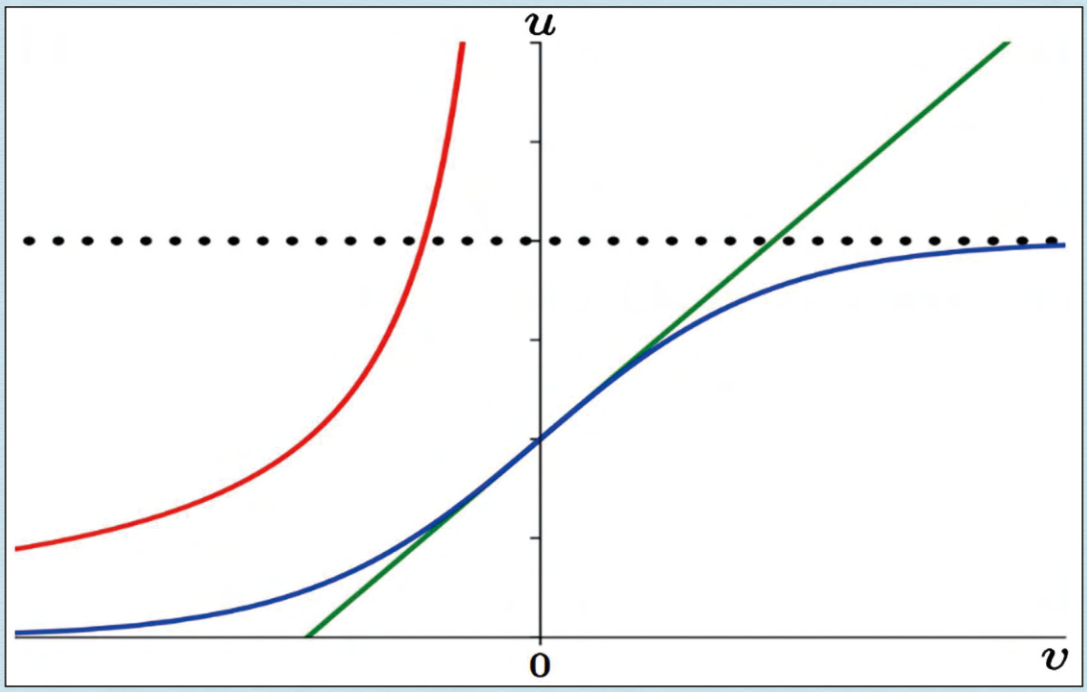
ISSN: 2153-1196 Volume 13, Number 4, April 2022

Scientific Research Publishing

Journal of Modern Physics



ISSN: 2153-1196
9 772153 119007 04

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

Journal Editorial Board

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

Editor-in-Chief

Editorial Board

Journal of Modern Physics, 2022, 13, 323-670
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Table of Contents

Volume 13 Number 4 April 2022

Journal of Modern Physics, 2022, 13, 323-670
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

The figure on the front cover is from the article published in Journal of Modern Physics, 2022, Vol. 13, No. 4, pp. 323-335 by Bartolomé Pons-Rullán.

Journal of Modern Physics (JMP)
Journal Information

SUBSCRIPTIONS

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

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

SERVICES

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

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

COPYRIGHT

PRODUCTION INFORMATION

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

Journal of Modern Physics, 2022, 13, 323-335
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Temporal No-Linearity: An Alternative to Dark Energy

Bartolomé Pons-Rullán

University of the Balearic Islands, Palma de Mallorca, Spain

Email: bart@bartolo.com.es

How to cite this paper: Pons-Rullán, B. (2022) Temporal No-Linearity: An Alternative to Dark Energy. *Journal of Modern Physics*, **13**, 323-335.
https://doi.org/10.4236/jmp.2022.134023

Received: November 22, 2021
Accepted: April 3, 2022
Published: April 6, 2022

Abstract

Mathematical compatibilities and constraints of a hypothetical 5D space-time with time referenced by two coordinates (3-2) have been revisited in detail in several recent papers. It has been prescribed from the GR the compatibility constraints of the FLRW metric in each temporal brane, to be restricted to Closed Universes, Smooth Initial Singularities, and "Open CTC". In a first view, this leads to leaving these works considering mathematical games discarded by the Standard Candles data. However, if time would be referred by two coordinates, they would not be linearly related, and it will be mathematically stated that space-time may not be flat in any case because time-like branes geometry will never be. If so, the time scale "lived" over a two-time dimension geodesic necessarily is not constant over its linear projection on one of both coordinates. Consequently, the correlations between Redshift and Distance Modulus—Distance Ladder—may be corrected by a synchronization function (if a no-linear two-time geodesic trajectory over a "warped temporal geometry" is linearly divided into constant segments, then their projections are not linear in any case). We apply an example of time-trajectory over the time slices, matching the Standard Candle data for a Closed Universe dominated by matter (>90%) in a bulk 3-2 configuration, with open temporal branes and smooth singularity. If time can be referred by two coordinates, then there is no need of Darkness to explain astrophysical data and Universe can be closed.

Keywords

Closed Universe, Multidimensionality, Time-Like Dimensions, No-Linear Time, Accelerated Expansion, Smooth Big Bang

1. Introduction

Every truth is certain only inside its paradigm. Data can certify consistency, even

they do not certify the exclusivity of an interpretation but in reference to its assumptions. The interpretation of the observational data on SNeIa & GRB, as an acceleration of Expansion, is true within its relativistic 3-1 paradigm.

Within this paradigm the Horizon's Problem, Dark Energy, Dark Matter, baryonic decay [1], relativistic determinism, the entropy of the black holes, galactic macro-singularities, and multiverses, can be considered, as consequences pending to solve by insisting with additional hypotheses, or also as symptoms of the need for a paradigm change, in any case, to which unification target also points.

A single temporal dimension kidnaps GR into determinism: if Relativistic Mechanics were considered from two or more time dimensions, the temporal branes would be nothing more than a frame of a statistical population of potential time trajectories, defined by temporal geometries (equivalent to microstates). If this temporal geometry would be flat and unique, then time trajectory would be linear, recovering 4D.

FLRW Metric, Cosmological Deceleration, Causality (CTC), Well-posedness, more symmetries for more equations, tachyons and phantom particles, dimensional compactification, have been often considered as "pathologies", reason for disinvestment in multitemporal theories. But alternatively, they may also be considered as limiting conditions that alternative interpretations of the astrophysical data available must meet. They can also be clues to finding time-geodesic function that eventually may converge with SR.

There had been alternative interpretations of SNeIa & GRB data: Effects that may influence the extrapolation of the maximum brightness (subclasses of SNeIa and influence of metallicity [2] [3]; Incoherent Light, [4]; Tired Light, Variable Speed of Light (VSL), [5]; and/or variable G at cosmological scale, MOG [6], MOND [7] [8]. Beside all those speculations, maybe Type Ia Supernovae are not compelling as standard candle that much as it has been said, like in 2014J [9], [10]. This paper will analyze if time can be referred to two no-linear related coordinates, a closed universe with smooth singularity and no-flat open temporal geometry is possible.

2. FLRW Metric

The application of the Campbell Theorem in its weak statement to a multitemporal configuration allows a unique solution in which any Friedmann-Lemaître-Robertson-Walker, -FLRW-, metric can be embedded in a 3-2 Ricci Flat Space, maintaining the symmetry group. Within the multidimensional paradigm, assuming Friedmann's Model, the detailed analysis in [11] of the metrics states the compatibility of relativistic equations with temporal branes in Universes of three spatial and two temporal dimensions (3-2), with certain restrictions. From [12]:

From a 5D = 3-2 Paradigm, FLRW metric as a constrain, states in [13]:

- The Universe must be closed: decelerated Scale Factor. For a generic equilibrium combination of matter, radiation, and the cosmological constant, with an adiabatic index

$$\omega_{\max} \leq 1/2, \text{ if } \epsilon = -1 \Rightarrow k = +1 \quad (1)$$

- The necessary and sufficient condition for the absence of particle horizons is a "Smooth Big Bang" (asymptotic to the *v*-axis): the proper time derivative of the expansion factor $R(t)$ must be finite and not diverge faster than R^{-2}. The Horizon Problem would be an answer to a wrong question (asymptotic to a *v*-axis parallel).
- Inside a 5D Universe, a temporal trajectory function $\tau = \Psi(t)$, such as the hyperbolic examples in Figure 1, cannot be constant or linear in any case. Any linear relationship between temporal coordinates -a straight line in the temporal plane u/v in Figure 1, by rotation and/or translation would reset the 4D configuration.
- Then 5D space-time can never be always flat, though there is a no-linear time geodesic trajectory. It can seem to be approximately flat on the asymptotic phases over the *v*-axis (in Figure 1, at the smooth BB process, or even in its tired old phase; in Figure 2, which represents another alternative $\Psi(t)$, not included on the Bona *et al.* paper, only in its stabilization process). From this assumption, if the Universe seems to be 4D at low redshift, it may be because we are in an asymptotic stabilization evolution, very similar to a horizontal straight line (dots in Figure 1). $\Psi(t) \approx$ constant.
- Any function that relates temporal coordinates, $\Psi(t)$, will represent a trajectory in the temporal geometry always longer than its projection on any axis, or any straight line $\Psi = a + bt$, (see Figure 1, with hyperbolic examples). A billion-year living over the trajectory τ for a 5D observer will be projected to less than a billion-year at the *v*-axis set as *t*; and the other way around, a year from a 4D assumption measure, *t*, will be translated to more than a year over τ if the observer does not synchronize clocks.

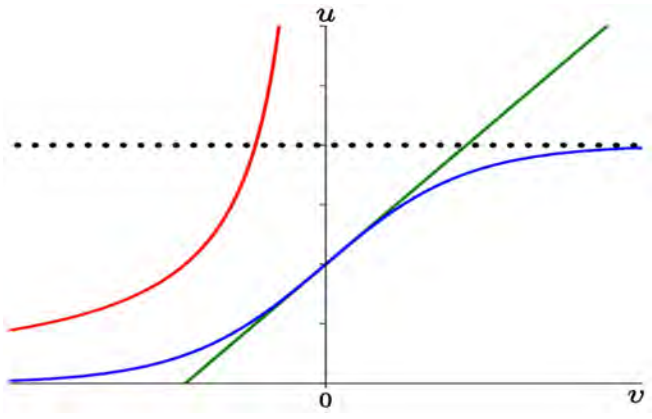


Figure 1. Timelines in the time plane of the bulk M-metric, each one leading to different FLRW projected metrics. The big-bang singularity is here the $u = 0$ line. The straight line corresponds to the standard pure radiation model (recovering 4D by rotation & translation). The other two lines correspond to FLRW models without a big bang, with an initial accelerating (inflationary) phase and a final decelerating phase. The deceleration is not apparent (see the text for the detailed calculation in [12]). In the hyperbolic tangent case, the inflexion point corresponds to a pure radiation phase.

$$t_2 - t_1 < \int_{t_1}^{t_2} \sqrt{1 + \left(\Psi'(t)\right)^2}\, \mathrm{d}t \tag{2}$$

From a Bulk of three space-two time dimensions, if we experience a single time trajectory at a constant rate taken over $\tau = \Psi(t)$; when we measure time with a constant clock, we state a constant "length" of a time-lapsed segment of $\Psi(t)$. This time-length cannot be projected linearly on a constant scale over the single time v-axis; which will mean that the moment in history in which an event occurs will be apparently further on time for the one who takes brightness and redshift as measures of a certain event from the 4D assumption (constantly projected v-time-scale), than for the 5D assumption (constant time-scale over the time trajectory $\Psi(t)$). In this case, we will have to apply a synchronization function:

$$z(t) = \frac{1 - R\left(t, \Psi(t)\right)}{R\left(t, \Psi(t)\right)} \tag{3}$$

The single time coordinate assumption drives to a constant Scale Factor, eventually modified by an extra accelerating rate to fit observational data. Though graphically, the measures will stand over the flat expectative, or also can be represented with a variable look back time scale shrinking z (**Figure 3**). From 3-2 perspective, a constant time over Ψ, would have not a constant projected t over the u-axis, which is precisely the assumption of 3-1 perspective, or the other way around, if we size assuming constant t, Scale Factor over Ψ would never be constant.

Suppose the Universe has a configuration of at least two temporal dimensions. In that case, the Distance Ladder assuming 4D would be affected by a correction due to a no linear decelerating Scale Factor, depending on the shape of $\Psi(t)$ and always maintaining the inequality for any function representing the association between both time coordinates. From 3-1 assumption, c 10 Bly ago, were "quicker" for our today's clock (length over v-axis), but from 3-2 assumption, constant for the same clock if used by then (length over $\Psi(t)$). As an analogy, a clock held by a near-c traveler would seem "slower than a second per second" from our clock point of view (if both "seconds" are from different clocks, one on the rocket and the other on the lab), being both clocks identical. In this case, they have a well-known synchronization function by SR.

In the opposite direction in the v axis, the farther from the BB smooth process, the more linear-like approximation in both hyperbolic examples of **Figure 1**. In a naive analogy 2-1D, from our clock point of view (t over v-axis), it could be imagined for an external observer as a time-bubble seeming to grow very fast at the beginning and slowing with time, or in look-back-time, accelerating growth of the Universe at high z (enlarging the apparent distance between z, because the shortest projected time by then). But for a clock measuring at each moment, evolving with the age of the Universe, it would seem to be a constant Expansion rate.

For an observer from 3-1 assumption, rules and clocks always size the same

c-length, even if he uses his rule and clock 10 By later. This is not true for a 3-2 assumption, and this do not mean that c is variable, but it may seem variable if the observer do not consider that his rules and clocks are subjective and dependent on the Scale Factor.

An observer conditioned by the 4D assumption (projected time trajectory over a constant scale at v-axis) would include a synchronization bias concerning a 5D observer, considering a constant scale over the time-like geodesic correspondence between time and redshift. Thus, adjusting the observational data of the SNeIa & GRB is only to choose a correct hypothesis of $\Psi(t)$, representing in its relation to the Distance Modulus, μ, an axis with redshift according to time scale on the trajectory, τ. Setting the lifetime of the Universe to 1:

$$\frac{1}{z_\tau + 1} = \int_{\frac{1}{z_t+1}}^{1} \sqrt{1 + \left(\Psi'(t)\right)^2}\, \mathrm{d}t \tag{4}$$

With z_τ, (τ), and z_t, (t) we do not mean to say there are two redshifts, but a different time scale transformation between z and look-back-time, depending on the 3-1 (t) & 3-2 (τ) paradigms: how old is z at the relative scale of the other dimensional assumption.

$$\frac{1}{z_\tau + 1} \,\&\, \frac{1}{z_t + 1} \tag{5}$$

Being z_τ, the redshift where the μ has to be considered from a 3-2 bulk perspective at the scale of z_t (as a translation of z on the time axis, but from each perspective, the size the same z in different time scales).

If time has two coordinates, maybe space is not always flat, but for sure, time is not flat and follows a non-linear trajectory in a warped temporal surface, but we use the projected time to site an event. By ergodicity, the 4D clock may overestimate the look-back-time-distance. It must be corrected because c is the distance over time events and may seem to happen further than they do if we assume constancy of the denominator. This will always happen for any $\Psi(t)$, but the bias quantification will depend on the time-geodesic shape. From this assumption, SNeIa & GRB data may not be interpreted as an Acceleration of Expansion but as an asymptotic stabilization of the second temporal coordinate in a matter-dominated Universe.

No-flat Universe in Expansion of the time-like coordinates are consistent with the equations of the GR in a decelerating Universe and the FLRW metric, guaranties Causality if $\tau = \Psi(t)$ is open and asymptotic. Although, before proposing a time geodesic trajectory shape -like those in **Figure 1**, that will make consistent the hypothesis with astrophysical data, a "pathology" remains unsolved.

3. Well-Posedness

Adding a dimension increases the mathematical cost, which is already very expensive in terms of symmetry assumptions. Even for 4D which implies ten equations, but 5D raises the number to 15 and needs more symmetries to determine

the system. With no extra conditions PDEs may become ultra-hyperbolic. We propose to add a temporal geodesic trajectory hypothesis, open-asymptotic-$\Psi(t)$, that must deductively postulate symmetries for including them into GR equations, maybe due to an extra conserved quantity, which will determine the system of differential equations.

That means the requirement for NLT to add a hypothesis, but by the way, the same happens to be in the Acceleration Interpretation, which also needs a supplement in 4D Paradigm: Dark Energy. In both cases, graphically acceleration or deceleration can alternatively be represented as a no-linear scale from constant z to t, to an apparent variable z to constant t, (Figure 3), shrinking v scale if there is some kind of energy in such a shape that drives to a tipping point from concave to convex at $z \approx 0.64$ *i.e.*, due to the radiation + matter dynamic as t^n ($n < 1$), while Dark Energy is supposed to be growing as e^t), or expanding the v-axis scale if there is some kind of gravitational brake.

Another way (fictitious but didactic) to formulate and visualize the time-v-axis variable scale in terms of corrected apparent look-back time to redshift, stretching the double-time geodesic over v-axis by an apparent non-constant time rate, which means an apparent VSL (not real in 5D, but this could be the interpretation from a 4D assumption). That means, that if we set as constant $8\pi G/c^4$, which relates the second derivative of the metric tensor $G_{\mu\nu}$, with the relativistic energy-momentum tensor, $T_{\mu\nu}$ has to remain constant, G would apparently be variable to the 4th power of the apparent change of c, just if so, from 4D assumption. This might have observable consequences in enormous stars and black hole sizes for high z.

That apparent $G(t)$ would be time-dependent only if we use our day clock and would also drive us to consider the no-baryonic Dark Matter interpretation, which would be interpreted from 3-2 bulk, as a consequence of clocks desynchronization (apparent G variation to the 4th power of c). In fact, observational data indicate that if extrapolated to the Transparency Event, the CMB peaks should be attributed, at least to a much greater extent than nowadays, to baryonic dark matter. But *"Dark matter had less influence in the early universe. Observations of distant galaxies carried out with the VLT suggest that they were dominated by ordinary matter"* [1].

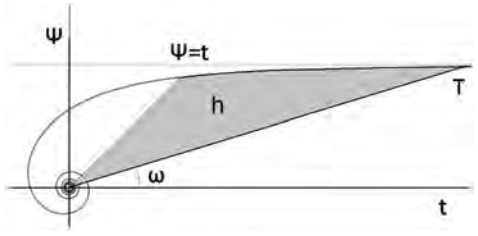


Figure 2. *v*/*t*; *u*/Ψ. Additional hypothesis for $\Psi(t)$ no linear time geodesic in a no flat temporal geometry, in polar coordinates $T = B/\omega$, being T, linear time as a 4D assumption. "Temporal rotation" is only a semantic analogy to describe a simplification for warped-time geometry, not a physical description.

Another "possible clue" to constrain the translation between a symmetry hypothesis and a temporal geodesic trajectory function proposition $\tau = \Psi(t)$, for well-posedness purposes, may be entropy. Entropy increases if the matter can clump together, releasing potential energy and creating clusters that further unbalance the contents. As the mass is finite, CMB entropy would be finite (to the square), proportional to a hypersurface (Holographic Principle, [14]). Wheeler&DeWitt already postulated an analogy with the Universe Wave Function [15].

Gas has a dynamic configuration in the position of all its molecules, although its entropy is calculated according to the number of possible configurations (besides they are used to measure, thermodynamics does not pretend that they all exist in Parallel Universes). Each macrostate would be constituted by equivalent configurations of temporal evolutions on two dimensions, gravitationally consistent of all the masses, and the destiny would not be written (Recurrence Theorem). However, its patterns would be predictable with some probability, and multitemporal would allow a GR with time-arrow.

Another clue to propose a $\Psi(t)$ hypothesis is the question of how a linear Bang transforms into a generical orbital and spinning movements (maybe Kolmogorov turbulence?). We know that at present, everything inside rotates, but the Universe itself seems not to have an Angular Momentum. But was it so at the Big Bang smooth process? Is it possible to consider an Angular Momentum Conservation to income a symmetry at the very BB moment, but not now? Observations about Universal Angular Momentum or Expansion recommend using a $\Psi(t)$ function that becomes flat after the BB smooth process: $t \approx \tau$ reaching stationary state in our days (shape-like hyperbolic tangent case in **Figure 1**).

Another possibility to elucubrate a synchronization function is to consider c as the initial Scale Factor, declining R with expansion due to gravitational friction and momentum conservation, and apply then Special Relativity -SR- synchronization between clocks at different R'(a clock for the observer at a lower R -now-, would size quicker apparent c at the early times). But applying backwards from present up to $z < 10$, will get a very low slope with a high error and with a far point at CMB, and it may not give more than the shape of the tail of an undefined function.

Those and more 5D well-posedness requirements, hints and elucubrations, could be an open frame to additional hypotheses taken as restrictions to be fulfilled for any temporal geodesic trajectory with a more or less known tail. Here are some possibilities to speculate about the determination of equations through an additional synchronization function Ψ. With those "clues", we will propose here an extra hypothesis to determine the system only as an example. But even the SNeIa & GRB data provide an observational shape that can be inversely fitted just approximating a statistical regression, to be understood lately through a conservation law or not. Even hyperbolical tangent examples, with appropriate parameters, will fit astronomical data, but why that shape? $\Psi(t)$ to determine PDEs remains open, and many propositions may fit data in closed universes

without Dark Energy.

4. First Shot

The range of possibilities is wide, and we have tried several "temporal geometry speculations" [16]. Between the many options, our "first shot" is to be consistent with the Gödel Universe plus an initial Novikov conjecture for a single event (Big Bang as a White Hole born from a "Higher Dimension Mother Black Hole", inheriting its entropy, mass, charge, and angular momentum, but losing a dimension) [17]. Several models even proposed a switching between space and time nature of coordinates inside a Black Hole.

By analogy with the family of geometric solutions for the movement of a particle according to a central force, inversely proportional to the cube -(3 coordinates)- to its distance to the origin (Cote's spiral), the First Shot additional function has been to test if a synchronization due to Hyperbolic Time Spiral with h proportional to constant Λ/Ω, it fits with data. We do not claim that this is the solution, but this one we will see it fits.

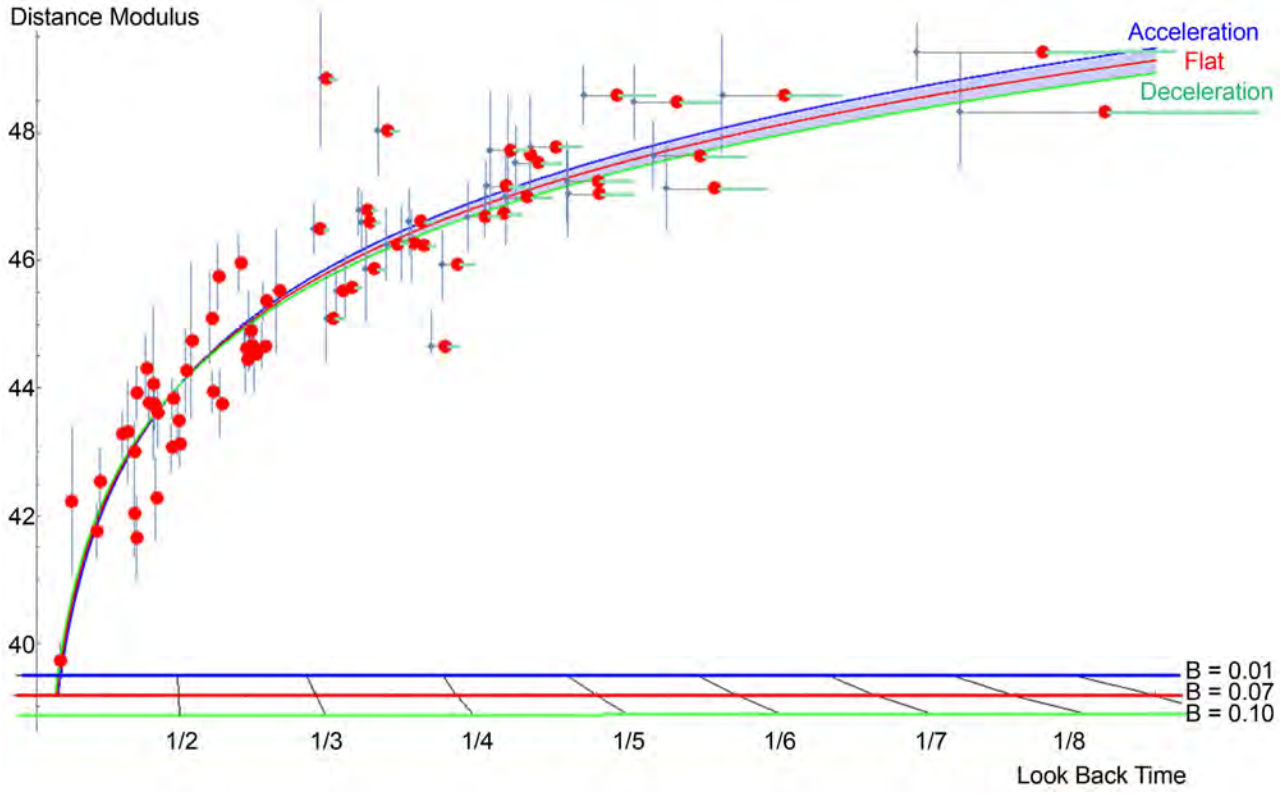


Figure 3. From (6) equation solution. Small-Blue dots with blue error bars are the GRB events over 4D time-scale, same as 5D with $B = 0.01$ (referred to the red axis), and though the blue line is its regression. Big-Red dots are GRB events over 5D time-scale with First Shot $\Psi(t)$ twisting hypothesis -$B = 0.07$ fitted with Flat Universe at constant-scale-, and the red line is its regression. Green regression is fitted with $B = 0.1$ ($\Lambda \approx 95\%$) over the red constant-scale v-axis. Filled in grey is the range of regressions from $B = 0.01$ up to 0.1. Blue regression at the blue scale is red-dashed, so it is green regression at the green scale. As a reference to fit, the dashed red line for a 4D Flat Universe Model radiation/matter: 0.5/0.5 with $\Psi' = H_0 = 69.6$ (same as the $2\Lambda/\Omega = 0.07$). So we can set the Legacy No Flat Time Energy hold in its warpness between 0.5 to 9.5%, from Open to Closed Universe, Matter Dominated > 90%.

This "first shot" is not a Rotational Universe proposal because it would need to be space over time (if we could approximate a simplification where time is warped and space near-flat, is it impossible to consider also the evolution of time over space?). Just for the sake of the analogy in losing a dimension (opening PDEs), we must refer to Silk [18], who formally demonstrated that the 4-1D rotational models presented density instability on space-like dimensions when perturbing along the axial axis but stabilized in the perpendicular plane of rotation. An additional space-like dimension would be unstable and would concentrate mass in the vicinity of a disk, folding into a small and constant value (thickness). Analogously to galaxies or planetary systems, it would concentrate time-space in a temporal brane orthogonal to the temporal axis: it would collapse and lose that dimension (orthogonal disc in **Figure 2**).

Assuming the analogy into this very particular "temporal geometry" synchronization, taken as a first shot example of the $\tau = \Psi(t)$ hypothesis, the assumption could be easily extended to 3-3, losing a third time-like dimension at the Big Bang asymptotic smooth process. All those justifications of the first shot origin are no more than the explanation of the why we have chosen this example, including some conservation law, and not any other with no fundamental motivation, like hyperbolic examples in **Figure 1**, but the proceeding is appliable to any other justification of a different hypothesis, that match requirements. This new example time shape would look as in **Figure 2**.

If we assume ergodicity between space and time expansion, conserved quantity h may be interpreted as proportional to Λ/Ω, and the slope Ψ', as the Hubble "Constant", H_o. Then, choosing an appropriate synchronization between t & τ, we will be able to approach the CMB event in 5D time-scale in terms of the gap between measurements based on the Distance Ladder and based on CMB: around 8%, [19] (Tail slope at that time, **Figure 2**).

Just inside the prospected margin of slope at the projected curves of **Figure 3**. It would be great but quite strange to fix $\Psi(t)$ in a First Shot synchronization, but either this shape will be discarded in the future; in our simulations, it fits surprisingly fine with scare data in high z. In any case, equation (4) translated to polar coordinates has been solved and the relation between survival time dimensions, t & τ, would be:

$$\left(\sqrt{B^2+1} + B\ln\left(1+Z\right)\frac{B+\sqrt{B^2+\frac{1}{\left(1+Z\right)^2}}}{B+\sqrt{B^2+1}} - \sqrt{B^2+\frac{1}{\left(1+Z\right)^2}}, \frac{c}{Z+1} \right) \tag{6}$$

Being $B = 2\Lambda/\Omega$, the shape parameter of the time-Cotes spiral, because ergodicity, proportional to radiation/matter.

From this perspective, the cycling phase of the spiral trajectory, while time changes from positive to negative values and back again (smooth and asymptotic Big Bang), began before 5% from our age, but as the 13.72 billion light-years are

sized from a 4D perspective; this does not mean after CMB of a 5D perspective (in this example around 8% higher scope, fitting with observations). By then, a second measured by a clock may seem -apparently, if we synchronize with nowadays clock- a very much shorter projection of the same second at the same clock (**Figure 2**, negative t & Ψ).

When we set this "first shot time-geodesic example", there is a statistical coincidence between both approaches (4D+ΛCDM & NLT + inverse time trajectory spiral) up to $z \approx 2$, and we have used a sample of higher z-GRB to fit data with hypothesis on a model for regression as $44 + b\ln(z)$. [20]. Without being conclusive, given the wide standard deviation of the scarce data at $z > 1$, in this example, it is confirmed that by moving the apparently higher z_τ in terms of the 4D time-scale, in the Distance Modulus/Look-Back-Time graph, with $B = 0.07$ (best-fitted tuning with a 4D Flat Universe model), the SNeIa & GRB Distance Modulus data are compatible either with accelerated or decelerated Universe, Matter Dominated > 90% (as we have assumed in the additional first shot hypothesis, proportionality between h & Λ/Ω to determine Ψ with the inverse spiral shape).

The more closeness with less matter might be interpreted as a stronger effect of the Λ on the shape of the time spiral over the Ω gravitational "brake" (friction). The 5D perspective allows open, flat, and closed configurations, but to seem Spatially-Flat, Matter ≈ 96.5% & Legacy Time Brane Warp Energy ≈ 3.5%. This does not completely withdraw Dark Energy, but makes unnecessary the acceleration to explain data, only because time is set in two coordinates instead of one and that implies time is not linear if it is not a single dimension.

Flat-like & Warp-like coordinate simplification diversifies the space & time nature of dimensions. Still warpness in a couple of coordinates might produce effects even in space-like dimensions while having lower strength. That may be observable maybe during the early BB process, probably not from our time position.

5. Conclusions

The analyzed papers by Bezares, Bona, Pons-Rullan and Vigano [12], confirm that a bulk 3-2 configuration is compatible with the relativistic equations, with the FLRW metric and with a smooth singularity for each temporal brane, without the need for compactification. The price for the change of the time-scale, the Universe must be closed, but precisely restrictions are fed back because another price to pay is warp-time coordinates that introduce a correction in the Scale Factor that closes the Universe to fulfill astrophysical data. According to this criterion, when translating the Redshift Scale to Look-Back-Time, the Closed Universe is compatible with the Standard Candles data and more than this: it fits quite well as shows the data in **Figure 3**.

Here we provide a procedure to test the additional symmetries needed to determine the equations through its time-like-geodesic. Still we do not claim to

have found the conservation laws that adjust the definitive correlation of astrophysical data. For example, we demonstrate that the procedure can give consistent results with a broad spectrum of hypotheses and that Dark Energy is not the only answer. We succeed in fitting the astrophysical data in a Space-Flat, Warped-Time, and Causal and Closed Universe with Smooth BB, as accurate as the Open Universe Dark Energy interpretation (red curve in **Figure 3**).

It may be said that this time brane warping evolution to produce a hyperbolic trajectory is highly speculative. Neither in the single constant time interpretation, nature and dynamics of Dark Energy are not known (Why Dark Energy is not qualified as a "pathology"?). However it would be unfair to demand an alternative interpretation to fix the definitive $\Psi(t)$. Meanwhile, if we consider there must be an additional Dark hypothesis for its dynamics, both interpretations would be at the same speculative level.

We might choose between inventing a 96% Universe with a not known energy and matter, or assume not extra time-like dimensions (other alternatives like variability in constants, anisotropy, magnetic forces,… or even misunderstanding of the SNeIa process, are not considered here), but the latter is not dark, we do not have a Horizon Problem; neither Hiperinflation as so: we can explain no-baryonic Dark Matter ($c\&G^4$), and through it we can understand the Giga black holes on the galaxies centrum, intermediate size black holes -GRB 950830- and the why we are going to size much more massive gravitational merging events in the near future); we can also fill the gap on H_o measurements on respect CMB, and opens GR to statistics of time trajectories or non-linearly warped branes.

As future surveys incorporate more events in $z > 2$ and better measurements of H_o, the regressions will be more accurate, and it will be possible to improve the determination of more universal symmetries with this tool. We will continue to test other $\Psi(t)$ additional hypotheses that adjust the data and incorporate R' and H'_o into the tuning. Still, we already anticipate that space-time ergodicity and a conserved Λ/Ω may be proven consistent, which will not mean there are no other candidates for temporary functions derived from it, particularly SR transformations between observers at different Scale Factor expansion rate.

This is not a rotational and/or a variable constant theory, but an anisotropic an inhomogeneous time hypothesis (which is confirmed by our experience and thermodynamics). Maybe space dimensions can be considered as homogeneous and isotropic, but time has a different nature. $\partial\psi/\partial t$ & $\partial\psi/\partial\tau$. Time rotation is not a movement but a shape: a graphical description by the analogy of an example of a temporal geodesic shape. The Speed of Light, Gravitational Constant or Fine Structure Constant remains "constant" for whenever a clock measures from its own time (not the same inside a rocket or at home), but apparently they are not from 4D time-scale if observers presume the same clock along time as his own clock, just because every clock measures the same length of time. Still our projected clock on a single linear time, do not measures the very far away time properly because it assumes parallel time-like-geodesics. A light-year is constant at the very beginning and now, but from our rules-based in wave-length and

clocks set in our days, apparently it is not. $\Psi(t)$ & $\Psi'(t)$, set the bias.

In our understanding, there is something clear: even if they are both highly speculative, No-Linear Time (Space Open, Flat or Closed Universe) fits data as well as Dark Energy (Open Universe) and offers a mathematical path forward. Maybe there is no Dark Energy or no baryonic Dark Matter (apparent variable c, G, H_o & a), or maybe there are no other time-like dimensions. Surely this "first shot" is not the right one. It is just an example in which its tail fixes data, but other time-geodesics with another conservation hypothesis may be, and we only want to mean that it is not an unquestionable true that the Universe is accelerating.

Conflicts of Interest

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

References

[1] Förster-Schreiber, N.M., Genzel, R., *et al.* (2017) *Nature*, **543**, 397-401. https://doi.org/10.1038/nature21685

[2] Cinabro, D., Kessler, R., Li, A., Miller, J. and Scolnic, D. (2016) *Monthly Notices of the Royal Astronomical Society*, **466**, 884-891. https://doi.org/10.1093/mnras/stw3109

[3] Baron, E., Brown, P.J., Milne, P., Roming, P.W.A. and Wang, L. (2015) *The Astrophysical Journal*, **809**, 37. https://doi.org/10.1088/0004-637X/809/1/37

[4] Dam, L.H., Heinesen, A. and Wiltshire, D.L. (2017) *Monthly Notices of the Royal Astronomical Society*, **472**, 835-851. https://doi.org/10.1093/mnras/stx1858

[5] Meng, X. and Zhang, P. (2014) *Modern Physics Letters A*, **29**, Article ID: 1450103. https://doi.org/10.1142/S021773231450103X

[6] Brownstein, J.R. and Moffat, J.W. (2006) *The Astrophysical Journal*, **636**, 721-741. https://doi.org/10.1086/498208

[7] Milgrom, M. (2015) *Canadian Journal of Physics*, **93**, 107. https://doi.org/10.1139/cjp-2014-0211

[8] Bekenstein, J.D. (2012) *Philosophical Transactions of the Royal Society A*, **369**, 5003-5017. https://doi.org/10.1098/rsta.2011.0282

[9] Perez-Torres, M.A., Lundqvist, P., Beswick, R.J., Bjornsson, C.I., Muxlow, T.W.B., Paragi, Z., Ryder, S., Alberdi, A., Fransson, C., Marcaide, J.M., Marti-Vidal, I., Ros, E., Argo, M.K. and Guirado, J.C. (2014) *The Astrophysical Journal*, **792**, 38. https://doi.org/10.1088/0004-637X/792/1/38

[10] Nielsen, J., Guffanti, A. and Sarkar, S. (2016) *Scientific Reports*, **6**, Article No. 35596. https://doi.org/10.1038/srep35596

[11] Bezares, M. and Bona, C. (2019) *Physical Review D*, **100**, Article ID: 043509.

[12] Bezares, M., Bona, C., Pons-Rullán, B. and Vigano, D. (2019) *Physical Review D*, **99**, Article ID: 043530.

[13] Bona, C. (2019) The Horizon Problem as a Clue: A Smooth Big Bang?

[14] Afshordi, N., Coriano, C., Delle Rose, L., Gould, E. and Skenderis, K. (2017) *Physical Review Letters*, **118**, Article ID: 041301. https://doi.org/10.1103/PhysRevLett.118.041301

[15] Hartle, J.B. and Hawking, S.W. (1983) *Physical Review D*, **28**, 2960-2975. https://doi.org/10.1103/PhysRevD.28.2960

[16] Pons-Rullán, B. (2017) *Journal of Physical Mathematics*, **8**, Article ID: 1000234.

[17] Novikov, I.D. (1965) *Soviet Astronomy. AJ*, **8**, 857-863.

[18] Silk, J. (1970) *Monthly Notices of the Royal Astronomical Society*, **147**, 13-19. https://doi.org/10.1093/mnras/147.1.13

[19] Freedman, W.L. and Madore, B.F. (2010) *Annual Review of Astronomy and Astrophysics*, **48**, 673-710. https://doi.org/10.1146/annurev-astro-082708-101829

[20] Capozziello, S., Cardone, V.F. and Dainotti, M.G. (2009) *Monthly Notices of the Royal Astronomical Society*, **400**, 775-790. https://doi.org/10.1111/j.1365-2966.2009.15456.x

Journal of Modern Physics, 2022, 13, 336-346
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Empirical Equation for a Fine-Structure Constant with Very High Accuracy

Tomofumi Miyashita

Miyashita Clinic, Osaka, Japan
Email: tom_miya@plala.or.jp

How to cite this paper: Miyashita, T. (2022) Empirical Equation for a Fine-Structure Constant with Very High Accuracy. *Journal of Modern Physics*, **13**, 336-346.
https://doi.org/10.4236/jmp.2022.134024

Received: March 3, 2022
Accepted: April 4, 2022
Published: April 7, 2022

Abstract

We proposed several empirical equations about the electromagnetic force and gravity. The main three equations were connected mathematically. However, these equations have small errors of approximately 10^{-3}. Therefore, we attempted to improve the accuracy. Regarding the factors of 9/2 and π, we used 4.48870 and 3.13189, respectively. Then, the errors become smaller than 10^{-5}. However, we could not show any reasons for these compensations. We noticed the following equations. $4.488855 = \sqrt{\frac{136.0113\times 4}{27}}$, $3.131777 = \frac{Rk}{(3\times 136.0113)^{1.5}}$. Then, we can explain the von Klitzing constant $Rk = 3.131777037\times 4.488855463\times 13.5\times 136.0113077$. It is well known that the von Klitzing constant can be measured with very high accuracy. We examined this equation for the von Klitzing constant in detail. Then, we noticed that 136.0113 should be uniquely determined. The von Klitzing constant is highly related to the fine-structure constant. After the examination of the numerical connections, we can explain the value of 137.035999081 as a fine-structure constant with very high accuracy. Then, we attempt to explain this value from Wagner's equation.

