universe to be $q_0 \sim -0.48$ for the flat spacetime (for a closed spacetime, q_0 is smaller, *e.g.* $q_0 = -0.6$) [35]. Riess et al. and Perlmutter et al. explained the acceleration of the universe by suggesting the big bang universe to be dominated by dark energy up to about $\sim 73\%$ [36–37]. In the black hole universe model, however, the universe accelerates because it inhales the outside matter in an increasing rate, *i.e.* a positive $\ddot{M} > 0$. To have $q_0 = -0.48$, the present black hole universe

only needs to inhale the outside matter in an increasing rate at $\ddot{M}_0 = -q_0 M_0 H_0^2 \sim 2.2 \times 10^{17}$ kg/s^2 (or about 220 solar masses per year square).

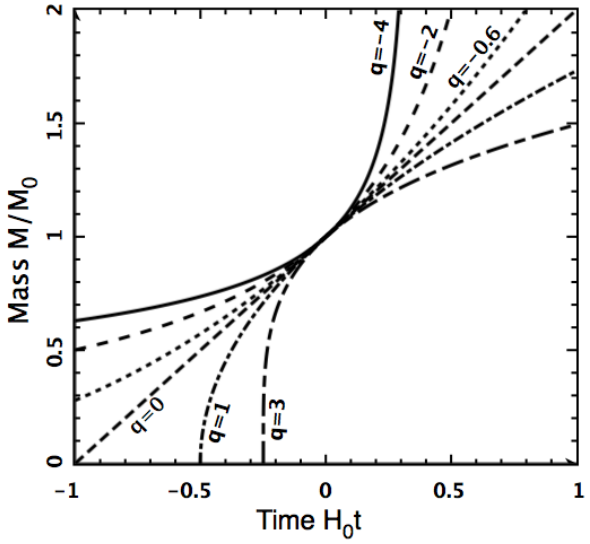

Fig. 5: Mass of black hole universe versus time with various deceleration parameters $q = -4, -2, -0.6, 0, 1, 3$.

For a constant acceleration expansion universe, the time-dependent mass can be analytically solved from (8) as [10],

$$\frac{M}{M_0} = [(q+1)H_0 t + 1]^{1/(q+1)}. \tag{9}$$

To quantitatively see how the mass $M(t)$ varies with time and depends on the deceleration parameter q, we plot in Fig. 5 according to (9) the mass as a function of time with various values of $q = -4, -2, -0.6, 0, 1, 3$. The lines with negative q are concave upward, which show that the mass increases in an increasing rate and the universe accelerates. The lines with positive q are concave downward, which show that the mass increases in a decreasing rate and the universe decelerates. For $q = 0$, the black hole universe accretes matter or increases its mass in a constant rate and thus expands uniformly.

The cosmological redshift of light from a source in an expanding spacetime is determined by

$$1 + Z = \frac{R_0}{R} = \frac{M_0}{M}. \tag{10}$$

The luminosity distance of the light source depends on the redshift as [10,38–39],

$$\begin{aligned} d_L &= (1+Z)M_0 \sin\left[\int_t^0 \frac{cdt}{M}\right] \\ &= (1+Z)R_0 \sin\left[\frac{c^3}{2GM_0H_0}\frac{1-(1+Z)^{-q}}{q}\right]. \end{aligned} \tag{11}$$

Here we have applied (9) and (10) to complete the integration. Eq. (11) reduces to the Hubble law, $H_0 d_L = cZ$, at $Z \ll 1$ [40] and perfectly explains the type Ia supernova measurements if the universe accretes matter in an increasing rate of $q = -0.6$ [34]) or $\ddot{M} \sim 3 \times 10^{17}$ kg/s^2 in average [10].

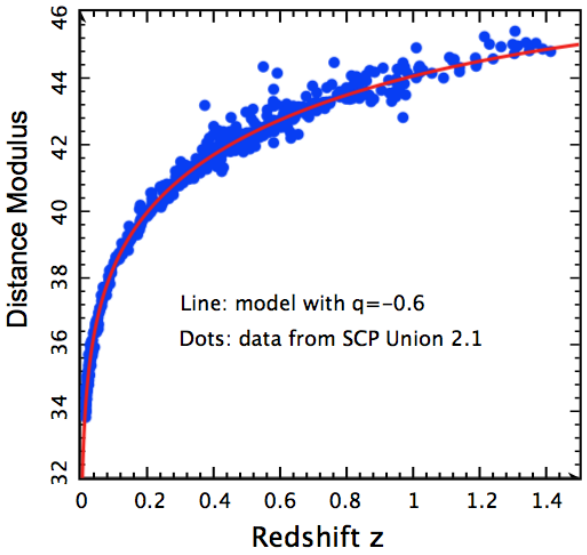

Fig. 6: Luminosity distance-redshift relation of type Ia supernovae. Blue dots are measurements credited by the Union2 compilation of 580 SNeIa data from Supernova Cosmology Project [41–42]. Red line is analytical results from this study with $q = -0.6$. The distance modulus is plotted as a function of the cosmological redshift.

Fig. 6 plots the luminosity distance-redshift relation (red line) along with the type Ia supernova measurements (blue dots), which are credited by the Union 2.1 compilation of 580 SNIA data from Supernova Cosmology Project [41–42]. In this plot the Hubble constant is chosen to be $H_0 = 70$ km/s/Mpc and the deceleration parameter is chosen to be $q = -0.6$. The distance modulus, which is defined by $\mu = 5 \log_{10} d_{\rm L} - 5$ with $d_{\rm L}$ in parsecs, is plotted as a function of redshift. The chi-square statistic is very close to unity [10]. Therefore, the black hole universe model can perfectly explain the measurements of type Ia supernovae without dark energy, which is needed to take $\sim 73\%$ in the big bang universe [36–37].

3.3 Temperature of spacetime and background radiation

The temperature inside a spacetime or black hole depends on the state and density of matter enclosed and hence depends on the radius or mass. The stellar physics has shown that a neutron star can reach trillions of Kelvin at the moment of its birth and then quickly cools down to hundred millions of Kelvin due to strong radiation and neutrino emissions. Since it is compact as a neutron star, a star-like black hole should also initially reach trillions of Kelvin but statically holds this hotness due to lack of significant emissions to the outside (the Hawking radiation is negligible). The thermal radiation inside a spacetime or black hole is the blackbody radiation

governed by the Planck law, from which one can derive the total energy of blackbody radiation inside a spacetime or black hole with radius R and temperature T to be

$$U_\gamma = \alpha R^3 T^4. \tag{12}$$

Here, the constant α is given by, $\alpha \equiv 32\pi^6 k^4/(45h^3c^3) \sim 3.2\times^{-15}$ J/m^3/K^4, with k the Boltzmann constant and h the Planck constant.

When a spacetime or black hole accretes matter and radiation from its outside, it becomes dynamic and expands. Considering that the gain of matter and radiation inside is equal to the loss of matter and radiation outside, we have [8]

$$\frac{dT}{dR} = -\frac{3T}{4R}\left[1 - \left(\frac{T_p}{T}\right)^4\right], \tag{13}$$

where T is the temperature inside and T_p is the temperature outside. This equation governs the thermal history of the black hole universe from its origin as a star-like black hole with several solar masses and growing through a supermassive black hole with billions of solar masses to the present state with hundred sextillions of solar masses. Since the temperature outside is always less than that inside, $T_p < T$, the temperature of a spacetime or black hole decreases with its radius. As the black hole universe grows in size from a star-like black hole to the present state, its temperature decreases from trillions of Kelvin to about 3 K [8]. The cosmic microwave background radiation (CMB) is explained as the blackbody radiation of the black hole universe – an ideal blackbody – rather than the fireball leftover of the big bang universe.

Considering the black hole universe to decrease its relative temperature in a rate slightly faster than the mother universe, we have [8]

$$T_p = aT^b, \quad \text{or} \quad T_p/T = aT^{b-1} \tag{14}$$

Here b is a constant slightly less than 1 and a can be derived from b according to the temperature and radius of the present universe (T_0 and R_0). Then, (13) can be analytically solved as

$$T = R^{-3/4}\left(a^4 R^{3-3b} + T_s^{4-4b} R_s^{3-3b}\right)^{1/(4-4b)}, \tag{15}$$

where the constant a is given by

$$a = \left[T_0^{4-4b} - \left(\frac{R_s}{R_0}\right)^{3-3b}\right]^{1/4}. \tag{16}$$

Choosing b appropriately (or slightly less than 1), we can completely determine the thermal history of the black hole universe that evolved from a hot star-like black hole with temperature T_s and radius R_s to the present universe with temperature T_0 and radius R_0. In Fig. 7, the temperature of the black hole universe is plotted as a function of the universe radius with $b = 0.93$. Here we have chosen $T_0 = 2.725$ K, $R_0 = 1.32 \times 10^{26}$ m, $R_s = 8.9$ km, and $T_s = 10^{12}$ K.

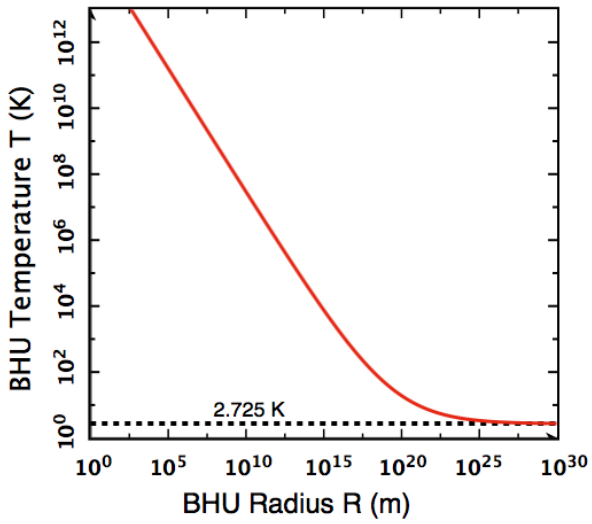

Fig. 7: The possible thermal history of the black hole universe. A hot star-like black hole with $T_s = 10^{12}$ K expands to the size of the present universe and cools down to ~ 2.725 K. The temperature line is curved by concaving upward and approaches ~ 2.725 K at the present time as the black hole universe expands to the present size.