Keywords

Fine-Structure Constant, Wagner's Equation

1. Introduction

We discovered Equation (1) [1] [2], which was seemed to be very simple.

$$\frac{Gm_p}{\frac{\lambda_p}{2}}\times 1\text{ kg} = \frac{9}{2}kT_c \tag{1}$$

Next, we discovered Equations (2) and (3) [3], which were seemed to be simple, too.

$$\frac{Gm_p^2}{\left(\frac{e^2}{4\pi\varepsilon_0}\right)} = \frac{4.5}{2\pi} \times \frac{m_e}{e} \times hc \times \left(1\frac{\mathrm{C}}{\mathrm{J\cdot m}} \times \frac{1}{1\,\mathrm{kg}}\right) \tag{2}$$

$$\frac{m_e c^2}{e} \times \left(\frac{e^2}{4\pi\varepsilon_0}\right) = \pi k T_c \times \left(1\frac{\mathrm{J\cdot m}}{\mathrm{C}}\right) \tag{3}$$

Equations (1), (2) and (3) were connected mathematically [3]. However, we could not establish the background theory. Then, we abandoned theoretical explanations and tried to reduce the error.

Our three equations have small errors of approximately 10^{-3} and 10^{-4} [3]. We attempted to reduce the error in previous reports by changing Factors 4.5, π and T_c [4] [5]. Regarding the value of T_c, 2.72642 K was used instead of 2.72548 K. Regarding the factors of 9/2 and π, we used 4.48870 and 3.13189, respectively. Then, the errors become smaller than 10^{-5}. Then,

$$\frac{Gm_p^2}{hc} = \frac{4.4887}{2} \frac{kT_c}{1\,\mathrm{kg}\times c^2} \tag{4}$$

Equation (4) is deduced from Equation (1).

$$\frac{Gm_p^2}{\left(\frac{e^2}{4\pi\varepsilon_0}\right)} = \frac{4.4887}{2\times 3.13189} \times \frac{m_e}{e} \times hc \times \left(1\frac{\mathrm{C}}{\mathrm{J\cdot m}} \times \frac{1}{1\,\mathrm{kg}}\right) \tag{5}$$

$$\frac{m_e c^2}{e} \times \left(\frac{e^2}{4\pi\varepsilon_0}\right) = 3.13189 \times kT_c \times \left(1\frac{\mathrm{J\cdot m}}{\mathrm{C}}\right) \tag{6}$$

4.48870 and 3.13189 are connected with the following equations.

$$\frac{m_p}{m_e} = \frac{1}{4.4887\times 3.13189} \frac{q_m}{e} \tag{7}$$

However, we could not provide any reasons behind 4.48870 and 3.13189. In this report, we proposed the following two values for compensation.

$$3.131777037(\Omega) = \frac{Rk}{\left(3\times 136.0113077\right)^{1.5}} \tag{8}$$

$$4.488855463 = \sqrt{\frac{136.0113077\times 4}{27}} \tag{9}$$

The rest of the paper is organized as follows. In Section 2, we present an empirical equation for the fine-structure constant with very high accuracy. In Section 3, we explained the value of the von Klitzing constant using these values to support our empirical equation. In Section 4, we propose and refine our three equations. For the value of T_c, we use 2.72652 K instead of 2.72548 K. The significant figures of the measured T_c are only 3 or 4 in the present study. Therefore, this assumption is valid. For the value of G, we use 6.674777×10^{-11}

$m^3 \cdot kg^{-1} \cdot s^{-2}$ instead of 6.6743×10^{-11} $m^3 \cdot kg^{-1} \cdot s^{-2}$. In Section 5, we discuss the meaning of our empirical equations.

2. Empirical Equation for a Fine-Structure Constant with Very High Accuracy

2.1. Symbol List

These Values Were Obtained from Wikipedia.

G: gravitational constant: 6.6743×10^{-11} ($m^3 \cdot kg^{-1} \cdot s^{-2}$)
(we used the compensated value 6.674777×10^{-11} in this report)
T_c: temperature of the cosmic microwave background: 2.72548 (K)
(we used the compensated value 2.72652 K in this report)
k: Boltzmann constant: 1.380649×10^{-23} ($J \cdot K^{-1}$)
c: speed of light: 299792458 (m/s)
h: Planck constant: $6.62607015 \times 10^{-34}$ (J·s)
ε_0: electric constant: $8.8541878128 \times 10^{-12}$ ($N \cdot m^2 \cdot C^{-2}$)
μ_0: magnetic constant: $1.25663706212 \times 10^{-6}$ ($N \cdot A^{-2}$)
e: electric charge of one electron: $-1.602176634 \times 10^{-19}$ (C)
q_m: magnetic charge of one magnetic monopole: $4.13566770 \times 10^{-15}$ (Wb)
(this value is only a theoretical value, $q_m = h/e$)
m_p: rest mass of a proton: $1.6726219059 \times 10^{-27}$ (kg)
(we used the compensated value $1.672621923 \times 10^{-27}$ kg in this report)
m_e: rest mass of an electron: $9.1093837 \times 10^{-31}$ (kg)
λ_p: Compton wavelength of a proton: 1.32141×10^{-15} (m)
Rk: von Klitzing constant: 25812.80745 (Ω)
Z_0: wave impedance in free space: 376.730313668 (Ω)
α: fine-structure constant: 1/137.035999081

2.2. Empirical Equation for a Fine-Structure Constant

Our empirical equation for the fine-structure constant is

$$\alpha = \frac{1}{\left(1+\frac{1}{1836.152654\times 3}\right)\times\frac{1836.152654}{13.5}+1} = \frac{1}{137.0359991} \tag{10}$$

where 1836.152654 is explained as follows:

$$\frac{m_p}{m_e} = 1836.152654 = 13.5\times 136.0113077 \tag{11}$$

where the measured m_p is $1.6726219059 \times 10^{-27}$ (kg). Therefore, the precise value is

$$\frac{m_p}{m_e} = 1836.152673 \tag{12}$$

Therefore, the error is

$$\text{Error} = \frac{1836.152673}{1836.152654} - 1 = 1.01977\times 10^{-8} \tag{13}$$

This value is sufficiently small. When compensation is needed, the value of m_p should be $1.672621923 \times 10^{-27}$ kg. This small error will be explained by a very difficult theory, such as the explanation for the Lamb shift, in the future. In the next section, we attempt to explain Equation (10) from the examination of the value of the von Klitzing constant.

3. Examination of the Von Klitzing Constant

3.1. Two Important Coefficients

Previously, we proposed the two coefficients 4.48870 and 3.13189 [4] [5]. Then, we thought that 4.48870 is the deviation from 4.5. 3.13189 is the deviation from π.

Then, we noticed the following equations.

$$4.48867 = \sqrt{\frac{136 \times 4}{27}} \tag{14}$$

In Equation (14), when we believe that 4.48867 should be a deviation from 4.5,

$$4.48867 = 4.5 \times \sqrt{\frac{17}{\left(\frac{3}{2}\right)^7}} \tag{15}$$

Next, we persevered for half a year and discovered the following equation.

$$3.13217(\Omega) = \frac{Rk}{(3 \times 136)^{1.5}} \tag{16}$$

In Equation (16), when we believe that 3.13217 should be a deviation from π,

$$\pi = 2\left(1+\frac{1}{3}\left(1+\frac{2}{5}\left(1+\frac{3}{7}\left(1+\frac{4}{9}\left(1+\frac{5}{11}\left(1+\frac{6}{13}\right)\right)\right)\right)\right)\right) = 3.132156732 \tag{17}$$

For convenience, Equation (7) is rewritten as follows:

$$\frac{m_p}{m_e} = \frac{1}{4.4887 \times 3.13189}\frac{q_m}{e} = \frac{1}{14.05809432}Rk \tag{18}$$

where 14.5809432 (Ω) is the resistance. Using new values of 4.48867 and 3.13217,

$$3.13217(\Omega) \times 4.48867 = 14.05926332(\Omega) \tag{19}$$

Therefore, the error between Equations (18) and (19) is

$$\text{Error} = \frac{14.05926332}{14.05809432} - 1 = 8.31552 \times 10^{-5} \tag{20}$$

In Equation (20), the error is too large. In Equation (18), we use 14.05926332 instead of 14.05809432.

$$\frac{m_p}{m_e} = \frac{1}{14.05926332}\frac{q_m}{e} = \frac{25812.80746}{14.05926332} = 1836 \neq 1836.1526 \tag{21}$$

The ratio between m_p and m_e becomes smaller than the measured value. This

problem can be solved using any other values instead of 136 in Equations (14) and (16), which will be explained in a later section.

3.2. Examination of the Relationship among 4.48866, 3.13217 and *Rk*

The relationship among 4.48866, 3.13217 and *Rk* is examined in this section. For convenience, Equations (14) and (16) are rewritten below:

$$4.48867 = \sqrt{\frac{136\times 4}{27}} \tag{22}$$

$$3.13217(\Omega) = \frac{Rk}{(3\times 136)^{1.5}} \tag{23}$$

From Equations (22) and (23),

$$Rk = 3.13217\times 4.48867\times 13.5\times 136 \tag{24}$$

Equation (24) can be deduced from Equation (21), directly. Again, from Equations (22) and (23),

$$\frac{3.13217}{4.48867} = \frac{\dfrac{Rk}{(3\times 136)^{1.5}}}{\sqrt{\dfrac{136\times 4}{27}}} = \frac{Rk}{(136)^2}\times\frac{1}{2} = 0.697794 \tag{25}$$

Therefore,

$$Rk = 2\times\frac{3.13217}{4.48867}\times(136)^2 \tag{26}$$

In Equation (26), *Rk* can be explained by 136. From Equation (22),

$$136 = \frac{13.5}{2}\times(4.48867)^2 \tag{27}$$

In Equation (27), 136 can be explained by 13.5. From Equation (23),

$$Rk = 3.13217\times(3\times 136)^{1.5} \tag{28}$$

In Equation (28), *Rk* can be explained by 3 and 136. From Equations (27) and (28),

$$Rk = 3.13217\times\left(3\times\frac{13.5}{2}\times(4.48867)^2\right)^{1.5} \tag{29}$$

In Equation (29), *Rk* can be explained by 3 and 13.5. Therefore,

$$Rk = \left(\frac{9}{2}\right)^3\times 3.13217\times(4.48867)^3 \tag{30}$$

Consequently, we proposed the four equations (Equations (24), (26), (28) and (30)) to explain *Rk*. The key numbers are 3, 13.5 and 136.

3.3. Equations (22) and (23) as a Function of the Value 136

We believe that Equations (22) and (23) are functions of 136. Then, we propose the following equations. The significant digits should be more than ten to ex-

plain the following equations.

$$4.488855463 = \sqrt{\frac{136.0113077 \times 4}{27}} \tag{31}$$

$$3.131777037 = \frac{Rk}{\left(3 \times 136.0113077\right)^{1.5}} \tag{32}$$

From Equations (31) and (32),

$$Rk = 3.131777037 \times 4.488855463 \times 13.5 \times 136.0113077 \tag{33}$$

Again, from Equations (31) and (32),

$$Rk = 2 \times \frac{3.131777037}{4.488855463} \times \left(136.0113077\right)^2 \tag{34}$$

From Equation (31),

$$136.0113077 = \frac{13.5}{2} \times \left(4.488855463\right)^2 \tag{35}$$

From Equation (32),

$$Rk = 3.131777037 \times \left(3 \times 136.0113077\right)^{1.5} \tag{36}$$

From Equations (35) and (36),

$$Rk = \left(\frac{9}{2}\right)^3 \times 3.131777037 \times \left(4.488855463\right)^3 \tag{37}$$

Consequently, we propose four equations (Equations (33), (34), (36) and (37)) to calculate *Rk*. Furthermore, 1836.152654 in Equation (11) can be explained. The key numbers are 3, 13.5 and 136.0113077. Thus, Equation (10) can be explained.

3.4. Relationship between *Rk* and Z_0

$$Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} = 2\alpha \times Rk \tag{38}$$

For example, from Equations (34) and (38),

$$Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} = \frac{4 \times 3.131777037}{4.488855463} \times \frac{\left(136.0113077\right)^2}{137.0359991} = 376.7303137 \tag{39}$$

However, Equation (10) explaining α is too complex to calculate Z_0. Therefore, an explanation for the electric constant and the magnetic constant will be published in future reports.

4. Explanation for Our Three Refined Empirical Equations

4.1. The Problem for the Measured Values of *G* and T_c

The significant figures of the measured T_c are only 3 or 4 in the present study. For the value of T_c, we use 2.72652 K instead of 2.72548 K.

https://en.wikipedia.org/wiki/Cosmic_microwave_background.

The measured public value of G was often changed. We strongly propose that the value of G should be measured in space. On the ground, the value of G is influenced by the centrifugal force due to the rotation of the earth. We use 6.674777×10^{-11} $m^3 \cdot kg^{-1} \cdot s^{-2}$ instead of 6.6743×10^{-11} $m^3 \cdot kg^{-1} \cdot s^{-2}$.

https://en.wikipedia.org/wiki/Gravitational_constant.

4.2. Our Three Refined Empirical Equations

Equation (1) is refined as follows:

$$\frac{Gm_p^2}{hc} = \frac{4.488855}{2}\frac{kT_c}{1\,\mathrm{kg}\times c^2} \tag{40}$$

$$\frac{Gm_p^2}{hc} = \frac{6.674777\times10^{-11}\times\left(1.6726219\times10^{-27}\right)^2}{6.626070\times10^{-34}\times2.9979246\times10^{8}} = 9.400600\times10^{-40} \tag{41}$$

$$\begin{aligned}\frac{4.488855}{2}\frac{kT_c}{1\,\mathrm{kg}\times c^2} &= \frac{4.488855}{2}\times\frac{1.3806490\times10^{-23}\times2.726516}{\left(2.9979246\times10^{8}\right)^2} \\ &= 9.400600\times10^{-40}\end{aligned} \tag{42}$$

Equation (41) is equal to Equation (42). Therefore, the compensation method is perfect.

Next, Equation (2) is refined as follows:

$$\frac{Gm_p^2}{\left(\frac{e^2}{4\pi\varepsilon_0}\right)} = \frac{4.488855}{2\times3.131777}\times\frac{m_e}{e}\times hc\times\left(1\frac{\mathrm{C}}{\mathrm{J}\cdot\mathrm{m}}\times\frac{1}{1\,\mathrm{kg}}\right) \tag{43}$$

$$\frac{Gm_p^2}{\frac{e^2}{4\pi\varepsilon_0}} = \frac{6.674777\times10^{-11}\times\left(1.672621923\times10^{-27}\right)^2}{\frac{\left(1.60217663\times10^{-19}\right)^2}{4\pi\times8.8541878\times10^{-12}}} = 8.094129\times10^{-37} \tag{44}$$

$$hc = 6.626070\times10^{-34}\times2.9979246\times10^{8} = 1.986446\times10^{-25} \tag{45}$$

$$\begin{aligned}\frac{4.488855}{2\times3.131777}\times\frac{m_e}{e}\times hc &= \frac{4.488855\times9.10938\times10^{-31}\times1.986446\times10^{-25}}{2\times3.131777\times1.602177\times10^{-19}} \\ &= 8.094129\times10^{-37}\end{aligned} \tag{46}$$

Equation (44) is equal to Equation (46). Therefore, the compensation method is perfect.

Next, Equation (3) is refined as follows:

$$\frac{m_e c^2}{e}\times\left(\frac{e^2}{4\pi\varepsilon_0}\right) = 3.131777\times kT_c\times\left(1\frac{\mathrm{J}\cdot\mathrm{m}}{\mathrm{C}}\right) \tag{47}$$

$$\begin{aligned}&\frac{m_e c^2}{e}\times\left(\frac{e^2}{4\pi\varepsilon_0}\right) \\ &= \frac{9.10938\times10^{-31}\times\left(2.9979246\times10^{8}\right)^2\times\left(1.60217663\times10^{-19}\right)^2}{1.60217663\times10^{-19}\times4\pi\times8.8541878\times10^{-12}} \\ &= 1.1789142\times10^{-22}\end{aligned} \tag{48}$$

$$\begin{aligned} 3.131777 \times kT_c &= 3.131777 \times 1.3806490 \times 10^{-23} \times 2.726516 \\ &= 1.1789142 \times 10^{-22} \end{aligned} \tag{49}$$

Equation (48) is equal to Equation (49). Therefore, our refined empirical Equations (40), (43) and (47) are perfect. However, we use the compensated value for T_c and G, which is slightly different from the measured value.

5. Discussion

The main purpose of this report is to show empirical equations. However, without any theories, this report can appear to be nothing more than numerology. Therefore, we wrote this section. Our arguments strongly depend on numerical connections rather than logical connections. Therefore, our discussions appear to be different from the majority's opinion. This does not mean that we ignore the majority's opinions. Using simple unknown ideas, our opinions may be compatible with the majority's opinion.

5.1. Consider the Value of 136.0113077

The value of 136.0113077 is related to the mass ratio between some particle and an electron. The rest mass of a pion (π^+) is 139.57 MeV. The mass of an electron is 0.51100 MeV. Therefore,

$$\frac{139.57 \text{ MeV}}{0.51100 \text{ MeV}} = 2 \times 136.57 \approx 2 \times 136.011 + 1 \tag{50}$$

When the pion consists of two quarks and an electron, the value of 136.011 is the mass ratio between a quark and an electron. This conclusion seems to disprove the concept of gluon. However, we notice that the concept of a gluon is useful for the mass ratio between a proton and an electron.

5.2. Consideration for the Value of 1836.152654

The value of 1836.152654 is related to the mass ratio between a proton and an electron.

$$\frac{m_p}{m_e} = 1836.152654 = 4.5 \times 3 \times 136.0113077 \tag{51}$$

When a proton consists of three quarks and the value of 136.011 is the mass ratio between a quark and an electron, there remains the mystery about the factor of 4.5. We believe that the factor of 4.5 derives from the degrees of freedom of the thermal energy due to the combined three quarks. Then, we must explain the value of 3.5 (=4.5 − 1).

According to Albert Einstein, $E = mc^2$. Therefore, the thermal energy should be related to the weightness. A hot mass should be heavier than a cold mass. Thus, we must consider what happens concerning the degree of freedom.

When the boundary between thermal energy and the rest mass energy becomes unclear, 77% (=3.5/4.5) of the rest mass energy related to the thermal energy from the degree of freedom seems to be binding energy. This concept of

binding energy may be compatible with the concept of gluons.

5.3. Consideration of the Theoretical Meaning of Equation (10)

The theory of loop quantum gravity is related to the theory of the lattice structure in space time. In the area of electrochemistry, the theory of vacancies in the lattice structure is governed by Wagner's equation. The fine-structure constant is the interference constant. When the interference constant for electrons is α, the transference number for electrons should be $1-\alpha$. Then, the transference of quarks is α, which explains the strong interaction [6]. According to the advanced Wagner model, the diffusion time for mixed electronic and ionic currents should decrease exponentially with distance [7]. When the diffusion time for mixed electrons and quark flux should decrease exponentially with distance, the Yukawa potential can be explained.

For convenience, Equation (10) is rewritten as follows:

$$\alpha = \frac{1}{\left(1+\frac{1}{1836.152654\times 3}\right)\times\frac{1836.152654}{13.5}+1} = \frac{1}{137.0359991} \tag{52}$$

Therefore,

$$\frac{1}{\alpha} = \left(1+\frac{1}{\frac{m_p}{m_e}\times 3}\right)\times\frac{\frac{m_p}{m_e}}{13.5}+1 = \frac{\left(\frac{m_p}{m_e}\times 3+1\right)\times\frac{m_p}{m_e}}{\frac{m_p}{m_e}\times 3\times 13.5}+1 = \frac{\left(\frac{m_p}{m_e}\times 3+1\right)}{3\times 13.5}+1 \tag{53}$$

So,

$$\frac{1}{\alpha} = \frac{1}{13.5}\times\frac{m_p}{m_e}+\frac{1}{3\times 13.5}+1 \tag{54}$$

From Equations (11) and (35),

$$\frac{1}{13.5}\times\frac{m_p}{m_e} = \frac{1836.152673}{13.5} = 136.0113077 = \frac{13.5}{2}\times\left(4.488855463\right)^2 \tag{55}$$

Therefore, from Equations (54) and (55),

$$\frac{1}{\alpha} = \frac{13.5}{2}\times\left(4.488855463\right)^2+\frac{1}{3\times 13.5}+1 \tag{56}$$

Equation (56) is one of the variations of Equation (52). From Equation (56), we can justify that the value of 4.488855463 is not coincidental.

5.4. Consideration of the Degree of Freedom inside Electrons

In Equation (47), the number 3.131777037 is the deviation from π. The degree of freedom inside an electron seems to be 2π. We believe that 2π is related to the spin of the electron. Angrick *et al.* confirmed that the spin of electrons cannot be ignored thermodynamically [8]. Furthermore, Aquino *et al.* discovered a new method for vector analysis [9]. We hope that the degrees of freedom in electrons will be clarified experimentally in detail.

6. Conclusions

Previously, we proposed several empirical equations to describe the relationship between an electromagnetic force and T_c [1] [2] [3]. Three equations were explained by the factors of 9/2 and π. However, these equations have small errors of approximately 10^{-3} and 10^{-4} [4] [5].

We tried to improve the accuracies by changing the value of 9/2 and π. Then, we persevered and discovered the following two equations.

$$4.48867 = \sqrt{\frac{136\times 4}{27}}$$

$$3.13217(\Omega) = \frac{Rk}{(3\times 136)^{1.5}}$$

Then, the above two equations were the function of 136. Using the value of 136.0113007,

$$3.131777037(\Omega) = \frac{Rk}{(3\times 136.0113077)^{1.5}}$$

$$4.488855463 = \sqrt{\frac{136.0113077\times 4}{27}}$$

Then, Rk can be calculated as follows.

$$Rk = 3.131777037\times 4.488855463\times 13.5\times 136.0113077$$

$$\frac{m_p}{m_e} = 1836.152654 = 13.5\times 136.0113077$$

Then, the empirical equation for the fine-structure constant with very high accuracy is

$$\alpha = \frac{1}{\left(1+\frac{1}{1836.152654\times 3}\right)\times\frac{1836.152654}{13.5}+1} = \frac{1}{137.0359991}.$$

Avoiding numerology, we discuss this equation. Then, we can justify that the value of 4.488855463 is not coincidental. The explanation for the electric constant and the magnetic constant will be published in future reports.

Conflicts of Interest

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

References

[1] Miyashita, T. (2020) *Journal of Modern Physics*, **11**, 1180-1192. https://doi.org/10.4236/jmp.2020.118074

[2] Miyashita, T. (2021) *Journal of Modern Physics*, **12**, 623-634. https://doi.org/10.4236/jmp.2021.125040

[3] Miyashita, T. (2021) *Journal of Modern Physics*, **12**, 859-869. https://doi.org/10.4236/jmp.2021.127054

[4] Miyashita, T. (2020) *Journal of Modern Physics*, **11**, 1559-1560. https://doi.org/10.4236/jmp.2020.1110096

[5] Miyashita, T. (2021) *Journal of Modern Physics*, **12**, 1160-1161. https://doi.org/10.4236/jmp.2021.128069

[6] Miyashita, T. (2018) *Journal of Modern Physics*, **9**, 2346-2353. https://doi.org/10.4236/jmp.2018.913149

[7] Duncan, K.L. and Wachsman, E.D. (2009) *Journal of The Electrochemical Society*, **156**, B1030. https://doi.org/10.1149/1.3158513

[8] Angrick, C., Braun, J., Ebert, H. and Donath, M. (2021) *Journal of Physics*: *Condensed Matter*, **33**, Article ID: 115501. https://doi.org/10.1088/1361-648X/abd338

[9] Aquino, F.W. and Wong, B.M. (2018) *Journal of Physical Chemistry Letters*, **9**, 6456-6462. https://doi.org/10.1021/acs.jpclett.8b02786

Journal of Modern Physics, 2022, 13, 347-367
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

A Re-Interpretation of Quasi-Local Mass

Edwin Eugene Klingman

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

How to cite this paper: Klingman, E.E. (2022) A Re-Interpretation of Quasi-Local Mass. *Journal of Modern Physics*, **13**, 347-367.
https://doi.org/10.4236/jmp.2022.134025

Received: March 8, 2022
Accepted: April 4, 2022
Published: April 7, 2022

Abstract

Because the equivalence principle forbids local mass density, we cannot formulate general relativistic mass as an integral over mass density as in Newtonian gravity. This century-old problem was addressed forty years ago by Penrose, and many papers have since extended the concept. Currently there is no satisfactory physical understanding of the nature of quasi-local mass. In this paper I review the key issues, the current status, and propose an alternative interpretation of the problem of local mass and energy density for gravity systems from an information perspective.

Keywords

Quasi-Local Mass, Curved Space-Time, Gravitational Energy, Stress-Energy Tensor, Encoding Information in Geometry, Schwarzschild Metric

1. Introduction

Except for general relativity, classical mechanics can be understood entirely within the context of the conservation of energy and momentum. Yet every general relativity text contains discussion of the fact that the energy of gravitating systems cannot be formulated in curved space-time. A recent Centennial paper [1] by Chen, *et al.* begins by stating:

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

The problem of how best to describe the energy-momentum and angular momentum in gravitating system suffers from the fact that "*It is known that these quantities cannot be given a local density.*" A century-long failure to solve the fundamental problem indicates confusion; the goal of this work is to provide necessary clarification of the issue.

In fact, a *second* fundamental issue surfaces in quasi-local mass, that of Ha-

miltonian-based physics, in which the dynamics of a system is described through total energy by Hamilton's equations of motion. This approach is completely equivalent to Newtonian mechanics but does not apply to general relativity. Noether, who in 1918 proved "*there is no covariant total energy-momentum density tensor for gravitating systems*", also provided the fundamental basis of most 20th century physics by linking conservation theorems to symmetry transformations. But there is no general translational symmetry in curved space, so this avenue is denied to physicists. Penrose observed, non-local gravitational field energy is formulable only in expressions used *at infinity* for an *asymptotically flat* space-time manifold. In other words, the problems of curved space are treated by backing off to a "flat space" formulation at infinity. Moreover, an attempt is made to measure the energy of a system by enclosing it with a membrane, or closed spacelike two-surface, and attaching this to an energy-momentum four vector. Little wonder that all major contributors to the quasi-local mass approach admit that there is no framework in which it is understood. As this paper was being completed, yet another approach was published in which Hamiltonians in *asymptotically flat spacetime in five spacetime dimensions*, focused on spatial infinity.

The plan of this paper is as follows:

Part I introduces coordinate systems and the underlying issue of curved space.

Part II introduces the concept of quasi-local mass, invented by Penrose forty years ago, and reviews Yau's treatment of the problem and the reason for the focus on "null infinity" and asymptotically flat space-time, then Bart's 2019 summary of the current status of quasi-local conserved quantities.

Part III develops a reinterpretation of quasi-local mass, based on an information theoretic approach to the subject.

Part I Introduction to coordinate systems and invariance

An axiom of physics is that coordinate systems have no effect on physical reality. When this is violated, physics becomes confusing. Despite that coordinates can have no effect on physics, *the equivalence principle* of general relativity yields a built-in contradiction: by transforming local coordinates the gravitational field can be banished. Weinberg [2] on the equivalence principle:

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

Einstein based his theory of gravity on the equivalence principle, which states that choice of the proper coordinate system makes gravity disappear; the basic problem of local density of energy-momentum. The principle is only approximate, applying when tidal forces can be ignored, but is generally treated as absolute. We later introduce an absolute principle, the principle of primordial self-interaction, but first we review coordinate system issues. We begin by noting that Euclidean space is Pythagorean in that distances in space and time are un-

limited.

Pythagorean distance in Euclidean *space* and *time*:

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

Euclidian four-space can thus be mapped onto all events in the universe, allowing us to label every event and relate any event to any other. These relations constitute our *physics* or *models of reality*. But what are the relations? They are intended to capture objective reality in some sense. Nozick, in *Invariance: the structure of the objective world* [3] observes that an objective fact is accessible from different angles, *i.e.*, "*an objective fact is invariant under various transformations.*"

The Galilean transformation describes a photon's position in 4-space with time $\vec{x} = \vec{v}t$ where velocity $\vec{v}$ can point in any direction and $\left|\pm\vec{v}\right| = c$, with $x = -ct$ representing a photon moving in the negative direction. Differentiating we have $\mathrm{d}\left(x - ct\right) = 0$ or $\mathrm{d}\left(x + ct\right) = 0$. After Hestenes, [4] we define geometric algebra multi-vector $X = ct + \vec{x}$ and its complement $\tilde{X} = ct - \vec{x}$ and show that $\mathrm{d}X\,\mathrm{d}\tilde{X} = c^2\mathrm{d}t^2 - \mathrm{d}\vec{x}^2 = 0$ is an *invariance* that holds for all distances $\mathrm{d}\vec{x}$ and corresponding duration $\mathrm{d}t$, independent of coordinate systems. If we denote the value $\mathrm{d}X\,\mathrm{d}\tilde{X}$ by $\mathrm{d}s^2$ we obtain the invariance relation for the photon equation of motion, where we use $\mathrm{d}\vec{x}^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$:

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

We interpret this as an *invariance relation* in Euclidian 4-space and call it the *Minkowski invariance*. But Minkowski did not conceive of this as an invariance relation; he believed it to be a description of "space-time", famously stating: "*space and time by itself are doomed to fade away… only a kind of union of the two will preserve an independent reality.*" The Lorentz transformation applied to this imagined spacetime rotates space into time and time into space. In *Energy-time theory*, developed in *Physics of clocks in absolute space and time* [5] this fundamental photon-based invariance (2) is used to derive classical Hamiltonian

$$H = E = \sqrt{m_0^2c^4 + c^2p^2}\ . \tag{3}$$

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

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} = \gamma = \frac{1}{\sqrt{1 - \left(v^2/c^2\right)}} \tag{4}$$

Derivation of this *time dilation* is based on Galilean transformations in absolute time and space and relativistic inertia $m = \gamma m_0$ with m_0 the rest mass. Per Lucas and Hodgson [6]: "*If we… retain Newtonian dynamics, and the Newtonian definition of velocity and acceleration, then we… still obtain relativistically correct results if we… allow mass to depend on the velocity.*"

In special relativity rest mass is defined in all inertial reference frames in relative motion, and the Lorentz transformation applied to Minkowski "space" and

"time". Yet the most solid experimental fact of special relativity, *time dilation*, is reproduced *exactly* in absolute space and time. Energy-time theory does *not* yield *length contraction* which Rindler [7] says will probably *never* be tested. Nor does it yield Einstein's *velocity addition law*, known to be violated in particle accelerators [8]. The physics of *clock slowing* in absolute time (*universal simultaneity*) and absolute space (with *preferred frame* defined by local gravity) follows from increased inertia of the clock mechanism and consequent decrease in acceleration of the restoring force common to all harmonic oscillators, by *exactly* a factor of γ. Energy-time physics (compatible with *all* relativistic experiments [9]) is based on Galilean transformation *in time and space*: *Minkowski is an invariance relation*.

Thus, Hestenes' multi-vector formulation of time and space leads to special relativity if viewed as *Minkowski spacetime* operated on by Lorentz transformation but leads to classical physics if treated as *Minkowski invariance*. The mathematical construct representing time and space is the same; the ontological assumptions determine what one does with it.

The principle of general covariance implies that coordinates are labels of space-time events that can be assigned completely arbitrarily. The only quantities that have physical meaning, the measurables, are *invariant* under coordinate transformations." Such invariance is used to derive the physics of absolute space and time in reference 5; the physics of *energy-time* rather than relativistic *space-time.* Yet Poisson and Will [10] state: "*All local aspects of gravity can be turned off by doing physics in a freely moving (coordinate) frame of reference. Gravity is not present in these frames.*"

In other words, one can mathematically *do away with* the local gravitational field and hence any associated energy, a problem that has never been solved. Most physicists are aware of Noether's seminal work on symmetry and conservation, but the reason Noether began investigations was "*to clarify the issue of gravitational energy.*" She proved "*there is no covariant total energy-momentum density tensor for gravitating systems*." MTW [11] discuss this as "*a consequence of the equivalence principle…*". A century of effort *has not solved the problem*.

2. Coordinates and Gravity as Curved Space-Time

Einstein's transformation of gravity into geometry followed Minkowski's idea of spacetime, and Einstein's own analysis of Maxwell-Hertz equations, based on his interpretation of Michelson-Morley's experimental results. Hertzian electrodynamics *is* Galilean invariant, but Einstein believed Lorentz invariance was required. Symmetry arguments imply that Lorentz's transformation in four-dimensional spacetime is not supported in 3-space plus absolute time, but these arguments were not advanced until long after Einstein's relativity was cast in stone. Following the historical approach, flat space is viewed as a limit to the general curved space metric:

$$ds^2 = g_{ij}dx^i dx^j, \tag{5}$$

where g_{ij} represents the curvature on a manifold in four-space with tangent vectors *i* and *j*. The geometric meaning of g_{ij} is quite clear, however the physical meaning is debatable. Most treatments seem to identify the term with gravity, however MTW state:

"*…nowhere has a precise definition of the term 'gravitational field' been given – nor will one be given. Many different mathematical entities are associated with gravitation: the metric, the Riemann curvature tensor, the Ricci curvature tensor, the curvature scalar, the covariant derivative, the connection coefficients, etc. … the terms 'gravitational field' and 'gravity' refer in a vague, collective sort of way to all of these entities.*"

This is backwards; a better statement of physics would be:

"*… all of these entities refer in a vague, collective sort of way to 'gravitational field' and 'gravity'.*"

In our treatment we view the metric g_{ij} as key. Beckwith [12] states: "*In general relativity metric $g_{ij}(\vec{x},t)$ is a set of numbers associated with each point which gives the distance to neighboring points, i.e., general relativity is a classical theory.*" Per Yau [13], in general relativity Einstein's equation is obtained by taking the variation of

$$\frac{1}{16\pi}\int R + \int \mathcal{L} \tag{6}$$

where R is the scalar curvature of the space-time and $\mathcal{L}$ is the Lagrangian of matter coupled to gravity. The gravitational interaction is described by means of space-time metric g_{ij} with signature (–, +, +, +); the metric of Minkowski space-time $\mathbb{R}^{3,1}$ (the vacuum with zero matter) being $ds^2 = g_{ij}dx^i dx^j = -dt^2 + dx^2 + dy^2 + dz^2$. The variational equation has the form

$$R_{ij} - \frac{1}{2}Rg_{ij} = T_{ij}. \tag{7}$$

We interpret this equation as follows: R_{ij} is curvature on a manifold, a Cartesian creation. The term R is a mean curvature over a manifold geometry, and T_{ij} is the physical stress energy tensor that induces the local curvature. In matter-free space $T_{ij} = 0$ and local curvature is essentially given by $R_{ij}/R = g_{ij}$, with the more global curvature represented by the smooth mean term *R*. When local stress energy is added, the smooth curvature is locally distorted with the difference in curvature represented by Equation (7). This interpretation will be useful for understanding quasi-local mass. The first solution to Einstein's equation is the Schwarzschild metric

$$ds^2 = -\left(1-\frac{2M}{r}\right)dt^2 + \left(1-\frac{2M}{r}\right)^{-1}dr^2 + dr^2\left(d\theta^2 + \sin^2(\theta)d\phi^2\right). \tag{8}$$

This and the Kerr metric both solve the Einstein equations with $T_{ij} = 0$ (vacuum).

Part II. Penrose and "quasi-local mass"

General relativity's incompatibility with local energy density of gravity has been known since Einstein formulated his equations, *i.e.*, the paradigmatic concept of energy conservation is *not* incorporated in Einstein's theory with respect to the energy of gravity itself. Energy tensor $T^{\mu\nu}$ describes the energy density of all *non-gravitational* fields; the gravitational field contributes only non-locally to the total energy. It is formulated only at infinity for an asymptotically flat space-time, while the energy of angular momentum exhibits an origin-dependence that is problematic even at flat space infinity. Non-local gravitational energy remained an unsolved problem, so Penrose in 1982 [14] proposed that "*energy-momentum is quasi-local: i.e., it is associated with a closed 2-surface…*". Of his new approach to the problem of local density, *quasi-local mass*, he noted: "*several problems of interpretation remain to be solved.*" The expression for total energy, momentum, or angular momentum surrounded by a closed surface, led to equations that, "*being highly over-determined, have no non-trivial solutions in the general curved space-time.*"

His attempts to formulate quasi-local angular momentum in terms of "twistor space", were based on complex analytic methods to solve problems in real differential geometry. According to Murray [15], in most cases the emphasis is on the *geometry* of the problem rather than *analysis*. Physical fields on Minkowski space are encoded into complex analytic objects on twistor space via the Penrose transform, which, per Murray, is an integral transform given by Whittaker and Watson. Penrose observes that "*we seem to require an analog*" of the infinity twistor, $I^{\alpha\beta}$. Based on several suggested definitions, and without the existence of the required analog, he used a relation from standard twistor theory, $m^2 = -A_{\alpha\beta}\bar{A}_{\alpha\beta}/2$, which Tod applied to a sphere of symmetry in Schwarzschild, Reissner-Nordstrom, and Friedmann-Robertson-Walker space-times, obtaining results related to the mass that would be surrounded by a sphere in flat space, of surface area equal to that of the quasi-local boundary, Ω, immersed in a fluid of density equal to that of the model. Yet the suggested definitions are not conformally invariant so they cannot be immediately applied at null infinity. Penrose ends by asking whether a *mass positivity theorem* can be proved at the quasi-local level, and whether one can obtain an *angular momentum flux law*. His approach to the problem of null local mass density in general relativity has survived for 40 years, despite the general level of confusion as to the physical interpretation of quasi-local mass.

3. Yau's Continuation of Penrose's Quasi-Local Mass

Following Equations (6) and (8) Yau considers the dynamics of scalar field $\Phi(t,r,\theta,\phi)$ in Kerr space-time whose propagation is described by a scalar wave equation. Since the Kerr space-time has a Killing vector $\partial/\partial t$, the Lagrangian $\int|\nabla\Phi|^2$ associated to the wave equation defines a local energy density which is positive except within the ergosphere, where it is negative, causing the energy method to breakdown. "*An important question for the wave equation is whether*

the wave will decay in time if initially it is not spread out. (…) *The problem of linear stability of Kerr is still open.*" The decay of the solution of the scalar wave equation is a special case of the decay of solutions to the Teukolsky equation which describes the linear stability of the Kerr black hole. A remarkable property of the Kerr metric is that the Teukolsky equation is separable by the ansatz

$$\mathrm{e}^{-i\omega t-\mathrm{i}k\phi}R(r)\Theta(\theta) \tag{9}$$

which Chandrasekhar observed "*has the aura of the miraculous.*" This is however compatible with the separation of gravity G and gravitomagnetic C field solutions of the Heaviside equations.

Most conservation theorems in physics derive from Noether's treatment of symmetry. For most mechanical systems the Lagrangian is invariant under translation of time and space; the resulting conserved quantity is a four-vector defining energy and linear momentum. However, most general relativity space-times do not admit translational symmetry, thereby frustrating the use of Noether's theorem. This is key to understanding Penrose's work: "*…one can define a total energy-momentum vector for an isolated physical system if there is asymptotic translational symmetry.*" When space-time is asymptotically flat, there is a space-like hypersurface which outside a compact set is diffeomorphic to $\mathbb{R}^3$ minus a ball and the metric g_{ij} has the form

$$g_{ij}=\delta_{ij}+\mathcal{O}\left(\frac{1}{r}\right). \tag{10}$$

In [16] I show that *the principle of self-interaction of the primordial field*, $\nabla\psi=\psi\psi$, leads to the linearized metric formulated in flat space, which is the goal of the asymptotic approach.