It is seen that the temperature line is concave upward and approaches ~ 2.725 K as the black hole universe expands to the present size. The initial temperature of the star-like black hole T_s is not critical to the present universe. The reason is because most matter and radiation are from the mother universe. This reason also explains why all other physical properties of the star-like black hole, including its size (or mass), angular momentum, and charge, and the evolution of the early universe are not critical to the present universe. Furthermore, the early process of material accretion and black hole mergers do not have significant leftover in the present universe.

The above explanation of the CMB of this universe requires a decreasing temperature outside, *i.e.* an expanding mother universe. To have an expanding mother universe and explain its CMB with a decreasing temperature, there needs an expanding grandmother universe, and so forth. Therefore, the entire space is eternal and infinite, containing everything with infinite layers (Fig. 8). Nothing can be outside the entire space. The star-like or supermassive black holes called child universes belong to the innermost layer. They are subspacetimes of our black hole universe (the second innermost layer) that we live in. Our black hole universe is a subspacetime of the mother universe (the third innermost layer). The mother universe may contain a great number of child universes that are parallel to (and hence sisters of) our black hole universe. Mathematically, we can use an infinite large family tree that contains infinite generations or an infinite large set that contains infinite subsets to represent the relationships among different generations of black hole universes. The outermost layer called grand universe is infinitely large in size, mass, and entropy but has zero limits for both the density and

absolute temperature.

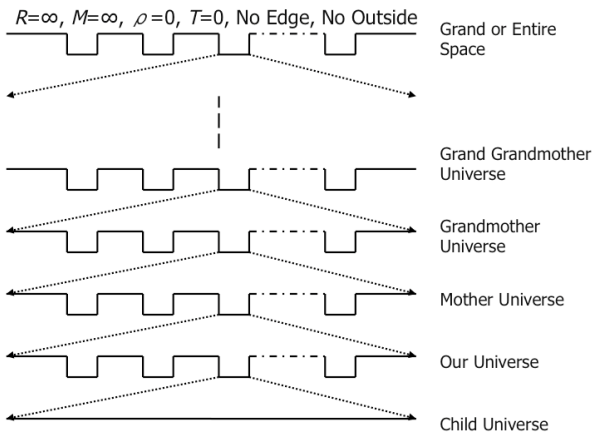

Fig. 8: The entire space with infinite layers or subspacetimess [8]. The bottom layer is a child universe or an empty spacetime. The child universe is a subspacetime of the universe in which we live in. Similarly, our universe is a subspacetime of the mother universe, and so on. The top layer is the entire space of all subspacetimes.

Each layer or black hole universe tends to absorb its outside matter and radiation and takes its outside space and expands outward. When our black hole universe expands to be one as large as the mother universe, the inside star-like and supermassive black holes will have merged and grown up into a black hole universe that is similar to the present one. This process is irreversible with neither a beginning nor an end. The evolution of black hole universe is iterative – beginningless and endless. When one black hole universe is expanded out, a new similar black hole universe is formed from inside child universes [7]. The black hole model of the universe is complete because it can address our universe not only at the present as well as its inside, but also in the past and future as well as its outside.

The total radiation energy inside the black hole universe is plotted in Fig. 9 as a function of the radius. It is seen that a young black hole universe with radius less than 10^{15} m or mass less than some hundred billions of solar masses remains the total radiation energy as a constant. This characteristic allows us to explain the activities and emissions of dynamic star-like and supermassive black holes observed in the universe.

3.4 Emissions of dynamic black holes

For a star-like black hole with several solar masses to grow through a supermassive black hole with billion solar masses, the temperature outside is negligibly lower than the temperature inside, *i.e.* $T_p \ll T$. In this case, (13) can be solved as

$$R^3T^4 = \text{Constant}, \tag{17}$$

which implies that the total radiation energy inside a spacetime or black hole with mass about billions of solar masses or

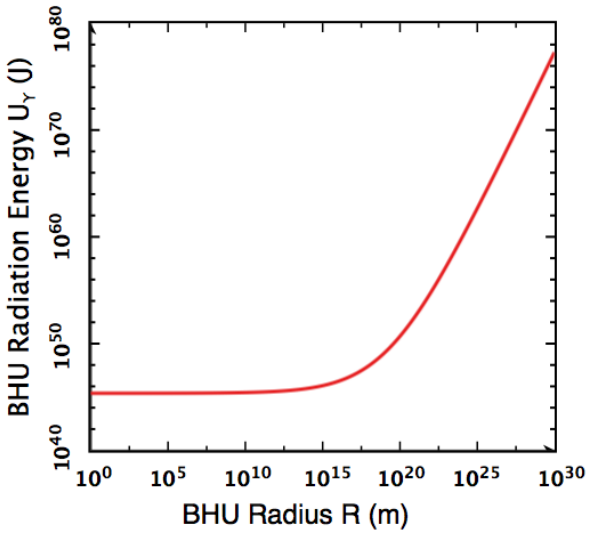

Fig. 9: Radiation energy of the black hole universe. As a hot star-like black hole with $T_s = 10^{12}$ K expands to the size of the present universe and cools down to ~ 2.725 K, its radiation energy first remains as a constant and then rapidly increases with radius when it grows into a supermassive black hole with radius greater than about thousand billions of kilometers or mass greater than about hundred billions of solar masses.

less remains the same amount as shown in Fig. 9. Therefore, accreting outside matter and radiation or merging with other black holes into a single one, a black hole not only becomes dynamic and expands but also intensively emits its inside hot and hence high-frequency blackbody radiation out of its horizon, which has been disturbed or broken by the accretion or merger in order for the total energy of its inside radiation to remain as a constant.

This emission mechanism of dynamic black holes can self-consistently explain the observed gamma ray bursts, X-ray flares from galactic centers, and quasar emissions as emissions of dynamic star-like, massive, and supermassive black holes, respectively (the details on these have been described in [9,11–12]). Dynamic star-like black holes with trillions of Kelvin radiate gamma rays and produce gamma ray bursts, while dynamic massive or supermassive black holes with millions to billions of Kelvin radiate X-rays such as X-ray emissions from quasars and X-ray flares from Sgr A* (a massive black hole at the Milky Way center). The energetic events associated with black holes are activities of child universes. The author has shown that a child universe with radius $R \gtrsim 10^{18}$ m or mass $M \gtrsim 3 \times 10^{14}$ solar masses does not emit [9], but can strongly attract and accrete its ambient matter including galaxies, which may help us to understand great attractors observed with thousand trillions of solar masses, *e.g.* the Norma Cluster. On the other hand, quasars if electrically charged may have a significant electric redshift as illustrated by [43]. The merger of star-like black holes if missing mass may release significant gravitational waves as recently detected by LIGO [32].

4 Discussion and conclusions

In addition to above issues that have been addressed in details in the early papers [7–12,39], the black hole model of the universe can also self-consistently illustrate various other problems of the universe such as why the redshifts of galaxies are quantized, how the galaxies and clusters are formed, why the expansion of the universe can be anisotropic, how the elements are synthesized, why the universe increases its entropy extremely without significantly increasing its disorder, how the heavy-ion enriched objects are formed in extremely deep fields or the young universe, what the great attractor is, why the voids exist, and so on. Preliminary results on some of them have been presented in a sequence of AAS (213^{rd}-215^{th}, 217^{th}, 219^{th}-224^{th}, 228^{th}) meetings and the details on all these problems will be addressed in future in full length papers.

The BHU stands on three bases, which are (1) GR of describing matter effect on spacetime, (2) CP of spacetime homogeneity and isotropy, and (3) SBHEP of spacetime black hole equivalence. We have not yet explored the quantum effect on this model to pop up baby universes and holes. In this model, baby or child universes are star-like and supermassive black holes, which are formed from stars and galaxies. To appropriately explain CMB, the entire space is favored to be infinite and eternal and includes infinite universes, which are layered hierarchically and evolved iteratively. Due to gravity and Jeans collapse criterion, matter forms stars, which then, if massive, end as black holes or child universes. A black hole, once formed, will grow and expand by accreting its ambient matter and merging with other black holes. A galaxy (usually including a massive black hole at its center), once most stars run out their fuels and died as dwarfs, neutron stars, and black holes, will eventually form a supermassive black hole (or quasar) by accreting all galactic matter and objects, and merging all stellar black holes into the massive black hole at the center. LIGO recently discovered the gravitational wave that confirms the existence of black holes and their merger [32]. A black hole universe can be considered to be originated or born from a star-like black hole (or child universe) without a big bang singularity, flatness, horizon, and magnetic monopole problems. It gradually grows or expands by accreting outside matter or merging with other black holes without dark energy and inflation problems. Each star-like black hole or supermassive black hole is usually rotating with significant angular momentum. But when many randomly rotating black holes merge to form a large universe like our present universe, the net angular velocity may be negligibly small. Inside a fully expanded or grown universe, objects formed from the collapse of matter (*e.g.* planets, stars, galaxies, clusters, *etc.*) can rotate globally. Gamow speculated that the rotations of these objects might be due to the cosmic rotation [44] and Godel obtained a cosmological solution of Einstein's field equation for rotating universes [45]. The black hole model of the universe is a model with multiverses (infinite or uncountable), which are hierarchically layered. It is different from other models of multiverse such that the many-world (or universes) interpretation of quantum physics proposed by [46] and the branes model of multiverse that suggested the visible 4D spacetime universe to be restricted inside a higher-dimensional space [47].