Yau: "*The total energy in general relativity cannot be obtained by integrating any local density along a hypersurface—the density would depend on first order differential differentiation of* g_{ij} *and there is a coordinate system where such quantities are zero at that point*", per the equivalence principle, violating the axiom of physics that physical quantities must be independent of choice of coordinate system. Yau proves that certain integrals over mean curvature are non-negative, then defines quasi-local mass in terms of the curvature of a closed space-like 2-surface in space-time, subject to conditions. He considers spherical symmetric space-time foliated by orbits of SU(2) and associates to an orbit the area $4\pi r^2$. The mean curvature vector of the orbit is $-2\nabla r/r$ where ∇ is with respect to the quotient Lorentzian (1, 1) metric. From curvature formulas he obtains the quasi-local mass of this orbit's sphere

$$M=r\left(1-\left|\nabla r\right|\right). \tag{11}$$

He compares this to Misner and Sharp's 1964 definition [17] of mass

$$m=\frac{r}{2}\left(1-\left|\nabla r\right|^2\right) \tag{12}$$

and obtains the relation

$$m = M - \frac{M^2}{2r} \tag{13}$$

such that at space-like infinity $M = m$. Yau closes by stating that he is still in process of deriving more properties of the quasi-local mass that he just introduced.

Yet another popular quasi-local mass is the Bartnik mass [18], defined as

$$m_B \Upsilon(\Sigma) = \frac{1}{8\pi}\int_\Sigma (H_0 - H)\,\mathrm{d}\Sigma \tag{14}$$

where Σ bounds a space-like hypersurface Ω and H_0 is the mean curvature of the unique isometric embedding of Σ into $\mathbb{R}^3$ and H is the mean curvature of Σ in Ω. Recall that we interpreted the variational Equation (7) as follows: R_{ij} is curvature on a manifold, while R is a mean curvature over a manifold geometry, and T_{ij} is the physical stress energy tensor inducing local curvature. The Bartnik mass (14) represents the mean curvature H of Σ in Ω, roughly corresponding to local curvature R_{ij} in four-space while H_0 is the mean curvature of the unique isometric embedding of Σ into $\mathbb{R}^3$, that is, Minkowski space, defined asymptotically. This is a merely heuristic explanation, since, as Penrose noted, "*several problems of interpretation remain to be solved.*"

Yau stated that the Misner and Sharp mass (12) is the same as the Hawking mass

$$m_{H(\Sigma)} = \sqrt{\frac{|\Sigma|}{16\pi}}\left(1 - \frac{1}{16\pi}\int_\Sigma |H|^2\,\mathrm{d}\Sigma\right), \tag{15}$$

where Σ is a spacelike 2-surface (the boundary over the region) and H is the mean curvature vector of Σ in the spacetime. If we assume that the H_0 term in the Bartnik mass is normalized such that

$$\frac{1}{16\pi}\int_\Sigma |H_0|^2\,\mathrm{d}\Sigma = 1, \tag{16}$$

we can transform the Hawking mass Equation (15) to more closely resemble Bartnik mass (14):

$$m_{H(\Sigma)} = \sqrt{\frac{|\Sigma|}{(16\pi)^3}}\left(\int_\Sigma \left(|H_0^2| - |H^2|\right)\mathrm{d}\Sigma\right). \tag{17}$$

4. Quasi-Local Conserved Quantities in General Relativity

Henk Bart's 2019 PhD dissertation [19] summarizes the current situation:

"*…there exists no general framework in which a definition of quasi-local energy is sufficiently understood.*"

Bart attempts to provide such a framework, noting that "*in general relativity, as a consequence of the equivalence principle, a local energy momentum tensor of the gravitational field does not exist.*" Yet, "*curvature of space-time is understood as a result of the presence of sources of energy. Energy in general relativity certainly exists, but not in a local sense. … defining a notion of quasi-local*

energy has proven to be surprisingly difficult…at the time of this writing the community does not agree on a set of pragmatic criteria that a notional quasi-local energy would have to satisfy."

After review of some energy basics, having to do with the equivalence of the Hamiltonian and Lagrangian formalisms, he introduces the concept of *asymptotically flat spacetimes*, while admitting that what one means by "asymptotes" is a matter of taste.

The issue, again, is that conservation of energy has been formulated in terms of Noether's theorem in which conservation is inherently linked to symmetry – conservation of energy and momentum are linked to time translation and space translation respectively. These translational symmetries are not generally available in curved space time and thus Noether's power tools are not available. This lies behind the attempt to asymptotically map curved space-time into flat space-time where Noether's theorem holds. So, Bart reviews *asymptotic symmetries at null infinity*, referred to as BMS symmetry after Bondi. He notes that the iso-symmetry group of flat space-time is the 10-dimensional Poincare group consisting of the Lorentz group of rotations and boosts and the translation group and stresses the fact that the asymptotic symmetry group is not the Poincare group, as might have been expected—instead *the asymptotic symmetry group is infinite dimensional*; a Hamiltonian at null infinity does not exist.

In attempting to provide a general framework for constructing quasi-local conserved quantities, he focuses on conserved quantities associated with asymptotic symmetries at null infinity, in the manner of Wald and Zoupas [20]. Bart also chooses the Einstein-Hilbert action supplemented with the Gibbons-Hawking-York boundary term

$$S = \frac{1}{16\pi}\int_M R + \frac{1}{8\pi}\int_{\partial M}\left(K - K_0\right) \tag{18}$$

where R as usual, denotes the Ricci scalar and K is the trace of the extrinsic curvature density of the (time-like) boundary ∂M. A Hamiltonian on M is defined in terms of Noether's charge two-form $Q[\xi]$ as follows:

$$H_X[\xi] = \int_C Q[\xi] - \xi \cdot B_X \tag{19}$$

representing a quasi-local conserved quantity defined on ∂M, with boundary conditions X. This equation is not yet well defined. B is a closed spacelike co-dimension two-surface in ∂M, on which H represents a quasi-local conserved quantity associated with a vector field that is tangent to ∂M. Much of Bart's effort involves an attempt to meaningfully define B subject to conditions X. He does not in detail investigate either uniqueness or existence of the proposed boundary. He does show that the asymptotic limit of the quasi-local BMS charges (his Equation (2.78)) agree with the BMS charges constructed by Wald and Zoupas at null infinity. He concludes that the BMS charge given by his Equation (2.78) satisfies certain criteria:

1) It vanishes on Minkowski space-time

2) It asymptotes to Bondi mass at null infinity.

3) On round spheres in the metric it is equal to the Misner-Sharp energy.

He then puts forward the possibility that the gravitational part of the zero mode BMS charge as constructed may be a useful definition of the quasi-local energy. This is his result.

Summary of quasi-local mass

Penrose attempted to address the problem by defining "quasi-local mass" (and angular momentum), and others during the past forty years have worked on this concept because

"*many important statements in general relativity make sense only with the presence of a good definition of quasi-local mass.*"

It is apparent that four decades of development of the concept of quasi-local mass has produced nothing that provides a physically meaningful interpretation of local energy density; forbidden by the equivalence principle. As seen, the primary tool of conservation, Noether's theorem, fails due to the general lack of translation symmetry in curved space-time. The approach to this has been to seek meaningful asymptotic symmetry at null infinity, although this introduces its own set of problems. In this context a Hamiltonian-like construct is defined and attempts to reproduce Noether-like results are made; nevertheless, no one today has any physically meaningful understanding of local energy density in general relativity. Baggott [21]: For Bohr, the Copenhagen interpretation obliges us to resist the temptation to ask: *but how does nature actually do that*?

"*And there lies the rub: for what is the purpose of the scientific theory if not to aid our understanding of the physical world*?"

It seems unlikely that any further work along the lines we have examined above will lead to such. Therefore, a re-interpretation of quasi-local energy seems appropriate. This is our focus in part III.

5. Problems with Einstein's Equations

A non-vanishing energy-density necessarily produces gravity, represented as curvature of space-time. Recall that Einstein's vision of gravity is *pure geometry*. His basic equation

$$G_{\mu\nu} = \kappa T^{(m)}_{\mu\nu} \tag{20}$$

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

$$-\kappa^{-1} G_{\mu\nu} + T^{(m)}_{\mu\nu} = 0 \tag{21}$$

as describing the vanishing sum of gravitation and material energy-momentum. However, this is generally *self-contradictory* if the solution is in curved space-

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

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

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

Physicists generally have no intuitive conception of *curved space-time*. Feynman [22] observed that gravity theory suffers because

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

Two possible solutions exist: the trivial solution is $G_{\mu\nu} = 0$; the stress-energy tensor is everywhere zero in all coordinate frames. This generally implies "flat space" but Vishwakarma has analyzed curvature in [23] in view of the fact that a proper energy-stress tensor of the gravitational field does not exist. He also discusses the Kasner solution, which I have treated in [24]. As noted in his title, this trivial solution involves "*a new paradigm in GR*".

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

Heaviside [25] extended Newtonian physics to include *field energy density* in flat space, and this is recognized as iteratively equivalent to Einstein's non-linear field equation. Feynman noted:

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

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

Part III Information Coding of Density as Geometry

Consider the transition from *physical-field-based* physics in flat space to *curved-space-time-geometry-based* physics? Despite the common perception that *Minkowski space-time* views space as *empty*, Einstein [26] came to realize that

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

Further, *physical fields are real and have energy*. Ohanian and Ruffini [27] "*The gravitational field may be regarded as the material medium sought by Newton; the field is material because it possesses energy density.*" Gravity is not viewed as *curved space-time*, but as a *field with energy density*. In fact, special relativity is based on pretending that gravity does not exist at all; and general relativity on pretending that gravitational energy does not exist "locally" – it can always be transformed away. Yet *coordinate systems*, curved or flat, *can have no effect on physical reality*. The *logical contradiction* that changes in coordinate system can make a physically real field vanish is hidden in the *Equivalence Principle* used to derive our premier theory of gravity: "*gravity as space-time curvature*".

Rather than view "information" as a physical entity [28], we treat information in its original *coding* perspective: there is *no energy density information encoded in flat space coordinates*. Coordinate systems can change with *absolutely no effect* on field energy density; and the energy distribution can change with absolutely no effect on the coordinate system. All information about the physical field is contained in the *energy density distribution*; *none in the flat space coordinates*!

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

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

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

Physical density-based reality is distinguished from *curved space-time-based reality* via inertia. In *Energy-time theory* the special relativity γ-factor applies to *inertial mass*. In *curved spacetime* real inertial force is replaced by the geodesic abstraction. However, Feynman, Weinberg, Padmanabhan and others insist that "*Curved space-time is not a necessary conception of gravity.*"

The primordial gravitational field is treated in *The Primordial Principle of Self-interaction*. Self-interaction equation $\vec{\nabla}\vec{\psi} = \vec{\psi}\vec{\psi}$ is used to derive Heavi-

side equations known to be equivalent under iterations to Einstein's field equations. A geometrical algebra multi-vector is $\vec{\psi} = \vec{G} + i\vec{C}$ where $\vec{G}$ is the gravitational field vector and $\vec{C}$ is the gravitomagnetic bivector and duality operator i converts $\vec{C}$ to $\vec{C}$. In the absence of circulation bivector $\vec{C}$ the solution to $\vec{\nabla}\vec{\psi} = \vec{\psi}\vec{\psi} \Rightarrow \vec{\nabla}\vec{G} = \vec{G}\vec{G}$ is $\vec{G} = 1/\vec{r}\left(\equiv \vec{r}/r^2\right)$. Hence $\vec{G}\vec{G} = 1/r^2$ is the energy density of the primordial field. The mass of the field inside a sphere of radius r is (modulo $4\pi/3$)

$$m \cong r^3(\rho) = r^3\left(\vec{G}\vec{G}\right) = r^3\left(\frac{1}{r^2}\right) = r \Rightarrow \frac{m}{r} \cong 1 \tag{22}$$

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

$$U(r) = W_{\infty r} = \int_{\infty}^{r} \vec{F}\cdot\mathrm{d}\vec{r} = \int \frac{gMm}{r^2}\mathrm{d}r = gMm\left[-\frac{1}{r}\right]_{\infty}^{r} = -\frac{gMm}{r} = m\phi \tag{23}$$

The gravitational energy per unit mass, ϕ describes potential energy at the point in question; it is interaction energy $m\phi$ that would exist if m units of mass were to be introduced at that point. This *test mass m* is presumed so small that its own gravitational potential need not be considered. Next add mass M to space and recall that the only point common to both flat space and curved space is the origin $(0,0,0)$, assumed coincident. The massive body M does not exist at a point but assume it is symmetrical and apply Birkhoff's *Shell theorem* to reduce M to an equivalent point mass. The gravitational energy (per unit mass) r distant from the origin is $\phi = -M/r$. The total gravitational energy is the sum of the primordial gravitational field plus the potential due to mass M located at the origin

$$\frac{m}{r} + \frac{M}{r} \Rightarrow 1 + \phi. \tag{24}$$

Unit cube $\mathrm{d}x\mathrm{d}y\mathrm{d}z$ with $\mathrm{d}x = 1, \mathrm{d}y = 1, \mathrm{d}z = 1$ contains energy density $1+\phi$. The negative sign of the potential is suppressed, but as seen in Equation (30) the term of interest is always positive, whether $(1+\phi)$ or $-(1+\phi)$. We normalize this density $1+\phi \to 1$ by choosing new variable coordinates $\mathrm{d}x' \neq 1, \mathrm{d}y' \neq 1, \mathrm{d}z' \neq 1$. Based on the equivalence principle one might assume that $1+\phi \to 0$. But local density that *cannot be defined* is not the same as local energy density disappearing; but it ceases to provide *physical information*. We normalize the energy in every volume element such that $\mathrm{d}x'\mathrm{d}y'\mathrm{d}z'$ always contains unit energy as seen on the right side of **Figure 1**.

In this way information has been transformed from the energy density of the field ϕ, in constant (informationless) coordinates of flat space, to the information

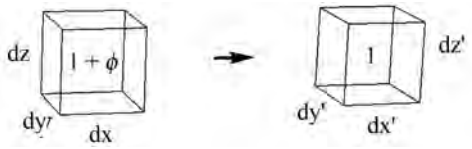


Figure 1. The primordial gravitational field has $\phi_0 = m/r = r/r = 1$ in flat space, $\mathrm{d}x, \mathrm{d}y, \mathrm{d}z = 1$. Mass M at the origin adds potential energy $\phi = M/r$. We transform to a unit density (no density information) in curved spacetime as shown in the cube at right, $\mathrm{d}x', \mathrm{d}y', \mathrm{d}z' \neq 1$.

of the (variable) coordinates containing a constant (informationless) energy density in curved space-time; consistent with the fundamental requirement that coordinate systems cannot affect physical reality. They label physical reality such that we all agree upon the points under discussion. The value 1 represents apparent energy density in curved space instead of the value *zero* seemingly implied by the equivalence principle. Thus, it is not required that energy density vanish, only that *information about the energy density must vanish*, in accordance with the statement that "*gravitational energy is not localizable*". So, we work not to get rid of physical energy density; but to get rid of energy density *information*. Energy density is normalized such that every local region in curved space contains exactly one unit of energy in a region bounded by $\mathrm{d}x', \mathrm{d}y', \mathrm{d}z' \neq 1$. Defining $\mathrm{d}x_i'$ to accomplish this goal transfers physical density information in flat space to curvature information of curved space. Every flat space differential has unit length; no information contained in a normalized unit of the coordinate system. In curved space, metric intervals are defined to normalize energy density of the field, removing information from the energy density, transferring it to the curved space metric.

6. Analyzing the Encoding Scheme

"Gravitational energy is not localizable" if *every local region* (in curved space-time) *contains a unit of energy indistinguishable from any other local region*, whereas the local metric at any point is distinguishable from the metric at any neighboring point. This is captured by the generalized Pythagorean for curved space-time:

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu . \tag{25}$$

Metric $g_{\mu\nu}$ transforms energy density information contained in ϕ into the local coordinate information contained in $g_{\mu\nu}$. Whereas d*x*d*y*d*z* coordinate differentials are independent of ϕ, the transformed coordinates are totally dependent on ϕ, *i.e.*,

$$\mathrm{d}x'\mathrm{d}y'\mathrm{d}z' \equiv \mathrm{d}x'(\phi)\mathrm{d}y'(\phi)\mathrm{d}z'(\phi) . \tag{26}$$

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

$$(1+\phi)\mathrm{d}x_i'(\phi) = \mathrm{d}x_i \tag{27}$$

in which case the key coordinate variable relation becomes

$$\mathrm{d}x_i'(\phi)=\frac{\mathrm{d}x_i}{1+\phi} \tag{28}$$

The consequent volume element becomes

$$\mathrm{d}x_1'(\phi)\mathrm{d}x_2'(\phi)\mathrm{d}x_3'(\phi)\equiv\frac{\mathrm{d}x_1}{1+\phi}\frac{\mathrm{d}x_2}{1+\phi}\frac{\mathrm{d}x_3}{1+\phi}. \tag{29}$$

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

$$\mathrm{d}x_i'(\phi)\mathrm{d}x_j'(\phi)\equiv\frac{\mathrm{d}x_i}{1+\phi}\frac{\mathrm{d}x_j}{1+\phi}. \tag{30}$$

In 3-space this is effectively a *time-slice* of Euclidian reality, $\mathrm{d}t\equiv 0$; using convention $i,j\in\{1,2,3\}$, $\mu,\nu\in\{0,1,2,3\}$. Four dimensional spacetime contributes to the difficulty of reconciling Einstein's theory with intuitive ideas of time and space. We consider 3-space energy density so we begin with the Pythagorean relation $\mathrm{d}s^2=g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu$. From Equations (25) and (30) we see that

$$g_{ii}=\frac{1}{(1+\phi)^2}=(1+\phi)^{-2}. \tag{31}$$

Consider Wheeler's remarks about a white dwarf star: "*It is small, but not terribly small; dense, but not terribly dense. Space-time is 'flat' within it…*" The term "flat" here means ϕ is small, approaching zero, allowing use of the binomial expansion $(1+\phi)^{-2}\cong 1-2\phi$. Yilmaz, [29] derives a metric proportional to $\mathrm{e}^{-2\phi}$ compatible with consensus GR for which $g_{ij}=1-2\phi$. This approximation is consistent with linearized equation $g_{\mu\nu}=\eta_{\mu\nu}+h_{\mu\nu}$ where $\eta_{\mu\nu}$ is the Minkowski metric with signature (+, −, −, −) and $h_{\mu\nu}=-2\phi$. Limiting ourselves to 3-space $(\mathrm{d}t\equiv 0)$ we obtain

$$\mathrm{d}s^2=\frac{-1}{1+2\phi}\left(\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2\right). \tag{32}$$

since the binomial expansion also yields $\frac{1}{(1+\phi)^2}=\frac{1}{1+2\phi}$. Factor h_{ii} represents diagonals term $\mathrm{d}x_i\mathrm{d}x_i$. In *A Primordial Space-Time Metric* the gravitomagnetic field of the dense region of field moving in the *z*-direction satisfies $h_{xy}-h_{yx}=0$. This angular momentum relation causes d*x*d*y* and d*y*d*x* terms to cancel, yielding Equation (32). The exact solution we've been considering is based on the mass *M* located at the origin where spherical symmetry implies $h_{xy}=h_{yx}=0$ (and cyclical iterations); also, compatible with Equation (32). The alternative form of Equation (32) yields $\mathrm{d}s^2=-(1-2\phi)\left(\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2\right)$. Observe that since $\phi=-gM/r$ we let $g=1$ and then rewrite the equation as

$$\mathrm{d}s^2=-\left(1-\frac{2M}{r}\right)^{-1}\left(\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2\right) \tag{33}$$

Inspection shows that Equation (33) is identical to the $\mathrm{d}r^2$ term in Schwarzschild metric Equation (8) when the Minkowski signatures are considered. GR textbooks can be consulted for derivation of the Schwarzschild metric surrounding mass *M* at the origin. The normalizing encoding process derives *ex-*

actly the three-space metric representing gravitational energy density equivalent information. Rindler explains the Schwarzschild metric spatial geometry in understandable terms:

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

$$\mathrm{d}l^2 = \frac{\mathrm{d}r^2}{1+2\phi} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\varphi^2\right) \tag{34}$$

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

In (34), we consider a *radial ruler*, [in which case $\mathrm{d}\theta = \mathrm{d}\varphi = 0$] and note that the rulers shrink by factor $(1+2\phi)^{1/2} \approx 1+\phi$; *exactly* the result for $\mathrm{d}x, \mathrm{d}y$ and $\mathrm{d}z$ that we achieved in Equation (28) by transforming from actual physical energy density in flat space (constant rulers) to the equivalent information encoded as geometry of curved space (variable metric). QED.

By comprehending that we are *encoding energy-density as geometry* [30] and drawing a few simple diagrams we transform from flat space energy density (physical reality) to curved space with almost no computation and obtain the standard Schwarzschild space metric. What about the *time-time* metric, g_{00}?

7. The Time Metric Associated with Curved 3-Space

Gravitational energy density *information* is banished in curved space by the equivalence principle, but physical information associated with the density of the field cannot be simply discarded—it must be transformed into the geometry of the variable metric coordinate systems. In this case the non-trivial solution represented by the Schwarzschild metric. The assumption of absolute space and time underlying the physics of inertial clocks is based on defining absolute time as *universal simultaneity*, hence our curved space solution holds at any time (slice) t with $\mathrm{d}t \equiv 0$. Thus the time dimension dropped out of Schwarzschild's metric solution reducing the problem to that of curved 3-space as derived in Section 6. Recall that the Minkowski invariant defined by Equation (2), $\mathrm{d}s^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 = 0$ is *invariant* because it does not vary with coordinate systems, hence $\mathrm{d}s'^2 = c^2\mathrm{d}t'^2 - \mathrm{d}x'^2 - \mathrm{d}y'^2 - \mathrm{d}z'^2 = 0$ and hence $\mathrm{d}s^2 = 0 = \mathrm{d}t^2 - \mathrm{d}r^2 = \mathrm{d}t'^2 - \mathrm{d}r'^2$. Equation (28) expresses coordinate differentials as $\mathrm{d}x_i$ and $\mathrm{d}x_i'$ therefore we use $\mathrm{d}s^2 = g^{\mu\nu}\mathrm{d}x_\mu \mathrm{d}x_\nu$. Hence, to obtain an expression for the time-time metric we use

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

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

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

Therefore, from Equation (35) we find $\frac{\mathrm{d}r^2}{\mathrm{d}t^2} = g^{00}(1+2\phi) = c^2$. The inverse

metric $g_{00}=1/g^{00}$ and for $c^2=1$ we obtain $g^{00}g_{00}=1$ and $g^{00}(1+2\phi)=1$. Therefore

$$g_{00}=1+2\phi \tag{37}$$

which is found to be in *exact agreement* with the Schwarzschild time-time metric, thus our flat space distribution of field energy density, $\phi(\vec{r})$ encodes energy density information as geo(metric) information of curved space and the corresponding time (the conjugate of energy).

8. Quasi-Local Mass Density in Encoded Geometry

Having derived exactly the Schwarzschild metric solution to Einstein's geometric formulation, in terms of the Heaviside equations in Euclidean space, we apply this to the interpretation of quasi-local energy distribution. The energy exists, but the density information has been normalized and now exists in the curvature metric: every $\mathrm{d}x'\mathrm{d}y'\mathrm{d}z'$ cell in the geometric formulation contains one unit of gravitational field energy; thus, gravitational fields in flat space, with physically real energy density, can be formulated in curved space with normalized field energy. All density information of the field has been transferred to the coordinate system metric, such that energy is normalized in *every* volume element in curved space. Schwarzschild's solution to Einstein's equations literally falls out of this approach, implying that Feynman, Weinberg, and others are correct in their claims that curved space-time is not a necessary conception of gravity.

Heaviside's equations derive straightforwardly from a primordial self-interaction principle, which, unlike the approximate equivalence principle, is an *absolute* principle of physics—Heaviside is not a *weak field approximation* but instead is valid for *all* field strengths. Misunderstanding this has led physicists to view Einstein's curved space formulation as the *true* theory of gravity, and flat space equations as only valid when fields are weak. The two are equivalent formulations but it is the misnomer "weak field approximation" that sticks in physicist's minds.

We conclude from this treatment that reality is Euclidian, not Riemannian. Einstein adopted a nonphysical geometric approach, even though physical gravity in flat space connects geometrical gravity and curved spaces *only at one point*, the shared origin of the two coordinate systems. It is not trivial that Feynman claims "*we have no idea how to go about writing down a proper* $T^{\mu\nu}$. Equally significant is the claim by Will and Poisson that Heaviside-based *weak field* methods work amazingly well in strong fields.

We are thus led to a re-interpretation of quasi-local energy density. Energy density physically exists (how could it not?) but information contained in the physical field is transferred to variable coordinates as described. Local volumes of space are defined such that every volume contains the same energy as every other volume, where the volumes are defined as $\mathrm{d}x'\mathrm{d}y'\mathrm{d}z'$ related to flat space volume's $\mathrm{d}x\mathrm{d}y\mathrm{d}z=1$ by Equation (28). Each side of a cube in curved space is $1/(1+M/r)$ that of a corresponding cube in flat space, leading to the situation

shown schematically in Figure 2, which shows the density-based volumes at 10 different equally spaced values of *r*. This technique is chosen because it is reasonably easy to calculate and reasonably easy to visualize. An attempt to draw the configuration at every point in space would be far more complicated and would add nothing to the comprehension of the meaning of quasi-local energy density.

Earlier we observed that Yau's calculation of quasi-local mass in spherical symmetric space-time foliated by orbits of SU(2) [Equation (11)] compared to Misner and Sharp's [and Hawking's] definition of mass [Equation (12)] yields relation (13)

$$m = M - \frac{M^2}{2r},$$

such that at spacelike infinity $m = M$. We now rewrite this as

$$\frac{m}{M} = 1 + \frac{\phi}{2}, \tag{38}$$

recognizing that the equality at space-like infinity corresponds (asymptotically) to the flat space density, while on the apparent horizon he obtains $M = 2m$, we ignore the factor of two and consider

$$\frac{m}{M} = 1 + \phi. \tag{39}$$

Unsurprisingly we find that quasi-local mass is formulated such that the key geometric encoding term $(1+\phi)$ is prominent. For a concept that has resisted physical interpretation for 40 years, this is about as far as we wish to push things.

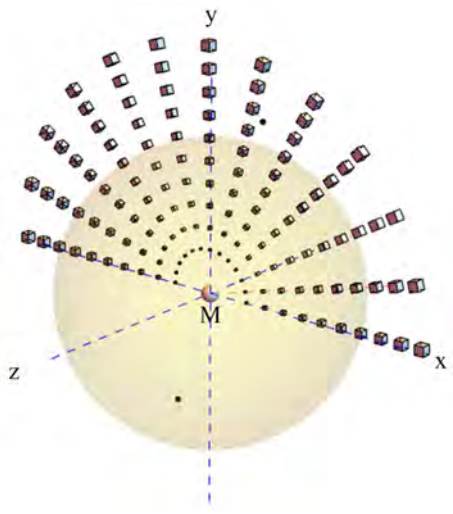


Figure 2. The graphical representation of quasi-local energy density based on the density information encoded geometrically. Every volume element in curved space holds exactly one unit of field energy. That is, the curved space volume elements are distributed in flat space to aid visualization. This approach yields *exactly* the Schwarzschild solution to Einstein's geometric formulation. One half of one slice through the spherical volume about mass *M* is shown, as the full 3D image would be too "busy" and add nothing to our understanding.

9. Conclusions

The *equivalence principle-based* banishment of local energy density is so intolerable that physicists and mathematicians proposed the idea that, while energy density did not exist locally, it could somehow be integrated inside an appropriately defined curved-space boundary. Penrose proposed this concept in 1982 but noted that *several problems of interpretation remain to be solved*. Thirty years later Yau closed his treatment saying he was still in process of deriving new properties of the quasi-local mass that he just introduced. A decade later Bart noted that *currently no framework exists in which a definition of quasi-local energy is sufficiently understood*. The approach taken by these three (and others) consisted essentially of attempting to formulate energy density as proportional to curvature, with emphasis on "quasi-local" density as the difference between local curvature and mean curvature. This was associated with "*asymptotic symmetries at null infinity*" to regain translational symmetry so that Noether's theorem could be applied, and meaningful Hamiltonians introduced. The conglomeration of concepts, none of which are physically meaningful, produced mathematical results that proved nothing, but led to hope that things could be worked out, if only we understood the problem a little better.

We here re-interpret gravity theory in such a way that the energy density of the field is physically real; its disappearance is explained by transferring density information from the field in information-less flat space coordinates to curvature information based on the Newtonian potential, with results equivalent to the Schwarzschild solution. **Figure 2** represents normalized curved space density displayed in flat space coordinates (corresponding to "null infinity"). It views each curved space volume element as the relevant boundary, inside of which a known quantity of field energy exists. The total energy is obtained by adding the relevant volume elements. The exact mathematical procedure to best accomplish this may be messy, but the physical understanding of gravitational field reality should be quite clear. Based upon our analysis of the gravitational field energy-momentum we concur with Feynman, Weinberg, and others: the concept of curved space-time is *not* necessary for gravity. Ontological arguments imply that both theories are not equally valid.

The *equivalence principle-based* disappearance of energy of the gravitational field when curved coordinate frames appear has confused general relativists since Einstein proposed his geometric approach. In *energy-time theory* the specification of field energy density in flat space determines the physics. If this information is removed, it must be replaced by some other source of information; the curved coordinates supply this information. The physical gravitational field does *not* vanish; its energy density is normalized: every corresponding volume element in curved space coordinates, $dx'dy'dz'$, contains exactly *one* unit of energy, yielding a metric *exactly* equal to Schwarzschild's solution. Gravitational field energy does *not* vanish in this approach; the normalization effects the transfer of *gravitational field information* to curved space coordinates.

Will and Poisson note that this *flat space* approach is useful for very strong fields, but they offer no reason for this fact. *The primordial principle of self-interaction* derives Heaviside's equations with no *weak field* assumptions; the equations work for *all* field strengths and represent a complete theory of gravity. This radically differs from the prevailing view that only curved space-time equations are complete.

The latest relevant paper dealing with quasi-local mass [31] illustrates the major issue: physicists have essentially given up on physically understanding reality in favor of mathematically elaborating upon well-established problems. It investigates the asymptotic structure of Einstein gravity in *five spacetime dimensions* (which no one understands, other than mathematically) by focusing on spatial infinity and using Hamiltonian techniques.

The implications of *the primordial principle of self-interaction* and of *encoding energy-density in geometry* are that curved-space formalisms are artifacts and gravitational field theory in Euclidean space represents physical reality. An ontological shift to a density-based formalism has significant consequences.

Conflicts of Interest

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

References

[1] Chen, C.-N., Nester, J.M. and Tung, R.-S. (2015) *International Journal of Modern Physics D*, **24**, Article ID: 1530026, arXiv:1507.07300v1. https://doi.org/10.1142/S0218271815300268

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

[3] Nozick, R. (2001) Invariance: The Structure of the Objective World. Belknap Press, Cambridge.

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

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

[6] Lucas, J.R. and Hodgson, P.E. (1990) Space-Time and Electromagnetics. Oxford University Press, Cambridge.

[7] Rindler, W. (1991) Introduction to Special Relativity. 2nd Edition, Oxford University Press, Cambridge.

[8] Cannoni, M. (2016) *International Journal of Modern Physics A*, **32**, Article ID: 1730002, arXiv:1605.00569v2. https://doi.org/10.1142/S0217751X17300022

[9] Klingman, E. (2018) Everything's Relative, or Is It. arXiv:1824.0424.

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

[11] Misner, Thorne, Wheeler (1973) Gravitation. Freeman, New York.

[12] Beckwith, A. (2014) Advances in High Energy Physics, 2014, Article ID: 230713, arXiv:1404.7167. https://doi.org/10.1155/2014/230713

[13] Yau, S.-T. (2011) Quasi-Local Mass in General Relativity. In: Mrowka, T. and Yau, S.-T., Eds., *Surveys in Differential Geometry*, Vol. 15, International Press, Boston, 421-433. https://doi.org/10.4310/SDG.2010.v15.n1.a12

[14] Penrose, R. (1982) *Proceedings of the Royal Society of London. A*, **381**, 53-63. https://doi.org/10.1098/rspa.1982.0058

[15] Murray, M. (2006) Twistor Theory. http://maths.adelaide.edu.au/michael.murray/twistors.pdf

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

[17] Misner, C. and Sharp, D. (1964) *Physical Review*, **136**, B571-B576. https://doi.org/10.1103/PhysRev.136.B571

[18] Bartnik, R. (1989) *Physical Review Letters*, **62**, 2346-2348. https://doi.org/10.1103/PhysRevD.61.084027

[19] Bart, H, (2019) Quasi-Local Conserved Quantities in General Relativity. Dissertation, Loyola Marymount University, Los Angeles. https://edoc.ub.uni-muenchen.de/25814/

[20] Wald, R. and Zoupas, A. (2000) *Physical Review D*, **61**, Article ID: 084027. https://doi.org/10.1103/PhysRevD.61.084027

[21] Baggott, J. (2020) Quantum Reality. Oxford University Press, Oxford.

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

[23] Vishwakarma, R. (2013) *Open Astronomy Journal*, **6**, 14-19. arXiv:1206.2795v2.

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

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

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

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

[28] Verlinde, E., (2011) *Journal of High Energy Physics*, **2011**, Article No. 29. https://doi.org/10.1007/JHEP04(2011)029

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

[30] Klingman, E. (2021) *Journal of Modern Physics*, **12**, 1190-1209. https://doi.org/10.4236/jmp.2021.129073

[31] Fuentealba, O., Henneaux, M., Matulich, J. and Troessaert, C. (2022) *Physical Review Letters*, **128**, Article ID: 051103. https://doi.org/10.1103/PhysRevLett.128.051103

Journal of Modern Physics, 2022, 13, 368-384
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Qubit or Qudet: Projections onto Reality

Edwin Eugene Klingman

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

How to cite this paper: Klingman, E.E. (2022) Qubit or Qudet: Projections onto Reality. *Journal of Modern Physics*, **13**, 368-384.
https://doi.org/10.4236/jmp.2022.134026

Received: March 4, 2022
Accepted: April 8, 2022
Published: April 11, 2022

Abstract

2022 is the centennial of an event which many consider to be a basis from which quantum mechanics can be derived—the Stern-Gerlach experiment of 1922—despite that the meaning of quantum theory is today an open question. Key is "the measurement problem", the need to measure quantum phenomena with classical equipment while the boundary separating quantum from classical is unknown. The mechanism of the SG-experiment is analyzed, and the *Qubit* nature normally projected onto the data is traced to *quantization of the detector*, labelled a *Qudet*. This novel interpretation should have downstream consequences, such as the SG-based interpretation of Bell's Theorem.

Keywords

Qubit, Qudet, Stern-Gerlach, Inhomogeneous Field, Measurement Problem, Ontology, Quantum Mechanics

1. Introduction

The experiment performed by Stern and Gerlach in 1922 can be used to derive the *theory of quantum mechanics* [1] [2], but the experiment was performed a year before de Broglie proposed his particle-wave model, $p = \hbar/\lambda$, that is the basis of quantum theory. Therefore, it was *not* understood in the way quantum theory is understood today. It was performed three years before Goudsmit and Uhlenbeck, based on spectroscopic studies, proposed ½-integral spin existed. Goudsmit's interpretation that every spin component, $s_i = \pm 1$, became embedded in quantum mechanics as the *projection postulate*, now known as the *qubit*. However, in 1928 Gordon established that the magnetic moment of the electron is due to the circulating flow of charge in the electron wave field; a 3D flow compatible with the Maxwell-Hertz field equations of classical electrodynamics; the same electromagnetic field as is the basis of quantum electrodynamics

(QED).

There is, a century later, still no complete quantum theoretic treatment of Stern-Gerlach, complicated further by *Jarzynski's Equality* relating work on a non-equilibrium thermodynamics system to the free energy of the system [3] [4]. In these terms Deissler investigated the fundamental question of whether a magnetic field does work on an atom [5]. Work is not an observable in the standard sense [6]; it is not represented by a Hermitian operator, thus is not an ordinary quantum observable: the number of possible values of work $W = E_f - E_i$ is typically larger than the dimension of the space of states; hence a Hermitian operator representing work cannot exist [7].

The "qubit" nature of spin is neither derived nor proved; it is assumed. The radical nature of this assumption, deriving largely from Goudsmit's *Projection Postulate*, is re-analyzed. We begin with derivation of constant classical spin from Poisson brackets, which spin will serve as the basis of a work-based analysis of Stern-Gerlach, leading to an *Energy-Exchange Theorem* underlying the dynamics. It is shown that dynamic spin stability in an inhomogeneous field leads to beam-splitting only if a gradient threshold is exceeded, associating quantization with the detector, hence *Qudet*. This *Qudet* interpretation is then contrasted with the standard *Qubit* interpretation and implications for Goudsmit's postulate are discussed.

2. Derivation of Constant Classical Spin from Poisson Brackets

The current state of understanding of particle dynamics is that the spin that physicists focus on is generally *not observable*. If spin is a *rotation* characterizing a particle, the rotating charge on the particle conceptually forms a current loop that induces a local magnetic moment which is the basis of indirect observation of particle spin. A magnetic moment traversing a nonuniform field is typically treated as a particle in a field. The dynamic equations of the field, expressed in geometric algebra, are $\left(\nabla + \partial_t\right)\boldsymbol{F} = \boldsymbol{J}$ where $\boldsymbol{F} = \boldsymbol{E} + i\boldsymbol{B}$ and $\boldsymbol{J} = \rho + \boldsymbol{j}$, the single equation equivalent to the four Maxwell electrodynamic equations in standard form [8]. Particle dynamics are prescribed by Poisson brackets: $\dot{\boldsymbol{x}} = \{\boldsymbol{x}, H\}$ where $\boldsymbol{x} \in \{\boldsymbol{r}, \boldsymbol{p}, \boldsymbol{s}\}$ and H is the Hamiltonian: $H = \frac{p^2}{2m} - \boldsymbol{\mu}\cdot\boldsymbol{B}$, $\boldsymbol{\mu} = \mu_0\boldsymbol{s}$. We analyze inhomogeneous magnetic field $\boldsymbol{B}(\boldsymbol{r})$ in the standard Stern-Gerlach model (Griffiths, [9]) $\boldsymbol{B} = \left(B_0 + \alpha x\right)\hat{i} + \left(-\alpha y\right)\hat{j}$ and in the quadrupole form $\boldsymbol{B} = \left(y\right)\hat{i} + \left(x\right)\hat{j}$. *Only precession can be solved for exactly.* In $\{\boldsymbol{r}, \boldsymbol{p}, \boldsymbol{s}\}$ -representation the particle at position $\boldsymbol{r}$ has momentum $\boldsymbol{p}$ and spin $\boldsymbol{s}$ and magnetic moment proportional to spin $\boldsymbol{s}$: $\boldsymbol{\mu} = \mu_0\boldsymbol{s}$. Point-based classical physics led to the Poisson bracket formalism in which the time derivative of the physical parameters is given by the generic

$$\dot{\boldsymbol{x}} = \{\boldsymbol{x}, H\} \tag{1}$$

where $\boldsymbol{x}$ is the physical entity/aspect of reality and H is the Hamiltonian of the

system. As suggested by Kauffman we solve the equations deriving from the Hamiltonian in magnetic field $\boldsymbol{B}(\boldsymbol{r})$:

$$\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}=\{\boldsymbol{r},H\}\Rightarrow\frac{\boldsymbol{p}}{m}\quad\text{velocity predicts position at a moment in time} \tag{2}$$

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t}=\{\boldsymbol{p},H\}\Rightarrow-\mu_0\nabla\left(\boldsymbol{s}\cdot\boldsymbol{B}(\boldsymbol{r})\right)\quad \textit{momentum}\text{ force} \tag{3}$$

$$\frac{\mathrm{d}\boldsymbol{s}}{\mathrm{d}t}=\{\boldsymbol{s},H\}\Rightarrow-\mu_0\boldsymbol{B}(\boldsymbol{r})\times\boldsymbol{s}\quad \textit{precession}\text{ force (torque)} \tag{4}$$

Given the *specified quadrupole B-field*, $\boldsymbol{B}(\boldsymbol{r})=\alpha(y,x,0)$, $\beta=B_0/r_0$ we calculate

$$\boldsymbol{s}\cdot\boldsymbol{B}=\beta\left(s_x,s_y,s_z\right)\cdot(y,x,0)=\beta\left(ys_x+xs_y\right)$$

and

$$\nabla\left(\boldsymbol{s}\cdot\boldsymbol{B}\right)=\beta\left(\partial_x,\partial_y,\partial_z\right)\left(ys_x+xs_y\right)=\beta\left(s_y,s_x,0\right).$$

From the *Poisson brackets* we obtain the equation of motion:

$$\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2}=-\frac{\mu_0}{m}\nabla\left(\boldsymbol{s}\cdot\boldsymbol{B}\right)=-\frac{\mu_0}{m}\beta\left(s_y,s_x,0\right) \tag{5}$$

i.e.

$$\ddot{x}=-\frac{\mu_0}{m}\beta s_y,\quad \ddot{y}=-\frac{\mu_0}{m}\beta s_x,\quad \ddot{z}=0 \tag{6}$$

while the *first order spin equation* is

$$\dot{\boldsymbol{s}}=\boldsymbol{B}(\boldsymbol{r})\times\boldsymbol{s} \tag{7}$$

which, for the specified quadrupole field yields

$$\boldsymbol{B}\times\boldsymbol{s}=\begin{vmatrix}\hat{i}&\hat{j}&\hat{k}\\ y&x&0\\ s_x&s_y&s_z\end{vmatrix}=\left(xs_z,-ys_z,\left(ys_y-xs_x\right)\right). \tag{8}$$

The most salient fact that emerges follows from:

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\left(s_x^2+s_y^2+s_z^2\right)=2\left(s_x\dot{s}_x+s_y\dot{s}_y+s_z\dot{s}_z\right)\\&=2\mu_0\beta\left(xs_zs_x-ys_zs_y+ys_zs_y-xs_zs_x\right)\equiv0\end{aligned} \tag{9}$$

which implies that *the magnitude of spin is conserved*:

$$\frac{\mathrm{d}}{\mathrm{d}t}|\boldsymbol{s}|^2=0\Rightarrow|\boldsymbol{s}|^2=\text{constant} \tag{10}$$

This result derives from *classical* equations of motion, not *quantum*, although it implies that the spin is quantized. In addition, we see from the third spin precession Equation (6c) that s_z is conserved. To relate spin physics to reality, we consider the Stern-Gerlach experiment.