The three bases of BHU (GR, CP, and SBHEP) with well-developed physics theories and laws such as the conservation of energy, Planck's radiation, and so on can derive some laws or regularities of the BHU such as the spacetime equilibrium, the spacetime expansion and acceleration, the conservation of blackbody radiation, the increase of entropy, and so on that regulate and govern the development and dynamics of black hole universes. These laws or regularities can help us to explain and describe the origin, structure, expansion, evolution, acceleration of the universe, CMB, quasar, Sgr A* X-ray flare, *etc.* and meantime to overcome problems such as the horizon, flatness, monopole, dark matter, dark energy, low initial entropy, redshift quantization, big bang, old objects in the young universe, entropy, and so on. The BHU does not have unknowns. Both the charge and angular momentum are zero ($Q = 0$ and $J = 0$). The mass M is the only or key parameter. The radius or scale factor R, the temperature T, and the entropy S are derived from M according to the relations given by (4), (13), and entropy equations of thermodynamics. Measuring density tells us the radius R, and thus T and S. Measuring the Hubble parameter H tells the rate of change in $\dot{M}$, and thus $\dot{R}$, $\dot{T}$, $\dot{S}$. Measuring the deceleration parameter q tells us the double rate of change in mass $\ddot{M}$, and thus radius $\ddot{R}$, temperature $\ddot{T}$, entropy $\ddot{S}$. Measuring CMB, supernovae, *etc.* also tells us R and thus M, T, S, so that finds how the universe expands, *e.g.* acceleration or not.

As a consequence, installing one more leg (or fundamental) – the Principle of Spacetime Black Hole Equivalence – to the cosmology, we can attempt to fully explain the universe and simply overcome the difficulties according to the well-developed physics without needing to make other hypotheses such as inflation, dark energy, and so on. The black hole model of the universe is robust by only needing one stroke (the single postulate or principle SBHEP) rather than relying on an increasing number of hypothetical entities (HEs) as done in the big bang model [5] to explain the universe and solve the cosmic problems.

Acknowledgements

This work was partially supported by the NSF HBCU-UP, REU (PHY-1263253), and AAMU Title III programs.

Submitted on August 1, 2016 / Accepted on September 9, 2016

References

1. Friedmann A. Über die Krümmung des Raumes. *Zeitschrift für Physik*, 1922, v. 10, 377–386.

2. Lemaître G. Expansion of the universe, A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae. *Monthly Notices of the Royal Astronomical Society*, 1931, v. 91, 483–490.
3. Robertson H. P. Kinematics and world-structure. *Astrophysical Journal*, 1935, v. 82, 284–301.
4. Walker A. G. On Milne's theory of world-structure. *Proceedings of the London Mathematical Society*, 1937, v. 42, 90–127.
5. Arp H. et al. An open letter to the scientific community - signed by scientists/engineers/researchers. *New Scientist*, 2004, May, 22.
6. Zhang T. X. A new cosmological model: Black hole universe. *American Astronomical Society 211st Meeting*, 2007, Abstract #152.04.
7. Zhang T.X. A new cosmological model: Black hole universe. *Progress in Physics*, 2009, v. 5 (2), 3–11.
8. Zhang T. X. Cosmic microwave background radiation of black hole universe. *Astrophysics and Space Science*, 2010, v. 330, 157–165.
9. Zhang T. X. Quasar formation and energy emission in black hole universe. *Progress in Physics*, 2012, v. 8 (3), 48–53.
10. Zhang T. X., Frederick C. Acceleration of black hole universe. *Astrophysics and Space Science*, 2014, v. 349, 567–573.
11. Zhang T. X. Gamma ray bursts and blak hole universe. *Astrophysics and Space Science*, 2015, v. 358., article.id. #14, DOI 10.1007/s10509-015-2409-1, 8 pp.
12. Zhang T. X., Wilson C. and Schamschula M. P. X-ray flares from Sagittarius A* and black hole universe, *Progress in Physics,*, 2016, v. 12 (1), 61–67.
13. Einstein A. Die grundlage der allgemeinen relativitatstheorie. *Annalen der Physik*, 1916, v. 354, 769–822.
14. Einstein A. Kosmologische betrachtungen zur allgemeinen relativitatstheorie. *Sitz. Preu. Akad. Wiss. (Berlin)*, 1917, Part 1, 142–152.
15. de Sitter W. Einstein's theory of gravitation and its astronomical consequence. *Monthly Notices of the Royal Astronomical Society*, 1917, v. 78, 3–28.
16. Friedmann A. Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. 1924, *Zeitschrift für Physik*, v. 21, 326–332.
17. Schwarzschild K. Ober das gravitationsfeld eines massenpunktes nach der Einsteinschen theorie. *Sitz. Pruess. Akad. Wiss.*, 1916, v. 1, 189–196.
18. Zhang T. X. Principle of spacetime and black hole equivalence. *American Astronomical Society 228th Meeting*, 2016, Abstract #403.08.
19. Misner C. W. Coley A. A., Ellis G. F. R., Hancock M. The isotropy of the universe. *The Astrophysical Journal*, 1968, v. 151, 431–457.
20. Misner C. W. Thorne K. S., Wheeler J. A. *Gravitation*, 1973, San Francisco: W.H. Freeman and Co., ISBN 0-7167-0344-0.
21. Guth A. H. Inflationary universe: A possible solution to the horizon and flatness problems. *Physical Review D*, 1981, v. 23, 347–356.
22. Brans C. H., Dicke R. H. Mach's principle and a relativistic theory of gravitation. *Physical Review*, 1961, v. 124, 925–935.
23. Sciama D. W. On the origin of inertia. *Monthly Notices of the Royal Astronomical Society*, 1953, v. 113, 34–42.
24. Davidson W. General relativity and Mach's principle. *Monthly Notices of the Royal Astronomical Society*, 1957, v. 117, 212–224.
25. Dicke R. H. Remarks on gravitation and cosmology. *Proceedings of the International Symposium for Theoretical Physics*, 1969, v. 1, 507–510.
26. Dicke R. H., Peebles P J. E. The big bang cosmology - enigmas and nostrums. *General Relativity*, 1979, 504–517.
27. Hughes J. P., Birkinshaw M. A. A measurement of the Hubble constant from the X-ray properties and the Sunyaev-Zeldovich rffect of CL 0016+16, *The Astrophysical Journal*, 1998, v. 501, 1–14.
28. Mauskopf P. D. *et al*. A determination of the Hubble constant using measurements of X-ray emission and the Sunyaev-Zeldovich effect at millimeter wavelengths in the cluster Abell 1835. *The Astrophysical Journal*, 2000, v. 538, 505–516.
29. Sandage A., Tammann G. A., Saha A., Reindl B., Macchetto F. D., Panagia N. The Hubble constant: A summary of the Hubble Space Telescope Program for the luminosity calibration of type Ia supernovae by means of Cepheids. *The Astrophysical Journal*, 2006, v. 653, 843–860.
30. Suyu S. H. *et al*. Dissecting the gravitational lens B1608+656. II. Precision measurements of the Hubble constant, spatial curvature, and the dark energy equation of state. *The Astrophysical Journal*, 2010, v. 711, 201–221.
31. Carroll S. M. Spacetime and geometry. An introduction to general relativity. San Francisco, CA, USA: Addison Wesley, 2004, ISBN 0-8053-8732-3.
32. Abbott B. P. *et al*. Observation of gravitational waves from a binary black hole merger. Physical Review Letters, 2016, v. 116, id.061102
33. Carter B. Black hole equilibrium states, in B. S. DeWitt and C. DeWitt, Black Holes (Les Astres Occlus), 1973, 57–214.
34. Hawking S. Black hole explosions? *Nature*, 1974, v. 248, 30–31.
35. Daly R. A. *et al*. Improved constraints on the acceleration history of the universe and the properties of the dark energy *The Astrophysical Journal*, 2008, v. 677, 1–11.
36. Riess A. G. *et al*. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astronomical Journal*, 1998, v. 116, 1009–1038.
37. Perlmutter S. *et al*. Measurements of Omega and Lambda from 42 high-redshift supernovae. *Astrophysical Journal*, 1999, v. 517, 565–586.
38. Weinberg S. *Gravitation and Cosmology*, Wiley: New York, NY, USA, 1980.
39. Zhang T. X. Key to the mystery of dark energy: Corrected relationship between luminosity distance and redshift. *Progress in Physics*, 2013, v. 5, 1–6.
40. Hubble E. P. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences of the United States of America*, 1929, v. 15, 168–173.
41. Amanullah R. *et al*. Spectra and Hubble Space Telescope light curves of six type Ia supernovae at $0.511 < z < 1.12$ and the Union2 compilation, *The Astrophysical Journal*, 2010, v. 716, 712–738.
42. Suzuki N. *et al*. The Hubble Space Telescope Cluster Supernova survey. V. Improving the dark-energy constraints above z ¿ 1 and building an early-type-hosted supernova sample, *The Astrophysical Journal*, 2012, v. 746, article id. 85, 24 pp.
43. Zhang T. X. Electric redshift and quasar. *The Astrophysical Journal Letters*, 2006, v. 636, L61–L63.
44. Gamow G. Rotating universe? *Nature*, 1946, v. 158, 549–549.
45. Godel K. An example of a new type of cosmological solutions of Einstein's field equations of gravitation. *Reviews of Modern Physics*, 1949, v. 21, 447–450.
46. Everett H. Relative state formulation of quantum mechanics, *Reviews pf Modern Physics*, 1957, v. 29, 454–462.
47. Rubakov V. A., Shaposhnikov M. E. Do we live inside a domain wall? *Physics Letters B*, 1983, v. 125, 136–138.

Attempt to Replicate Cahill's Quantum Gravity Experiment to Measure Absolute Velocity

Jay R. Seaver

Energy Matters Foundation, PO BOX 2588, Longmont, CO 80502, USA
E-mail:jay@energy-matters.org

In December 2015 in a laboratory in Longmont, Colorado, USA, I attempted to repeat the experiments of Reginald T. Cahill for detecting dynamical space by using reverse biased Zener diodes as quantum tunnelling devices whose tunnelling currents are modulated by the motion of dynamical space relative to the earth. I successfully produced the correlated signals of the same frequency and amplitude as Cahill has produced in his laboratory in Adelaide, Australia. But I determined that rather than being disturbances in space, these signals were merely transient responses to local electromagnetic disturbances which appeared to be correlated due to the identical natural frequencies of the two detectors. This paper is a report on those experiments.

1 Introduction

Recent papers by Cahill [1–3] discuss gravity wave detection using reverse-biased Zener diodes as "quantum gravitational wave detectors". In December 2015, in Longmont, Colorado, I built these quantum wave detectors using the identical parts and schematic as Cahill in order to confirm his measurements. I consulted with Cahill via email to make sure they were exactly as he designed them.

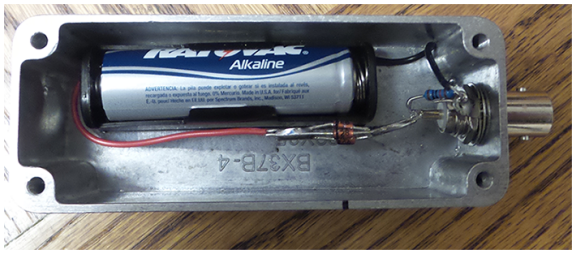

Fig. 1: Inside of Quantum Detector used in experiment.

Figure 1 shows a photograph of the inside one of the quantum detectors used in the experiment. It consists of a parallel connected array of three 3.0 V 1N4728 Zener diodes serially connected to a 1.5 V battery and a 10 kOhm sense resistor. The voltage across the sense resistor goes through a BNC connector and a 3 ft. RG58 coax cable to the AC-coupled input of a LeCroy 1 GHz bandwidth Digital Sampling Oscilloscope (DSO). The schematic and a picture of the inside of Cahill's detector can be seen in Figure 1 of [1].

Figure 2 shows the detector after it has been sealed up inside an aluminum case and connected to the coax cable that goes to the DSO input.

In my correspondence with Cahill in December 2015, he was kind enough to take some additional measurements and send me the oscilloscope pictures of the correlated quantum waves he is detecting in his laboratory in Adalaide, Australia.

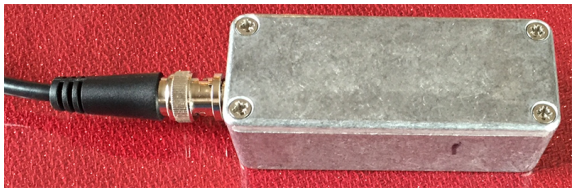

Fig. 2: Enclosed Quantum Detector used in experiment.

I was able to capture nearly identical correlated signals in my laboratory in Longmont, Colorado. However, upon further investigation of these signals I determined that they were of local origin and that the frequency of the waveforms was tied to the resonant frequency of the detector-cable system. I have concluded that the "correlation" Cahill sees is only an apparent correlation because the circuits of the two detectors, when excited by an external disturbance, produce nearly the same transient response due to their being nearly identical circuits with nearly the same natural frequency. An external disturbance, such as a nearby static discharge is required to excite the transient response. The correlated signals start out in phase but slip with time because the two resonant frequencies are not exactly the same. The measured phase difference is simply a function of how much time elapses from the moment of excitation until the scope triggers and captures the waveforms.

This paper documents the experiments I performed and my reasoning for coming to the above conclusions.

2 Cahill's data

Figures 3 and 4 show data from Dec. 13 and 14, 2015, taken by Cahill in his laboratory in Adelaide, Australia, and sent to me via email as an example of what the current fluctuations from the gravity waves look like for collocated detectors. Similar pictures of gravity waves in his detectors can be seen in [1–3]. Notice that the frequencies in these two

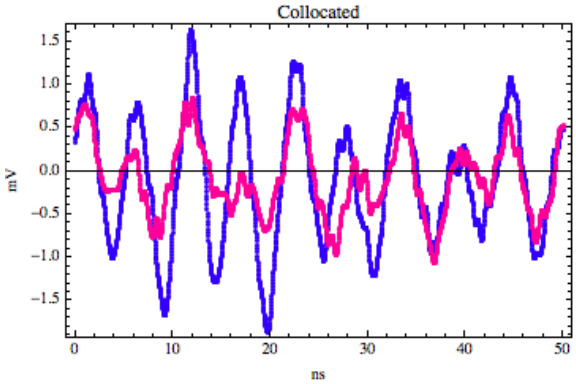

Fig. 3: Cahill Dec 13 Typical Data with detectors collocated.

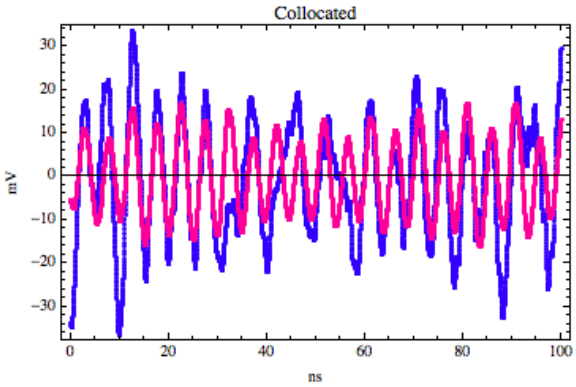

Fig. 4: Cahill Dec 14 Typical Data with detectors collocated.

plots are near 200 MHz and appear almost like a tone superimposed on noise. This seemed odd to me given that Cahill in his papers says that the frequency spectrum of the fluctuations in the detectors has a 1/f amplitude relationship. The reason the tone seen in his data does not show up in the frequency spectrum plots, is because they occur so infrequently. Most of the time the current fluctuations are at very small quiescent levels that look like random, uncorrelated noise. This quiescent current is disturbed at random periods by bursts of energy at mostly a single frequency, which are the waveforms captured in Cahill's pictures. Because these energy bursts are short with long periods of time between them, they have little effect on the Fourier transform over a wide frequency band — hence the 1/f relationship without evidence of these tones.

3 My experimental data

Figure 5 is a photo of my oscilloscope on Dec. 11, 2015, showing the quiescent signal from the detectors. Notice that there is little, if any, evidence of correlation between the two waveforms. The scale is 2 mV per division vertically and 10 ns per division horizontally. Notice also that the peak-to-peak fluctuations are typically less than 1 mV.

On December 11, when I took the picture in Figure 5, I was unable to detect any signals except the quiescent current.

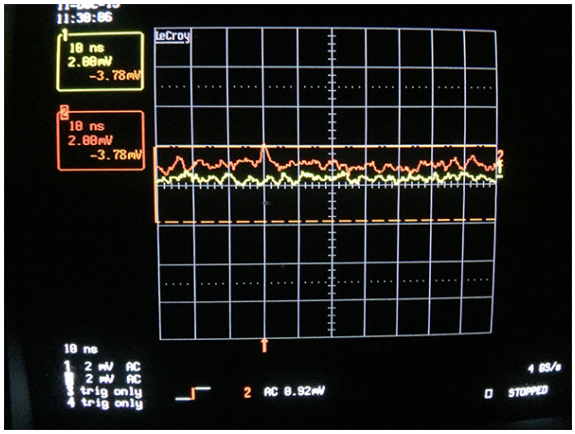

Fig. 5: Quiescent waveforms of collocated detectors.

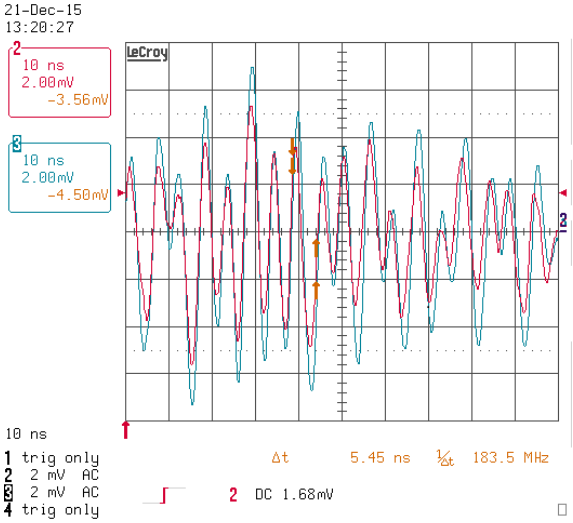

Fig. 6: Burst of energy from collocated detectors in my laboratory on Dec. 21, 2015.

I tried various orientations of the detectors but gave up after a few hours of searching. After communicating with Cahill via email, he sent me the pictures of Figures 3 and 4 showing me what he was seeing in his laboratory. I then went back into my laboratory on Dec 21 and set up my oscilloscope to trigger on any signals above 1.5 mV. After several minutes, I suddenly got a large burst of energy at about 200 MHz just like Cahill. This is shown in Figure 6.

The fundamental frequency of this waveform is highly correlated between the two detectors. However, I noticed a subtle difference between the two waveforms that should not be there if they are truly being modulated by the same source. The phase of the two waveforms is nearly perfectly aligned on the left side of the screen but it is drifting apart as one moves towards the right side of the screen. This is what one would see if two different, but nearly identical frequencies were observed. It is not what one would see from a single modulating

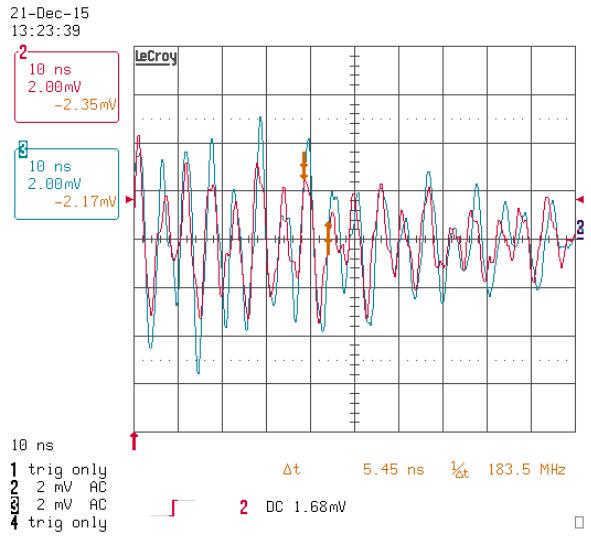

Fig. 7: Another captured waveform from collocated detectors in my laboratory on Dec. 21, 2015.