3. Physics of Stern-Gerlach and Analysis of Work

Angular momentum of particles at the quantum level is notoriously hard to meas-

ure. *Spinning charge* induces a magnetic dipole proportional to angular momentum; as this dipole interacts with magnetic fields, spin experiments are based on measuring magnetic dipoles as a surrogate for spin. Angular momentum $\boldsymbol{\lambda}$ of an atom is often called an observable, but is generally unobservable; a magnetic dipole moment associated with charged particle motion $\boldsymbol{\mu} = \mu_B \boldsymbol{\lambda}$ where μ_B is the Bohr magneton, 0.927×10^{-20} erg/gauss. Erg-per-gauss relates to energy of a dipole in a magnetic field, and this energy, often indirectly observable via photon emission and absorption, is the physical observable. To observe $\boldsymbol{\lambda}$, one must establish a reference frame, via a magnetic field, $\boldsymbol{B}$. Jared Stenson, focused on representation, [10] notes that: "…*the Stern-Gerlach experiment has become axiomatic in modern physics… it ties together classical and quantum systems of thought. It seems to explain a clearly quantum result in terms of almost purely classical concepts.*" The experiment sends neutral particles with a spin-based magnetic moment through magnetic fields. Two cases are significant: particles in a *constant field* precess—a spin $\boldsymbol{\lambda}$, moving in a constant field simply precesses and is not deflected. But the same particles moving through an inhomogeneous field are deflected by the gradient force $\boldsymbol{F} = \nabla(\boldsymbol{\mu} \cdot \boldsymbol{B})$ with $\boldsymbol{\mu}$, the magnetic moment (proportional to λ), represented by quantum operator $\hat{\sigma}$, and $\boldsymbol{B}(\boldsymbol{x})$ the magnetic field in direction $\boldsymbol{b}$. The data are from the iconic postcard Stern and Gerlach sent to Bohr announcing their discovery of the spin-dependent "splitting", shown in **Figure 1(b)**. Per Messiah [11]:

"*The appearance on the screen of a more or less spread-out distribution of impacts indicates that the atoms are not all in the same initial condition and that the dynamical variables defining the initial states are statistically distributed over a somewhat extended domain.*"

For vertical x and horizontal z, the x-component of the magnetic field is B_x at the south end of the dipole. The field component at the north end depends on the gradient and the projections of $\boldsymbol{\lambda}$ on respective axes. The force exerted on the magnet is the sum of forces acting on each end:

$$F_x = \hat{m}(\boldsymbol{\lambda} \cdot \nabla B_x) = \boldsymbol{\mu} \cdot \nabla B_x \Rightarrow \boldsymbol{F} = \nabla(\boldsymbol{\mu} \cdot \boldsymbol{B}). \tag{11}$$

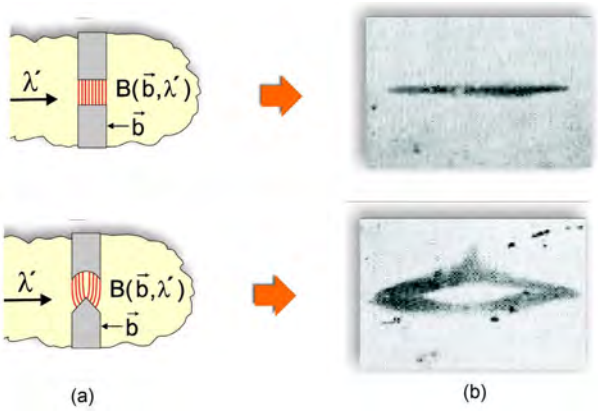


Figure 1. (a) Apparatus with no gradient. (b) Stern-Gerlach apparatus with gradient.

The force on the dipole is proportional to magnetic dipole moment, $\boldsymbol{\mu}$, the gradient of B_x and the cosine of the angle between $\boldsymbol{\mu}$ and ∇B_x. In addition to this force on the magnetic moment due to the inhomogeneous field, the moment precesses in a homogeneous field. Magnetic force on a charged particle is orthogonal to the path, and hence does no work, but this is not the case for an uncharged magnetic dipole. The field gradient force acting on the magnetic dipole is exerted over a finite distance, $F \cdot x$, yet no work is done on the system (KE = kinetic energy): Normally

$$\int_0^x \mathrm{d}(KE) = \text{'work'}, \tag{12}$$

but no work is done if precession energy (PE) is exchanged with (changed into) kinetic energy (KE) as described in the following *Energy-Exchange theorem*. Assume the particle is deflected from the z-axis ($x = 0$) to a distance x from the z-axis and precession energy *PE decreases* to compensate the increase in kinetic energy of translation until precession energy is exhausted at q.

$$\int_0^x \mathrm{d}(KE) + \int_0^q \mathrm{d}(PE) = \sum \Delta E = \text{'work'} \tag{13}$$

To show the energy exchange explicitly we break the kinetic energy integral from 0 to x into two integrals, from 0 to q and from q to x:

$$\int_0^x \mathrm{d}(KE) + \int_0^q \mathrm{d}(PE) = \underbrace{\int_0^q \mathrm{d}(KE) + \int_0^q \mathrm{d}(PE)}_{\text{NO WORK}} + \underbrace{\int_q^x \mathrm{d}(KE)}_{\text{WORK}} \equiv \Delta E \tag{14}$$

Work ($W = \Delta E$) is defined as the *change in energy*; we observe that no work is done from zero to q as the total energy change is zero. After the moment aligns with the local magnetic field at position q (where precession has ended) work is done from q to x. In this model no work is done *while the spin is in process of alignment* with the local field, but once aligned, the field performs work on the particle. The dynamics (which caused Bell to despair [12]) are even more complicated than Deissler assumed as there are now *two* phases: a phase in which no work is done by the field, due to internal compensation, followed by a phase in which the field performs work on the particle, as seen in **Figure 2**.

In an SG-apparatus, a magnetic moment traverses a non-uniform magnetic field, experiencing a gradient-based deflecting force. Uniform magnetic fields do *not* do work on particles, and inhomogeneous magnetic fields do not do work on a *spin-less atom*, but analysis of a quasi-particle defined by electron spin, local magnetic field, and precession energy shows that spin models of particles in an inhomogeneous field will *exchange energy between local energy modes* to compensate for changes in the local field. We are thus motivated to describe the energy transfers between eigenmodes, in which case an *Energy-Exchange theorem* is basic.

4. The Energy-Exchange Theorem

In a physical system possessing two energy modes, M_0, M_1, *coupled to a*

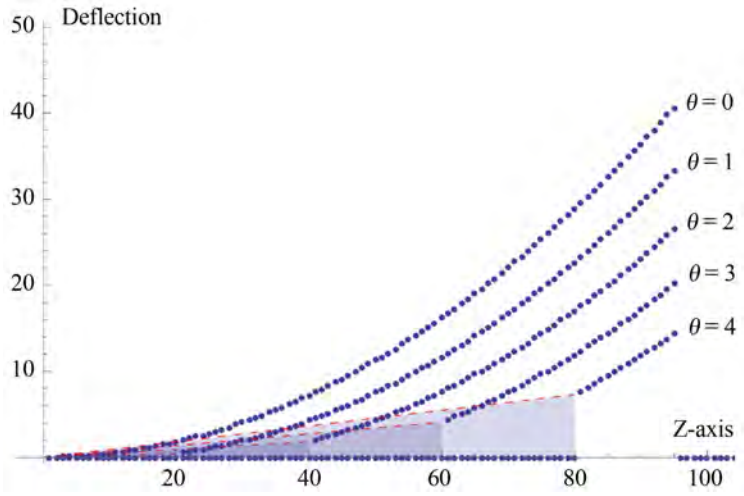


Figure 2. The θ-dependence for small angles is shown. Shaded areas below the dashed red lines represent the alignment process during which the initial spin enters the field with angle θ and proceeds to align with the local field. After alignment the gradient force accelerates all particles equally, seen as dotted blue lines. Scales have been chosen to enhance the different deflections. Actual scales depend upon field strength, field gradient, length of travel, and geometry.

common variable θ, *if energies of the modes are not separated by a quantum gap* $\Delta\varepsilon > 0$, *then if the common variable changes,* $\mathrm{d}\theta/\mathrm{d}t \neq 0$, *the modes will exchange energy.*

Assume that total energy is $\varepsilon = \varepsilon_0 + \varepsilon_1$ when $H_i|\psi\rangle = \varepsilon_i|\psi\rangle$ and that total energy $H = H_0 + H_1$ is conserved:

$$\begin{aligned} &\frac{\mathrm{d}H}{\mathrm{d}t} = 0 \Rightarrow \frac{\mathrm{d}H_0}{\mathrm{d}t} + \frac{\mathrm{d}H_1}{\mathrm{d}t} = 0 \Rightarrow \frac{\mathrm{d}H_0}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}t} + \frac{\mathrm{d}H_1}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}t} = 0 \\ &\Rightarrow \left(\frac{\mathrm{d}H_0}{\mathrm{d}\theta} + \frac{\mathrm{d}H_1}{\mathrm{d}\theta}\right)\frac{\mathrm{d}\theta}{\mathrm{d}t} = 0 \end{aligned} \tag{15}$$

Since $\mathrm{d}\theta/\mathrm{d}t \neq 0$ then $\frac{\mathrm{d}H_0}{\mathrm{d}\theta} = -\frac{\mathrm{d}H_1}{\mathrm{d}\theta}$ and energy flows between mode M_0 and M_1. QED

In the Stern-Gerlach experiment a neutral silver atom enters the inhomogeneous field with *z*-axis momentum and its intrinsic magnetic moment $\boldsymbol{\mu}$ at angle θ to the local B-field in the *x*-direction. The force of the *x*-directed field gradient $\boldsymbol{F} = \nabla(\boldsymbol{\mu}\cdot\boldsymbol{B})$ accelerates the particle in the *x*-direction while making no change to the initial *z*-momentum. The change in kinetic energy due to particle acceleration is compensated by the precession energy to conserve local energy: $E_{in} = E_{out}$

$$-\boldsymbol{\mu}\cdot\boldsymbol{B} = mv_x^2/2 - |\boldsymbol{\mu}||\boldsymbol{B}| \tag{16}$$

where *z*-momentum energy has been canceled for both sides. The change in translational kinetic energy, $mv_x^2/2$ is thus $mv_x^2/2 = \mu B(1-\cos\theta)$, where $\mu = |\boldsymbol{\mu}|$, $B = |\boldsymbol{B}|$. Applying the constant acceleration formula, $v_f^2 = v_i^2 + 2ax$, as an approximation, yields. $m(ax) \cong \mu B(1-\cos\theta)$ and hence an approximate θ

-dependent component of deflection

$$x = \frac{\mu B}{ma}\left(1-\cos\theta\right). \tag{17}$$

The distance x is the amount of deflection from the z-axis that a particle with initial angle of precession θ experiences when the particle becomes aligned with the local field. The relation $x = f(\theta)$ is such that deflection x is determined by *initial* angle θ as well as field strength and field gradient. The *Energy-Exchange Principle* constrains local dynamics until locally available energy is exhausted. A quasi-particle with internal degrees of freedom traversing a non-uniform field locally conserves energy over its N-degrees-of-freedom. A precessing, translating, magnetic moment has two such degrees of freedom, but finite compensation mechanisms, when exhausted, can no longer accommodate changes in the local field. At such time the local gradient begins delivering power. So, a local field gradient will not impart energy as work to the quasi-particle, either atom or precessing spin, *until* precession energy has been converted to deflection energy. If precession energy vanishes ($\theta = 0$) the local gradient drives the translation of the particle.

Rate of change of precession

An unanswered question concerning a dipole in an inhomogeneous magnetic field is the rate at which the angle of precession changes; how long it takes for the angle of precession to reach zero, *i.e.*, for the dipole to align with the local field. If it's not instantaneous, the rate is a matter of geometry and field strength and strength of gradient. If it were instantaneous, the entire travel through the Stern-Gerlach device would experience the maximum force and the deflection would be to one spot, regardless of the initial angle; corresponding to Bell's model, deflection measurements would yield ±1. A finite decay time is angle dependent and the spread of deflection is a function of the initial angle. Absent means of calculating the decay rate of the precession angle to determine the time of alignment, the question *is* answerable experimentally; the original Stern-Gerlach data on the iconic postcard exhibits the spread of deflections expected from energy exchange. From *the Energy-Exchange theorem* we conclude that precession energy is exchanged with kinetic energy of translation; the particle is accelerated by force $\boldsymbol{F} = \nabla(\boldsymbol{\mu}\cdot\boldsymbol{B})$. While the rate of precession may be independent of the particle mass, the rate of acceleration is not. Since $\boldsymbol{F} = m\boldsymbol{a} = \nabla(\boldsymbol{\mu}\cdot\boldsymbol{B})$ and $\boldsymbol{a} = \nabla(\boldsymbol{\mu}\cdot\boldsymbol{B})/m$ acceleration is a function of the gradient of the B-field, the angle of precession, and the mass of the particle. Based on these dependencies we take the derivatives of $mv^2/2 = \mu B(1-\cos\theta)$. Maximum energy of translation occurs when $\theta = 0$, corresponding to maximum velocity v, thus maximum velocities occur for small angles θ. From $\cos\theta \approx 1-\theta^2/2+\cdots$ in (17):

$$mv^2/2 = \mu B\left(1-\cos\theta\right) \xrightarrow{\theta\to 0} \underbrace{mv^2}_{\text{FINI}} = \underbrace{\mu B\theta^2}_{\text{INIT}}. \tag{18}$$

With $k = \sqrt{\mu/m}$ we obtain $v \approx k\sqrt{B}\theta$ hence $\dfrac{\mathrm{d}\theta}{\mathrm{d}t} \approx \dfrac{1}{k\sqrt{B}}\dfrac{\mathrm{d}v}{\mathrm{d}t}$ where from

Equation (11) $\frac{\mathrm{d}v}{\mathrm{d}t} \sim \frac{\mu}{m}\frac{\partial B}{\partial x}$, so:

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} \approx \frac{k}{\sqrt{B}}\frac{\partial B}{\partial x} \quad . \tag{19}$$

The Asymmetric Approximation

A finite decay rate implies a θ-dependent spread of SG deflections, in contrast to the single data point expected if alignment were instantaneous. The asymmetric treatment displays the second-order effects of the stronger and weaker regions of gradient. The combination of these classical effects is shown in Figure 3 overlapping the real Stern-Gerlach data from the iconic postcard, on which we have overlaid the gradient-producing magnetic field geometry. We consider only $\partial B/\partial x$, *i.e.*, vertical deflection; incompatible with Maxwell's $\nabla \cdot \boldsymbol{B} = 0$. Inclusion of a y-axis term provides the left/right displacements shown in the data but adds no new physics insight to the model.

A typical asymmetric Stern-Gerlach x-deflection for random initial angle θ is next calculated by rewriting Equation (17) to accommodate the asymmetry via β and $\beta/3$:

$$x = +1 + \beta\cos\theta \qquad \theta < \pi/2 \tag{20}$$

$$x = -1 - (\beta/3)\cos\theta \qquad \theta > \pi/2 \tag{21}$$

The scaling value $\beta = 4$ is chosen to yield a best match to the iconic postcard data. Actual deflection depends on the strength of the field, the strength of the gradient, the initial angle, the velocity, the length of the SG-magnet and the distance from the magnet to the detection screen. We effectively normalize the deflection by choosing the deflection due to the magnet-to-screen travel to be +1 or −1. The values can be approximated via appropriate strengths and geometries. If we vary the spin angle θ randomly and plot the vertical deflections based on Equations (20) and (21) we obtain the asymmetric distribution; the (red) calculated random data points are scaled and overlaid on the SG iconic data. The match between calculated data and experimental data is extremely good.

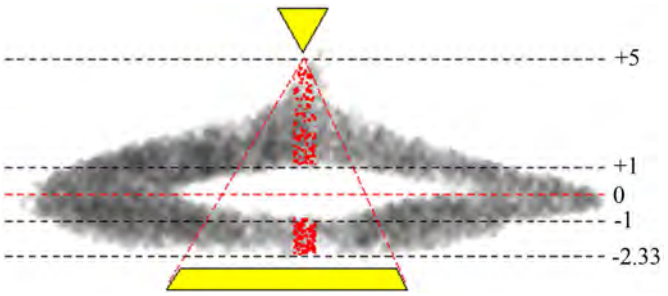


Figure 3. The θ-based model of spin deflection in the Stern-Gerlach experiment is driven with random spin vectors and trajectories are calculated. The red dots representing individual particles are overlaid on the gray SG data from the iconic postcard. The scale has been chosen to facilitate the overlay. The Stern-Gerlach magnets are diagrammed in yellow. Horizontal spreading of the red data points is for illustrative clarification and is not physically meaningful.

5. Analysis of Dynamic Spin Stability

To address Bell's statement about "*compass needles pointing in the wrong direction*" being "*not dynamically sound*" requires making use of Deissler's observation about spin and precessional rotation and requires an asymmetry foreign to Bell's analysis, best understood by comparison with the assumptions of first-order energy exchange, namely

$$\begin{array}{ll} \text{Symmetric} & \text{Asymmetric} \\ E_{in} = E_{out} & E_{in} = E_{out} \\ B_{in} = B_{out} & B_{out} = B_{in} + \Delta B \end{array} \tag{22}$$

First-order approximation energy-exchange model with instantaneous decay of precession leads to Bell-like measurement spectrum of ±1. Our finite decay model leads to a symmetric spread spectrum. A second-order gradient-based approximation yields an asymmetric energy-exchange model. To develop the asymmetric model, we explicitly consider "spin-up" and "spin-down" cases.

Spin down $\boldsymbol{\theta \le \pi/2}$

Dipole moment $\boldsymbol{\mu}$ is the product of electronic charge and spin, so the negative electron implies $\boldsymbol{\mu = -s}$ as shown in Figure 4 and spin-down precession (CCW) is opposite the rotation of the spin (CW). As the force on the dipole accelerates the particle to the region of stronger B-field, $B \to B + \Delta B$, the (CCW) precession frequency increases, opposing the intrinsic CW spin and decreasing kinetic rotational energy:

$$\left\{\begin{matrix}\text{cw}\\ \text{spin}\end{matrix}\right\} + \left\{\begin{matrix}\text{ccw} + \Delta\text{ccw}\\ \text{precession}\end{matrix}\right\} = \left\{\begin{matrix}\text{rot. energy}\\ \Downarrow\end{matrix}\right\}$$

As $\boldsymbol{B}$ increases, the $-\boldsymbol{\mu}\cdot\boldsymbol{B}$ energy becomes more negative. If energy is conserved, translational kinetic energy $mv_x^2/2$ becomes more positive as the particle is accelerated upward. This agrees with the decrease in rotational kinetic energy as shown. We next consider the opposite spin.

Spin up $\boldsymbol{\theta \ge \pi/2}$

A particle with spin up is accelerated into a region of weaker local magnetic field so the precession frequency (energy) decreases. Since the (CCW) spin and (CCW) precession are in the same direction, as shown in Figure 5, the net rotational energy decreases:

$$\left\{\begin{matrix}\text{ccw}\\ \text{spin}\end{matrix}\right\} + \left\{\begin{matrix}\text{ccw} - \Delta\text{ccw}\\ \text{precession}\end{matrix}\right\} = \left\{\begin{matrix}\text{rot. energy}\\ \Downarrow\end{matrix}\right\}$$

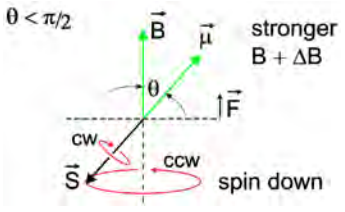


Figure 4. Spin $\boldsymbol{s}$ and magnetic moment $\boldsymbol{\mu}$ in inhomogeneous magnetic field $\boldsymbol{B}$ with spin down.

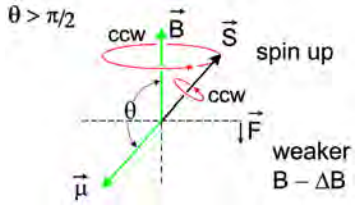


Figure 5. Spin $\boldsymbol{s}$ and magnetic moment $\boldsymbol{\mu}$ in inhomogeneous magnetic field $\boldsymbol{B}$ with spin up.

Thus as $\boldsymbol{B}$ decreases the $\boldsymbol{\mu}\cdot\boldsymbol{B}$ energy becomes smaller (less positive) and the translational energy $mv_x^2/2$ grows as the rotational energy decreases. Hence for all angles of the initial dipole with respect to the magnetic field

$$\begin{matrix}\left\{\begin{matrix} I\omega^2/2 \\ \Downarrow \end{matrix}\right\} + \left\{\begin{matrix} mv_x^2/2 \\ \Uparrow \end{matrix}\right\} = \begin{matrix}\text{energy} \\ \text{balance}\end{matrix} \\ \text{rotational} \quad \text{translational}\end{matrix}$$

With this detailed analysis of the spin dynamics we now reanalyze the energy balance equation $E_{in} = E_{out}$. Since there is no change in the *z*-component of velocity, we know that it cancels and we also assume that the incoming *x*-velocity is zero, thus we retain our energy exchange equation $-\boldsymbol{\mu}\cdot\boldsymbol{B}_{in} = mv_x^2/2 - |\boldsymbol{\mu}||\boldsymbol{B}_{out}|$ adapted from Equation (6). We again consider two cases $\theta \le \pi/2$ and $\theta \ge \pi/2$. Our initial energy exchange analysis explicitly conserved energy, $E_{in} = E_{out}$, but simplistically assumed $B_{in} = B_{out}$. A better approximation, $B_{out} = B_{in} \pm \Delta B$ takes a necessary gradient into account. We can now explore the energy exchange spin dynamics in an explicit gradient formulation. It is again necessary to treat the spin up and spin down cases separately, as the $\cos\theta$ term yields a sign change between these two cases.

For $\theta < \pi/2$ the dipole experiences a positive force toward a region where the field is stronger. The initial angle between the spin and the local field is θ and the final aligned state has zero angle. Begin with local conservation of energy: $E_{in} = E_{out}$ and $-\boldsymbol{\mu}\cdot\boldsymbol{B}_{in} = mv_x^2/2 - |\mu||B_{out}|$. Let

$$B_{out} = B_{in} + \Delta B \quad \text{where} \quad \Delta B = \int_0^x \frac{\mathrm{d}B}{\mathrm{d}x}\mathrm{d}x \tag{23}$$

$$mv_x^2/2 = |\mu||B_{out}| - \boldsymbol{\mu}\cdot\boldsymbol{B}_{in} = |\mu|(B_{in} + \Delta B) - \boldsymbol{\mu}\cdot\boldsymbol{B}_{in} \tag{24}$$

$$mv_x^2/2 = |\mu||B_{in}|(1-\cos\theta) + \mu\Delta B\,. \tag{25}$$

All terms are positive and thus the kinetic energy is positive as required. The $\mu\Delta B$ additive energy represents greater kinetic energy (and correspondingly greater deflection) than the simpler symmetric approximation.

For $\theta > \pi/2$ the dipole experiences a force toward a weaker region of the field. Here the cosine is negative, so $\boldsymbol{\mu}\cdot\boldsymbol{B}$ is negative and energy $-\boldsymbol{\mu}\cdot\boldsymbol{B}$ is therefore positive. Again $E_{in} = E_{out}$ but now $-\boldsymbol{\mu}\cdot\boldsymbol{B}_{in} = mv_x^2/2 - |\boldsymbol{\mu}||B_{out}|(\cos\pi)$. Taking signs into account, this becomes:

$$|\boldsymbol{\mu}||\boldsymbol{B}_{in}||\cos\theta| = mv_x^2/2 + |\mu||B_{out}| \tag{26}$$

$$mv_x^2/2 = |\boldsymbol{\mu}||\boldsymbol{B}_{in}||\cos\theta| - |\mu||B_{out}| \tag{27}$$

But, due to the gradient, $B_{out} < B_{in}$ so $B_{out} = B_{in} - \Delta B$ and thus,

$$\begin{aligned} mv_x^2/2 &= |\boldsymbol{\mu}||\boldsymbol{B}_{in}|\left(|\cos\theta| - 1\right) + \mu\Delta B \\ \Rightarrow \Delta B &> \left|B\left(|\cos\theta| - 1\right)\right| \end{aligned} \tag{28}$$

The term $\left(|\cos\theta| - 1\right)$ is *always* negative, so the requirement of positive kinetic energy of translation $mv_x^2/2$ demands the $\mu\Delta B$ term be sufficiently positive to overcome the negative term. Thus Equation (28) implies a threshold gradient, *below which the Stern-Gerlach apparatus will not work*. Equations (28) show the (integral of) the gradient to be a function of θ and of translational velocity *v*:

$$\Delta B = B\left(1 - |\cos\theta|\right) + \frac{1}{2}\left(\frac{v^2}{k^2}\right) \Rightarrow \frac{mv^2}{2} > 0 \tag{29}$$

The above analysis enables *a new physical conclusion*, which is the existence of a velocity-dependent threshold of inhomogeneity required for Stern-Gerlach apparatus. Below this threshold the atomic beam will not split, in accord with Bell's statement that "the compass needles" would be pointing in the wrong direction and hence "not dynamically sound". Below threshold the apparatus only detects one beam distribution for random θ input atoms. It becomes a *quantum detector* or *Qudet* only when the gradient threshold is exceeded. Above the threshold the classical atomic beam is quantized or split into spin-dependent beams.

6. The Qubit Interpretation of Stern-Gerlach

The above analysis of Stern-Gerlach in terms of magnetic moments traversing inhomogeneous magnetic fields forces particles with random spin orientation to align with the 3D field and makes intuitive physical sense. The approximate equations derived from the analysis produce an output distribution that readily maps onto the iconic SG "post card" results. We compare this interpretation of SG with the quantum interpretation, yet according to *Quantum Theory at the Crossroads*, [13]:

"*The interpretation of quantum theory is probably as controversial as it has ever been … the meaning of quantum theory is today an open question.*"

We adopt the Stanford interpretation of the meaning of quantum mechanics presented by *Stanford Physics Department head*, Leonard Susskind, in his video lecture. Texts by Feynman, Sakurai, and Townsend say that spin in a Stern-Gerlach experiment leads to the quantum formalism. For Susskind *qubits* are fundamental and spin is incidental; he introduces *a "qubit" or quantum bit rather than physical spin and defines it as a two-state system or "a bit"*. In his qubit-based intro to quantum mechanics [14] he claims we should not expect to "understand" the qubit in any classical sense: At *minute 11*, *second 17*, written [11:17], he states:

"*Quantum mechanics is about things that your evolutionarily developed neural structure is not prepared to deal with.*"

At [21:00] "*the simplest possible system in quantum mechanics… I'm going to call it a qubit—a system that can have two states.*" It has a mathematical degree of freedom, σ, and an equivalent arrow symbolism shown in **Figure 6**.

At [24:05] "*in addition to the system, it has an apparatus. The apparatus is a black box oriented with a "This side up" label, containing a window in which measurement data is displayed (+1 or −1) connected to a detector that senses the qubit. In QM, we cannot ignore the apparatus, but we don't care how it works. The experiment is to measure the state of the qubit*: ↑ or ↓."

But from the preceding sections it is clear that "*we should care how it works.*" The physics of the Stern-Gerlach *Qudet* experiment is the theory of "*how it works*" and teaches us that any random spin measured by the system will align itself with the *Qudet*, or *quantized detector*.

"*A device that measures is also a device to prepare a system in a given state.*" [29:10] After measurements confirm that the state is ↑ we turn the detector "upside down" and we find the state ↓. We can do so repeatedly. [32:00] "*by turning the detector over, we get the opposite answer.*" What is this telling us about the nature of the system? "… *we're learning that whatever the system is, this particular qubit has a sense of directionality to it. … The qubit has spatial directionality associated with it. … a sense of orientation … We might begin to suspect that it's a vector.*"

Thus, the idea is that we can, as stated at [34:00], "*detect the value of a component of a (spin) vector along the (vertical) direction*" and then [by rotating the detector] "*measure not the vertical component of the vector but the horizontal component of the vector*". But the idea that a *constant* spin vector with components s_x, s_y and s_z can be measured sequentially and components s_x, s_y and s_z determined is *counterfactual to reality*; the SG detector requires a non-uniform magnetic field or else the result of the experiment is null; and in a non-uniform magnetic field all three spin components s_x, s_y and s_z will *change*! The SG physics is such that spin will align *with* or *against* the local B-field, thus yielding ±1 results *regardless* of the original orientation of the spin.

Key to the qubit interpretation is the implication that the same *component* can be detected multiple times. If the component in the x-direction is ↑, then

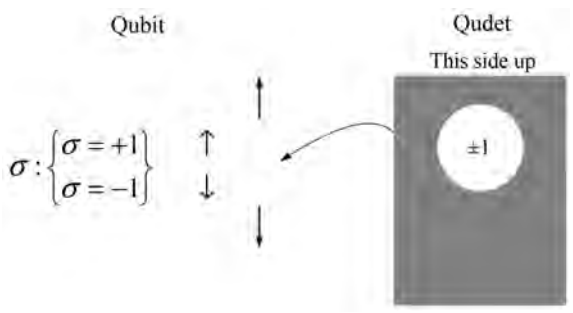


Figure 6. Schematic diagrams of (a) *qubit* and (b) *qudet*.

when measured again with the apparatus upside-down, we detect the x-component as $\downarrow$. Finally, Susskind lays the *Qudet* on its side, and explains that, if we prepare an up-state $\uparrow$ and test it using the apparatus on its side, we expect to get zero, but we always get ±1, in agreement with Goudsmit's *projection postulate*. But if the apparatus causes the physical system represented by the qubit to *align* or *anti-align with* the detector, then the quantum (±1 on *any* axis) argument vanishes; the classical (±1 on *one* axis) argument prevails; the *one* axis being the current B-field direction of the SG apparatus.

This counterfactual reasoning, based on the physically nonsensible idea that we can actually measure *each* of the spin components s_x, s_y and s_z via *three sequential measurements*, is the basis of the quantum *qubit*. Goudsmit interpreted Stern-Gerlach *splitting* of atomic beams as based on half-integral spin. If Goudsmit had simply proclaimed that

"*The projection of spin on one axis is* ±1," 〈the $\boldsymbol{B}$ axis〉 this would have been compatible with classical physics *and* with the facts of the SG experiment. Instead, his counterfactual projection postulate "*The projection of spin on any axis is* ±1", has led to countless apologies to the effect that viewing spin as a "*little arrow pointing in some direction*" is naïve, and "*any attempt to visualize it classically will badly miss the point.*" SG-measurement of any axis affects *other axes*, rendering meaningless the "preservation of $\boldsymbol{s}$-components" and subsequent comparison of such with another 3D direction. Only *one axis* is relevant: the physical B-field local axis. Irrespective of original spin, measured spin is parallel or anti-parallel to the local axis after the measurement. After exiting the magnetic field, spin is preserved until a next measurement, at which time the spin is rotated to the current local *x*-axis. The orthogonal axes vanish.

7. Analysis of Classical Qudet vs. Quantum Qubit

At [57:30] I ask Susskind about the logic of his presentation:

"*You've postulated a two-state system and a detector, but what would be the difference in logic in saying your system was a variable because your detector is a two-state device. It seems like you get the same output.*"

Susskind asks me to repeat the question.

"Y*ou've postulated a two-state system: +1 and −1. But then you have a detector that always produces plus or minus one. If you postulated a variable* [system] *to start with, your detector still seems like it would still give you +1 or −1.*"

Susskind: (hand to ear) "*Seems like what?*"

"*You're really saying your detector is a two-state system that you can orient, and that doesn't seem to imply* (*that your system is a two-state system*)."

Susskind: "*you're absolutely right, that at some point we have to come back to this and say the detector is a system and it has states and understand the combined system as a quantum system composed of two quantum systems.*

The standard approach is that the universe is quantum, but all measurement devices are treated as classical; we do not know where the boundary between

quantum and classical resides. Yet, from the above, we can alternatively understand the entire thing as *two classical systems*: the classical spin and the classical SG apparatus. We know the *apparatus* to be a 1D system, the *Qudet*; we do not know that the *system* is 1D, a *qubit*; this is an *assumption*, *not a conclusion based on the facts.* The *qubit* nature of the particle is an artifact, imposed by the 1D apparatus.

Susskind has published a book derived from the video lecture [15]. On page 6: a measurement results in $\sigma_z = +1$ based on the (*Qudet*) apparatus pointing in the *z*-direction. When the apparatus is flipped without disturbing the previously measured spin the new measurement is $\sigma_z = -1$, thus "*turning the system over changes* $\sigma = +1$ to $\sigma = -1$. *If we are convinced that spin is a vector, it has 3 components* $\sigma_x, \sigma_y, \sigma_z$. *When the apparatus measures the x-component, we expect* $\sigma_x = 0$ *but we find that* $\sigma_x = \pm 1$." Since the apparatus measures only one direction, "*one simply cannot know the components of spin along the two different axes…*" Hence, "*there is something different about the state of a quantum system and the state of a classical system.*"

Let us examine the state of a classical system. Assume we have a gyroscope pointing in an arbitrary direction in space. The gyroscope is spinning. If we overlay an (*x*,*y*,*z*)-coordinate system on the gyroscope, and assume (for simplicity) that the gyro spin $\boldsymbol{s}$ is in the *xz*-plane at angle θ with respect to the *z*-axis, there is no *y*-component of the gyro-spin, $s_y = 0$, $s_z = |\boldsymbol{s}|\cos\theta$, $s_x = |\boldsymbol{s}|\sin\theta$. The magnitude of spin is $|\boldsymbol{s}| = \sqrt{\boldsymbol{s}\cdot\boldsymbol{s}} = \sqrt{s_x^2 + s_y^2 + s_z^2} = \sqrt{|\boldsymbol{s}|^2 \sin^2\theta + 0 + |\boldsymbol{s}|^2 \cos^2\theta} = |\boldsymbol{s}|$. Why is this different from electron spin? It is different because our measurement device is the eyeball measuring photons reflected from the gyro-axis. The photons are quantum objects, but they do *not* disturb the gyro-spin axis; they are too weak. As we have seen the Qudet *does* disturb all axes, aligning the spin with the quantized apparatus. So, the *Qudet* is responsible for the quantum aspect of the spin of the atom. Yet, per Susskind, "… *knowing a quantum state means knowing as much as can be known about how the system was prepared.*" However, if the system was prepared by a *Qudet*, then all we can know about the quantum state is that *it is the state of the Qudet*!

Many authors are quite clear about the central role of the measurement apparatus: [16] Amri *et al.* state: "*In the quantum world, the measurement process plays a central role, as it leads to an unavoidable and strong modification of the system which is measured.*" Similarly, Gondran and Gondran, [17] state: "*The result of the Stern-Gerlach experiment is not the measure of the spin projection along the z-axis, but the orientation of the spin either in the direction of the magnetic field gradient, or in the opposite direction.*" Devereaux (2015) argues that "*a comparison with double slit experiments of photons or electrons is misleading because in the case of SG a real energy transfer takes place which destroys any superposition.*" Franca (2009) *proposed an explanation of the alignment of the magnetic moment as a purely classical phenomenon.* Kauffmann

(2015) remarked that "… *your view of measurement as a gross transformation of the quantum system rather than as a mere sampling of it that leaves it essentially undisturbed is helpful in dispelling some of that confusion*."

Yet according to Schmidt-Bocking *et al.* [18], "*for most physicists this was a "miraculous interaction" between moving atoms and the SG apparatus. … the physical mechanism responsible for the alignment of the silver atoms remained and remains a mystery*. […] *the trajectories of the atoms can be calculated using equations of motion from classical physics, but the rotation of the magnetic moments into well-defined orientation remain a puzzle*." Even for physicists of the caliber of Julian Schwinger: "*it is as though the atoms emerging from the oven have already sensed the direction of the field of the magnet and have lined up accordingly… we must learn to live with it*." And Feynman: "*That a beam of atoms whose spins would apparently be randomly oriented gets split up into separate beams is most miraculous… instead of trying to give you a theoretical explanation, we will just say that you are stuck with the result*." Bell noted that phenomena of the kind exhibited by Stern-Gerlach "*made physicists despair of finding any consistent space-time picture of what goes on on the atomic and subatomic scale*."

From the literature it seems that most physicists expect the atomic spin to exhibit Larmor precession with fixed precession angle θ and expect the atom to move under the force of the inhomogeneous magnetic field. Yet per Carrega *et al.* [19] "*A deep understanding … of energy exchange in strongly coupled systems … may have a profound impact … from a fundamental point of view*." Of this aspect, Navascues and Popescu [20] state: "*Traditional descriptions of the measurement process and quantum mechanics typically overlook the energy exchange between the system under consideration and the measurement device carrying it.*"

8. Conclusions: Projections onto Reality

Pusey, Barrett, and Rudolph [21] quote Jaynes:

"*Our* [quantum] *formalism* [was] *all scrambled up by Heisenberg and Bohr into an omelet that nobody has seen how to unscramble. …if we cannot separate the subjective and objective aspects of the formalism, we cannot know what we're talking about; it is that simple.*"

A major conceptual part of the omelet is due to Goudsmit's 1925 statement that "*The projection of spin on any axis is* ±1." We cannot geometrically *picture*, hence not *imagine*, a physical spin or quantized entity whose projection on *any* axis is ±1. However [22]: "…*when we measure a particle's component of spin in a* [any] *particular direction in a Stern-Gerlach experiment, it is the general belief that we are not measuring a pre-existing property*."

There are physical situations, such as electron spin in magnetic domains, in which broken symmetry causes spins to align with each other. In such cases the two-state formalism is useful. But if the iconic postcard data represents actual

physics occurring in an inhomogeneous magnetic field, our model accurately depicts the spin dynamics expected from the SG experiment. The *energy-exchange model* treats free energy that is *not* accounted for in standard quantum treatments.