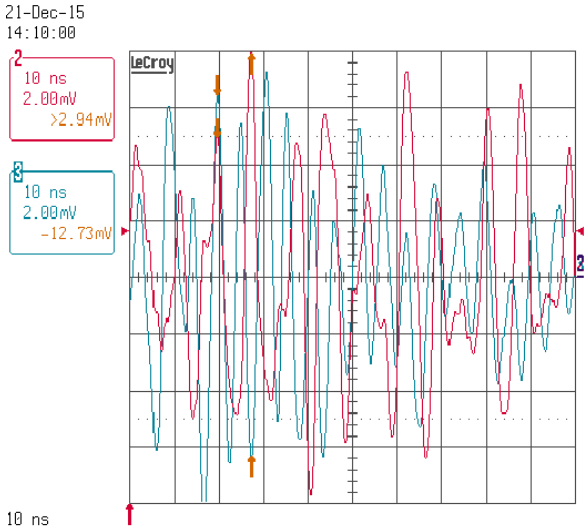

Fig. 8: Waveforms from collocated detectors with different cable lengths.

source observed on two different detectors.

I set the scope up to capture another signal and got the waveform shown in Figure 7. Notice again the same effect. The phase is aligned on the left and is slowly drifting apart as it moves to the right. If collocated detectors were being excited by the exact same gravity waves, the phase between them would not drift. At this point I realized that something was not right. My first suspicion was that my two "identical" detectors were not quite identical, but had natural frequencies in their circuits (including the cables) that were not quite the same. They were being excited by some external signal, but the actual response I was seeing was not a gravity wave, but simply the transient response of each of these circuits as they resonated at their not-quite-equal natural frequencies.

To test this theory, I replaced the cable on one of the detectors with a longer cable. I now had a 3 ft. coax cable on the detector going to channel 3 (blue) of the scope and a 5 ft. cable on channel 2 (red). The result was the waveforms shown in Figure 8. This shows very clearly that the red waveform has a fundamental frequency significantly lower than the blue waveform. I had proof positive that these 200 MHz energy bursts were not from 200 MHz gravity waves.

But there was still the question of what caused the excitation of the circuits to start with. Could it be Cahill's gravity waves that provide the initial excitation? Or was the source of local origin? My next experiment was to separate the detectors by a few millimeters to see if the phases of the two waveforms would start out with an initial phase difference. This is what would happen if they were being exited by passing through Cahill's gravity waves due to the velocity of space past the earth. Cahill asserts that the velocity of the earth is about 500 km/sec which represents about 2 ns/mm in phase shift if the detectors are directly aligned with this velocity. If they are not aligned, an even larger phase shift per mm would be observed. I saw no change in the phase relationship between the two signals as the detectors were moved relative to each other. The waveforms remained in phase at the beginning and drifted with time. This indicates an initial excitation disturbance moving at the velocity of light — not 500 km/sec.

In [4], Vrba noted that the battery, diode and resistor circuit form an electromagnetic wave sensing loop having a substantial cross section. Although the circuit is enclosed inside an aluminum box that shields electric fields, it is not a perfect shield. It will highly attenuate an electromagnetic wave, but with the oscilloscope set to its most sensitive level of 2 mV per division, even an attenuated signal could still be large enough to be detected.*

As I was pondering how to identify the source of the initial excitation, I noticed something very interesting. My oscilloscope would not trigger unless there were people in the laboratory. If everyone left and there was no nearby human activity, the signals would remain at their quiescent (< 1.4 mV) level and the scope would never trigger. But once nearby human activity resumed the scope would begin triggering again every few minutes. It didn't take long to find a correlation between static discharges from human activity and the energy bursts in the scope. By experimenting, I found that I could generate a frequency burst that would trigger the scope from as far away as 20 meters by shuffling my shoes on the car-

*Although not reported in this paper, I designed a second experiment using an architecture similar to Vrba's. It included 200× amplification with a bandwidth of 10 MHz to detect even smaller signals. The resonant disturbances disappeared, which left only the random noise. Visually examining these waveforms, I saw no evidence of correlated signals. The raw data files are available upon request at the email address given above for anyone desiring to perform a more sophisticated search for correlation between the waveforms.

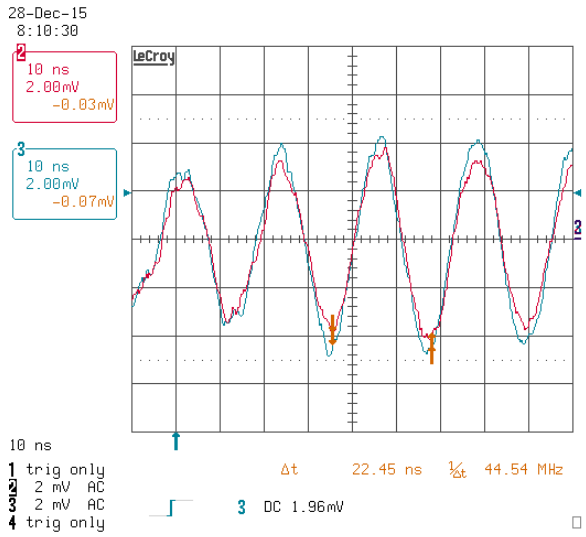

Fig. 9: Waveforms from collocated detectors with battery bypassed with 1 nF capacitor.

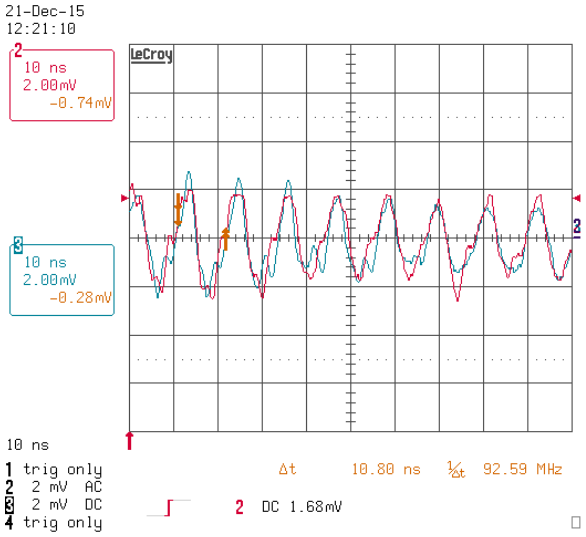

Fig. 10: Waveforms from collocated disconnected coax cables.

pet and touching something metallic. The waveforms looked identical in amplitude and frequency to those above.

As an additional proof that the detected frequency was entirely determined by the circuit, I made 2 modifications to the circuit. First, I put a 1 nF capacitor across the battery to provide a low impedance path for high frequencies. It caused the frequency of the transient response to drop to below 50 MHz as shown in Figure 9. I then removed the quantum detector entirely, and just left the two 3-foot, collocated coax cables disconnected. The waveforms are shown in Figure 10. These results further strengthen the argument that the frequency of the waveforms are determined entirely by the circuit itself.

4 Conclusions

After attempting to repeat the gravity wave detection experiments of Cahill using reverse biased Zener diodes as quantum tunnelling devices, I found no evidence of current fluctuations due to anything but normal random noise or local disturbances followed by a transient oscillation at the natural frequency of the detector circuits. The so-called correlation of the signals between detectors was merely an apparent correlation due to the fact that the circuits have natural frequencies that are nearly identical. This was proven by changing the natural frequencies of the circuits and showing that the frequency of the "gravity waves" changed to the new frequency.

The initial excitation of the circuits was shown to be from local sources — not disturbances in "dynamical space" as proposed by Cahill. The detectors exhibited no evidence of being excited by anything but uncorrelated random noise unless nearby human activity was generating static discharges. No evidence of any correlated signals between detectors was ever seen at any frequency other than the natural frequency of the detector circuits (superimposed on noise and/or reflections in the cables).

Whether Cahill has ever detected disturbances due to dynamical space, I cannot say. But I am satisfied that in Longmont, Colorado in December of 2015, there was no evidence that dynamical space was detectable using the Zener diode circuit Cahill has proposed in his papers.

Acknowledgements

I am indebted to Reg Cahill for his assistance in getting the quantum detectors built properly, providing advice on how to detect the signals and in supplying me with examples of the signals he obtained in his laboratory. It is unfortunate that my experiments ended up contradicting his papers rather than confirming them.

Submitted on September 15, 2016 / Accepted on September 17, 2016

References

1. Cahill R.T. Quantum Gravity Experiments. *Progress in Physics*, 2011, v. 11(4), 317–320.
2. Cahill R.T. Nanotechnology Quantum Detectors for Gravitational Waves: Adelaide to London. *Progress in Physics*, 2013, v. 4(4), 57–62.
3. Cahill R.T. Gravitational Wave Experiments with Zener Diode Quantum Detectors: Fractal Dynamical Space and Universe Expansion with Inflation Epoch. *Progress in Physics*, 2014, v. 10(3), 131–138.
4. Vrba A.L. Reservations on Cahill's Quantum Gravity Experiment. *Progress in Physics*, 2015, v. 11(4), 330.