Another popular approach is to interpret qubits as "quantum information" which can exist as both a 0 and a 1. Treatment of such is beyond the scope of this paper, however Siegfried ends an article on this approach by saying "*So it seems likely that the meaning and power of quantum information, represented by the qubit, has yet to be fully understood and realized.*" [23]

Quantum theory is probabilistic, not predictive. What does *testable by experiment* mean? Generally, it means that a theory *can fit the data*. But if our experiment must yield at least two states, and the theory provides four extra states that only theoretically exist, how do we *fit the data*? Recall Fermi: "*with 5 points I can fit an elephant.*" The assumption that three spin axes can be measured, when, in fact, only one axis, the *Qudet* axis, can be measured (prepared) implies imaginary data that must be handled properly. This is achieved in quantum mechanics by probabilistically forcing the "average" measurement of the other axes to zero, or to $\cos\theta$ if the *Qudet* is rotated an angle θ with respect to the prepared spin. In short, quantum theory projects non-existent spin axes onto reality and formalistically "fits the data", at the cost of ending up with a theory of reality believed to be "*beyond our evolutionarily developed neural structure*".

Acknowledgements

The author acknowledges helpful discussions with Steven Kenneth Kauffman, Otis Monty Frost, and Richard Zacher during the development of this model. It was also helpful that Leonard Susskind presented on video a development of the qubit in such a manner that it becomes obvious that it is the detector that is quantized, as shown herein.

Conflicts of Interest

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

References

[1] Townsend, J. (2012) A Modern Approach to Quantum Mechanics. 2nd Edition, University Science Books, Mill Valley.

[2] Sakurai, J. (1994) Modern Quantum Mechanics. Addison-Wesley Publishing, Reading.

[3] Jarzynski, C (1997) *Physical Review Letters*, **78**, 2690-2693. https://doi.org/10.1103/PhysRevLett.78.2690

[4] Crooks, G. (2009) *The Journal of Chemical Physics*, **130**, Article ID: 107101. https://doi.org/10.1063/1.3080751

[5] Deissler, R. (2008) *Physical Review E*, **77**, Article ID: 036609. https://doi.org/10.1103/PhysRevE.77.036609

[6] Talkner, P., Lutz, E. and Hanggi, P. (2007) *Physical Review E*, **75**, Article ID:

050102(R). https://doi.org/10.1103/PhysRevE.75.050102

[7] Roncaglia, A., Cerisol, F. and Paz, J. (2014) *Physical Review Letters*, **113**, Article ID: 250601. https://doi.org/10.1103/PhysRevLett.113.250601

[8] Arthur, J. (2011) Understanding Geometric Algebra for Electromagnetic Theory. IEEE Press, John Wiley & Sons, Hoboken. https://doi.org/10.1002/9781118078549

[9] Griffiths, D. (2005) Introduction to Quantum Mechanics. 2nd Edition, Pearson Prentice-Hall.

[10] Stenson, J. (2005) Representations for Understanding Stern-Gerlach Effect. Thesis, Brigham Young University.

[11] Messiah, A. (1961) Quantum Mechanics. John Wiley and Sons, New York.

[12] Bell, J (1987) Speakable and Unspeakable in Quantum Mechanics. Cambridge University Press, Cambridge

[13] Bacciagaluppi, G. and Valentini, A. (2009) Quantum Theory at the Crossroads. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139194983

[14] Susskind, L. (2012) The Theoretical Minimum Quantum Mechanics. http://freevideolectures.com/Course/3151/The-Theoretical-Minimum-Quantum-Mechanics

[15] Susskind, L. (2014) Quantum Mechanics. Basic Books, New York.

[16] Amri, T., Laurat, J. and Fabre, C. (2011) *Physical Review Letters*, **106**, Article ID: 020502. https://doi.org/10.1103/PhysRevLett.106.020502

[17] Gondron, M. and Gondron, A. (2013) Measurement in the de Broglie-Bohm Interpretation: Double-Slit, Stern-Gerlach and EPR-B. arXiv:1309.4757v3.

[18] Schmidt-Bocking, H., Schmidt, L., Jürgen Lüdde, H., Trageser, W. and Sauer, T. (2016) *The European Physical Journal H*, **41**, 327-364. https://doi.org/10.1140/epjh/e2016-70053-2

[19] Carrega, M., Solinas, P., Sassetti, M. and Weiss, U. (2016) *Physical Review Letters*, **116**, Article ID: 240403. https://doi.org/10.1103/PhysRevLett.116.240403

[20] Navascués, M. and Popescu, S. (2014) *Physical Review Letters*, **112**, Article ID: 140502. https://doi.org/10.1103/PhysRevLett.112.140502

[21] Pusey, M.F., Barrett, J. and Rudolph, T. (2012) *Nature Physics*, **8**, 475-478. https://doi.org/10.1038/nphys2309 http://arxiv.org/pdf/1111.3328v2.pdf

[22] Leifer, M. (2014) *Physical Review Letters*, **112**, Article ID: 160404. https://doi.org/10.1103/PhysRevLett.112.160404

[23] Siegfried, T. (2017, July) Birth of the Qubit. *Science News*.

Journal of Modern Physics, 2022, 13, 385-409
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Some Ideas of James Watt in Contemporary Energy Conversion Thermodynamics

Jovan Mitrovic

Stuttgart, Germany

Email: mitrovic@tebam.de

How to cite this paper: Mitrovic, J. (2022) Some Ideas of James Watt in Contemporary Energy Conversion Thermodynamics. *Journal of Modern Physics*, **13**, 385-409.
https://doi.org/10.4236/jmp.2022.134027

Received: March 11, 2022
Accepted: April 8, 2022
Published: April 11, 2022

Abstract

James Watt contributed significantly to the development of the thermodynamics of energy conversion as a science. Several of his ideas are now integral part of thermodynamics, but Watt as their creator is not mentioned. This paper presents some of Watt's concepts of energy conversion, including his thermodynamic analysis of the Newcomen steam engine that marks the beginning of thermal engineering. The analysis illuminated the causes of the enormously high heat losses in the installation and showed the ways for their reduction. This led him to a new conception of the steam engine with a separate condenser. Not less important was Watt's determination of some physical properties of water and steam used as the working substance. In the experiments he observed the decrease of the latent heat of steam with increasing temperature and its disappearance at very high temperature led him to postulate the existence of a thermodynamic critical state of water. He introduced the work associated with volume change into thermodynamics and illustrated it graphically. Several of Watt's numerous ideas deserve to be included into the history of the thermodynamics of energy conversion but they are rarely mentioned in the scientific literature. Arguably the most important is the First Law of Thermodynamics, which he introduced in his 1769 patent and related works in 1774 and 1778.

Keywords

Thermodynamics, Energy Conservation and Conversation, Mechanical Equivalent of Heat, James Watt, Volume Work, First Principle of Thermodynamics, Latent Heat, Steam Expansion and Condensation

1. Introduction

In trying to understand the coal consumption of the Newcomen steam engine as

a function of its power, James Watt (1736-1819) believed that he had first to understand how the engine worked. This step was impossible, however, without an understanding of the processes involved in energy conversion and the knowledge of the physical properties of water and steam, used in the engine as the working substance. On pp. 7-8 [1] Watt summarized the state of knowledge regarding steam and the steam engine at the time he started his work on a model of Newcomen's engine, sometime prior to 1765. In particular, he mentions steam condensation in contact with, and heat communication to, cold bodies. Watt knew that water can boil at temperatures below 100˚F (≈37.8˚C) and that evaporation causes cooling of an unheated evaporating liquid and the bodies in contact with it. In addition, Watt himself conducted experiments to determine the specific heats of some substances in relation to that of water, the quantity of water which could be evaporated by consumption of a given quantity of coal, the steam pressure at different temperatures, and so on (see also pages 311 & 395 in Farey [2]). Of particular importance for the modification of the engine were knowledge of the elasticity of steam and the steam pressure at various temperatures.

In a thermodynamic analysis of the Newcomen engine[1] Watt arrived at the conclusion that heat losses caused useless condensation of saturated steam and have to be prevented in an economically conceived installation. He turned his attention to the boiler and especially to the steam cylinder. The steam cylinder was last stage in the conversion of heat into mechanical work, which served alternately as a steam condenser. The heat losses of the steam cylinder were therefore of central importance for Watt.

General aspects of Watt's thermodynamics have been considered in [3], including a discussion of why Watt has generally been excluded from the history of thermodynamics (as opposed to mechanical engineering). In the present paper we analyze some of Watt's ideas about the thermodynamics of energy conversion. Many of these ideas were actually published by Watt in his patents or as drawings. The content of these publications has generally been ignored by those who did not consider technical drawings and sketches to be a means of imparting knowledge. This is not the case in the present paper. To fully understand the meaning of Watt's drawings our present knowledge will be involved. However, the paper must not be qualified as the Whig history.

Our analysis will encompass the most important of Watt's ideas on the conversion of one energy form into another (particularly of heat into mechanical work), energy conservation, the separate condenser, the interpretation and determination of physical quantities, (particularly the latent heat of steam by using mass and heat balances), and the work associated with change of volume (during the expansion of the steam within the cylinder). We also sketch his formulation of the **First Law of Thermodynamics**. In the Appendix, Conversion of mechanical energy in heat, the attention is focused on Robert Boyle as the pioneer (1675) of energy conversion.

[1]When we speak of the Newcomen engine, we mean its model.

The topic of the present work limits its scope in a way that a comprehensive overview of the literature is neither necessary nor possible. In addition, James Watt's science is considered prehistoric. This, however, does not justify his firm exclusion from the History of Thermodynamics which actually created a gap in the development of this science. Of many works devoted to the history of thermodynamics, the recently published paper by W. M. Saslow [4] discusses several important details. The content of the present paper is limited to the Watt's historical ideas.

2. Watt's Thermo-Energetic Analysis of the Steam Engine

For much of the 18th century, the Newcomen steam engine was the main facility for converting heat into mechanical work. Its poor thermomechanical characteristic aroused Watt's interest in this machine. Watt thoroughly analysed the thermodynamics of a model of the engine and uncovered the reasons for its low performance. We next summarize the results Watt deduced from the analysis because they provide insight into the state of knowledge prior to Watt.

2.1. Actual-to-Minimum Steam Consumption

Watt modified the construction of the steam boiler and determined the amount of steam leaving it ([2], Chap. V, p. 309). This gave him the total production of steam. In the next step he determined the amount of steam required to run the engine which he assumed to be the amount of steam required to fill the cylinder with one stroke of the engine.

Denoting the mass of steam leaving the boiler by $M_S = \rho_S V_S$ and the steam quantity required for filling the cylinder by $M_{Smin} = \rho_S V_{CYL}$, Watt found

$$\frac{M_S}{M_{Smin}} = \frac{V_S}{V_{CYL}} = 4, \tag{1}$$

where M and V denote, respectively, the mass and volume of steam (subscript S) and the subscript CYL refers to the steam cylinder.

According to Equation (1) the mass of steam generated in the boiler was 4 times larger than the minimum mass required. In other words, only 1/4 of the steam leaving the boiler was effectively used to perform work while 3/4 had been wasted. On the basis of this relationship, Watt estimated the potential saving of coal or the "increase" in coal energy to be 75%. From this result he concluded that any useless vapour condensation had to be prevented. Being the cause of this condensation, heat losses in the entire engine's installation had to be avoided. This insight was important, not only for the further development of steam engines, but also for the economy of coal consumption.

2.2. Causes of High Steam Losses

1) Geometric similarity

Watt found that the boiler of the model Newcomen engine could not supply the cylinder with sufficient steam to keep the engine in operation. In order to

understand the enormous waste of steam, Watt compared the geometries of the model and the real steam engine [5], p. 99. He calculated, in both cases, the ratio of the surface area A to the volume V of the cylinder occupied by steam. Watt did not state any formula he used, but his description suggests the following expression,

$$A/V = (4/d) + (2/L) \sim 1/d \ , \quad L \gg d \ , \tag{2}$$

where d is the diameter and L is the length of the cylinder. The ratio A/V may be viewed as the area of heat transfer surface per unit volume of steam occupying the cylinder. Watt's calculation showed that the model (small d) possessed much larger ratio A/V than the real engine, and the two cylinders were not geometrically similar. (It is not clear whether Watt actually had the geometric similarity criterion in mind when he compared the steam cylinders). The geometric disproportion was unfavorable for the model regarding fuel (coal) economy. Taking the other parameters to be the same, a cylinder with a smaller diameter d provides a larger ratio A/V and withdraws more heat from the steam per unit volume, thus causing a relatively larger heat loss.

2) Material properties

In addition to the geometric disproportion, Watt also considered the effects of the material properties of the cylinder on the heat losses. The cylinder of the model was made of brass, a material with thermal conductivity, approximately double that of the cast iron used in large engines. Relatively larger heat losses across the cylinder wall of the model engine were thus unavoidable. In this context the large mass density of the cylinder materials (metal) should also be mentioned. Watt tried to use other materials for the cylinder, like wood, but the disadvantages outweighed the advantages and the idea was not pursued further.

3) Mode of operation

In the Newcomen engine, the steam cylinder was alternately heated and cooled, and Watt realised that this was the main, though not the only, reason for the high heat losses and wasted steam condensation. In addition, he also realised that the heat supplied to the cylinder was partly transferred across the cylinder wall to ambient as a heat loss. For this reason, he enveloped the steam cylinder of the engine that he developed later with a layer of saturated steam. This measure reduced the heat losses almost completely.

In conclusion, Watt realized the heat losses of the Newcomen engine to be decisive regarding the useless consumption of fuel whereas the mechanical losses (friction and inertia) were of less importance for the economy of the engine. He tried to reduce the heat loss by insulating the engine parts, an idea that later proved to be very fruitful. In our terminology, Watt's ideas on heat losses are synonymous with a reduction of thermal irreversibility of the system. Sadi Carnot used the temperature difference in his theoretical discussion of thermal reversibility. According to Newton's law of cooling, the heat flow (loss) across the system boundary is proportional to the temperature difference, so Carnot would have arrived at the same results if he had worked in terms of the heat loss instead. The no-

tion that Watt's ideas were included in Carnot's analysis, cannot be ruled out.

3. Physical Properties of Water and Steam

3.1. The Latent Heat of Steam

The term latent heat was introduced by Joseph Black (1728-1799) who was, apparently, the first to recognize that heat is consumed or released during the change of phase of a substance even though the temperature remains constant. For example, the conversion of water to steam at constant pressure consumes heat that cannot be detected by a thermometer but is regained when the phase transition is reversed and the steam condensed. With respect to temperature, the action of the heat remained ***hidden*** and Black referred to it as ***latent heat***. Independently of Black, James Watt also realized that heat consumption or production occurred during the phase change, particularly with water evaporation and steam condensation. Here we discuss Watt's understanding and his experimental determination of the latent heat of steam.

Watt's first, rather crude, experiments were performed in 1765 ([1], p. 10, Footnote) and these were followed by more precise measurements in 1781 ([1], p. 6, Experiment I). In the experiments, Watt used the method of direct condensation of saturated steam in initially subcooled water of known mass and temperature. He introduced into the subcooled water a stream of saturated steam which, according to Farey, [2], p. 312,

… *mixed with, and condensed in, that water, which received all the heat of the steam, till it became boiling hot, and could condense no more* : *the water in the jar was then found to have gained about one-sixth part of its weight, by the addition of the condensed steam; whence it appeared that one pound of water, in the state of steam, can heat six pounds of water from 52 deg. to 212 deg.*

Due to the steam condensation both the mass and the temperature of the water increased and this allowed Watt to calculate the latent heat. Watt's method is best illustrated if we neglect thermal losses in the experiments and repeat his evaluation. Assuming isobaric conditions in the experiments, the following energy balance:

$$m_S h_{vl} = m_W c_{pW} \Delta T_W \tag{3}$$

holds, where m denotes the mass of initially subcooled water (subscript W) and saturated steam (subscript S) condensed in the water; c_{pW} is the average specific heat capacity of the water in the temperature range ΔT_W (52˚F to 212˚F) covered by the experiments, and h_{vl} is the specific latent heat of steam condensation. Watt did not write the energy balance using symbols, but effectively applied Equation (3) in his calculations.

Using Watt's experimental data,

$$m_W / m_S = 6\,, \quad \Delta T_W = (212 - 52)^{\circ}\mathrm{F} = 160^{\circ}\mathrm{F} = 88.89^{\circ}\mathrm{C}\,,$$

and the NIST (US National Institute of Standards and Technology) value[2] for the

[2]In [1] Watt mentions the specific heat of some metals in comparison to water; for water, he set c_{pW} = 1 as a reference value, and measured the latent heat in degrees Fahrenheit.

specific heat capacity of water, $c_{pW} = 4200.0\ \mathrm{J/(kg \cdot K)}$, Equation (3) gives,

$$h_{vl} = (m_W / m_S) c_{pW} \Delta T_W = 6 \times 4200.0 \times 88.89 = 2240.0\ \mathrm{kJ/kg}.$$

This value Watt obtained 250 years ago from a simple but ingenious idea is only 0.77% smaller than the actual standard (NIST) value of 2257.4 kJ/kg. The deviation is most probably caused by heat losses to the surroundings and to the pan containing the water in Watt's experiments. The remarkable agreement underlines the depth of Watt's investigative ability and the rigour of his analysis.

Watt adopted Joseph Black's method for expressing the latent heat of evaporation (equal to latent heat of steam condensation) as the temperature rise of the liquid phase in degrees Fahrenheit (˚F) which would be caused by the absorption of that heat without evaporation. This method follows from Equation (3). Watt obtained a latent heat of 950 (degree F) at a saturation temperature of 212˚F and of 1000 (degree F) at a saturation of 70˚F ([1], pp. 6-7). Watt therefore concluded that the latent heat decreases with increase of temperature.

It is instructive to compare Watt's method with the one used by Joseph Black. Three years prior to Watt's experiments, in 1762, Black [6] conducted experiments on the latent heat of evaporating water. He heated and evaporated a quantity of water in an open pot and measured the time required to heat the subcooled water from the initial to the boiling temperature and also the time needed to completely evaporate the saturated water. From a comparison of these times, he obtained the latent heat of boiling water. Black did not use any heat balance and his method is affected by the condition of heat transfer to the heated and evaporating water. Compared to Watt's experiments of 1765 and 1781, Black's method was not scientifically rigorous. This is possibly the reason why Watt did not use it in his experiments. Watt was therefore the first scientist to determine precisely the latent heat of water.

3.2. The Total Heat of Saturated Steam

Using the latent heat, Watt defined the total heat of steam as the sum of the sensible heat required to raise the water temperature from the reference value (e.g. 32˚F = 0˚C) to the saturation temperature at the prevailing pressure and the latent heat required to evaporate the water at that pressure. For an easier discussion, we shall express this idea analytically.

Denoting the total heat of saturated steam by h_S, we can express Watt's idea by the equation,

$$h_S = c_{pW} (T_S - T_R) + h_{vl} \tag{4}$$

where the subscripts R and S refer, respectively, to the reference and the saturation state of water; (T_R is usually taken to be zero, $T_R = 0^{\circ}\mathrm{C}$). The first term that results from the temperature difference is sensible heat, since it can be detected by a thermometer; h_{vl} is the latent heat. Today, Equation (4) defines the specific enthalpy of saturated steam. In Watt's time Equation (4) was known as ***Watt's law of latent heat***; see e.g. [7] and [8]. Watt was inclined to believe that the total

heat of steam h_S was independent of pressure which was an acceptable approximation given the range of pressure variation covered by his experiments.

Watt's Equation (3) is valid only when the originally subcooled water is heated up to the saturation temperature T_S. Its validity can be extended to any temperature $T_1 < T_S$ by using the total heat of steam according to Equation (4). The mass and heat balances then give:

$$m_S\left(c_{pW}\left(T_S - T_1\right) + h_{vl}\right) = m_W c_{pW}\left(T_1 - T_W\right), \tag{5a}$$

or

$$h_{vl} = \left(\left(m_W / m_S\right)\left(T_1 - T_W\right) - \left(T_S - T_1\right)\right)c_{pW}. \tag{5b}$$

The change in the total heat of steam (left hand side of Equation (5a)) covers the heating of subcooled water m_W from T_W to T_1. For $T_1 = T_S$, Equation (5a) becomes identical to Equation (3). Setting $c_{pW} = 1$, as Watt actually did, Equation (5a) results in Watt's original scheme for calculating the latent heat [1], p. 7; in this case, Equation (5b) becomes identical to the equation Capecchi [9] deduced from the Watt's calculation table.

3.3. The Thermodynamic Critical State of Water

The interaction of the sensible heat and latent heat as parts of the total heat of steam was important for Watt regarding the coexistence of the two phases (steam and liquid). Watt understood these components to undertake reciprocal alterations (at constant total heat) when the pressure is varied; an increase of pressure decreases the latent heat. He linked the coexistence of the phases (water and steam) along the saturation line with the latent heat, and from his experiments he observed a decrease of latent heat with increasing pressure. At a sufficiently high pressure, he expected the latent heat to disappear while water underwent some remarkable changes. Accordingly, Watt specified in 1782 a new state of the steam-water system, today known as the *thermodynamic critical state*, [10], p. 5, see below. The temperature and pressure at which the latent heat becomes zero determine a point that is now known as the *critical point*, and which we might also call the Watt's point.

Watt stated this idea in a letter to J.-A. De Luc (Jean-André De Luc, Swiss-English scientist, 1727-1817), dated 13 December 1782. In an extract from this letter ([10], p. 4), we read:

Dr. Priestley has made a most surprising discovery, which seems to confirm my theory of water's undergoing some very remarkable change ***at the point where all its latent heat would be changed into sensible heat****, which must follow from the diminution of the latent heat, as the sensible heat increases, probably at or near 1200° of Fahrenheit.* (emphasis added).

The letter was written at a time when the composition of water was still being discussed. Watt was involved in this discussion and his letter, printed in Muirhead book: *Correspondence of the Late James Watt on his Discovery of the Theory of the Composition of Water* [10], could actually be associated with the

constitution of water. However, based on Watt's experiments on the "common" latent heat of steam that decreases as the pressure and temperature increase, the quotation is concerned with the vapour-liquid phase transition. Watt stated the total heat of steam at which the latent heat would disappear to be nearly 1200 (degree F), independent of pressure. According to his own experiments at atmospheric pressure, (Watt [1], p. 7, Miller [8], Capecchi [9]), the total heat of steam should lie between 1134 (degree F) and 1177 (degree F) which is very near to the value he estimated earlier. His Equation (4), too, gives for the total heat of steam ($c_{pW}=1$ and $h_{vl}=T_R=0$) the value $h_S=1200$ (degree F), approximately. In 1853 Charles Siemens [11] published a review paper stating the total heat of water at the critical point to be 1180 (degree F), which is only 20 (degree F) below the value of 1200 (degree F) which Watt had estimated seven decades earlier, in 1782. These additions support the notion that Watt was referring to the critical state of the steam-water system, and not the state of thermal decomposition of steam at much higher temperature.

Watt explained the decrease and disappearance of latent heat in a letter of 13th December 1782 sent to Dr. Black, stating ([10], p. 5)

… *that, as steam parts with its latent heat as it acquires sensible heat, or is more compressed, that when it arrives at a certain point it will have no latent heat, and may, under proper compression, be an elastic fluid nearly as specifically heavy as water; at which point I conceive it will again change its state and become something else than steam or water. My opinion has been that it would then become air*; …

This quote refers to a steam-water system below the critical temperature where the latent heat is finite and decreases as the critical point is approached, whereupon it disappears. Watt's comparison of water at this point with air could be interpreted as the decomposition of water into its constituents. Actually, however, the comparison seems to be symbolical. Watt chose air for comparison to emphasise the dramatic changes in the state of water at the critical point which, according to Watt, was neither steam nor ordinary liquid. He mentions a large steam elasticity which he did not expect at such a large steam density. Today we can express Watt's idea on elasticity at the critical point by the relationship $\partial p/\partial v=0$, which implies that a small variation of pressure p causes a very large variation of the specific volume v.

Some authors ascribe the detection of the critical point to Cagniard de la Tour, who in 1822/3 published two papers dealing with critical phenomena. Berche *et al.* [12] mention James Watt's decrease of the latent heat with increasing temperature, but not its disappearance at a certain (critical) temperature. Nonetheless, the above facts show conclusively that Watt's idea on the critical state of water preceded the works of Cagniard de la Tour by four decades.

Figure 1 illustrates the position of the Watt (critical) point in a T, p-diagram; its coordinates for water are approximately: $T_{CR}\approx 647.0$ K, $p_{CR}\approx 22.0$ MPa. Along the Watt-curve the two phases exist in equilibrium, while in the supercritical region, $T>T_{CR}$ and $p>p_{CR}$, the system is in a kind of single-phase

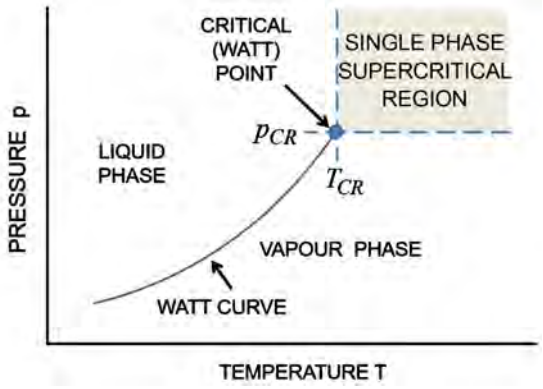


Figure 1. Position of the thermodynamic critical (Watt's) point of a pure fluid substance in the T, p-diagram. Starting at low values of p and T, the latent heat decreases along the curve, vanishing at the temperature $T = T_{CR}$.

state. In this region the system shows remarkable changes, as stated by Watt in 1782. Watt's experiments extended from a pressure of 0.15 inches of Hg to 22 inches of Hg and from 30 inches of Hg to 82 inches of Hg and hence only covered a small region of the whole pressure range.

4. Watt's Energy Conversion Principle

4.1. Watt's Patent of 1769: The Separate Condenser

With this patent Watt devised and introduced a separate condenser and substantially changed the conception of the Newcomen steam engine.[3] Here we provide a short explanation for the need of the condenser which Watt himself formulated in his patent of 1769 [13] [14]. Watt states in the first paragraph of the patent (the letters a), b) and emphasis added):

***First**, that vessel in which **a) the powers of steam are to be employed to work the engine**, which is called the cylinder in common fire engines, and which I called the steam vessel, must during the whole time the engine is at work kept **b) as hot as the steam that enters it**, first, by enclosing it in a case of wood or any other materials that transmit heat slowly; secondly, by surrounding it with steam or other heated bodies; and, thirdly, by suffering neither water or any other substance colder than the steam to enter or touch it during that time.*

According to Watt, a), the power of steam drives the engine; he does not mention any heat explicitly in this context because the first energy conversion step has been completed by the steam production in the boiler. In addition, he was not convinced that the word heat alone would adequately encompass the ability of steam to do work. He sought to include also the pressure into considerations, but first without much success. Watt's idea is understandable: a multiplication of the pressure by the specific volume of the fluid, shifts the physical property

[3]In 1974 A. J. Pacey [15] stated the invention of the Newcomen atmospheric steam engine (1712), the patent of James Watt for a condenser separate from the engine's cylinder (1769), and Watt's further patent for the expansive use of steam (1782) to be the outstanding events of the eighteenth century for the historian of thermodynamics. These Watt patents are important for the development of thermodynamics as a science.

pressure to a higher quality, namely energy. Watt's intuition turned out to be justified by his invention of the pressure indicator.

Watt required, b), the steam cylinder to be kept at the same temperature as the steam entering it throughout the whole time the engine is at work. This means the complete thermal decoupling of the cylinder from the surroundings and the avoidance of all heat losses. He also specified the ways how this could be achieved.

Watt's patent does not allow for any temperature changes in the cylinder wall, neither temporally nor spatially. This condition also excludes the alternating heating of the cylinder (filling with saturated steam) and its cooling with water (steam condensation) as in Newcomen's engine. Consequently, the whole of Newcomen's conception of the steam engine collapses with the Watt's patent.

Watt's formulation was not an ***ad hoc*** statement but was based on a thermodynamic inspection of the Newcomen engine and a detailed analysis of the heat losses, starting with the boiler. According to Watt's ideas, the first step of energy transformation in his cycle occurs with water evaporation in the boiler where the heat is mostly consumed for steam production (increase of total heat, Equation (4)) at constant pressure. Thence heat is partly transformed in mechanical (potential) energy of steam, $p\left(v_S - v_L\right)$, and partly dissipated to ambient as heat loss. Watt seemingly understood the product pv_S as the power of the steam in his patent of 1769, which would justify his attempt to include the pressure in the thermodynamic analysis of the engine.

According to Watt's analysis, the latent heat of steam was irrelevant for the conception and introduction of the separate condenser. Consequently, unjustified are also speculations in the literature linking the role of Joseph Black and his doctrine of latent heat of steam with Watt's invention. Watt's guiding idea for the conception of a separate condenser was not the latent heat but the suppression of all unintended heat losses which caused useless steam condensation and a reduction of the power delivered.

Figure 2 shows a drawing of the steam engine as described in Watt's patent of 1769. The steam cylinder (a), enveloped by a layer of steam, is connected to the boiler (not shown in the drawing). This construction almost completely prevented heat losses to the environment. The separate condenser (h), situated below the steam cylinder, is immersed in cooling water. In comparison to Newcomen's installation, the Watt's arrangement increased the performance of the engine substantially. The figure contains all the important details of the working principle of a steam engine and can be read like a book on thermodynamics. In this respect it is comparable to the Galileo's book of nature.

4.2. The Equivalence of Heat and Work

For some 40 years, around the transition from the 18th to the 19th century, Watt's steam engine was the main industrial facility for converting thermal energy (heat) into mechanical work. Watt practised this kind of energy conversion several decades before the "official" development of thermodynamics began in the first half of the 19th century. On page 337 John Farey [2] states (see also p. 330):

In 1778, when Mr. Watt first established his engine, his proposals were to raise 500,000 cubic feet of water 1 foot high by the consumption of one hundred weight (=112 lbs) of Wednesbury coals.

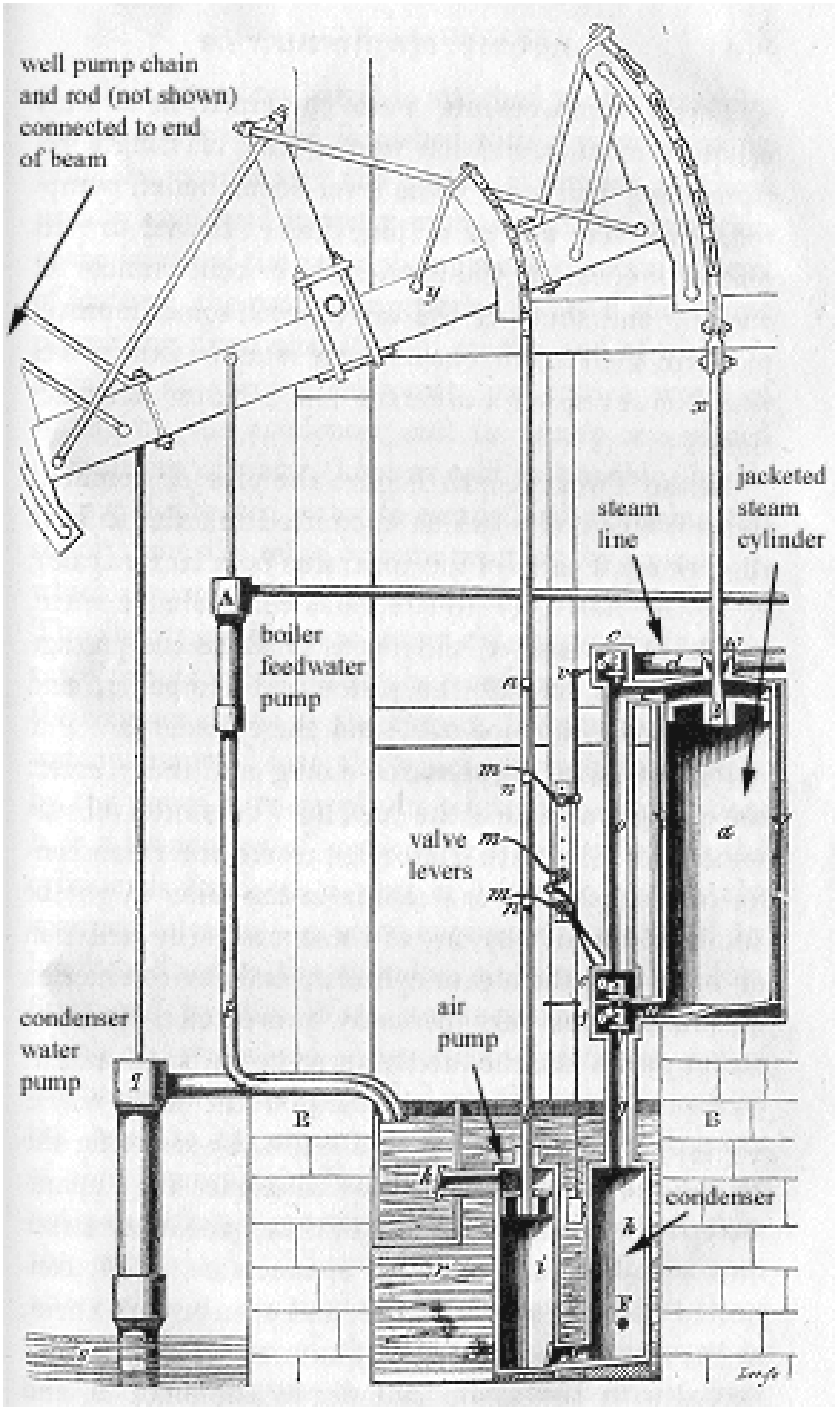


Figure 2. Watt's steam engine with separate condenser (h) as described in his patent of 1769, built 1774; the water boiler is not shown in the drawing. The drawing in Stuart's book ([5], p. 114) is without text; C. Lyra, Michigan State University, presents the figure with text but without mentioning the author, https://www.egr.msu.edu/~lira/supp/steam/wattengine.htm.

This proposal is important for understanding the equivalence and mutual convertibility of different energy forms. Its greatest value consists in showing the profound connection between heat that disappears and reappears as work. This interconnection became later a pillar of the foundation of the dynamical theory of heat. In order to easier overview the proposal's content, we shall reduce it to a relation of physical quantities, assuming the term consumption to mean the burning of coal and the production of heat. Watt's proposal then expresses the energy balance of a set of real processes spanning three modes of energy conversion (chemical energy of coal → heat → mechanical energy of steam → useful mechanical work), taking into account all the energy losses of the engine's installation.

Multiplying the mass, M, of coal consumed in the reaction by the heat of this reaction, H, Watt's proposal can be cast in the quantitative energy equation:

$$H \cdot M = V \cdot \rho \cdot g \cdot h + E_{loss} \tag{6}$$

(Heat = Work + Losses)

(V-volume, ρ-mass density, h-elevation height of water,
g-acceleration of gravity, E_{loss}-sum of all energy losses).

Equation (6) specifies the quantity of heat (coal) required to produce a quantity of mechanical work by using Watt's steam engine. The quantity E_{loss} encompasses all losses of the energy conversion process, *i.e.*, friction and heat losses, including also the heat of steam condensation in the condenser. Consequently, Watt's proposal not only states the equivalence of heat and work but also measured the heat in mechanical units.[4] The conversion of energy from one form into the other, does not contradict to the idea of its conservation, hence the Watt's proposal expresses also the principle of energy conservation. On its basis Cardwell [16] guessed that a given amount of heat must correspond to a given amount of work. Because of the statement's clarity, it is possible that later investigations into the mechanical equivalent of heat, such as those of Joule [17], have been motivated by the Watt's proposal.

4.3. The First Law of Thermodynamics

Equation (6) leads to an important relationship of thermodynamics if the energy (heat) losses are assumed—as is usually done in publications—to arise only from steam condensation in the condenser and set, $E_{loss} = Q_C$, where Q_C denotes that wasted heat. The product $H \cdot M$ is equal to the heat Q_B, the working substance receives in the boiler. Denoting, in addition, by W the elevation work, $V\rho gh = W$, Equation (6) takes the form

$$Q_B = W + Q_C . \tag{7}$$

The heat Q_B is converted in the steam cylinder into the work W and wasted heat Q_C. This equation is well-known from literature as the ***First Principle of Thermodynamics***. It's formulation by James Watt in 1778 preceded by more

[4]Equation (6) does not contain the coefficients of energy equivalence which would be necessary if the terms were measured in different units.

than 6 decades the next generation of similar works, starting with R. Mayer and J. P. Joule in the 1840s. Collins [18] names it *the Mayer-Joule Principle*. The meaning of the terms Q_B and W are mentioned in the Watt's proposal, while the term Q_C follows from the fact that steam engine was used for energy conversion. In his 1769 patent, Watt described the principle of the steam engine and identified the heat of condensation of the steam Q_C as wasted heat [3].

Equation (7) is suitable for quantifying the consequence of the idea that heat is transported from the boiler to the condenser without consumption, $Q_B = Q_C$, as S. Carnot and W. Thomson (Lord Kelvin) (and apparently also R. Clausius) originally thought; in this case no work is won, $W = 0$. If the work W is supplied to the system instead of Q_B, then with $Q_B = 0$, the Watt's Equation (7) becomes $W + Q_C = 0$, expressing the energy balance of the Joule's padlewheel experiment where Q_C represents the change of the internal energy of the fluid.

For other formulations of the first law of Thermodynamics, Equation (7), the reader is referred to common books on Thermodynamics. Here we only show the version with the internal energy which was motivated by an anonymous reviewer. We consider a cylinder filed with gas that is enclosed with a movable piston and take Q_B to be the heat supplied to the gas, W is the work done by the gas (expansion). Then Q_C is not wasted as in the steam engine but it remains in the gas representing the change of internal gas energy, $Q_C = \Delta U$. Hence: $Q_B = W + \Delta U$.

Energy splitting in the steam engine: The energy splitting is an important step of the energy conversion process. This step is illustrated in **Figure 3**. It shows a section of **Figure 2** surrounding the steam cylinder. The splitting of the energy Q_B according to Equation (7) takes place on the piston as it moves down the cylinder. Literally, the piston acts as a permeable device, allowing only the mechanical energy $p\mathrm{d}V$ of the steam to leave the cylinder and do the work W, $W = \int p\mathrm{d}V$.

The transfer of the work from the piston to the water pump q takes place via the vertical rod x, **Figure 2**, and the massive rocker (beam), which can perform small angular movements in the plane of the drawing around its suspension point. The heat Q_C leaves the cylinder mainly as latent heat of steam; after the condensation in the separate condenser, it becomes absorbed by cooling water and wasted as sensible heat.