LETTERS TO PROGRESS IN PHYSICS

Some Insights on the Nature of the Vacuum Background Field in General Relativity

Patrick Marquet

18 avenue du Président Wilson, 62100 Calais, France
E-mail: patrick.marquet6@wanadoo.fr

In our publication "Vacuum Background field in General Relativity" (*Progress in Physics*, 2016, v. 12, issue 4, 313–317) we introduced a kind of "relic" field permanently filling the empty space-time. This proved to be a necessary ingredient to formulate a true vector describing the gravitational field arising from matter, in contrast to the awkward pseudo-tensor usually suggested to ensure the conservation of the energy-momentum in the field equations. In this short paper, we give this field a mathematical description in terms of geodesics.

The background field that persists in vacuum devoid of any matter or energy, finds a physical meaning if we consider the *Landau-Raychaudhuri equation* for a congruence of non-intersecting timelike unit vectorial field $\boldsymbol{X}$, $(X_a X^a = 1)$, i.e.:

$$R^a{}_b\, X_a X^b = -{}^\circ X^a{}_{;a} - \omega_{ab}\,\omega^{ab} + \sigma_{ab}\,\sigma^{ab} + {}^\circ\theta + \frac{1}{3}\,\theta^2, \quad (1)$$

where $^\circ$ means differentiation with respect to proper time τ. In the scalar ζ which is the *Lagrangian density of the vacuum background field*

$$\zeta = \sqrt{-g}\,\nabla_a\,\kappa^a \quad (2)$$

we set up

$$\nabla_a\,\kappa^a = \theta^2, \quad (3)$$

where θ is the *space time volume expansion* characterizing this background field through

$$\theta = X^a{}_{;a} = h^{ab}\theta_{ab} \quad (4)$$

with the expansion tensor $\theta_{ab} = h^c_a h^d_b\, X_{(c;d)}$ ($h_{ac} = g_{ac} - X_{ac}$ is the projection tensor).

The formula of $X_{(c;d)}$ can be regarded as measuring the rate of change of the space-time 4-volume, i.e. either expansion if positive, or contraction if negative.

The form (3) has been chosen so as to preserve the integrity of the gravity tensor equation irrespective of the sign of θ.

In our case (absence of energy/matter), the background field obviously follows a contraction process (negative expansion) of space-time, and the Landau-Raychaudhuri reduces to

$${}^\circ\theta = R^a{}_b\, X_a X^b - \sigma_{ab}\,\sigma^{ab} - \frac{1}{3}\,\theta^2 < 0 \quad (5)$$

(since the vorticity tensor ω^{ab} induces expansion, while the shear tensor σ^{ab} induces contraction and the geodesic equation ${}^\circ X^a{}_{;a}$ is zero).

$R^a{}_b\, X_a X^b$ (sometimes referred to as the Raychaudhuri scalar) is always positive ensuring that the Strong Energy Condition (SEC) is not violated when energy/matter is there. Therefore, we are left with the inequality

$${}^\circ\theta \leqslant \frac{1}{3}\,\theta^2. \quad (6)$$

Integrating it with respect to the proper time τ yields

$$\theta^{-1} \geqslant \theta_1^{-1} + \frac{1}{3}\,\tau\,, \quad (7)$$

where θ_1 is the initial value which can be positive to start with, but very soon after the short expansion, it is followed by re-collapse. The mathematical fate of (timelike) geodesics is a final focusing to a caustic ($\theta \to -\infty$) after a finite proper time of at most

$$\tau \leqslant \frac{3}{\theta_1} \quad (8)$$

after the measurement of the initial value.

Such a state is called *geodesic incompleteness* which is a notion introduced by Hawking-Penrose, to describe (or not !) a geodesics path of observers through space-time that can only be extended for a finite time as measured by an observer travelling along one.

Presumably, at the end of the geodesic, the observer has fallen into a "kink" or encountered some other pathology at which the laws of General Relativity breakdown.

As Landau pointed out, in a *synchronous comoving frame of reference* attached to a homogeneous fluid, such a singularity can be removed by the introduction of a pressure which tends to substantiate our space-time contraction hypothesis.

All these contribute to our impossibility to give a full description of the vacuum background field. In a sense, this marks the lowest horizon level of the space-time.

Submitted on August 31, 2016 / Accepted on September 30, 2016

References

1. Marquet P. Vacuum background field in General Relativity. *Progress in Physics*, 2016, v. 12, issue 4, 313–317.

2. Kramer D., Stephani H., Hertl E., MacCallum M. Exact Solutions of Einstein's Field Equations. Cambridge University Press, Cambridge, 1979.
3. Marquet P. The generalized warp drive concept in the EGR theory. *The Abraham Zelmanov Journal*, 2009, v. 2, 261–287.
4. Raychaudhuri A.K. Relativistic cosmology. I. *Physical Review*, 1955, v. 90, issue 4, 1123–1126.
5. Dadhich N. Derivation of the Raychaudhuri equation. arXiv: gr-qc/0511123v2.
6. Natario J. Relativity and singularities — a short introduction for mathematicians. arXiv: math.DG/0603190.
7. Kar S., SenGupta S. The Raychaudhuri equations: a brief review. arXiv: gr-qc/0611123v1.

Conservation Laws and Energy Budget in a Static Universe

Yuri Heymann

3 rue Chandieu, 1202 Geneva, Switzerland. E-mail: y.heymann@yahoo.com

The universe is characterized by large concentrations of energy contained in small, dense areas such as galaxies, which radiate energy towards the surrounding space. However, no current theory balances the loss of energy of galaxies, a requirement for a conservative universe. This study is an investigation of the physics nature might use to maintain the energy differential between its dense parts and the vacuum. We propose time contraction as a principle to maintain this energy differential. Time contraction has the following effects: photons lose energy, while masses gain potential energy and lose kinetic energy. From the virial theorem, which applies to a system of bodies, we find that the net energy resulting from the gain in potential energy and the loss in kinetic energy remains unchanged, meaning that the orbitals of stars in galaxies remain unaffected by time contraction. However, each object in a galaxy has an internal potential energy leading to a surplus of energy within the object. This internal energy surplus should balance with the energy radiated at the level of a galaxy. We illustrate this principle with a calculation of the energy balance of the Milky Way.

1 Introduction

We are in a universe governed by energy fluxes and exchanges either in the form of waves or particles in motion. Energy flows in space allow life to exist. The universe is characterized by vast concentrations of energy confined in small spaces such as galaxies in the immensity of a surrounding vacuum. Supermassive black holes at the center of galaxies contain a large portion of this energy. However, we do not understand how such energy segregation came into existence. Most of the energy in the universe radiates outward from these dense galaxies. The supermassive black holes at the center of galaxies may be the cosmic embryos that give rise to the birth of the stars and planets. Massive particles and atoms are attracted by gravitation to the dense points of the universe, a process which maintains the segregation between the vacuum and the dense parts. Because galaxies radiate a large amount of energy, they appear to have energy deficits. Here we investigate the physics of how the energy difference between the vacuum and the dense parts of the universe is maintained.

Many profound questions related to this issue have not yet been answered. Most notably, how did the galaxies come into existence? Do the galaxies have a life time? About 90% of galaxies are dwarf galaxies, and most are elliptical or lenticular in shape. Large spiral galaxies such as the Milky Way are the minority. What are the conditions for galaxies to form stars? For a galaxy to form a spiral it must rotate rapidly. We have observed powerful jets of particles ejected from galaxy central supermassive black holes in the direction of the axis of rotation of the galaxy. These jets, together with a vortex in the black hole, supposedly induce the galaxy to rotate, and then form arms and spirals of stars. A galaxy which has few stars radiates less energy than a galaxy forming stars in abundance. Without a doubt, the lives of galaxies should be considered among the greatest mysteries in the universe.

Nowadays, many people consider the static model of the universe outdated. Nevertheless, we believe there is a lesson to learn when considering the energy balance of the universe. After all, energy conservation is a cornerstone of physics. The elusive dark energy encourages us to inspect the energy balance of the universe from a different angle, in a static universe.

2 The entropic principle

The entropic principle in a thermal context is regarded as an indicator of the effectiveness or usefulness of a particular quantity of energy. Mixing a hot supply of energy with a cold one produces a mix of intermediate temperature, which is less effective. If we apply this principle at the level of the universe, it will eventually lead to the so-called “heat death of the universe”, when the outbound and inbound energy fluxes of galaxies reach an equilibrium that should stay at low temperature provided that the universe does not maintain its present energy differential between the vacuum and its denser parts. The inbound energy flow from cosmic radiations is much lower than the outbound flow radiating from a galaxy, giving galaxies the appearance of an energy deficit. Present theories do not permit us to balance this deficit.

3 Photon-particle interactions

We could conceive of a wind of particles that sweeps the remnant undulating energy in the vacuum of the universe in something like the Compton effect and brings it back to the denser parts of the universe to enrich the galactic gas and nebulae where new stars are formed. This scenario appears to be very unlikely as the inbound flux of cosmic rays is very low, and known interactions between low-energy photons and particles do not subtract energy to the photons. In Thomson scattering, the scattered photon energy is left at the same level, and

an increase of the scattered photon energy is obtained in the photon-particle interaction of the Sunyaev-Zeldovich effect. Compton scattering, which subtracts energy from the photons, is known to occur for high-energy light sources such as X-rays and gamma rays. Furthermore, there is no evidence that cosmic rays come from outside the galaxy, although most cosmic rays originate from outside the solar system [1]. Cosmic rays are composed primarily of high-energy protons and atomic nuclei. Some cosmic rays originate from supernovae [2]; however, this is not the only source of cosmic rays. Active galactic nulei also ought to produce cosmic rays [3].

Compton scattering is an interaction between photons and charged particles such as electrons [4,5]. During this interaction, part of the photon energy is transferred to the recoiling electron. The scattering of the photons produces a blurring effect of light.

Thomson scattering intervenes between photons having much lower energies compared to the mass energy of the particle [6–8]. This interaction occurs between free charged particles and photons. Thomson scattering is an elastic scattering, meaning that the energies of the particles and photons remain unchanged in this interaction. However, the wave is scattered, producing a blurring effect. This interaction produces polarization of light in the direction of its motion. The cosmic microwave background ration (CMBR) is linearly polarized and as such must have undergone Thomson scattering.

The Sunyaev-Zel'dovich effect is an interaction occurring between the CMBR and high-energy particles, which produces an inverse Compton effect [9]. It is the result of high-energy electrons transferring some of their energy to the photons. This interaction is observed in the hot gases contained in galaxy clusters, which change the frequency of the CMBR.

The images of galaxies we observe in the sky are not blurred, meaning a priori that no photon-particle interactions occur for these wave frequencies. For all these reasons we dismiss photon-particle interaction as a mechanism to regulate photon energy in the vacuum.

4 Stationary waves

Stationary waves, also called standing waves, are formed by the superposition of two waves of the same amplitude and frequency moving in opposite directions [10]. The result of this interference is a wave with no net propagation of energy. The locations at which the amplitude of the wave intersect with the x-axis are fixed points called the nodes, and the part of the wave contained between two nodes oscillates upside down in a given amplitude range. Because of the vibration of the standing wave, some energy would be stored in the vacuum, but with no energy being transmitted. Because of the isotropy of the universe we can assume that for every wave there exists another wave of same frequency and amplitude moving in the opposite direction. Standing waves may cause an accumulation of energy in the vacuum, but do not explain redshifts. Nevertheless, we would still need additional mechanisms to regulate the energy budget of galaxies and of the universe as a whole.

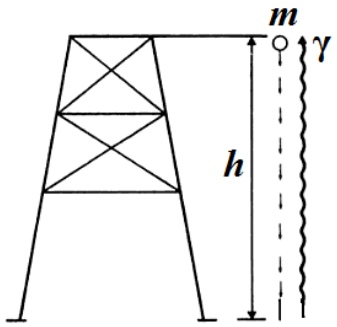

Fig. 1: A photon climbs up to a heigh h. Then, the photon is converted at the top of the tower into a mass m, and falls back to the ground. Perpetual motion is created unless the photon loses energy while climbing in the gravitational field.

5 Time contraction

5.1 Gravitational redshift and potential energy

Another way to look at the problem of energy budget in the universe is by considering gravitational redshift, a phenomenon based on the principle of energy conservation. Einstein imagined the following thought experiment. Let us consider a photon moving away from the ground surface in the direction of the sky up to a given height h. At this height, the photon is converted into mass according to $E = mc^2$, and then falls back to the ground (see Figure 1).

In this system there is an apparent gain of energy from the time the photon left the ground to the time when the mass came back to its initial position due to the potential energy gain when the photon moved upwards. This energy gain, of course is paradoxical. In terms of energy conservation, when considering the energy of a photon, we associate it with the potential energy of its virtual mass counterpart. In order to maintain the system at constant energy, the photon must lose energy when moving away from a mass in a gravitational field, which causes a redshift. The reciprocal is also true: when a photon moves towards a mass in a gravitational field, it is blueshifted. Another solution of the gravitational redshift is obtained with general relativity using the Schwarzschild metric. Both methods give similar solutions that converge asymptotically when the gravitational field is weak.

The gravitational redshift from mass-energy equivalence, which stems from special relativity, is derived as follows. By converting the photon energy into a rest mass we get $E = h\nu = mc^2$. The gravitational potential energy is:

$$U = -\frac{GMm}{r} = -\frac{GMh\nu_0}{rc^2}\,, \qquad (1)$$

where ν is the light-wave frequency, G the gravitational constant, M the mass producing the gravitational field, r the distance between the center of gravity of the mass M and the photon, c the speed of light, and h the Planck constant.

Hence, the frequency change of a photon of frequency ν_0 moving relative to a gravitational mass is $h\nu = h\nu_0\left(1 - \frac{GM}{rc^2}\right)$. Therefore, we get:

$$\frac{\nu}{\nu_0} = 1 - \frac{GM}{rc^2}\,. \tag{2}$$

The equation of the gravitational redshift from general relativity with the Schwarzschild metric is obtained from the equation [21]:

$$\delta\tau = \left(1 - \frac{2GM}{rc^2}\right)^{\frac{1}{2}} \delta t\,, \tag{3}$$

where $\delta\tau$ is the proper time interval, and δt the Schwarzschild time interval.

Because the light wavelength can be expressed as a function of the time interval, $\lambda = c\delta\tau$, we get the gravitational redshift

$$\frac{\nu}{\nu_0} = \left(1 - \frac{2GM}{rc^2}\right)^{\frac{1}{2}}\,, \tag{4}$$

where ν is the light-wave frequency, G the gravitational constant, M the mass producing the gravitational field, r the distance between the center of gravity of the mass M and the photon, and c the speed of light.

For weak gravitational fields, we can use the Taylor approximation $(1 - x)^{\frac{1}{2}} \approx 1 - \frac{x}{2}$ when x is small; hence, we obtain the same equation as the gravitational redshift obtained from mass-energy equivalence.

From general relativity, moving away from the ground surface at increasing altitude causes the clock to tick more rapidly, meaning that time is contracting as in the dichotomous cosmology presented in [11–13]. Based on the principle of time contraction in a static universe, we are able to derive Etherington's distance-duality equation [12]. This principle as an explanation of cosmological redshift is worth considering. One way to look at the problem of photon and matter energy is by linking time with energy, meaning that time contraction is causing both a decrease in the photon energy and an increase in the potential energy of a mass. If this is valid in a gravitational field, does it hold in general? From the mass-energy equivalence, there is an implicit duality between photon and mass, in which energies appear to be indissociable from one another.

Emmy Noether proved a theorem according to which every differentiable symmetry of the action of a physical system has a corresponding conservation law. From the Noether theorem, the law of conservation of energy follows from time homogeneity, meaning the Lagrangian is time-translation invariant. Time is preponderant in energy conservation. In special relativity we learn that time dilation has a direct effect on the energy balance between reference frames. In general relativity, the flow of time and gravitational potential are directly linked. This is a very simple principle that nature could use to regulate energy fluxes in the universe. Accordingly, time contraction would allow maintenance of the energy differential between the vacuum and the massive parts of the universe.

5.2 Effect of time contraction on the photon energy and the energy of a mass

In the dichotomous cosmology [12], we found that the time-contraction factor is expressed as $\gamma_t = \exp(-H_0 t)$. Therefore, the energy of the photon decreases according to an exponential law of the form:

$$E_{\text{photon}}(t) = E_0 \exp(-H_0 t)\,, \tag{5}$$

where H_0 is the Hubble constant, E_0 the initial photon energy, and t the time.

Because the gain in potential energy is in the same proportion as the photon energy loss from mass-energy equivalence, the gravitational potential energy of a mass shall increase according to the law:

$$U_{\text{mass}}(t) = U_0 \exp(-H_0 t)\,, \tag{6}$$

where U_0 is a negative potential energy at time zero, H_0 the Hubble constant, and t the time.

We still need to quantify the effect of time contraction on the kinetic energy of a mass. As time contracts, a clock is ticking more rapidly, and an object in motion appears to slow down. The apparent velocity of an object decreases in direct proportion to the time-contraction factor. Because the kinetic energy is expressed as $K = \frac{1}{2}mv^2$, the kinetic energy of a mass decreases by the square of the time-contraction factor. Hence, the kinetic energy of a mass decreases according to the law:

$$K_{\text{mass}}(t) = K_0 \exp(-2H_0 t)\,, \tag{7}$$

where K_0 is the kinetic energy at time zero, H_0 the Hubble constant, and t the time.

These are the laws that we propose regulate the energy budget of the universe.

Let us show that for a star in orbit in a galaxy, its orbital radius remains unchanged under time contraction. The total energy of the star with respect to other bodies in the galaxy is expressed as follows:

$$E_{\text{tot}}(t) = U + K = U_0 \exp(-H_0 t) + K_0 \exp(-2H_0 t)\,. \tag{8}$$

Let us take the time derivative of E_{tot}; therefore, we get:

$$\frac{dE_{tot}}{dt}(t) = -H_0\, U_0 \exp(-H_0 t) - 2H_0\, K_0 \exp(-2H_0 t)\,. \tag{9}$$

We evaluate this expression at $t = 0$, hence:

$$\frac{dE_{tot}}{dt} = -H_0\, U_0 - 2H_0\, K_0\,. \tag{10}$$

From the virial theorem, which applies to stable systems composed of many bodies, we get:

$$2K_0 + U_0 = 0\,, \tag{11}$$

where K_0 is the kinetic energy and U_0 the potential energy between the bodies.

From (10) and (11), we obtain $\frac{dE_{tot}}{dt} = 0$. Therefore, the total energy of a star in orbit remains unchanged under time contraction, meaning its orbital radius is not affected. This is the condition required to have stable galaxies in the universe.

The virial theorem only considers the potential energy between the bodies of the system. Because each object in a galaxy, either solid or fluid, has an internal potential energy, and that the kinetic energy inside a solid or fluid at rest is negligible, there is a surplus of potential energy from (6). This surplus of potential energy is converted into internal energy within the object. This is the principle we propose to balance the energy radiated by galaxies.

5.3 Energy balance of the Milky Way

From this principle, we would expect that the surplus of internal potential energy due to time contraction, at the level of a galaxy, balances with the outbound radiation flux. Let us do a rough estimation for the Milky Way. The luminosity of the Milky Way is estimated to be about $3.8{\times}10^{10} L_\odot$ [15], with the Sun radiating about 4.6×10^{26} watts, leading to an overall radiation of about $1.74{\times}10^{37}$ watts. We need to estimate the sum of the internal potential energy of each object contained in the Milky Way.

For a spherical solid, the internal potential energy is given by the equation [14]:

$$U_{\text{sphere}} = -\frac{3GM^2}{5R}\,, \tag{12}$$

where G is the gravitational constant, M the mass, and R the radius of the sphere.