4.4. Flow Sheet of Watt's Energy Conversion Process

The drawing of Watt's installation in **Figure 2** is too complex for a quick overview of the whole energy conversion process. On the basis of this drawing and Watt's patent of 1769, a simple flow sheet, shown in **Figure 4**, will be used for a brief description. The main elements are the water boiler, the steam cylinder, the separate condenser and the feed pump. These components constitute the Watt's cycle. The working substance changes its potential (state) mainly in the boiler and in the steam cylinder; the changes in the feed pump and the condenser are

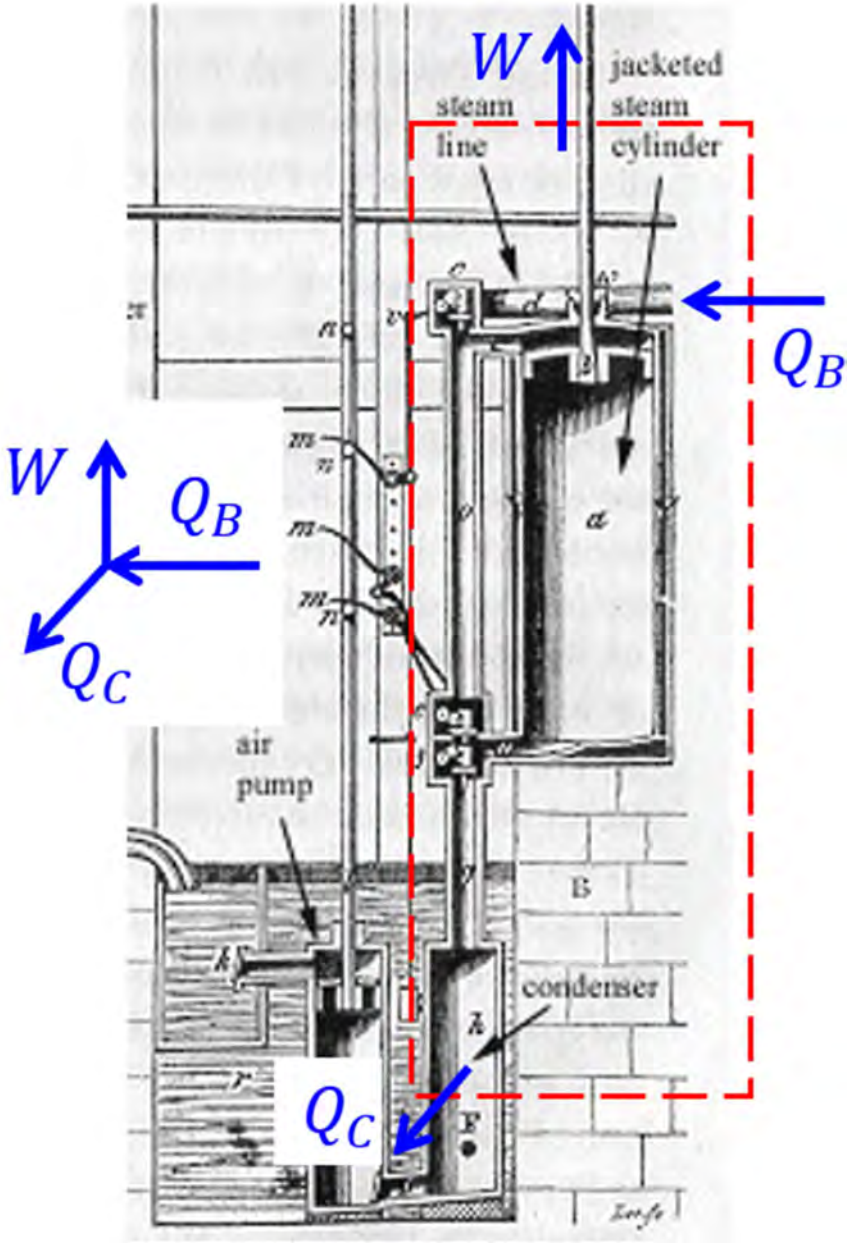


Figure 3. Watt's illustration of the first law of Thermodynamics as energy splitting in the steam engine (supplemented section of Figure 2 surrounding the steam cylinder).

not separately illustrated. With this assumption the change of potential in the boiler is equal and opposite to that in the steam cylinder. The term potential used here shall account for Watt's idea stated in his patent of 1769 that the power of steam drives the engine [3].

The conversion of energy in the Watt's cycle is best illustrated by the Newton analogy. Newton studied the interaction of elementary particles of matter in terms of forces, assuming that the interaction force is repulsive at one particle distance but attractive at the other. The change in force is continuous as the distance changes continuously. Using his algebra, Newton said [19]:

… And as in Algebra, where affirmative Quantities vanish and cease, there negative ones begin; so in Mechanicks, where Attraction ceases, there a repulsive Virtue ought to succeed. …

If we replace the positive and negative quantities in Newton's algebra by different forms of energy, we can conclude that where one form of energy decreases and disappears, the other form of energy emerges and increases, while the total energy of the systems remains conserved.

The flow sheet of the Watt's stem engine in **Figure 4** constitutes the skeletons of several modern energy conversion plants. For instance, replacing in Watt's installation the steam cylinder by a steam turbine with an electric-generator, it becomes identical to the flow sheet of a contemporary power plant operating with saturated steam. The flow sheet (on the left) not only shows the implementation of Equation (7) but also the temperature range covered by the steam engine. The steam has the highest temperature in the boiler and the lowest in the condenser. Watt recommends keeping the temperature in the condenser as low as possible (depending of cooling water). He did not limit the temperature in the boiler directly, but for safety reasons (boiler explosion) via the steam pressure, which he determined himself as a function of the temperature. Because of $T = f(p)$ he clearly defined the temperature of the working substance in the boiler and thus also the temperature range of the steam engine. This temperature range Watt defined corresponds exactly to the definition that Sadi Carnot published in his 1824 dissertation. Carnot's definition is known in the literature as Carnot's postulate, while James Watt is never mentioned as the originator of this idea. Further details are given in [3].

4.5. Watt's Patent of 1782: The Steam Expansion

In this patent, titled Certain New Improvements Upon Steam Or Fire Engines…, Muirhead [20], we consider only the specification No. 1:

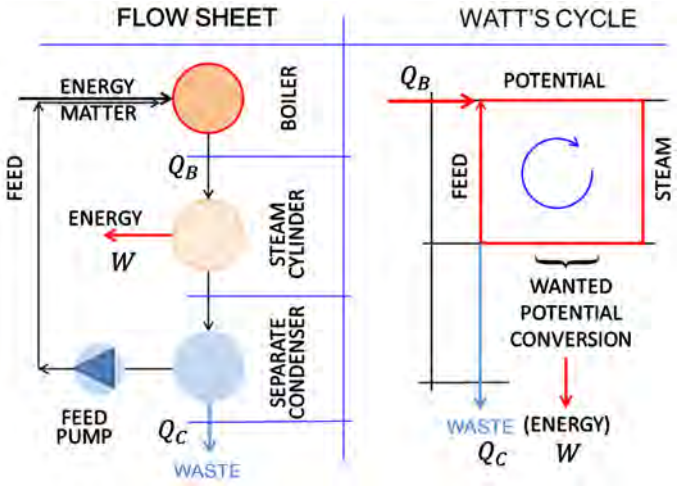


Figure 4. Simplified flow sheet of Watt's energy conversion process. For the meaning of Q_B, W and Q_C see Equation (7).

The use of steam on the expansive principle: *together with various methods or contrivances*, (*six in number, some of them comprising various modifications*), *for equalising the expansive power*.

Preparing the reader for an easier understanding of the expansive action of steam in the cylinder, Watt described in the patent first the effect of a constant steam pressure acting on the top of the piston moving downwards. He denoted this effect as the power of the elastic steam force. Watt's reasoning resulted in the definition of a new important quantity; the work done by the steam due to change of volume; next we illustrate some steps of Watt's idea.

1) The work due to change of steam volume at constant pressure

Watt states that, if the power of the elastic steam force were employed to act upon the piston through the whole length of its stroke, and to work a pump connected to the stroke as is usual in steam engines, it would raise through the whole length of its stroke a column of water whose weight should be equal to several pounds per square inch of the area of the piston, besides overcoming all the friction and inertia of the water and the parts of the engine. Watt utilized the increase of volume of steam in the cylinder at constant pressure to perform work. The later became known as the volume work of steam.

Following Watt's description, the steam pressure p acting on the piston, multiplied by the piston area A determines the force F acting on the piston,

$$p \cdot A = F \,, \tag{8}$$

which, multiplied by the length L of the stroke, gives the work W performed by the steam,

$$W = F \cdot L = p \cdot A \cdot L = p \cdot V \,, \tag{9}$$

$V = A \cdot L$ being the volume of the steam in the cylinder. If the piston travels a small distance dL, the work done by the steam is

$$\mathrm{d}W = p \cdot \mathrm{d}V \,, \quad \mathrm{d}V = A \cdot \mathrm{d}L \,. \tag{10}$$

These expressions were known to Watt; the last one is important when calculating the work of expanding steam. Thus far the volume work of steam at constant pressure according to Watt's idea.

2) The expansion work at constant temperature

For calculating the work of steam expanding at constant temperature, Watt devised a thought experiment which consisted in cutting off the steam filling the cylinder at certain point, before the whole cylinder becomes full of steam, and allowing the supplied steam to expand in the cylinder. **Figure 5** illustrates the core of the Watt's idea which was included into the patent specification of 1782. It shows the calculated pressure distribution along the cylinder length if cutting off of the steam supply occurs at the steam volume occupying 1/4 of the cylinder volume. The supplying steam is nearly at atmospheric pressure.

Watt showed in the patent that the specific power of the steam engine is not reduced if the steam admitted into the cylinder does not fill its entire volume. The subsequent steam expansion was described by equation of an ideal gas. The

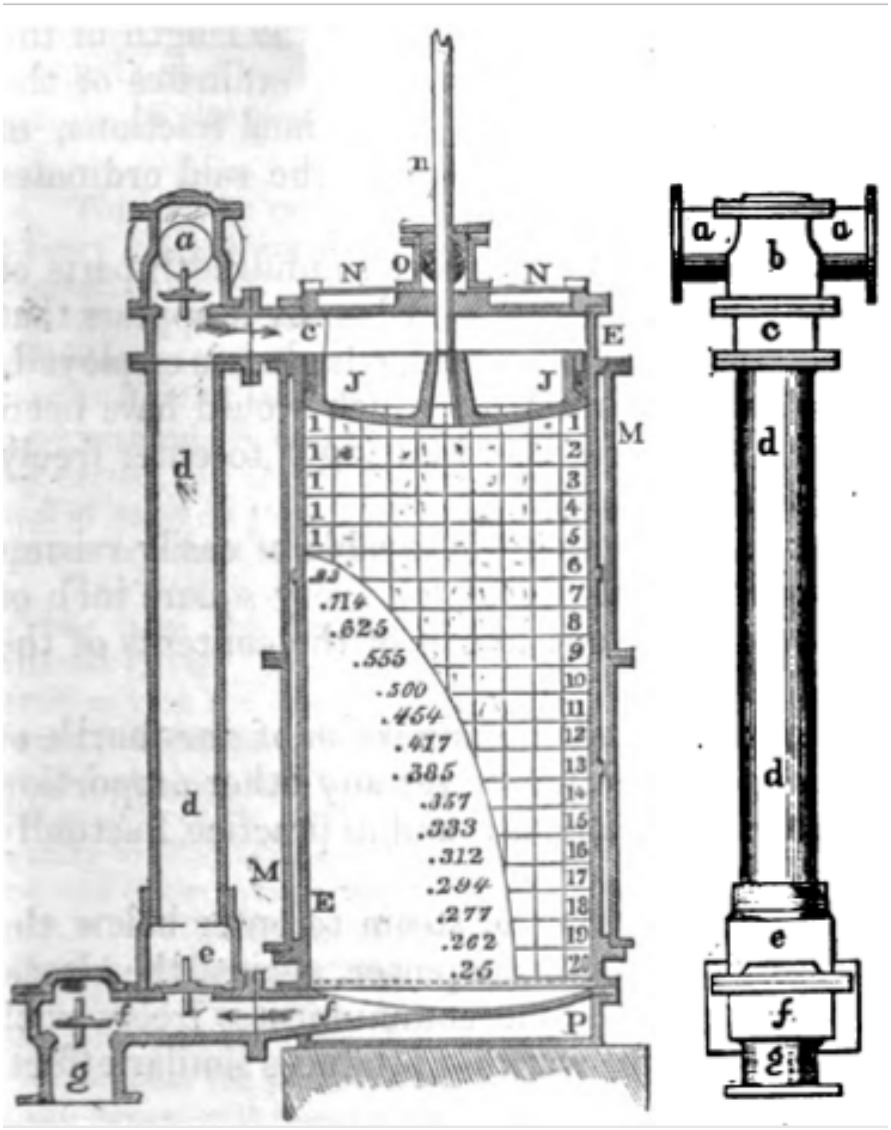


Figure 5. Isothermal steam expansion according to Watt's calculations as part of his patent specification of 1782. Pressure distribution along the cylinder when steam admission cut-off occurs at a steam volume in the cylinder equal to 1/4 of the cylinder volume, Farey [2], p. 347.

curve in **Figure 5** corresponds to this equation, the numbers along the curve represent the calculated pressure divided by the pressure of supplied steam prior to expansion. Every two neighboring horizontal lines symbolize circular surfaces which together with the cylinder circumference define an element of steam volume V_k, the index k referring to the k-th element (of total 20 identical elements). The volume work W_k of the k-th element is given by

$$W_k = p_k \cdot V_k, \tag{11}$$

which is basically the Equation (10). Since the volume V_k of the elements does not change along the cylinder, we can write

$$V_k = l_k \cdot A_k = l \cdot A = const$$

$$W_{exp} = \sum W_k = \sum \left(p_k \cdot V_k \right) = l \cdot A \cdot \sum p_k \tag{12}$$

Setting in Equation (12) $l = L/K$, whereby K (=20) denotes the total number of compartments, and L the length of the stroke, Figure 5, we get the expansion work

$$W_{exp} = L \cdot A \cdot \left(\sum p_k \right) / K = V_{CYL} \cdot p_{av} \tag{13}$$

$$p_{av} = \left(\sum p_k \right) / K$$

Equation (13) expresses the Watt's calculation of the volume work of expanding steam. Its derivation coincides with the description Watt gave in his patent. Of the equations reported here he mentions only the sum of pressures of expanding steam, $\sum p_k$. Obviously, Watt applied the numerical integration to evaluate the expansion work. Farey's book provides some details of Watt's calculations, which are not given in the original patent of 1782, Farey [2], p. 341.

According to Watt's calculations the product of the average steam pressure p_{av} and the volume of the cylinder V_{CYL} is less with, than without, the steam expansion. However, dividing these products with the masses of steam taken from the boiler gives a larger value with, than without, the steam expansion. In other words, the work of steam per unit of mass of steam consumed is greater with the steam expansion.

Watt's assumption of an isothermal steam expansion leads to the weakest pressure change and the maximum benefit of the expansion process. Watt was aware of this correlation and tried to realize the isothermal conditions in practice. He surrounded the steam cylinder by a layer of steam having the same temperature, Figure 2. As ingenious as this measure may seem with regard to the thermal insulation of the cylinder, it would not have been possible to achieve isothermal, but rather an adiabatic steam expansion.

The benefits of steam expansion have been studied by several authors aiming at optimization of cutting-off point of the steam admission to the engine working expansively. As example we mention the investigations by Clark [21], published in 1852. Figure 6 taken from Clark's publication shows the indicator diagrams obtained at various process conditions. The experiment N°1 provides apparently the largest expansion work, corresponding to the largest indicator area.

Watt's patent on steam expansion (1782) is a logical extension of his patent of 1769, where he already used steam expansion in the working step of the engine. By connecting the working cylinder filled with steam to the separate, empty and cold condenser, Figure 2, the steam expanded and enabled the piston to descend. Regarding the acceptance of Watt's steam expansion method in engineering practice and its use, particularly by Sadi Carnot, the reader is referred to the publication by Fox [22].

4.6. Condensation of Expanding Steam

In addition to the improvement of the engine performance by steam expansion, we shall also notice an important Watt's contribution to the foundations of

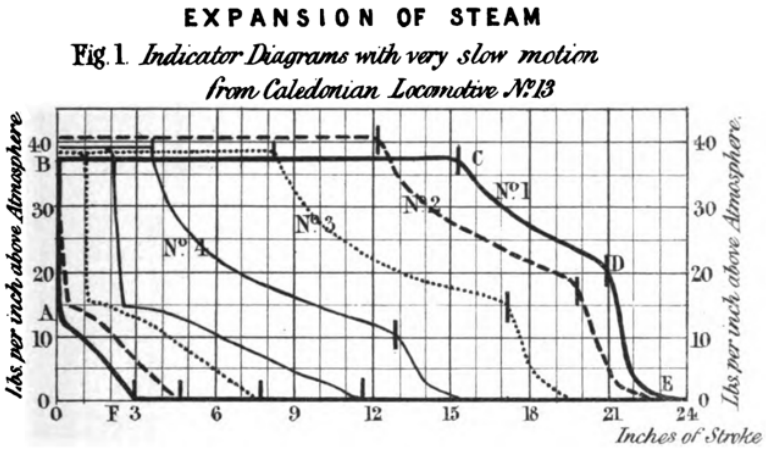


Figure 6. Effect of steam cut-off on the shape of the indicator diagram of expansively working steam engine: taken from Clark [21]. A—the *point of steam admission*; C—the *point of cutting-off, or suppression*; D—the *point of exhaust, or release*; F—the *point* of *compression*; also, the portion of the stroke described while the line AB is traced, is the period of *pre-admission*.

thermodynamics. Watt experimentally demonstrated the condensation of expanding steam in his laboratory at Soho, Birmingham. The occasion was the visit of Henry Cavendish and Charles Blagden on their "*Philosophical Tours*" in 1785. In this context, Miller [23] cites from the Cavendish's notebook:

Mr. Watt mentioned, that having found that some steam is condensed in the cylinder of the Steam engine, tho' surrounded with steam, he made an experiment to discover what happened. He threw steam into a Glass vessel close at the top (*By making it communicate with that part of the cylinder of this fire engine in which there is alternately Steam and Vacuum*) *and found that upon making the Vacuum some of the steam condensed on the sides of the Glass Vessel, and having heated the sides of this Vessel, so that none could condense upon them, he observed the condensation take place, so as to render the Steam visible in the middle of the Glass vessel* [. . .] *a cloud began to appear at each Vacuum* […]

Since the condensation requires a steam cooling, Watt was the first to realize that **expansion of a steam causes its cooling**. The condensation occurred also in the steam bulk, without influence of a solid surface, which is today known as homogeneous phase transition. Formation of viable nuclei of the new phase (tiny water droplets) under such conditions requires certain metastability of mother-phase, which is measured in terms of steam subcooling. In 1788, three years after Watt's experiments, Erasmus Darwin [24] (grandfather of Charles Darwin) published a paper on air cooling by expansion. Next experiments in this area were performed by Joule [25] in 1845, more than 5 decades after Watt's experiments. Watt's name is neither mentioned in Darwin's nor Joule's paper, nor in any other publication on cooling of gases by expansion.

The pressure diagram of expanding steam in **Figure 5** did not satisfy its originator (Watt) because it does not provide any information on the expansion dynamics. The science at that time could not provide any information on this point

and Watt was again required to devise a measuring device, this time a device that should indicate the steam pressure in the cylinder as function of the piston position. Thus, he later built and used in the experiments an instrument—the pressure indicator—for measurements of the pressure change in the steam cylinder.

4.7. The Watt's Indicator Diagram

In 1854 Rankine [26] published a paper and in 1859 a section in his book [27] (p. 301) on the expansive action of heat, including a *p, v*– diagram that corresponds to the famous diagram generated by the Watt's pressure indicator. The accompanying Rankine's text reads (1854):

The first application of a geometrical diagram to represent the expansive action of Heat was made by James Watt, when he contrived the well-known Steam-Engine Indicator, subsequently altered and improved by others in various ways. As the diagram described by Watt's Indicator is the type of all diagrams representing the expansive action of heat, its general nature is exhibited in Figure 1.

Figure 7 shows the shape of Watt's indicator diagram published as Figure 1 by Rankine in 1854. Rankine states that the area encircled by the closed curve in the *p, v*-coordinates (Figure 7) represent the useful energy provided by steam, without clearly mentioning James Watt as originator of this idea, see Watt patent of 1782. Actually, Rankine is inclined to acknowledge Sadi Carnot and Emile Clapeyron as the originators of the *p, v*-diagram, but he hesitated to do this, because he did not accept their material notion of heat. However, Watt was the first to define the volume (expansion) work and to identify the encircled area in the *p, v*–coordinates as work performed by expanding steam in the cylinder, see Figure 5. With the *p, v*-diagram Watt introduced graphics into thermodynamics and provided a possibility to look in the steam cylinder and to visualize the energy conversion.

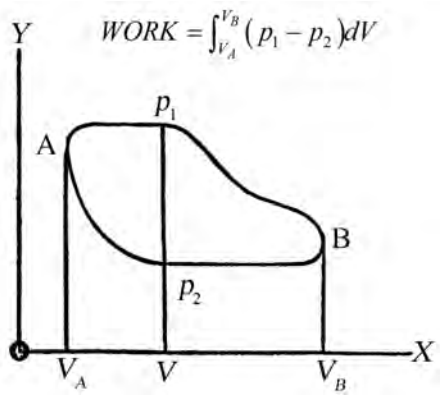


Figure 7. Rankine's sketch of Watt's indicator diagram (1854) supplemented by Watt's equation for expansion work, *Y* means pressure, *X* means the volume. (Rankine frequently stated the expression for calculating the expansion work but not clearly its origin. Already in his patent of 1782 Watt used such an expression, Figure 4. To show the details of his calculations, Watt expressed the integration by summation of elemental rectangles of the encircled area (numerical integration). We have included the integral form of Watt's expression as written by Rankine).

Watt used the pressure indicator and created several *p, v*–diagrams; from these he obtained the performance of his engines long before the next generation of scientists (including Carnot, Rankine, …) began to work in this field. As quoted above, Rankine acknowledges this too. Carnot also studied Watt's works and because of the clarity of Watt's ideas called him "***the famous Watt***" (Gillispie and Pisano [28]). Comparing Carnot's thermodynamics with Watt's thermodynamics, shown e. g. in **Figure 2**, we see that Carnot actually describes the process that Watt developed, published as drawings and used in practice for decades.

Lervig [29] extensively examined Sadi Carnot's studies of the works of James Watt, filling almost 50 pages. Attention was drawn to Carnot's involvement in calculating the steam engine's power. The heat of steam condensation was problematic because it wasn't clear whether to ignore or include it in the calculations. This is the quantity Q_C in our Equation (7) which gives an unambiguous answer: setting $Q_C = 0$ gives $Q_B = W$ and the heat Q_B would be completely converted into work, which is, however, thermodynamically impossible.

Based on Watts's contributions in the field of thermodynamics of energy conversion, one would expect his name to be at, or very close to, the central position regarding the history of thermodynamics. This is far from being the case! Watt is not mentioned in this context, and if the unit of power does not bear his name, James Watt, one of the originators of thermodynamics as a science, would long be forgotten.

5. Conclusions

James Watt's works associated with the development of the steam engines are at the origins of energy conversion thermodynamics. His patent of 1769, usually considered to be an improvement of steam engine, actually describes an energy conversion process of general validity. In this patent, Watt defined the range of steam temperature for running the engine, some 60 years prior to Sadi Carnot. The patent has brought thermal engineering to the scientific level. With the second, 1782 patent, Watt formulated and introduced the volume and expansion work of steam and enriched the general understanding of energy conversion.

Watt's works on latent heat of steam guided him to the conclusion that the latent heat of steam decreases with increasing temperature, becoming zero at certain temperature while the steam-water system assumes a new state which later was termed the thermodynamic critical state. Besides of latent heat of steam, he defined the total heat of steam, today known as the enthalpy of saturated steam. He was the first to observe condensation of expanding steam and to visualize energy conversion by using diagrams.

The most important Watt's contribution to the development of thermodynamics is probably his formulation of the first law of this discipline in 1778. However, the Watt's formulation went unrecognized and—as far as the author is aware—was first described in the present paper.

This paper supplements the previous work [3] and demonstrates the outstanding contributions of James Watt to understanding of the thermal processes. However, Watt's works were mostly considered to be prehistorical, which apparently motivated researcher to exclude him from the history of thermodynamics.

Acknowledgements

The author acknowledges the comments of Prof. A. Smyk on earlier draft of the paper.

A referee has carefully analyzed the content of the manuscript and stated several important questions.

Conflicts of Interest

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

References

[1] Watt, J. (1818) The Articles Steam and Steam-Engines. John Murray, London.

[2] Farey, J. (1827) A Treatise on the Steam Engine: Historical, Practical, and Descriptive. David & Charles, London.

[3] Mitrovic, J. and Smyk, A. (2021) *Voprosy istorii estestvoznaniia i tekhniki* (*Studies in the History of Science and Technology*), **42**, 397-442. (in Russian). https://doi.org/10.31857/S020596060016355-2

[4] Saslow, W.M. (2020) *Entropy*, **22**, Article No. 77. https://doi.org/10.3390/e22010077

[5] Stuart, R. (1824) A Descriptive History of the Steam Engine. J. Knight and H. Lacey, London.

[6] Black, J. and Robison, J. (1803) Lectures on the Elements of Chemistry. Vol. 1, Mundell and Son, Philadelphia, 151-152.

[7] Cardwell, D.S.L. (1971) From Watt to Clausius: The Rise of Thermodynamics in the Early Industrial Age. Cornell University Press, Ithaca, 161.

[8] Miller, D.P. (2004) *History of Science*, **42**, 333-360. https://doi.org/10.1177/007327530404200304

[9] Capecchi, D. (2021). Epistemology and Natural Philosophy in the 18th Century: The Roots of Modern Physics. Chapter 5, Springer Nature, Cham, 473-549. https://doi.org/10.1007/978-3-030-52852-2

[10] Watt, J., Muirhead, J.P., Murray, J. and Constable, T. (1846) Correspondence of the Late James Watt on his Discovery of the Theory of the Composition of Water. W. Blackwood & Sons, London.

[11] Siemens, C., (1853) *Minutes of the Proceedings of the Institution of Civil Engineers*, **12**, 571-590.

[12] Berche, B., Henkel, M. and Kenna, R. (2009) *Journal of Physical Studies*, **13**, Article No. 3201, arXiv:0905.1886[physics.hist-ph]. https://doi.org/10.30970/jps.13.3001

[13] Watt, J. (1769) Method of Lessening the Consumption of Steam and Fuel in Fire Engines. https://en.wiktionary.org/wiki/File:James_Watt_Patent_1769_No_913.pdf

[14] Miller, D.P. (2006) *History of Technology*, **27**, 43-76. https://doi.org/10.1080/01445340500290144

[15] Pacey, A.J. (1974) *The British Journal for the History of Science*, **7**, 135-145. https://doi.org/10.1017/S000708740001311X

[16] Cardwell, D.S.L. (1967) *The British Journal for the History of Science*, **3**, 209-224. https://doi.org/10.1017/S0007087400002661

[17] Joule, J.P. (1850) *Philosophical Transactions of the Royal Society*, **140**, 61-82. https://doi.org/10.1098/rstl.1850.0004

[18] Collins, M.W. (2015) *WIT Transactions on State of the Art in Science and Engineering*, **89**, 231-236. https://doi.org/10.2495/978-1-84564-149-8/009

[19] Newton, I., Bernard Cohen, I., Einstein Albert (1730) Opticks: Or a Treatise of the Reflections, Refractions, Inflections and Colours of Light. Querry 31. 4th Edition, Printed for Sam. Smith, and Benj. Walford, London, 295.

[20] Muirhead, J.P. (1854) The Origin and Progress of the Mechanical Inventions of James Watt. Vol. III, 55-87, Murray, London.

[21] Clark, D.K. (1852) *Proceedings of the Institution of Mechanical Engineers*, **3**, 60-88. https://doi.org/10.1243/PIME_PROC_1852_003_011_02 https://books.google.bj/books?id=1GEAAAAAMAAJ&hl=fr&source=gbs_book_other_versions_r&cad=3

[22] Fox, R. (1970) *Notes and Records of the Royal Society of London*, **24**, 233-253. https://doi.org/10.1098/rsnr.1970.0016 http://www.jstor.org/stable/531291

[23] Miller, D.P. (2011) *The International Journal for the History of Engineering & Technology*, **81**, 129-150. https://doi.org/10.1179/175812110X12869022260231

[24] Darwin, E. (1788) *Philosophical Transactions of the Royal Society*, **78**, 43-52. https://doi.org/10.1098/rstl.1788.0005

[25] Joule Esq, J.P. (1845) *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **26**, 369-383. https://doi.org/10.1080/14786444508645153

[26] Rankine, W.J.M. (1854) *Philosophical Transactions of the Royal Society of London* (1776-1886), **144**, 115-175. https://doi.org/10.1098/rstl.1854.0007

[27] Rankine, W.J.M. (1869) A Manual of the Steam Engine and other Prime Movers. 4th Edition, Griffin, Bohn, and Company, London, 301.

[28] Gillispie, C.C. and Pisano, R. (2013) Lazare and Sadi Carnot: A Scientific and Filial Relationship. Springer, Dordrecht, 167. https://doi.org/10.1007/978-94-017-8011-7

[29] Lervig, P. (1985) *The British Journal for the History of Science*, **18**, 147-196. https://doi.org/10.1017/S000708740002210X

[30] Thompson, B. (1798) *Philosophical Transactions of the Royal Society of London*, **88**, 80-102. https://doi.org/10.1098/rstl.1798.0006

[31] Boyle, R. (1675) Experiments, Notes, &c., about the Mechanical Origine or Production of Divers Particular Qualities: Among Which Is Inserted a Discourse of the Imperfection of the Chymist's Doctrine of Qualities; Together with Some reflections upon the hypothesis of Alcali and Acidum. E. Flesher, London. https://doi.org/10.5962/bhl.title.65740

[32] Leibniz, G.W. (2017) Essay in Dynamics. Jonathan Bennett, Bowen Island, §20, 6. https://www.earlymoderntexts.com/assets/pdfs/leibniz1695b.pdf

[33] Spector, M. (1975) *Studia Leibnitiana*, **7**, 135-144.

[34] Keeton, M.T. (1941) *Philosophy of Science*, **8**, 304-319. https://doi.org/10.1086/286706

[35] Valenti, P. (1979) *Fusion Magazine*, **3**, No.3, 27-46.

Appendix: Conversion of Mechanical Energy in Heat

The first artificial conversion of mechanical energy in thermal energy (heat) is thought to have been achieved with the creation of fire based on mechanical friction. Usually, Graf Rumford [30] (Benjamin Thompson, 1753-1814) is credited with the idea to be the first who conducted experiments in this area of energy conversion. In the experiments (… *boring of canon in the workshops of the military arsenal in Munich* …) conducted in 1798 he tried to answer the question: ***From whence comes the heat actually produced in the mechanical operation above stated?***

More than a century prior to Rumford, in 1675, Robert Boyle (1627-1691)[5] describes several methods of mechanical heat generation [31]. Despite the fact that Boyle's work provides excellent insights into the kinetic nature of heat, it is not mentioned in relevant publication. In Sect. II, p. 40, **Of the Mechanical Origin or Production of Heat,** Boyle notes:

… *the nature of heat consists mainly, if not only, in that Mechanical affection of matter we call Local motion mechanically modified*, …

On page 59, the experiment VI, he discusses the production of heat:

… *when a Smith does hastily hammer a nail or such like piece of iron, the hammer'd metal will grow exceeding hot, and yet there appears not anything to make it so, save the forcible motion of the hammer which impresses a vehement and variously determin'd agitation of the small parts of the Iron; which being a cold body before becomes … exceedingly hot*, …

Boyle recognized two important properties of heat:

1) Heat can be understood as motion of smallest parts of bodies,

2) Heat is produced by mechanical action (hammering of iron piece).

Consequently, heat cannot be of substantial origin, Boyle concluded, it is the magnitude of the process.

Boyle's work of 1675 precisely answers the Rumford's question stated in 1798. As follows from b), mechanical action produces heat. By incorporating the principle *cause equals effect* (G.W. Leibniz, 1646-1716): *There is neither more nor less power in an effect than in its cause* [32]), one arrives directly at a relationship between the mechanical work and heat, and Boyle could have determined the heat equivalent to the mechanical work that produced it from the following balance: Motion of gross body (kinetic energy of hammer) generates local motion (kinetic energy) of small parts of hammered piece. Identifying the kinetic energy of local motion as heat results in conversion of mechanical (kinetic) energy in heat.

The contacts of hammer and the piece of iron in Boyle's experiments can be viewed as a succession of inelastic impacts of two bodies. This immediately allows application of Leibniz's idea of conservation of energy – *vis viva* of 1692 in

[5]Robert Boyle was one of the founders of The Royal Society of London for Improving Natural Knowledge, established in 1660. The motto of the Society was: *Nullius in verba*, which could mean *Don't take anyone's words for granted*. The intention was to support experiments in scientific research.

an inelastic impact [33]. Since one body (hammered iron piece) is at rest, its kinetic energy (*vis viva*) is zero during the impact process. Also, the kinetic energy of the hammer after the impact is zero, which means that its pre-impact *vis viva* is completely transformed in local motion (heat) of iron during impact.

Actually, Leibniz assumed (in 1692) that *vis viva* in an inelastic collision is not conserved and that the apparent loss of this quantity is caused by its distribution among the small elements of impacting bodies, which is in agreement with the Boyle's experiment published in 1675, more than two decades earlier. This loss of *vis viva* is of local character and it is limited to the actual impact. Even in this case there is actually not a loss of *vis viva*, as Leibniz says: *For that which is absorbed by the minute parts is not absolutely lost for the total force of the concurrent bodies* [33], p. 144). Keeton provides a detailed discussion on the meaning of the terms conservation and conversion of energy [34].

In this context I shall mention a publication by Valenti of 1979 [35] which I happened to come across after the manuscript has been completed. Valenti states and credits Leibniz with the following balance equation:

$$\begin{aligned} &\text{vis viva consumed by machine} \\ &= \text{useful work}\left(\text{height a given quantity of wateris raised}\right) \\ &\quad + \text{heat lost in overcoming friction} \\ &\quad + \text{heat lost to superfluous cooling} + \text{other inefficiencies} \end{aligned} \tag{A1}$$

The equation expresses the equivalence of *vis viva* of fundamental elements of expanding steam and work done by steam in a steam engine. According to Leibniz, the work required to raise a heavy body to a certain height corresponds to the *vis viva* that the body acquires when falling from rest that height (Galileo's model). He suggested measuring the *vis viva* by the steam temperature, which thus also includes the vapor pressure into the balance. This equation shows the transformation of *vis viva* and its preservation (in other forms) in the whole process of the transformation. According to Boyle model sketched above, we can take the ***vis viva*** analogously to the heat of the steam; then the content of the Equation (A1) becomes identical to Equation (6) in the main text. However, Equation (A1) does not seem to exist in Leibniz's works which are accessible to the author of the present paper. Its existence does not emerge from Watt's works either. The actual author of Equation (A1) remains at present unknown.

Journal of Modern Physics, 2022, 13, 410-413
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Invariance of the Electromagnetic Field Vectors Obtained in Course of the Lorentz Transformation Characteristic for the Relativistic Theory

S. Olszewski

Institute of Physical Chemistry, Polish Academy of Sciences, Warsaw, Poland
Email: solszewski@ichf.edu.pl

How to cite this paper: Olszewski, S. (2022) Invariance of the Electromagnetic Field Vectors Obtained in Course of the Lorentz Transformation Characteristic for the Relativistic Theory. *Journal of Modern Physics*, **13**, 410-413.
https://doi.org/10.4236/jmp.2022.134028

Received: March 7, 2022
Accepted: April 15, 2022
Published: April 18, 2022

Abstract

The invariance of several new component electromagnetic-field vectors with respect to the Lorentz transformation has been demonstrated in the paper. The formalism of the classical relativistic mechanics has been applied in examining both the time-square variable of the field, as well as the square-values of the position coordinates of a moving particle.

Keywords

Lorentz Transformation, Electromagnetic Field Vectors

1. Introduction

The aim of the paper is to demonstrate the invariance of some new components of the electromagnetic field with respect to the Lorentz transformation characteristic for the special relativity. Well-known results of this kind were obtained a time ago for the mechanical parameters (see e.g. [1]). In more recent calculations—see [2], the invariance of the difference of two coordinate squares, say the time t, and one of the Cartesian coordinates of position, say x, has been found:

$$t^2 - x^2 = t'^2 - x'^2 \tag{1}$$

The parameters t and x entering (1) have been coupled by the Lorentz transformation giving t' and x':

$$t' = \frac{t - vx}{\sqrt{1 - v^2}} \tag{2}$$

and

$$x' = \frac{x - \nu t}{\sqrt{1-\nu^2}}. \tag{3}$$

Here

$$\nu = v/c \tag{3a}$$

where v is the actual velocity of the system directed along the coordinate x and c is a speed of light. The properties of the other coordinate transformations are

$$y = y' \tag{4}$$

and

$$z = z'. \tag{5}$$

In the present paper a special interest can be attributed to the electric and magnetic vector fields, viz.

$$\boldsymbol{E} \text{ and } \boldsymbol{H} \tag{6}$$

and their Lorentz transformations. A list of such transformations is given in [3].

Perhaps the best known results concerning $\boldsymbol{E}$ and $\boldsymbol{H}$ are [3]:

$$EH = \text{invariant}, \tag{7}$$

$$H^2 - E^2 = \text{invariant}. \tag{8}$$

The aim of the paper is approached in two steps. In the first one—a less accurate one—we identify the variables t and x examined in [2] with some special components of the electromagnetic fields.

Another, more accurate calculation, makes a reference of the electromagnetic field vector components submitted to the motion without a reference to t and x; see Section 3. In this case only two pairs of field components submitted to the motion are examined.

2. New Invariant Relations Presented for the Vectors *E* and *H*

Relations in [3] being similar to (4) and (5) are:

$$E_x = E'_x, \tag{9}$$

$$H_x = H'_x. \tag{10}$$

When (9) and (10) are combined with (4) and (5), we obtain

$$E_x \to E'_x = y \to y', \tag{11}$$

$$H_x \to H'_x = z \to z'. \tag{12}$$

In the next step (see in ([3], Section 23)) we can put

$$x' = H_y, \tag{13}$$

$$x = H'_y, \tag{14}$$

$$t' = E_z, \tag{15}$$

$$t = E'_z, \tag{16}$$

and examine transitions

$$t \to t', \tag{17}$$

$$x \to x'. \tag{18}$$

Other identifications can be

$$H_y' = t, \tag{19}$$

$$E_z' = x, \tag{20}$$

$$H_y = t', \tag{21}$$

$$E_z = x'. \tag{22}$$

A combination of the Equations (13)-(16) leads to the difference

$$\begin{aligned} t'^2 - x'^2 = E_z^2 - H_y^2 &= \left(\frac{t - \nu x}{\sqrt{1-\nu^2}}\right)^2 - \left(\frac{x - \nu t}{\sqrt{1-\nu^2}}\right)^2 \\ &= \frac{\left(E_z' - \nu H_y'\right)^2}{1-\nu^2} - \frac{\left(H_y' - \nu E_z'\right)^2}{1-\nu^2} \\ &= \frac{1}{1-\nu^2}\left[E_z'^2 - 2E_z'H_y'\nu + \nu^2 H_y'^2 - H_y'^2 + 2E_z'H_y'\nu - \nu^2 E_z'^2\right] \\ &= \frac{1}{1-\nu^2}\left[\left(1-\nu^2\right)E_z'^2 - \left(1-\nu^2\right)H_y'^2\right] \\ &= E_z'^2 - H_y'^2 = t^2 - x^2. \end{aligned} \tag{23}$$

A similar calculation can be done on the basis of (19) - (22):

$$\begin{aligned} t'^2 - x'^2 = H_y^2 - E_z^2 &= \left(\frac{t - \nu x}{\sqrt{1-\nu^2}}\right)^2 - \left(\frac{x - \nu t}{\sqrt{1-\nu^2}}\right)^2 \\ &= \frac{1}{1-\nu^2}\left[\left(H_y' - \nu E_z'\right)^2 - \left(E_z' - \nu H_y'\right)^2\right] \\ &= \frac{1}{1-\nu^2}\left[H_y'^2 - 2H_y'E_z'\nu + \nu^2 E_z'^2 - E_z'^2 + 2E_z'H_y'\nu - \nu^2 H_y'^2\right] \\ &= \frac{1}{1-\nu^2}\left[\left(1-\nu^2\right)H_y'^2 - \left(1-\nu^2\right)E_z'^2\right] \\ &= H_y'^2 - E_z'^2 = t^2 - x^2. \end{aligned} \tag{24}$$

In effect beyond of (7) and (8) we obtained two pairs of the electromagnetic field vectors which remain invariant upon the action of the Lorentz transformations:

$$1)\quad t'^2 - x'^2 = E_z^2 - H_y^2 = E_z'^2 - H_y'^2 = t^2 - x^2 \tag{25}$$

and

$$2)\quad t'^2 - x'^2 = H_y^2 - E_z^2 = H_y'^2 - E_z'^2 = t^2 - x^2. \tag{26}$$

3. Two Components of the Electromagnetic Field Vectors Taken to Calculations

Expressions (25) and (26) can be considered only as an approximate result because the dimensions of t'^2, x'^2 or t^2, x^2 differ from dimensions of

$E_z'^2, H_y'^2, E_z^2, H_y^2$ of the field counterparts. In order to get precise results we take into account the field components entering (25) and (26).

On the basis of [3], Section 23 we have

$$E_y = \frac{E_y' + \nu H_z'}{\left(1-\nu^2\right)^{1/2}}, \tag{27}$$

$$H_y = \frac{H_y' - \nu E_z'}{\left(1-\nu^2\right)^{1/2}}, \tag{28}$$

$$E_z = \frac{E_z' - \nu H_y'}{\left(1-\nu^2\right)^{1/2}}, \tag{29}$$

$$H_z = \frac{H_z' + \nu E_y'}{\left(1-\nu^2\right)^{1/2}} \tag{30}$$

On that basis because of (9) and (10) we obtain

$$\begin{aligned} \boldsymbol{E}^2 - \boldsymbol{H}^2 &= E_y^2 + E_z^2 - H_y^2 - H_z^2 \\ &= \left(E_y'^2 + 2\nu E_y' H_z' + \nu^2 H_z'^2 + E_z'^2 - 2\nu H_y' E_z' + \nu^2 H_y'^2\right)\frac{1}{1-\nu^2} \\ &\quad -\left(H_y'^2 - 2\nu H_y' E_z' + \nu^2 E_z'^2 + H_z'^2 + 2\nu H_z' E_y' + \nu^2 E_y'^2\right)\frac{1}{1-\nu^2} \\ &= \left(E_y'^2 + E_z'^2 + \nu^2 H_z'^2 + \nu^2 H_y'^2 - H_y'^2 - H_z'^2 - \nu^2 E_z'^2 - \nu^2 E_y'^2\right)\frac{1}{1-\nu^2} \\ &= E_y'^2 + E_z'^2 - H_y'^2 - H_z'^2. \end{aligned} \tag{31}$$

This result-together with (9) and (10) proves the invariance of the difference

$$\boldsymbol{E}^2 - \boldsymbol{H}^2 \tag{32}$$

upon the Lorentz transformation.

Acknowledgements

The paper is dedicated to the memory of blessed Pier-Georgio Frassati suddenly deceased in Italy in 1925.

Conflicts of Interest

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

References

[1] Sommerfeld, A. (1949) Mechanik. 4th Edition, Akademische Verlagsgesellschaft, Leipzig.

[2] Susskind, L. and Friedeman, A. (2019) Special Theory of Relativity and Field Theory. Pruszynski Media, Warsaw (in Polish).

[3] Landau, L.D. and Lifshits E.M. (1948) Field Theory. 2nd Edition. OGIZ, Moscow (in Russian).

Journal of Modern Physics, 2022, 13, 414-431
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Ab Initio Calculation of Accurate Electronic and Transport Properties of Zinc Blende Gallium Antimonide (zb-GaSb)

Yacouba Issa Diakite[1], Yuriy Malozovsky[2], Cheick Oumar Bamba[3], Lashounda Franklin[2], Diola Bagayoko[2]

[1]Department of Studies and Research (DSR) in Physics, Center of Calculation, Modeling and Simulation (CCMS), College of Sciences and Techniques (CST), University of Sciences, Techniques, and Technologies of Bamako (USTTB), Bamako, Mali
[2]Department of Mathematics and Physics (DMP), Southern University and A & M College, Baton Rouge, LA, USA
[3]Louisiana State University (LSU), Baton Rouge, LA, USA
Email: yacouba_idiakite@yahoo.fr

How to cite this paper: Diakite, Y.I., Malozovsky, Y., Bamba, C.O., Franklin, L. and Bagayoko, D. (2022) Ab Initio Calculation of Accurate Electronic and Transport Properties of Zinc Blende Gallium Antimonide (zb-GaSb). *Journal of Modern Physics*, **13**, 414-431.
https://doi.org/10.4236/jmp.2022.134029

Received: February 8, 2022
Accepted: April 16, 2022
Published: April 19, 2022

Abstract

This article reports the results of our investigations on electronic and transport properties of zinc blende gallium antimonide (zb-GaSb). Our ab-initio, self-consistent and non-relativistic calculations used a local density approximation potential (LDA) and the linear combination of atomic orbital formalism (LCAO). We have succeeded in performing a generalized minimization of the energy, using the Bagayoko, Zhao and Williams (BZW) method, to reach the ground state of the material while avoiding over-complete basis sets. Consequently, our results have the full physical content of density functional theory (DFT) and agree with available, corresponding experimental data. Using an experimental room temperature lattice constant of 6.09593Å, we obtained a direct band gap of 0.751 eV, in good agreement with room temperature measurements. Our results reproduced the experimental locations of the peaks in the total density of valence states as well as the measured electron and hole effective masses. Hence, this work points to the capability of ab-initio DFT calculations to inform and to guide the design and the fabrication of semiconductor based devices—provided a generalized minimization of the energy is performed.

Keywords

Gallium Antimonide, BZW Method, Self-Consistent Calculation, Density Functional Theory, Band Gap

1. Introduction

Gallium antimonide (GaSb) is one of the semiconductor compounds of the III-V family, derived from gallium and antimony; its stable crystal structure is zinc blende. It is a direct gap semiconductor which has the possibility of being p- or n-type doped, with good mobility. It also has a significant electro-optical potential in the infrared domain [1]. GaSb has emerged in recent years as a very technologically attractive semiconductor, given its applications in high-efficiency thermo-photovoltaics, mid-infrared lasers, photodetectors, high-speed electronic devices, and non-linear optics. Such applications derive from several interesting properties of the material, such as the high hole mobility (850 - 10,800 cm^2/Vs), the low carrier effective masses and the small direct band gap value [2]. The importance of its technological applications makes it an extensively studied semiconductor, both theoretically and experimentally.

Table 1 shows a list of results from previous DFT and other calculations of the electronic properties of zb-GaSb. Of the ab-initio local density approximation (LDA) calculations, the first six (6) reported negative band gaps, while the next five (5) found a zero gap. The following twenty-nine (29) results obtained with an ab-initio or ad hoc LDA potential range from 0.05 eV to 1.20 eV. Ad hoc potentials have no predictive capability, as the results depend on the values of the various parameters used in their construction. The computational approaches for the calculations in the table are primarily the full potential linearized augmented plane wave (FP-LAPW). Some authors have used the projected augmented wave (PAW), Gaussian-type orbitals (GTO) or the standard pseudopotential and plane wave method. These approaches are different implementations of the linear combination of atomic orbitals (LCAO).

The calculations with ab-initio, generalized gradient approximation (GGA) potentials follow the LDA ones in the table. The resulting, calculated band gap values of zb-GaSb are not satisfactory. Indeed, of the first thirteen (13) ab-initio GGA results for the band gap, the four (4) are negative and the remaining 9 are zero. The following twenty-seven (27) results in the table, obtained with GGA ab-initio potentials, range from 0.06 eV to 1.015 eV. Five (5) of these results were obtained with the Engel and Vosko GGA which tends to overestimate band gaps. One of these results, 0.726 eV for an indirect band gap, disagrees qualitatively with experiment that reports a direct band gap.

The above ab-initio calculations are followed in the table by several calculations with ad hoc potentials. In particular, two (2) of these calculations, with an empirical pseudopotential approach, reported band gaps of 0.62 eV and 0.715 eV. Two (2) of the calculations used a version of the modified Becky and Johnson potential and eight (8) used a hybrid potential HSE06 or HSE. Five (5) calculations, with Green's function and the dressed Coulomb approximation (GWA), were not DFT calculations. Their results are between 0.51 eV and 0.85 eV, a wide range. The results of other types of ad hoc potentials such as TPSS, HISS, B3LYP and Christensen are presented in the table. We recall that the

Table 1. Calculated band gaps (E_g, in eV) of zinc blende GaSb, along with pertinent lattice constants in Angstroms, and experimental values reported in the literature.

Computational Formalism	Potentials (DFT and others)	a (Å)	E_g (eV)
PW-PP	LDA	6.089	−0.348 [a]
PW-PP	LDA	6.043	−0.168 [b]
FP-LAPW	LDA	6.107	−0.165 [c]
FP-LAPW	LDA	6.096	−0.13 [c]
Ab initio pseudopotential	LDA		−0.10 [d]
FP-LAPW	LDA		−0.09 [e]
FP-LMTO	LDA	6.091	0.00 [f]
FP-LAPW + lo	LDA		0.00 [g]
FP-LAPW	LDA	6.0619	0.00 [h]
Ab initio pseudopotential	LDA	6.0959	0.00 [i]
Ab-initio-DFT in Plane wave	LDA		0.00 [j]
LMTO	vLB (LDA + vBL)		0.05 [k]
Hybrid exchange methodology	LSDA		0.09 [l]
PAW	LDA	6.065	0.13 [m]
Hybrid functional method	US-LDA		0.23 [m]
NMTO	LDA	6	0.25 [k]
Generalized density-functional theory (GDFT)	LDA	6.12	0.259 [n]
PW-PP	LDA	6.08	0.398 [o]
FP-LAPW	sX-LDA + SOC		0.43 [e]
LMTO	LDA		0.52 [k]
PW-PP	LDA	5.981	0.547 [p]
Ab initio pseudopotential	LDA	5.938	0.5885 [i]
FP-LAPW	DOS-mBJ-LDA		0.60 [q]
Sx- FLAPW	Sx-LDA		0.67 [r]
FP-LAPW + lo	mBJLDA	6.2086	0.68 [s]
Projector-augmented-wave (PAW)	MBJLDA		0.73 [t]
Linear-muffin-tin orbital within atomic-spheres approximation (LMTO-ASA)	LDA		0.74 [u]
Projector-augmented-wave (PAW)	MBJLDA		0.75 [t]
OLCAO	LDA		0.80 [v]
FLAPW	LDA	6.07181	0.80 (indirect) [w]

Continued

PW-PP (with correction)	LDA	6.043	0.81 [b]
Projector-augmented-wave (PAW)	$MBJLDA_{bgfit}$		0.82 [t]
FLAPW	LDA	6.07181	0.82 (indirect) [w]
FP-LAPW + lo	mBJLDA	6.1268	0.90 [s]
Generalized density-functional theory (GDFT)	LDA	5.95	0.925 [n]
FP-LAPW	mBJ-LDA		0.94 [q]
NMTO	vLB	6	0.94 [k]
FP-LAPW	mBJ-LDA	6.0619	1.036 [h]
FLAPW	LDA	6.07181	1.10 [w]
FP-LAPW + lo	mBJLDA	6.0449	1.20 [s]
Projector-augmented-wave (PAW)	GGA-PBE + SOC	6.203	−0.477 [x]
Projector-augmented-wave (PAW)	GGA-PBE	6.203	−0.259 [x]
FP-LAPW	Relativistic GGA	6.22	−0.35 [y]
Projector-augmented-wave (PAW)	GGA-PBE		−0.11 [t]
PAW	GGA	6.228	0.00 [m]
Hybrid functional method	US-GGA		0.00 [m]
FP-LAPW	WC-GGA		0.00 [z]
FP-LAPW	WC-GGA	6.107	0.00 [a']
FP-LAPW	PBEsol-GGA		0.00 [h]
FP-LAPW + lo	WC-GGA	6.115	0.00 [b']
PAW	PBE	6.237	0.00 [m]
middle-range screened hybrid exchange functional	PBE		0.00 [c']
Projector-augmented-wave (PAW)	PBE		0.00 [d']
middle-range screened hybrid exchange functional	PBE_{sol}		0.06 [c']
FP-LAPW + lo	GGA		0.07 [e']
FP-LAPW	GGA-PBE		0.11 [r]
FP-LAPW	GGA	6.09593	0.17 [f']
PAW	GGA-PBE	6.22	0.20 [g']
PW-PP	GGA-PBE	6.26	0.208 [o]
PW-PP	GGA-PW91	6.259	0.217 [o]
FP-LAPW	GGA-PBE	6.2145	0.220 [h']

Continued

PW-PP	GGA-RPBE	6.26	0.236 [o]
FP-LAPW	EV-GGA	6.1231	0.35 [h]
PW-PP	GGA-TPSS	6.259	0.352 [o]
FP-LAPW + lo	EV-GGA		0.361 [g]
FP-LAPW + lo	EV-GGA		0.396 [b']
FP-LAPW	GGA	6.193	0.4 [i']
PW-PP	GGA-PBEsol	6.09	0.428 [o]
PW-PP	GGA-AM05	6.091	0.482 [o]
NMTO	PBE	6	0.51 [k]
PW-PP	GGA-RTPSS	6.09	0.513 [o]
Hybrid functional method	GGA HSE06	6.141	0.62 [m]
PW-PP	PBE-GGA	6.06	0.63 [j']
FP-LMTO	GGA		0.726 (indirect) [k']
FP-LAPW	mBJ-GGA		0.734 [h']
FP-LAPW	PBE-GGA with SOI	6.118	0.812 [l']
FP-LAPW	mBJ-GGA	6.1231	0.844 [h]
FP-LAPW	non-relativistic EV-GGA	6.24	1.00 [y]
FP-LAPW	PBE-GGA without SOI	6.118	1.008 [l']
PW-PP	GGA-MBJ		1.015 [o]
GW method	GW		0.62 [n]
semiempirical	LDA	6.096	−0.38 [m']
middle-range screened hybrid exchange functional	Tao, Perdew, Staroverov and Scuseria (TPSS) Potential		0.09 [c']
FP-LAPW and LMTO			0.2 [n']
middle-range screened hybrid exchange functional	Henderson, Izmaylov, Scuseria and Savin (HISS) Potential		0.31 [c']
Projector-augmented-wave (PAW)	GW_0-PBE		0.51 [d']
Projector-augmented-wave (PAW)	HSE06 + SOC	6.124	0.614 [x]
Plane-wave pseudopotential			0.54 [o']

Continued

Projector-augmented-wave (PAW)	HSE		0.68 [d']
middle-range screened hybrid exchange functional	HSE		0.70 [c']
Empirical pseudopotential	Empirical pseudopotential	6.118	0.715 [q']
Projector-augmented-wave (PAW)	HSE06		0.72 [t]
Hybrid exchange methodology	HSE		0.72 [l]
FP-LAPW	MBJ		0.763 [r']
Projector-augmented-wave (PAW)	$GW^{TC\text{-}TC}$-HSE		0.77 [d']
Projector-augmented-wave (PAW)	HSE06	6.124	0.782 [x]
Projector-augmented-wave (PAW)	$GWΓ^{1}$-HSE		0.80 [d']
semiempirical	LDA + C	6.096	0.81 [m']
Hybrid exchange methodology	B3LYP		0.81 [l]
Projector-augmented-wave (PAW)	$HSE06_{\alpha}$ + SOC	6.124	0.819 [x]
Projector-augmented-wave (PAW)	HSE_{bgfit}		0.82 [t]
Projector-augmented-wave (PAW)	$G_0W_0{}^{TC\text{-}TC}$		0.85 [t]
FP-LAPW	mBJ		0.876 [a']
Projector-augmented-wave (PAW)	Christensen		0.88 [t]
	Experiment		
		300 K	0.67 [r']
		300 K	0.72 [3']
		300 K	0.729 [t']
		300 K	0.724 ± 0.005 [u']
		300 K	0.725 [v']
		300 K	0.73 [w', x']
		Room T	0.75 [y']
		Low T	0.78 [r']
		Low T	0.80 [n]
		Low T	0.809 ± 0.005 [z'']
		Low T	0.81 [a'']

Continued

	4.2 K	0.812 [b”]
	Low T	0.82 [l]
	Low T	0.822 [v’]

[a] Ref [3], [b] Ref [4], [c] Ref [5], [d] Ref [6], [e] Ref [7], [f] Ref [8], [g] Ref [9], [h] Ref [10], [i] Ref [11], [j] Ref [12], [k] Ref [13], [l] Ref [14], [m] Ref [15], [n] Ref [16], [o] Ref [17], [p] Ref [18], [q] Ref [19], [r] Ref [20], [s] Ref [21], [t] Ref [22], [u] Ref [23], [v] Ref [24], [w] Ref [25], [x] Ref [26], [y] Ref [27], [z] Ref [28], [a’] Ref [29], [b’] Ref [30], [c’] Ref [31], [d’] Ref [32], [e’] Ref [33], [f’] Ref [34], [g’] Ref [35], [h’] Ref [36], [i’] Ref [37], [j’] Ref [38], [k’] Ref [39], [l’] Ref [40], [m’] Ref [41], [n’] Ref [42], [o’] Ref [43], [p’] Ref [44], [q’] Ref [45], [r’]Ref [46], [s’] Ref [47], [t’] Ref [48], [u’] Ref [49], [v’] Ref [50], [w’] Ref [51], [x’] Ref [52], [y’] Ref [53], [z’] Ref [54], [a”] Ref [55], [b”] Ref [56].

calculations in this paragraph, which used ad hoc DFT potentials, have no predictive capabilities.

The theoretical results discussed above and shown in **Table 1** disagree with each other and with the experimental values of the band gap. Seven (7) of these experimental values, in the last 14 rows of **Table 1**, range from 0.67 eV to 0.75 eV, for eight (8) measurements at room temperature; and seven (7) other values are from 0.78 eV to 0.822 eV for low temperature experiments. The apparent, experimental agreement on a band gap around 0.81 eV, for low temperatures, and around 0.73 eV, for room temperature, contrasts strongly with the case of theoretical studies with 24 negative values or zero for the band gap of zb-GaSb. The disagreement between the theoretical values of the ab-initio DFT calculations and between these values and the accepted experimental values is the main motivation for this work. The importance of a correct, calculated band gap resides in the need to produce accurate optical and dielectric properties, densities of states, and effective masses. These quantities, among others, cannot be correctly calculated using an incorrect, theoretical band gap.

Our motivation to resolve the discrepancies noted above is further supported by the fact that the previous works of our group accurately described or predicted properties of semiconductors, using ab-initio DFT potentials [57] [58]. This feat was made possible by our use of the Bagayoko, Zhao and Williams (BZW) method or its enhancement by Ekuma and Franklin (BZW-EF). Bagayoko explained that these methods seek out and reach the ground state of a material without using over-completes basis sets. To our knowledge, none of the calculations in **Table 1** used successive, self-consistent calculations, with increasingly large, augmented basis sets, to reach verifiably the ground state of the material. The attainment of the ground state is required by the second DFT theorem and is essential for the results of a DFT calculation to have the full, physical content of DFT and to be consistent with experiment [57].

After this introduction, focused on the importance of the material and the previous theoretical and experimental studies, we describe in Section 2 our method of calculation and the details relating to the replication of our work. Section

3 is devoted to the presentation and discussion of our results for the band structure and the band gap, as well as the total and partial densities of states, and the effective masses of electrons and holes. A relatively succinct conclusion is set out in Section 4.

2. Computational Method

We used in this work the local density approximation (LDA) potential of Ceperley and Alder [59], as parameterized by Vosko, Wilk and Nusair [60]. We have applied the linear combination of atomic orbitals. Gaussian functions constitute the radial parts of these orbitals. We used a software package that was developed and refined over a decade at the US Department of Energy's Ames Laboratory, Ames, Iowa [61]. Our non-relativistic calculations utilized an experimental lattice constant, at room temperature. We applied the BZW method to perform the linear combination of atomic orbitals (LCAO). Our method allows the concomitant and self-consistent solution of the two coupled equations. The first is that of Kohn and Sham [62] while the second equation generates the density of charge—using only the wave functions of the occupied states. The second equation can be considered as a constraint on the Kohn-Sham equation (KS). Many other details on the BZW or BZW-EF method are widely explained in the previous articles [57] [63]-[69]. We describe the generalized minimization of the energy below.

Briefly, our approach uses successive and consistent computations, with basis augmented sets to perform a generalized energy minimization, as required by the second DFT theorem. If two consecutive calculations lead to the same occupied energies, these energies are local minimum the next, consecutive calculations lowers some of the occupied energies. However, if the third calculation produces the same occupied energies as the previous two, then these occupied energies have reached their absolute minima, in other words, the ground state. The necessary and sufficient criterion for stopping the computation is to have three consecutive calculations that produce identical, occupied energies.

In accordance with the above, our calculations necessarily begin with a small basis set which accommodate all the electrons of the system under study. Calculation II follows, with a basis set comprising that of Calculation I plus an orbital representing an excited state. We compare the occupied energies of the two self-consistent calculations: invariably, some occupied energies from calculation II are lower than their corresponding values from calculation I. We augment the basis set of calculation II, with an orbital, to perform calculation III. We compare the occupied energies from calculations II and III. We continue this process until three consecutive calculations produce the same occupied energies, indicating that we reached the ground state. Among these three (3) consecutive calculations, only the first, which has the smallest basis set, provides the DFT description of the ground state of the material [57] [58] [63]-[69]. The basis set of this calculation is the optimal basis set. The optimal basis set, once self-consistent

is achieved, leads to the ground state charge density of the material. Likewise, basis sets which are larger than the optima one, and which contain the optimal one, lead to the ground state charge density upon reaching self-consistency. As explained by our group in the publications referenced above and others, the use of these larger basis sets also lowers certain unoccupied energies: these lowered unoccupied energies are not due to a physical interaction in the Hamiltonian which has not changed from its value obtained with the optimal basis set. Incidentally, since the occupied energies do not change once, we reach the optimal basis set, the further lowering of some unoccupied energies because of using larger basis sets containing the optimum is one plausible explanation of the almost universal underestimation of band gaps and energy gaps by mainstream calculations [70]. Indeed, these calculations, to date, have used a single basis set which was deliberately chosen to be large in order to ensure completeness. Most often, these basis sets were over-complete for the description of the ground state [57].

Details relevant to the replication of our work follow. The crystal structure of GaSb is Zinc Blende [1]. We have used a measured value of the lattice constant at room temperature, 6.09593 Å. The self-consistent calculations for Ga^{3+} and Sb^{2-} provided the orbitals for the solid-state calculations. Gaussian functions are in the radial parts of atomic wave functions. We employed a set of even-tempered Gaussian exponents, with a minimum of 0.148 and a maximum of 0.93061×10^5 in atomic units, for Ga^{3+}. Nineteen (19) Gaussian functions were used for s and p orbitals, and 17 for d orbitals. Likewise, to describe Sb^{3-}, the exponents in Gaussian functions ranged from 0.125 to a maximum of 0.8×10^5. The numbers of Gaussian functions for s and p orbitals are 22 and that of the d orbitals is 19. A mesh 60 k points, with appropriate weights in the irreducible Brillouin zone, was used in the iterations for self-consistency. The error in calculating the valence charge was -0.0027647 for 44 electrons, or -6.28×10^{-5} per electron. The convergence criterion of the iterations is to have a difference not greater than 10^{-5} between the values of the potentials for two consecutive iterations. The number of iterations for this convergence was around 60. With the method described above and the associated calculation details, we studied zb-GaSb as presented below.

3. Results and Discussions

We present the successive calculations in **Table 2**. Columns 1, 2 and 3 of the table indicate respectively the number of a calculation, the valence state orbitals for Ga^{3+} and the valence state orbitals for Sb^{3-}. Columns 4 and 5 show the total number of valence functions and the direct band gap calculated at the Γ point, respectively. In this table, the occupied energies from calculations IV, V and VI are identical, within our calculation uncertainty of 5 meV. Therefore, Calculation IV provides the DFT description of zb-GaSb, in accordance with the above description of our method. In particular, the occupied energies from this

Table 2. The successive, self-consistent calculations of the BZW method for zb-GaSb.

Calculation number	Trial function for valence states of Ga^{+3}	Trial function for valence states of Sb^{-3}	No. of functions	Band Gap at Γ (in eV)
I	$3s^2 3p^6 3d^{10} 4s^2 4p^0$	$4s^2 4p^6 4d^{10} 5s^2 5p^4$	52	1.629
II	$3s^2 3p^6 3d^{10} 4s^2 4p^0 5s^0$	$4s^2 4p^6 4d^{10} 5s^2 5p^4$	54	0.564
III	$3s^2 3p^6 3d^{10} 4s^2 4p^0 5s^0 4d^0$	$4s^2 4p^6 4d^{10} 5s^2 5p^4$	64	0.421
IV	**$3s^2 3p^6 3d^{10} 4s^2 4p^0 5s^0 4d^0 5p^0$**	**$4s^2 4p^6 4d^{10} 5s^2 5p^4$**	**70**	**0.751**
V	$3s^2 3p^6 3d^{10} 4s^2 4p^0 5s^0 4d^0 5p^0 5d^0$	$4s^2 4p^6 4d^{10} 5s^2 5p^4$	80	0.795
VI	$3s^2 3p^6 3d^{10} 4s^2 4p^0 5s^0 4d^0 5p^0 5d^0 6s^0$	$4s^2 4p^6 4d^{10} 5s^2 5p^4$	82	0.644

calculation are those for the ground state of the material. The resulting charge density is that of the ground state of zb-GaSb.

We show in Figure 1 the calculated electronic energy band structure of zb-GaSb, from calculation IV. The direct band gap calculated at the Γ point is 0.751 eV. This value is in excellent agreement with the experimental value of 0.75 eV, at room temperature 0.75 eV [53].

Various characteristics of the electronic band structure can be explained in more detail using the table of calculated eigenvalues at high points symmetry in the Brillouin zone and figures for the total and partial densities states. Figure 2 and Figure 3 show the total state density (DOS) and the partial densities of state (pDOS), respectively, as derived from the bands in Figure 1. Many parts of our calculated density of state are very close to corresponding experimental values obtained by X-ray photoemission spectroscopy measurements [71]. Figure 14 of the paper of Ley *et al.* [71] reported the peak positions of HIT, I2, I1, S1, HIB, PII, HIIIT and PIII, respectively, at −0.8 eV, −1.7 eV, −2.1 eV, −3.4 eV, −3.7 eV, −6.4 eV, −9.2 eV and −10.0 eV. According to our work, the corresponding values are respectively −0.675 eV, −1.611 eV, −2.080 eV, −3.192 eV, −3.661 eV, −6.354 eV, −9.165 eV, and −9.809 eV. While the widths of the different groups of valence bands can be extracted from the total density of states, they are more easily derived from the contents of Table 3 which shows the calculated eigenvalues at high symmetry points in the Brillouin zone.

The total valence band width is 15.862 eV. The width of the lowest laying group of bands is 0.106 eV. The widths of the middle and upper most groups valence bands are respectively 2.651 eV and 6.847 eV. These widths, derived from Table 3, can also be estimated using the content of the figure below for TDOS. A purpose of this table is to be able to make comparisons with possible future experimental measurements of X-ray, ultraviolet (UV) or other spectroscopy.

According to our calculated pDOS in Figure 3, the lowest lying group of valence bands comes almost entirely from Ga d. The middle group of valence bands is largely from Sb s, with a small contribution from Ga s. The upper most group of valence bands is unmistakably dominated by Sb p, Ga p, and Ga s, with a very small contribution of Sb s.

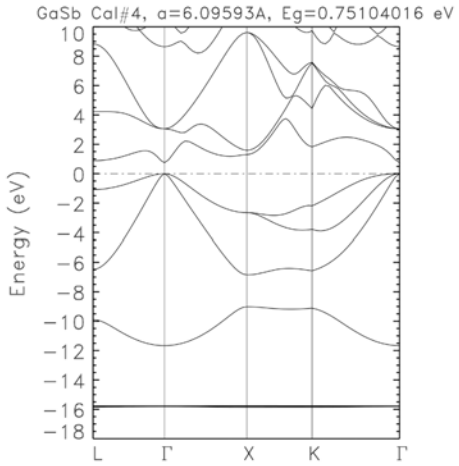


Figure 1. Calculated, band structure of zinc blende gallium antimonide (zb-GaSb), as obtained by the BZW method from Calculations IV. The calculated direct band gap is 0.751 eV.

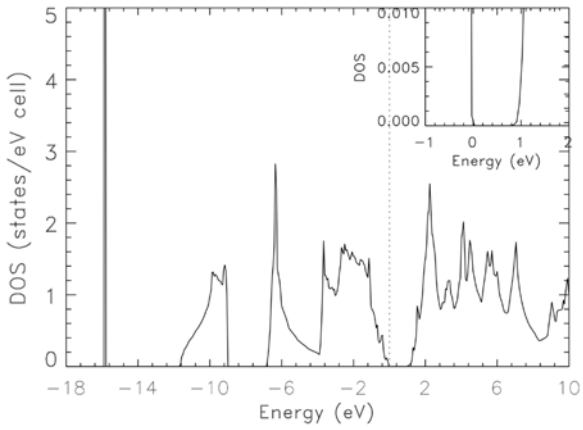


Figure 2. The calculated, total density of states of zb-GaSb, obtained from the energy bands in **Figure 1**. The zero on the horizontal axis indicates the position of the Fermi level.

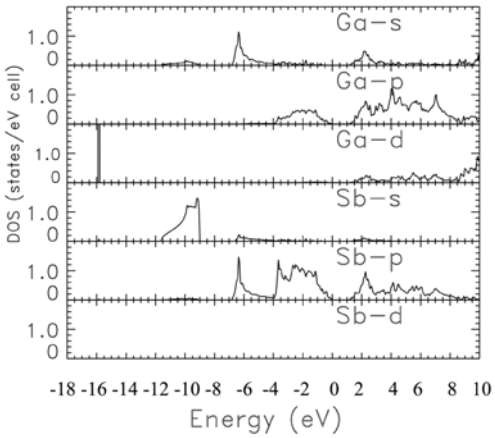


Figure 3. Calculated, partial densities of states (pDOS) for zb-GaSb, as derived from the bands in **Figure 1**. The zero on the horizontal axis indicates the position of the Fermi level.

Table 3. Calculated, electronic energies of zb-GaSb at high symmetry points in the Brillouin Zone, as obtained with the optimal basis set of Calculation IV. We used the experimental lattice constant of 6.09593 Å. This table is partly to enable comparisons with future room temperature, experimental and theoretical results.

L-point	Γ-point	X-point	K-point
8.836	3.053	9.612	7.590
4.221	3.053	9.612	7.556
4.221	3.053	1.591	4.436
0.885	0.751	1.302	1.827
−1.080	0.000	−2.651	−2.187
−1.080	0.000	−2.651	−3.785
−6.480	0.000	−6.847	−6.568
−9.894	−11.665	−9.014	−9.110
−15.759	−15.759	−15.756	−15.758
−15.759	−15.759	−15.767	−15.764
−15.814	−15.824	−15.803	−15.805
−15.814	−15.824	−15.803	−15.806
−15.850	−15.824	−15.862	−15.861

We used the electron band structure from Calculation IV (in **Figure 1**) to calculate the effective masses of the electron, at the bottom of the conduction band, and of the holes, at the top of the valence band. We present our results, as well as other theoretical and experimental values, in **Table 4**. The effective masses of the electrons are indicated by me while those of the heavy and light holes are respectively noted as m_{hh} and m_{lh}. The first column shows the effective masses with the specific directions in which they are calculated. It is gratifying to note that our calculated effective masses are in general agreement with experience [55] [74] [75] [76], and with parts of some previously calculated ones [12] [26] [72] [73]. Out of the nine (9) effective masses in **Table 4**, the only one for which there is a difference between an experimental value and our result is m_{hh} (Γ-L) for which we found 0.682 m_0 while the experiments report from 0.24 m_0 to 0.59 m_0 [74] [75] [76].

For Reference [a], atheo and aexp indicate results obtained with a theoretical and an experimental lattice constant.

The above agreement between our calculated effective masses and the corresponding, experimental values indicates the correct rendering, by our calculations, of the curvatures of the bands at the conduction band minimum (CBM) and the valence band maximum (VBM).

A discussion of our results, compared to previous ones, follows. We first recall that our main motivation for this work was the resolution of the glaring disagreement between measured band gaps of approximately 0.81 eV and 0.73 eV,

Table 4. Calculated, effective masses for zb-GaSb (in units of the free electron-mass, m_0): m_e indicates an electron effective mass at the bottom of the conduction band; m_{hh}, and m_{lh} represent the heavy and light hole effective masses, respectively. Theo.: theory, expt.: experiment.

	Our work	Theo. a_{theo} [a]	Theo. a_{exp} [a]	Theo. [b]	Theo. [c]	Theo. [d]	Expt. [e]	Expt. [f]	Expt. [g]	Expt. [h]
m_e (Γ-L)	0.041	0.036	0.008		0.045	0.045				
m_e (Γ-X)	0.041	0.035	0.004		0.045		0.039 average	0.042 ± 0.001 average		0.039 ± 0.001 average
m_e (Γ-K)	0.040	0.035	0.006		0.045					
m_{hh} (Γ-L)	0.682	0.646	0.747	0.551	0.710	0.55		0.40 ± 0.16	0.36 ±0.03	0.44 ± 0.15
m_{hh} (Γ-X)	0.280	0.272	0.219	0.231	0.247	0.38	0.250	0.29 ±0.09	0.26 ± 0.04	
m_{hh} (Γ-K)	0.397	2.074	2.274		0.500	0.45		0.36 ± 0.13	0.38 ± 0.04	
m_{lh} (Γ-L)	0.035	0.030	0.008	0.047	0.045				0.052 ± 0.004	
m_{lh} (Γ-X)	0.036	0.035	0.004	0.052	0.051	0.049	0.044	0.042 ± 0.002	0.052 ± 0.005	0.038 ± 0.002
m_{lh} (Γ-K)	0.038	0.031	0.006		0.046				0.052 ± 0.004	

[a] Ref [11], [b] Ref [72], [c] Ref [26], [d] Ref [73], [e] Ref [55], [f] Ref [74], [g] Ref [75], [h] Ref [76].

for low and room temperatures, respectively, and the results of dozens of ab-initio theoretical studies that have consistently underestimated them. In particular, 24 LDA or GGA ab-initio studies reported negative numbers or zero for the bandgap of zb-GaSb. The section describing our method points to a reason that previous, ab-initio DFT calculations generally underestimated the measured values: none of these calculations performed a generalized minimization of the energy to reach the ground state in a verifiable way, as required by the second DFT theorem. Therefore, while the output of such a calculation is self-consistent, it is a stationary state among an infinite number of such states. There is an infinite number of basis sets which can lead to self-consistent (*i.e.*, stationary) results. Therefore, the chances for a single basis set calculation to produce the ground state energies while avoiding over-complete basis sets are practically zero. On the other hand, our calculations have explicitly reached the ground state and avoided the use of over-complete basis sets.

Our results, due to our strict adherence to the conditions of validity of the DFT, have the entire physical content of the DFT. They agree with the available, corresponding, experimental data, as was the case for the band gap at room temperature (*i.e.*, 0.75 eV). Our work also made it possible to obtain the experimentally identified locations of the peaks in the total density of valence states.; This agreement points to the overall correct description of the ground state band structure by our calculations. In addition, this overall description of the bands is further indicated by our correct rendering of noted curvatures of the bands, in accordance with our results for the effective masses.

The results of Reference 12 for the electron effective masses confirmed our contention that as the bottom of the conduction band is spuriously lowered,

when using over-complete basis sets, the calculated electron effective masses fall below the corresponding, experimental ones. With band gap of 0.588 eV, these authors [12] found electron effective masses of 0.036, 0.035, and 0.035 m_0; for the much smaller band gap of zero, these electron effective masses are respectively 0.008, 0.004, and 0.006 m_0. Both sets of values are lower than the corresponding ones from our calculation.

Our truly ground state calculations [57] did not need to invoke self-interaction correction of the derivative discontinuity of the exchange correlation energy in order for their outcomes to agree with available, corresponding, experimental ones. The same was true for several previous works of our group [57] [58] [63]-[69].

4. Conclusion

We have performed ab initio, self-consistent calculations of electronic energy bands, total and partial densities of states and of effective masses for zb-GaSb. With the method of Bagayoko, Zhao and Williams (BZW), we performed a generalized minimization of the energy to reach the ground state, verifiably, without using over-complete basis sets. Our work reproduced not just the correct, experimental room temperature band gap, but also the locations of the peaks in the total density of states, and the electron and holes effective masses. With this overall accuracy, our calculated results can inform and guide the design and fabrication of semiconductor based devices, as envisioned by the Materials Genome Initiatives (MGI).

Acknowledgements

Our work was funded in part by the US Fulbright Fellowship Program, the Malian Ministry of Higher Education and Scientific Research, The US National Science Foundation (NSF Award No. HRD-2009765), the Louisiana Optical Network Initiative (LONI), at Southern University and A & M College in Baton Rouge (SUBR), and LaSPACE.

Conflicts of Interest

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

References

[1] Majeed, S. and Al-Rawi, B.K. (2021) *IOP Conference Series*: *Materials Science and Engineering*, **1095**, Article ID: 012003. https://doi.org/10.1088/1757-899X/1095/1/012003

[2] Samajdar, D.P., Das, U., Sharma, A.S., Das, S. and Dhar, S. (2016) *Current Applied Physics*, **16**, 1687-1694. https://doi.org/10.1016/j.cap.2016.10.010

[3] Wang, J.W., Zhang, Y. and Wang, L.-W. (2015) *Physical Review B*, **92**, Article ID: 045211.

[4] Wang, J.W. and Zhang, Y. (2014) *Journal of Applied Physics*, **116**, Article ID:

214301. https://doi.org/10.1063/1.4903063

[5] Wei, S.-H. and Zunger, A. (1989) *Physical Review B*, **39**, 3279-3304. https://doi.org/10.1103/PhysRevB.39.3279

[6] Zhu, X.J. and Louie, S.G. (1991) *Physical Review B*, **43**, 14142-14156. https://doi.org/10.1103/PhysRevB.43.14142

[7] Rhim, S.H., Kim, M. and Freeman, A.J. (2005) *Physical Review B*, **71**, Article ID: 045202. https://doi.org/10.1103/PhysRevB.71.045202

[8] Caid, M. and Rached, D. (2020) *Materials Science-Poland*, **38**, 320-327. https://doi.org/10.2478/msp-2020-0027

[9] Ahmed, R., Fazal-E-Aleem, Hashemifar, S.J., Rashid, H. and Akbarzadeh, H. (2009) *Communications in Theoretical Physics*, **52**, 527. https://doi.org/10.1088/0253-6102/52/3/28

[10] Hadjab, M., Bennacer, H., Berrah, S., Abid, H. and Ziane, I.M. (2015) Density Functional Approach to Study Structural and Electronic Properties of III-Sb Semiconductors by Modified Becke-Johnson Potential. In: *The First National Conference on Electronics and New Technologies* (*NCENT*'2015), May 19-20, M'Sila, 1-4.

[11] Karazhanov, S.Zh. and Lew Yan Voon, L.C. (2005) Ab Initio Studies of the Band Parameters of III-V and II-VI Zinc-Blende Semiconductors. *Semiconductors*, **39**, 161-173. https://doi.org/10.1134/1.1864192

[12] Johnson, K.A. and Ashcroft, N.W. (1998) *Physical Review B*, **58**, 15548-15556. https://doi.org/10.1103/PhysRevB.58.15548

[13] Datta, S., Singh, P., Chaudhuri, C.B., Jana, D., Harbola, M.K., Johnson, D.D. and Mookerjee, A. (2019) Simple Correction to Bandgap Problems in IV and III-V Semiconductors: An Improved, Local First-Principles Density Functional Theory. https://lib.dr.iastate.edu/mse_pubs/301

[14] Tomic, S., Montanari, B. and Harrison, N.M. (2008) *Physica E*, **40**, 2125-2127. https://doi.org/10.1016/j.physe.2007.10.022

[15] Bakulin, A.V. and Kulkova, S.E. (2014) *Russian Physics Journal*, **57**, 996-999. https://doi.org/10.1007/s11182-014-0336-1

[16] Remediakis, I.N. and Kaxira, E. (1999) *Physical Review B*, **59**, 5536. https://doi.org/10.1103/PhysRevB.59.5536

[17] Castano-Gonzalez, E.E., Sena, N., Mendoza-Estrada, V., Gonzalez-Hernandez, R., Dussan, A. and Mesa, F. (2016) *Semiconductors*, **50**, 1280-1286. https://doi.org/10.1134/S1063782616100110

[18] Wang, S.Q. and Ye, H.Q. (2002) *Physical Review B*, **66**, 235111-235118.