Let us consider the estimated mass of the Milky Way including dark matter to be about 1.39×10^{42} kg or $7{\times}10^{11}$ $M_\odot$ [16]. In [17] we show that the dark matter of a spiral galaxy is due to a correction coefficient applied to Newton's force in a disk. Hence, we need an estimate of the total baryonic mass of the Milky Way, which is approximatly one seventh of the apparent mass or about $1.87{\times}10^{41}$ kg. The mass of the cental supermassive black hole Sagittarius A* is about $4.0{\times}10^6$ solar masses [18], and its radius about 31.6 solar radii. Hence, the potential energy of Sagittarius A* from (12) is $-1.15{\times}10^{53}$ joules. We have used a gravitational constant G of $6.67{\times}10^{-11}$ $\mathrm{m^3\,kg^{-1}\,s^{-2}}$.

Because the majority of stars in the Milky Way are red dwarfs, and due to other dense objects such as neutron stars, white dwarfs, and black holes, the average radius of objects in the Milky Way is lower than the radius of the Sun. An estimate of 100 million neutron stars in the Milky Way was obtained by estimating the number of stars that have gone supernova [19]. Let us assume that these 100 million neutron stars in the Milky Way have an average mass of 1.35 solar masses. From the density of neutronium, we can infer that the radius of such a neutron star would be about 15 km. Therefore, from (12), the internal potential energy of those 100 million neutron stars all together is $-1.92{\times}10^{54}$ joules. According to [20] there are about 10 million black holes in the Milky Way. Let us assume that these 10 million black holes have an average mass of ten solar masses and a radius of 45 km. The radius of a black hole is computed from the "photon sphere" which is 1.5 times the Schwarzschild radius. The internal potential energy of those 10 million black holes all together is $-6.42{\times}10^{54}$ joules from (12). Let us assume there are 2 billion white dwarfs having an average mass of half a solar mass and a radius equal to the radius of the earth. The internal potential energy of those 2 billion white dwarfs is $-1.28{\times}10^{52}$ joules. Let us assume there are 200 billion stars lefts (mainly red dwarfs) having an average radius of 0.3 solar radii and average mass of $9.36{\times}10^{29}$ kg. The internal potential energy of those 200 billion stars all together is $-3.34{\times}10^{52}$ joules from (12).

Adding together the potential energies of Sagittarius A*, the 100 million neutron stars, the 10 million black holes, the 2 billion white dwarfs, and the 200 billion stars, the overall internal potential energy of the Milky Way is estimated to be about $-8.49{\times}10^{54}$ joules. The densest objects, although not the most numerous, contribute the greatest share of to the internal potential energy of the Milky Way. For this reason, black holes and neutron stars are responsible for most of the Milky Way's internal potential energy. The calculations for the internal potential energy of objects in the Milky Way are summarized in Table 1.

When multiplying the overall internal potential energy of the Milky Way by the Hubble constant of $H_0 = 2.16\times10^{-18}$ per second (corresponding to 67.3 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$), we obtain a surplus of internal energy of $1.83{\times}10^{37}$ watts. We compare this value with the estimate of the energy radiated of $1.74{\times}10^{37}$ watts. Of course this is a crude estimate, but from our calculations the internal energy surplus of the Milky Way is the same order of magnitude as the energy radiated by the galaxy.

Compact objects such as black holes and neutron stars are known to produce highly energetic jets emitted at relativistic velocities along their axis of rotation. We propose that the surplus of potential energy of compact objects is released to the galaxy through these jets. These jets might be made of neutrons that undergo beta decay to form protons, electrons and antineutrinos.

6 Conclusion

According to the entropic principle in a thermal context, mixing a hot source with a cold source produces a mix of average

Table 1: Internal potential energy of objects in the Milky Way

Object	Number	Mass	Radius	Potential energy
Sagittarius A* (central black hole)	1	$4.0\times10^6\ M_\odot$	2.2×10^7 km	-1.15×10^{53} joules
Black holes	10 million	$10\ M_\odot$	45 km	-6.42×10^{54} joules
Neutron stars	100 million	$1.35\ M_\odot$	15 km	-1.92×10^{54} joules
White dwarfs	2 billion	$0.5\ M_\odot$	6.30×10^3 km	-1.28×10^{52} joules
Remaining stars (mainly red dwarfs)	200 billion	$0.47\ M_\odot$	2.09×10^5 km	-3.34×10^{52} joules
Total	—	—	—	-8.49×10^{54} joules

temperature that is less useful from a mechanical standpoint. The universe is based on energy fluxes and exchanges, and galaxies radiate a large amount of energy. For the universe to be conservative there must be a mechanism to balance the energy deficit of galaxies, otherwise it will lead to the so-called "heat death of the universe". We analyzed photon-particle interactions, and concluded that such interactions cannot regulate the energy budget of the universe. We propose time contraction as a principle to regulate the energy balance in the universe, which would decrease photon energy, increase the potential energy of a mass, and decrease the kinetic energy of a mass. From the virial theorem, which applies to systems of bodies, we find that the net energy resulting from the gain in potential energy and loss in kinetic energy remains unchanged, meaning that the orbitals of stars in galaxies remain unaffected by time contraction. However, each object in a galaxy has an internal potential energy leading to a surplus of energy within the object. At the level of a galaxy, this internal energy surplus should balance with the energy radiated. We illustrated this principle with a calculation of the energy balance of the Milky Way.

Submitted on September 28, 2016 / Accepted on September 29, 2016

References

1. Gaisser T.K. Cosmic Rays and Particle Physics. Cambridge University Press, 1990, p. 1.
2. Ackermann M., Ajello M., Allafort A., Baldini L. et al. Detection of the characteristic Pion-decay signature in supernova remnants. *Science*, 2013, v. 339, no. 6121, 807–811.
3. HESS Collaboration. Acceleration of petaelectronvolt protons in the Galactic Centre. *Nature*, 2016, v. 531, 476–479.
4. Fuchs M., Trigo M., Chen J., Ghimire S., Shwartz S. et al. Anomalous nonlinear X-ray Compton scattering. *Nature Physics*, 2015, v. 11, 964–970.
5. Motz J.W., and Missoni G. Compton scattering by K-Shell Electrons. *Physical Review*, 1961, v. 124, 1458.
6. Gell-Mann M. and Goldberger M.L. Scattering of low-energy photons by particles of spin 1/2. *Physical Review*, 1954, v. 96, no. 5, 1433–1438.
7. Glenzer S.H., Alley W.E., Estabrook K.G., De Groot J.S., Haines M.G., Hammer J.H., Jadaud J.-P., MacGowan B.J., Moody J.D., Rozmus W., Suter L.J., Weiland T.L., and Williams E.A. Thomson scattering from laser plasmas. *Physics of Plasmas*, 1999, v. 6, 2117.
8. Moore C.I., Knauer J.P., and Meyerhofer D.D. Observation of the transition from Thomson to Compton scattering in multiphoton interactions with low-energy electrons. *Physical Review Letter*, 1995, v. 74, 2439.
9. Sunyaev R.A., and Zel'dovich Ya.B. The observation of relic radiation as a test of the nature of X-ray radiation from the clusters of galaxies. *Comments on Astrophysics and Space Physics*, 1972, v. 4, 173–178.
10. Born, M. Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light. Pergamon, 6th Revised ed., 1980.
11. Heymann Y. A Monte Carlo simulation framework for testing cosmological models. *Progress in Physics*, 2014, v. 10, no. 4, 217–221.
12. Heymann Y. A derivation of the Etherington's distance-duality equation. *International Journal of astrophysics and Space Science*, 2015, v. 3, no. 4, 65–69.
13. Heymann Y. The dichotomous cosmology with a static material world and expanding luminous world. *Progress in Physics*, 2014, v. 10, no. 3, 178–181.
14. Kittel C., Knight W.D. and Ruderman M.A. Mechanics. New York, McGraw-Hill, 2nd ed., 1973, 268–269.
15. Flynn C., Holmberg, J., Portinari L., Fuchs B., and JahreißH. On the mass-to-light ratio of the local Galactic disc and the optival luminosity of the Galaxy. *Monthly Notices of the Royal Astronomical Society*, 2006, v. 372, 1149–1160.
16. Eadie, G. M., and Harris, W. E. Bayesian Mass estimates of the Milky Way II: The dark and light sides of parameter assumptions. arXiv: 1608.04757 [astro-ph.GA].
17. Heymann Y. Dark matter, the correction to Newton's law in a disk. *Progress in Physics*, 2016, v. 12, no. 4, 347–352.
18. Boehle, A., Ghez, M., Schödel, R., Meyer, L., Yelda, S. et al. An improved distance and mass estimate for Sgr A* from a multistar orbit analysis. arXiv: 1607.05726 [astro-ph.GA].
19. Camenzind, M. Compact Objects in Astrophysics: White Dwarfs, Neutron Stars and Black Holes. Springer, 2007.
20. Plait P. Death from the Skies!: The Science Behind the End of the World. The Penguin Press, Reprint edition, 2009, p. 302.
21. Moore T.A. A General Relativity Workbook. University Science Books, 2012, Chapter 9.

Progress in Physics is an American scientific journal on advanced studies in physics, registered with the Library of Congress (DC, USA): ISSN 1555-5534 (print version) and ISSN 1555-5615 (online version). The journal is peer reviewed and listed in the abstracting and indexing coverage of: Mathematical Reviews of the AMS (USA), DOAJ of Lund University (Sweden), Scientific Commons of the University of St.Gallen (Switzerland), Open-J-Gate (India), Referential Journal of VINITI (Russia), etc. Progress in Physics is an open-access journal published and distributed in accordance with the Budapest Open Initiative: this means that the electronic copies of both full-size version of the journal and the individual papers published therein will always be acessed for reading, download, and copying for any user free of charge. The journal is issued quarterly (four volumes per year).

Electronic version of this journal: http://www.ptep-online.com

Postal address:
Department of Mathematics and Science, University of New Mexico,
705 Gurley Avenue, Gallup, NM 87301, USA