[19] Singh, A.K., Chandra, D., Kattayat, S., Kumar, S., Alvi, P.A. and Rathi, A. (2019) *Semiconductors*, **53**, 1584-1592. https://doi.org/10.1134/S1063782619160267

[20] Geller, C.B., Wolf, W., Picozzi, S., Continenza, A., Asahi, R., Mannstadt, W., Freeman, A.J. and Wimmer, E. (2001) *Applied Physics Letters*, **79**, 368. https://doi.org/10.1063/1.1383282

[21] Camargo-Martinez, J.A. and Baquero, R. (2012) *Physical Review B*, **86**, Article ID: 195106. https://doi.org/10.1103/PhysRevB.86.195106

[22] Kim, Y.-S., Marsman, M. and Kresse, G. (2010) *Physical Review B*, **82**, Article ID: 205212. https://doi.org/10.1103/PhysRevB.82.205212

[23] Puska, M.J., Lanki, P. and Nieminen, R.M. (1989) *Journal of Physics*: *Condensed Matter*, **1**, 6081. https://doi.org/10.1088/0953-8984/1/35/008

[24] Huang, M.Z. and Ching, W.Y. (1993) *Physical Review B*, **47**, 9449.

https://doi.org/10.1103/PhysRevB.47.9449

[25] Khanin, D.V. and Kul'kova, S.E. (2005) *Russian Physics Journal*, **48**, 70-77. https://doi.org/10.1007/s11182-005-0086-1

[26] Bastos, C.M.O., Sabino, F.P., Sipahi, G.M. and Da Silva, J.L.F. (2018) *Journal of Applied Physics*, **123**, Article ID: 065702. https://doi.org/10.1063/1.5018325

[27] Briki, M., Abdelouhab, M., Zaoui, A. and Ferhat, M. (2009) *Superlattices and Microstructures*, **45**, 80-90. https://doi.org/10.1016/j.spmi.2008.12.022

[28] Hassan, F.E. and Postnikov, A.V. (2010) *Journal of Alloys and Compounds*, **504**, 559. https://doi.org/10.1016/j.jallcom.2010.05.161

[29] Oumelaz, F., Nemiri, O., Boumaza, A., Ghemid, S., Meradji, H., Bin Omran, S., El Haj Hassan, F., Rai, D.P. and Khenata, R. (2017) *Indian Journal of Physics*, **92**, 705-714. https://doi.org/10.1007/s12648-017-1157-1

[30] El Haj Hassan, F., Breidi, A., Ghemid, S., Amrani, B., Meradji, H. and Pages, O. (2010) *Journal of Alloys and Compounds*, **499**, 80-89. https://doi.org/10.1016/j.jallcom.2010.02.121

[31] Lucero, M.J., Henderson, T.M. and Scuseria, G.E. (2012) *Journal of Physics*: *Condensed Matter*, **24**, Article ID: 145504. https://doi.org/10.1088/0953-8984/24/14/145504

[32] Hinuma, Y., Gruneis, A., Kresse, G. and Oba, F. (2014) *Physical Review B*, **90**, Article ID: 155405. https://doi.org/10.1103/PhysRevB.90.155405

[33] Drablia, S., Meradji, H., Ghemid, S., Labidi, S. and Bouhafs, B. (2009) *Physica Scripta*, **79**, Article ID: 045002. https://doi.org/10.1088/0031-8949/79/04/045002

[34] Malsawmtluanga, A., Lalnunpuia, Ralte, R.L. and Pachuau, Z. (2014) *Indian Journal of Scientific Research and Technology*, **2**, 108-111.

[35] Tahini, H.A., Chroneos, A., Murphy, S.T., Schwingenschogl, U. and Grimes, R.W. (2013) *Journal of Applied Physics*, **114**, Article ID: 063517. https://doi.org/10.1063/1.4818484

[36] Salehi, H., Badehian, H.A. and Farbod, M. (2014) *Materials Science in Semiconductor Processing*, **26**, 477-490. https://doi.org/10.1016/j.mssp.2014.05.020

[37] Al-Douri, Y. and Reshak, A. (2011) *Applied Physics A*, **104**, 1159-1167. https://doi.org/10.1007/s00339-011-6400-6

[38] Othman, M.S.H., Mishjil, K.A. and Habubi, N.F. (2012) *Chinese Physics Letters*, **29**, Article ID: 037302.

[39] Benaline Sheeba, V. and Louis, C.N. (2016) *International Journal of Scientific Research and Innovations*, **1**, 1-5.

[40] Ali, M.A., Aleem, H., Sarwar, B. and Murtaza, G. (2020) *Indian Journal of Physics*, **94**, 477-484.

[41] Wei, S.-H., Nie, X.L., Batyrev, I.G. and Zhang, S.B. (2003) *Physical Review B*, **67**, Article ID: 165209. https://doi.org/10.1103/PhysRevB.67.165209

[42] Reshak, A.H. (2005) *The European Physical Journal B*, **47**, 503-508. https://doi.org/10.1140/epjb/e2005-00364-3

[43] Wang, S.Q. and Ye, H.Q. (2002) *Journal of Physics*: *Condensed Matter*, **14**, 9579-9587. https://doi.org/10.1088/0953-8984/14/41/313

[44] Bouerissa, N., Baaziz, H. and Charifi, Z. (2002) *Physica Status Solidi* (*b*), **231**, 403-410. https://doi.org/10.1002/1521-3951(200206)231:2<403::AID-PSSB403>3.0.CO;2-6

[45] Gazhulina, A.P. and Maryche, M.O. (2015) *Journal of Alloys and Compounds*, **623**, 413. https://doi.org/10.1016/j.jallcom.2014.11.028

[46] Lide, D.R. (2004) CRC Handbook of Chemistry and Physics: A Ready-Reference Book of Chemical and Physical Data. CRC Press, Boca Raton.

[47] Adachi, S. (1987) *Journal of Applied Physics*, **61**, 4869. https://doi.org/10.1063/1.338352

[48] Iyer, S., Hegde, S., Abul-Fadl, A., Bajaj, K.K. and Mitchel, W. (1993) *Physical Review B*, **47**, 1329-1339. https://doi.org/10.1103/PhysRevB.47.1329

[49] Munoz, M., Wei, K. and Pollak, F.H. (1999) *Physical Review B*, **60**, 8105-8110. https://doi.org/10.1103/PhysRevB.60.8105

[50] Dutta, P.S. and Bhat, H.L. (1997) *Journal of Applied Physics*, **81**, 5821. https://doi.org/10.1063/1.365356

[51] Vladimir, A. and Serget, R. (2002) *Physical Status Solidi* (*b*), **244**, 243-248.

[52] Steiner, T.D. (2004) Semiconductor Nanostructures for Optoelectronic Applications. Artech House, Boston. https://doi.org/10.1108/sr.2004.24.3.320.3

[53] Arévalo, F., Saavedra, R. and Paulraj, M. (2008) *Journal of Physics*: *Conference Series*, **134**, Article ID: 012019. https://doi.org/10.1088/1742-6596/134/1/012019

[54] Munoz, M. and Pollak, F.H. (2000) *Physical Review B*, **62**, 16600-16604.

[55] Vurgaftman, I., Meyer, J.R. and Ram-Mohan, L.R. (2001) *Journal of Applied Physics*, **89**, 5815. https://doi.org/10.1063/1.1368156

[56] Dutta, P.S., Koteswara Rao, K.S.R., Bhat, H.L. and Kumar, V. (1995) *Applied Physics A*, **61**, 149-152. https://doi.org/10.1007/BF01538381

[57] Bagayoko, D. (2014) *AIP Advances*, **4**, Article ID: 127104. https://doi.org/10.1063/1.4903408

[58] Diakite, Y.I., Traore, S.D., Malozovsky, Y., Khamala, B., Franklin, L. and Bagayoko, D. (2017) *Journal of Modern Physics*, **8**, 531-546. https://doi.org/10.4236/jmp.2017.84035

[59] Ceperley, D.M. and Alder, B.J. (1980) *Physical Review Letters*, **45**, 566. https://doi.org/10.1103/PhysRevLett.45.566

[60] Vosko, S.H., Wilk, L. and Nusair Can, M. (1980) *Canadian Journal of Physics*, **58**, 1200-1211. https://doi.org/10.1139/p80-159

[61] Harmon, B.N., Weber, W. and Hamann, D.R. (1982) *Physical Review B*, **25**, 1109-1115. https://doi.org/10.1103/PhysRevB.25.1109

[62] Kohn, W. and Sham, L.J. (1965) *Physical Review*, **140**, A1133-A1138. https://doi.org/10.1103/PhysRev.140.A1133

[63] Bagayoko, D, Zhao, G.L., Fan, J.D. and Wang, J.T. (1998) *Journal of Physics*: *Condensed Matter*, **10**, 5645-5655. https://doi.org/10.1088/0953-8984/10/25/014

[64] Zhao, G.L., Bagayoko, D. and Williams, T.D. (1999) *Physical Review B*, **60**, 1563. https://doi.org/10.1103/PhysRevB.60.1563

[65] Ekuma, C.E., Jarrell, M., Moreno, J. and Bagayoko, D. (2013) *Physics Letters A*, **377**, 2172-2176. https://doi.org/10.1016/j.physleta.2013.05.043

[66] Franklin, L., Ekuma, C.E., Zhao, G.L. and Bagayoko, D. (2013) *Journal of Physics and Chemistry of Solids*, **74**, 729-736. https://doi.org/10.1016/j.jpcs.2013.01.013

[67] Bagayoko, D. and Franklin, L. (2005) *Journal of Applied Physics*, **97**, Article ID: 123708. https://doi.org/10.1063/1.1939069

[68] Diakite, Y.I., Traoré, S.D., Malozovsky, Y., Khamala, B., Franklin, L. and Bagayoko, D. (2015) *The African Review of Physics*, **10**, 315-327.

[69] Bamba, C.O., Inakpenu, R., Diakite, Y.I., Franklin, L., Malozovsky, Y., Stewart, A.D.

and Bagayoko, D. (2017) *Journal of Modern Physics*, **8**, 1938-1949. https://doi.org/10.4236/jmp.2017.812116

[70] Bagayoko, D. (1983) *International Journal of Quantum Chemistry*, **17**, 527.

[71] Ley, L., Pollak, R.A., McFeely, F.R., Kowalczyk, S.P. and Shirley, D.A. (1974) *Physical Review B*, **9**, 600-621. https://doi.org/10.1103/PhysRevB.9.600

[72] Huang, M.Z. and Ching, W.Y. (1985) *Journal of Physics and Chemistry of Solids*, **46**, 977. https://doi.org/10.1016/0022-3697(85)90101-5

[73] Higginbothan, C.W., Pollak, F.H. and Cardona, M. (1968) Band Structure and Optical Constants of InSb, InAs, and GaSb: The k.p Method. *Proceeding of IX International Conference on the Physics of Semiconductors*, Moscow, Vol. 1, 57.

[74] Reine, M., Aggarwal, R.L. and Lax, B. (1972) *Physical Review B*, **5**, 3033-3049. https://doi.org/10.1103/PhysRevB.5.3033

[75] Stradling, R.A. (1966) *Physics Letters*, **20**, 217-218. https://doi.org/10.1016/0031-9163(66)90330-1

[76] Filion, A. and Fortin, E. (1973) *Physical Review B*, **8**, 3852-3860. https://doi.org/10.1103/PhysRevB.8.3852

Journal of Modern Physics, 2022, 13, 432-441
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Hamilton-Jacobi and Lagrange Formulations of Relativistic Quantum Mechanics Wave Equations with Solutions with Only-Positive and Only-Negative Kinetic Energies

Luis Grave de Peralta[1,2*], Arquímedes Ruiz-Columbie[3]

[1]Department of Physics and Astronomy, Texas Tech University, Lubbock, TX, USA
[2]Nano Tech Center, Texas Tech University, Lubbock, TX, USA
[3]Wind Energy Program, Texas Tech University, Lubbock, TX, USA
Email: *luis.grave-de-peralta@ttu.edu

How to cite this paper: de Peralta, L.G. and Ruiz-Columbie, A. (2022) Hamilton-Jacobi and Lagrange Formulations of Relativistic Quantum Mechanics Wave Equations with Solutions with Only-Positive and Only-Negative Kinetic Energies. *Journal of Modern Physics*, **13**, 432-441.
https://doi.org/10.4236/jmp.2022.134030

Received: March 27, 2022
Accepted: April 18, 2022
Published: April 21, 2022

Abstract

Using the Hamilton-Jacobi and the Lagrange formalisms, a pair of relativistic quantum mechanics equations are obtained by abduction. These equations, in contrast with the Klein-Gordon and other relativistic quantum mechanics equations, have no solutions with both positive and negative kinetic energies. The equation with solutions with only positive kinetic energy values describes a spin-0 particle of mass m, which is moving at relativistic speeds in a scalar potential. The wavefunctions and the energies corresponding to the associated antiparticle can be obtained by solving the other equation, which only has solutions with negative kinetic energy values.

Keywords

Quantum Mechanics, Relativistic Quantum Mechanics

1. Introduction

The best-known relativistic quantum mechanics equations are the Klein-Gordon and Dirac equations [1] [2]. Both equations have solutions corresponding to a particle with mass m with both positive and negative kinetic energy values [1] [2]. Classical particles only can have positive kinetic energy values. The same occurs for the Schrödinger equation, which is the best-known non-relativistic quantum mechanics wave equation [3] [4]. Therefore, the existence of solutions of relativistic wave equations, that seems to correspond to a particle with nega-

tive kinetic energy, was a puzzle that conducted to the theoretical prediction of the existence of antiparticles [1] [2] [3] [4]. The existence of antiparticles is today a well-known experimental fact [1] [2] [3] [4]. Today, it is well-know that antiparticles also have positive kinetic energy values [1] [2] [3] [4]; nevertheless, the concept of antiparticle remains associated to the solutions of the Klein-Gordon and Dirac equations corresponding to negative values of the kinetic energy [1] [2].

Recently, an interesting relativistic quantum mechanics wave equation that does not have solutions corresponding to a particle with negative kinetic energy, has attracted our attention [5] [6] [7]. The authors of this work have been studying for a while the so-called general Grave de Peralta (gGP) equation [5] [6] [7] [8] [9]:

$$i\hbar\frac{\partial}{\partial t}\Psi = -\frac{\hbar^2}{2\mu}\nabla^2\Psi + V\Psi, \text{ with } \mu = \frac{1+\gamma}{2}m. \tag{1}$$

This is a Schrödinger-like, but relativistic quantum mechanics equation, that describes the quantum states of a particle moving in a scalar potential (V) with an effective mass $\mu > 0$, which depends on the parameter γ [9]. When $\gamma = 1$, then $\mu = m$, the relativistic invariant mass of the particle; thus, Equation (1) coincides with the Schrödinger equation [3] [4]. Equation (1) is the Grave de Peralta (GP) equation when γ is the classical Lorentz factor of the special theory of relativity [5] [6] [7] [10]:

$$\gamma = \sqrt{1+\frac{p^2}{m^2c^2}}. \tag{2}$$

In Equations (1) and (2), c is the speed of the light in vacuum, $\hbar$ is the reduced Plank constant, and $p = |\boldsymbol{p}|$, where $\boldsymbol{p}$ is the particle's three-dimensional linear momentum. Another practical choice for the parameter γ was proposed by Poveda [8]:

$$\gamma = \langle\Psi|\hat{\gamma}|\Psi\rangle = \sqrt{1+\frac{\langle\Psi|\hat{p}^2|\Psi\rangle}{m^2c^2}}, \text{ with } \hat{\gamma} = \sqrt{1+\frac{\hat{p}^2}{m^2c^2}}. \tag{3}$$

The quantum operators in Equations (1) and (3) are [1]-[9]:

$$i\hbar\frac{\partial}{\partial t} = \hat{H}' = \hat{H} - mc^2, \ \hat{p} = -i\hbar\nabla, \ \hat{K} = \frac{\hat{p}^2}{(\gamma+1)m} = -\frac{\hbar^2}{2\mu}\nabla^2. \tag{4}$$

In Equation (4), the energy operator (H') is associated to the total energy of the particle after discounting the energy associated to its mass (mc^2), and the kinetic energy operator (K) includes the correct relativistic relation with the linear momentum operator [5] [6] [7] [8]. Consequently, in the Poveda's approach, γ is the average value of the Lorentz factor in the quantum state Ψ [8]. Notably, Equation (1) reduces to well-known quantum wave equations in both the non-relativistic and ultra-relativistic limits. In the ultra-relativistic limit $p \gg mc$ ($\gamma \gg 1$), Equation (1) with $V = 0$ reduces to the following Weyl-like equation for a free spin-0 particle [11]:

$$i\hbar\frac{\partial}{\partial t}\Psi = \hat{p}c\Psi \Leftrightarrow \frac{1}{c}\frac{\partial}{\partial t}\Psi = -\nabla\Psi. \tag{5}$$

Previous works has focused in stablishing the existing relationship between Equation (1) and the best-known relativistic quantum mechanics wave equations [5] [6] [7] [8] [9] [11]. Taking advantage of the formal similitude between the gGP and Schrödinger equations, Equation (1) have been solved, following the same mathematical procedures required for solving the same problems using the Schrödinger equation, for a free particle [7], confinement of a quantum particle in box [5] [6] [7] [8], reflection by a sharp quantum potential [5], tunnel effect [5], the harmonic oscillator [9], and the Hydrogen atom [12] [13]. This allowed the extension to the relativistic domain of well-known results previously obtained using the Schrödinger equation [3] [4].

In this work, the attention in focused on the relationship of Equation (1) with well-stablished theoretical methods. We show here, for the first time, how Equation (1) and the following equation [14]:

$$i\hbar\frac{\partial}{\partial t}\Psi = \frac{\hbar^2}{(\gamma+1)m}\nabla^2\Psi + V\Psi \tag{6}$$

can be obtained by abduction using the Lagrange and Hamilton-Jacobi formalisms [1] [15] [16]. Equation (6) can be rewritten as Equation (1) but with $\mu < 0$. Equation (6) is a complementary gGP-like equation recently reported [14]. In contrast with Equation (1), Equation (6) only has solutions with negative kinetic energy values [14]. This work is then directed to explore the theoretical foundations of Equation (1) for any real value of μ.

The rest of this paper is organized in the following way. In Section 2, for self-reliance purposes, a summary is presented about how Equations (1) and (6) can be formally obtained. Then, in Sections 3 and 4, for the first time, we present how these equations can be obtained by abduction using Hamilton-Jacobi and Lagrange formalisms. Finally, the conclusions of this work are given in Section 5.

2. Poirier-Grave De Peralta Equations

How to obtain Equations (1) and (6), by applying a formal first quantization procedure on a Lorentz-invariant equation, which involves the energy-momentum 4-component vector (p^μ), has been discussed before [8] [14]. In short, we can start from the following Lorentz-covariant equation, which gives the module square of the energy-momentum 4-component vector of a free classical particle with mass m, energy E, and 3-component linear momentum $\boldsymbol{p}$ [1]:

$$p^\mu p_\mu = g^{\mu\nu}p_\nu p_\mu = m^2c^2. \tag{7}$$

In Equation (7), $g^{\mu\nu}$ is the metric tensor and p^μ and p_μ are the covariant and contravariant energy-momentum 4-component vector, respectively [1]:

$$p_\mu = \left\{\frac{E}{c}, -\boldsymbol{p}\right\} = \left\{\frac{E}{c}, -p_x, -p_x, -p_z\right\},\quad p^\mu = \left\{\frac{E}{c}, \boldsymbol{p}\right\} = \left\{\frac{E}{c}, p_x, p_x, p_z\right\}. \tag{8}$$

The Lorentz-covariant Equation (7) can be transformed in the following Lo-

rentz-covariant quantum wave equation:

$$\hat{p}^{\mu}\hat{p}_{\mu}\Psi = g^{\mu\nu}\hat{p}_{\nu}\hat{p}_{\mu}\Psi \Leftrightarrow -\hbar^2 g^{\mu\nu}\frac{\partial}{\partial x^{\nu}}\frac{\partial}{\partial x^{\mu}}\Psi \Leftrightarrow \left(\frac{\hat{H}^2}{c^2}-\hat{p}^2\right)\Phi = m^2c^2\Phi. \quad (9)$$

by making the following formal first-quantization substitution [1]:

$$E \to \hat{H} = i\hbar\frac{\partial}{\partial t}, \ \ \boldsymbol{p} \to \hat{p} = -i\hbar\nabla. \quad (10)$$

Equation (9) is the Klein-Gordon equation for a free spin-0 particle with mass m [1] [2]. We can proceed in the same way for obtaining the Klein-Gordon equation for a particle moving in a scalar potential V, but we should substitute E by $E-V$ and H by $H-V$ in Equations (8) and (9), respectively. We then obtain [1] [2]:

$$\left(g^{\mu\nu}\frac{\partial}{\partial x^{\nu}}\frac{\partial}{\partial x^{\mu}}+\frac{m^2c^2}{\hbar^2}\right)\Phi = 0 \Leftrightarrow \left[\left(\hat{H}-V\right)^2 - m^2c^4\right]\Phi = -c^2\hbar^2\nabla^2\Phi. \quad (11)$$

Due to its relative importance [1] [2] [3] [4], we will limit the scope of this work to study the wave equations of a spin-0 particle with mass moving in a scalar potential. The general case including the presence of an external magnetic field will be discussed in a separated work. Equation (11) can be rewritten in the following way [8]:

$$\left[\left(\hat{H}-V\right)-mc^2\right]\left[\left(\hat{H}-V\right)+mc^2\right]\Phi = c^2\hat{p}^2\Phi. \quad (12)$$

Using the following identities [8] [9]:

$$\hat{K} = \frac{\hat{p}^2}{\left(\hat{\gamma}+1\right)m} = -\frac{\hbar^2}{\left(\hat{\gamma}+1\right)}\nabla^2 = \left(\hat{\gamma}-1\right)mc^2 = \sqrt{\hat{p}^2c^2+m^2c^2}-mc^2. \quad (13)$$

When $K = \left(E-V\right)-mc^2 > 0$, then $\left(E-V\right)+mc^2 = \left(\gamma+1\right)mc^2$; therefore, Equation (12) can be rewritten as [5]:

$$\left[\left(\hat{H}-V\right)-mc^2\right]\Phi = -\frac{\hbar^2}{\left(\hat{\gamma}+1\right)}\nabla^2\Phi. \quad (14)$$

However, when $K = \left(E-V\right)+mc^2 < 0$, then $\left(E-V\right)-mc^2 = -\left(\gamma+1\right)mc^2$; therefore, Equation (12) can be rewritten as [14]:

$$\left[\left(\hat{H}-V\right)+mc^2\right]\Phi = \frac{\hbar^2}{\left(\hat{\gamma}+1\right)}\nabla^2\Phi. \quad (15)$$

Consequently, Equation (12) reduces to:

$$i\hbar\frac{\partial}{\partial t}\Phi_{\pm} = \begin{cases} -\dfrac{\hbar^2}{\left(\hat{\gamma}+1\right)}\nabla^2\Phi_{+} + V\Phi_{+} + mc^2\Phi_{+}, & \text{if } K > 0 \\ +\dfrac{\hbar^2}{\left(\hat{\gamma}+1\right)}\nabla^2\Phi_{-} + V\Phi_{-} - mc^2\Phi_{-}, & \text{if } K < 0 \end{cases} \quad (16)$$

The Poirier-Grave de Peralta (PGP) equations are obtained from Equation (16) after introducing the wavefunctions $\Psi_{\pm}$ in the following way [8]:

$$i\hbar\frac{\partial}{\partial t}\Psi_{\pm}=\begin{cases}-\dfrac{\hbar^2}{\left(\hat{\gamma}+1\right)}\nabla^2\Psi_{+}+V\Psi_{+},\ \text{with}\ \Psi_{+}=\Phi_{+}\mathrm{e}^{-\frac{i}{\hbar}mc^2t}\ \text{if}\ K>0\\ +\dfrac{\hbar^2}{\left(\hat{\gamma}+1\right)}\nabla^2\Psi_{-}+V\Psi_{-},\ \text{with}\ \Psi_{-}=\Phi_{-}\mathrm{e}^{+\frac{i}{\hbar}mc^2t}\ \text{if}\ K<0\end{cases}\tag{17}$$

Using Equation (13), Equation (17) can be rewritten as:

$$i\hbar\frac{\partial}{\partial t}\Psi_{\pm}=\begin{cases}+\sqrt{\hat{p}^2c^2+m^2c^2}\Psi_{+}+V\Psi_{+}-mc^2\Psi_{+},\ \text{with}\ \Psi_{+}=\Phi_{+}\mathrm{e}^{-\frac{i}{\hbar}mc^2t}\ \text{if}\ K>0\\ -\sqrt{\hat{p}^2c^2+m^2c^2}\Psi_{-}+V\Psi_{-}+mc^2\Psi_{-},\ \text{with}\ \Psi_{-}=\Phi_{-}\mathrm{e}^{+\frac{i}{\hbar}mc^2t}\ \text{if}\ K<0\end{cases}\tag{18}$$

The top equation in Equation (18) is related to the so called "spinless Salpeter" equation, a relativistic equation widely used to describe hadrons as bound states of the constituent quarks [17] [18] [19] [20]. Note the operator $i\hbar\partial/\partial t-mc^2$ in the top equation in Equation (16) is the operator H' in Equations (1) and (4). Different approximations may be used for substituting in Equation (17) the operator γ by a parameter γ [8] [9], thus transforming Equation (17) in Equation (1) for any real value of μ. For instance, we could use the Poveda's approach that considers γ as the average value of the operator γ in the quantum state $\Psi_{\pm}$ (Equation (3)). Alternatively, we could use the Grave de Peralta's approach that considers γ as the classical Lorent's factor of special theory of relativity.

The energies $E'_{+}=E_{+}-mc^2$ and the wavefunctions Ψ_{+}, that can be obtained by solving the top equation in Equation (17), correspond to a spin-0 quantum particle with mas $m_p=m$, energies $E_p=E'_{+}$, wavefunctions $\Psi_p=\Psi_{+}$, and kinetic energies $K_p=K_{+}=E'_{+}-V>0$ [5] [6] [7] [8] [9] [11] [12] [13]. Assuming that γ is a parameter, if the particle is a charged particle with charge $q_a=q$, and V is a potential produced by others charged particles surrounding the particle, then V change of sign when the sign of all the particles producing the potential change of sign (C-transformation). Consequently, the top equation in Equation (17) can be obtained from the bottom equation (and vice versa) by making the charge-parity-time (CPT) symmetry operations [1]: $t\to -t$ (T-transformation), $\boldsymbol{r}\to -\boldsymbol{r}$ (P-transformation), and $V\to -V$ (C-transformation); *i.e.* [14]:

$$\begin{aligned}&i\hbar\frac{\partial}{\partial(-t)}\Psi_{-}=\frac{\hbar^2}{\left(\gamma+1\right)}\nabla^2\Psi_{-}+\left(-V\right)\Psi_{-}\\ &\Leftrightarrow i\hbar\frac{\partial}{\partial t}\Psi_{+}=-\frac{\hbar^2}{\left(\gamma+1\right)}\nabla^2\Psi_{+}+V\Psi_{+}.\end{aligned}\tag{19}$$

In resemblance of the hole theory for the Klein-Gordon and Dirac equations [1] [14], this means that if the top equation in Equation (17) describes the states of a charged particle moving in a scalar potential V, which is produced by others charged particles surrounded the particle; then, the bottom equation in Equation (17) describes the states of the associated antiparticle moving in a scalar potential $-V$, which is produced by the corresponding charged antiparticles surrounded the antiparticle (matter-antimatter indistinguishability). This implies

that the energies $E'_{-} = E_{-} + mc^2$ and the wavefunctions Ψ_{-}, that can be obtained by solving the bottom equation in Equation (17), are related to the corresponding energies ($E_a = -E'_{-}$) and wavefunctions of the antiparticle associated to the spin-0 particle [14].

3. Hamilton-Jacobi Formulation of the gGP Equations

The relativistic Lagrangian corresponding to a classical particle of mass m moving in a scalar potential V with speed v is given by the following expression [15]:

$$\mathcal{L} = -\frac{mc^2}{\gamma} - V, \text{ with } \gamma = \frac{1}{\sqrt{1-\left(\frac{v}{c}\right)^2}} = \sqrt{1+\frac{p^2}{m^2c^2}}. \tag{20}$$

The relativistic Lagrangian defined by Equation (20) is not written in a covariant form, but written in a preferred inertial frame, while V is a scalar potential depending only on position, which is related to a conservative force in the Lorentz inertial frame under consideration [15]. The corresponding particle's relativistic Hamiltonian, written in the preferred inertial frame, is then given by [15]:

$$H - V = v\frac{\partial \mathcal{L}}{\partial v} - \mathcal{L} = \sqrt{p^2c^2 + m^2c^4} = \gamma mc^2. \tag{21}$$

Therefore:

$$\left(H - V\right)^2 = p^2c^2 + m^2c^4. \tag{22}$$

Note that Equation (22) is not equivalent to Equation (21) because Equation (21) gives only one of the two possible solutions of Equation (22):

$$H - V = \begin{cases} v\dfrac{\partial \mathcal{L}}{\partial v} - \mathcal{L} = \sqrt{p^2c^2 + m^2c^4} = \gamma mc^2, & \text{if } H - V > mc^2 \\ -\left(v\dfrac{\partial \mathcal{L}}{\partial v} - \mathcal{L}\right) = -\sqrt{p^2c^2 + m^2c^4} = -\gamma mc^2, & \text{if } H - V < -mc^2 \end{cases} \tag{23}$$

While a classical particle only can have non-negative values of the kinetic energy ($H - V - mc^2$), both cases in Equation (23) are possible for quantum particles. The relativistic Hamilton-Jacobi equation for a classical particle is [16]:

$$\left(\nabla S\right)^2 - \frac{1}{c^2}\left[\frac{\partial S}{\partial t} + V\right]^2 + m^2c^4 = 0. \tag{24}$$

It can be obtained from Equation (22) using the following relations [16] [17]:

$$\boldsymbol{p} = \nabla S, \quad H = -\frac{\partial S}{\partial t}. \tag{25}$$

In Equations (24) and (25), S is the action, which is defined by the following expression [16]:

$$S\left(r,t\right) = S_o + \int_{\delta(r_o,t_o)}^{(r,t)} \mathcal{L}\mathrm{d}t. \tag{26}$$

The integral in Equation (26) should be evaluated along the extremal curve δ

joining the initial and final positions at the initial and final times; the final and initial positions and the final time should be considered variables while the initial position is fixed [16]. Equation (24) can be rewritten as:

$$\left[\left(\frac{\partial S}{\partial t}+V\right)+mc^2\right]\left[-\left(\frac{\partial S}{\partial t}+V\right)+mc^2\right]\Phi=c^2\left(\nabla S\right)^2\Phi. \tag{27}$$

The formal first quantization procedure described in Section 2 corresponds, in the Hamilton-Jacobi formalism, to the so-called Schrödinger *ansatz* [16]:

$$\begin{aligned} &S\left(r,t\right)=-i\hbar Ln\left[\Phi\left(r,t\right)\right] \\ &\Rightarrow\frac{\partial S}{\partial t}\Phi=-i\hbar\frac{\partial\Phi}{\partial t}=\hat{H}\Phi \text{ and } \nabla S\Phi=-i\hbar\nabla\Phi=\hat{p}\Phi. \end{aligned} \tag{28}$$

This transform Equation (24) in Equation (11) and Equation (27) in Equation (12). As it was shown in the previous Section, from Equations (11) and (12) follows Equation (1). We have then found how to obtain Equation (1), for any real value of μ, from the relativistic Lagrangian corresponding to a classical particle of mass m moving in a scalar potential V (Equation (20)). A more heuristic analysis seems pertinent here. It has been showed above how to obtain the gGP equation (Equation (1)): from a proper Hamilton-Jacobi equation that enunciates a regularity in a classical wave's phase (Equations (24) and (27)) to an *ansatz* expression (Equation (28)) that allows the construction of an equation for the whole quantum wave (Equations (11) and (12)), and then the identification of the quantum operators involved. Perhaps engineer-minded people would identify this heuristic path as reverse engineering, something logicians prefer to call an abduction (the best possible explanation). In any case, the Schrödinger-like wave approach is a step back to continuousness, to a partial differential equation defined on the classical configuration space and abducted from the Hamilton-Jacobi framework that may offer a new way to approach complementarity, understanding it not only as a wave-particle duality but also as a tight-connected regularity via phase's consistency. Notice that the same Schrödinger's *ansatz* (Equation (28)) provides a unique bridge between classical mechanics and quantum mechanics independently of their relativistic or non-relativistic character.

4. Lagrange Formulation of the gGP Equations

Quantum mechanics equations can also be obtained from a Lagrange density that depends on the fields Ψ and Ψ^* [1]. In this Section, for clarity, we will denote the quantum fields corresponding to the ±gGP equations as Ψ_+ and Ψ_-, thus they satisfice Equation (1) with $\mu_+ > 0$ and $\mu_- < 0$, respectively. Therefore, the quantum field $\Psi_\pm^*$ satisfices the same equations but with i changed by $-i$. Assuming that γ is a parameter, it is easy to show that the Lagrange densities corresponding to the ±gGP equations are given by the following expressions:

$$\mathcal{L}_+=-\frac{\hbar^2}{2m}\left(\nabla\Psi_+\right)\cdot\left(\nabla\Psi_+^*\right)-\frac{\left(\gamma+1\right)}{2}\Psi_+V\Psi_+^*-\frac{\left(\gamma+1\right)}{2}i\hbar\Psi_+\frac{\partial\Psi_+^*}{\partial t}. \tag{29}$$

And:

$$\mathcal{L}_{-}=-\frac{\hbar^2}{2m}\left(\nabla\Psi_{-}\right)\cdot\left(\nabla\Psi_{-}^{*}\right)-\frac{\left(\gamma+1\right)}{2}\Psi_{-}V\Psi_{-}^{*}+\frac{\left(\gamma+1\right)}{2}i\hbar\Psi_{-}^{*}\frac{\partial\Psi_{-}}{\partial t}. \tag{30}$$

This means, the quantum fields $\Psi_{\pm}$ and $\Psi_{\pm}^{*}$ must satisfice the following Euler-Lagrange (field or wave) equations [1]:

$$\begin{aligned}\frac{\partial}{\partial t}\left[\frac{\partial\mathcal{L}_{\pm}}{\partial\left(\frac{\partial\Psi_{\pm}}{\partial t}\right)}\right]&=\frac{\partial\mathcal{L}_{\pm}}{\partial\Psi_{\pm}}-\nabla\left[\frac{\partial\mathcal{L}_{\pm}}{\nabla\Psi_{\pm}}\right],\\ \frac{\partial}{\partial t}\left[\frac{\partial\mathcal{L}_{\pm}}{\partial\left(\frac{\partial\Psi_{\pm}^{*}}{\partial t}\right)}\right]&=\frac{\partial\mathcal{L}_{\pm}}{\partial\Psi_{\pm}^{*}}-\nabla\left[\frac{\partial\mathcal{L}}{\nabla\Psi_{\pm}^{*}}\right].\end{aligned} \tag{31}$$

A straight substitution of Equations (29) and (30) in Equation (31) demonstrates that Equations (29) and (30) gives the correct Lagrange densities corresponding to the ±gGP equations. For instance, substituting Equation (29) in Equation (31), we obtain:

$$\begin{aligned}&\frac{\partial\mathcal{L}_{+}}{\partial\left(\frac{\partial\Psi_{+}}{\partial t}\right)}=0\Rightarrow\frac{\partial}{\partial t}\left[\frac{\partial\mathcal{L}_{+}}{\partial\left(\frac{\partial\Psi_{+}}{\partial t}\right)}\right]=0,\\ &\frac{\partial\mathcal{L}_{+}}{\partial\left(\frac{\partial\Psi_{+}^{*}}{\partial t}\right)}=-\frac{\left(\gamma+1\right)}{2}i\hbar\Psi_{+}\Rightarrow\frac{\partial}{\partial t}\left[\frac{\partial\mathcal{L}_{+}}{\frac{\partial\Psi_{+}^{*}}{\partial t}}\right]=-\frac{\left(\gamma+1\right)}{2}i\hbar\frac{\partial\Psi_{+}}{\partial t}.\end{aligned} \tag{32}$$

$$\begin{aligned}\frac{\partial\mathcal{L}_{+}}{\partial\Psi_{+}}&=-\frac{\left(\gamma+1\right)}{2}i\hbar\frac{\partial\Psi_{+}^{*}}{\partial t}-\frac{\left(\gamma+1\right)}{2}V\Psi_{+}^{*},\\ \frac{\partial\mathcal{L}_{+}}{\partial\Psi^{*}}&=-\frac{\left(\gamma+1\right)}{2}V\Psi_{+}.\end{aligned} \tag{33}$$

And:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\nabla\Psi_{+}}=-\frac{\hbar^2}{2m}\nabla\Psi_{+}^{*}\Rightarrow\nabla\left[\frac{\partial\mathcal{L}}{\nabla\Psi_{+}}\right]&=-\frac{\hbar^2}{2m}\nabla^2\Psi_{+}^{*},\\ \frac{\partial\mathcal{L}}{\nabla\Psi_{+}^{*}}=-\frac{\hbar^2}{2m}\nabla\Psi_{+}\Rightarrow\nabla\left[\frac{\partial\mathcal{L}}{\nabla\Psi_{+}^{*}}\right]&=-\frac{\hbar^2}{2m}\nabla^2\Psi_{+}.\end{aligned} \tag{34}$$

Therefore:

$$\begin{aligned}-i\hbar\frac{\partial\Psi_{+}^{*}}{\partial t}&=-\frac{\hbar^2}{\left(\gamma+1\right)m}\nabla^2\Psi_{+}^{*}+V\Psi_{+}^{*},\\ i\hbar\frac{\partial\Psi_{+}}{\partial t}&=-\frac{\hbar^2}{\left(\gamma+1\right)m}\nabla^2\Psi_{+}+V\Psi_{+}.\end{aligned} \tag{35}$$

As expected, in the non-relativistic limit the Lagrange density corresponding

to the +gGP (Equation (29) with $\gamma = 1$) is also the Lagrange density corresponding to the Schrödinger equation [1].

5. Conclusion

We presented how to obtain a pair of relativistic quantum mechanics equations by abduction from the Hamilton-Jacobi and Lagrange formalisms. In contrast with Klein-Gordon and others relativistic quantum mechanics equations, the +gGP equation (Equation (1) with $\mu > 0$) has only solutions with positive kinetics energies, while the −gGP equation (Equation (1) with $\mu < 0$) has only solutions with negative kinetics energies. Therefore, solving the +gGP equation allows for finding the wavefunctions $\Psi_p = \Psi_+$ and energies ($E_p = E'_+$) of a spin-0 particle of mass m, which is moving in a scalar potential at relativistic speeds. Complementarily, solving the −gGP equation allows for finding information about the energies ($E_a = -E'_-$) and wavefunctions corresponding to the associated antiparticle.

Conflicts of Interest

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

References

[1] Greiner, W. (2000) Relativistic Quantum Mechanics: Wave Equations. 3rd Edition, Springer-Verlag, New York. https://doi.org/10.1007/978-3-662-04275-5

[2] Strange, P. (1998) Relativistic Quantum Mechanics: With Applications in Condensed Matter and Atomic Physics. Cambridge University Press, New York. https://doi.org/10.1017/CBO9780511622755

[3] Griffiths, D.J. (1995) Introduction to Quantum Mechanics. Prentice Hall, New York.

[4] Davydov, A.S. (1965) Quantum Mechanics. Pergamon Press, Oxford.

[5] Grave de Peralta, L. (2020) *Results in Physics*, **18**, Article ID: 103318. https://doi.org/10.1016/j.rinp.2020.103318

[6] Grave de Peralta, L. (2020) *European Journal of Physics*, **41**, Article ID: 065404. https://doi.org/10.1088/1361-6404/aba7dc

[7] Grave de Peralta, L. (2020) *Journal of Modern Physics*, **11**, 196-213. https://doi.org/10.4236/jmp.2020.112012

[8] Grave de Peralta, L., Poveda, L.A. and Poirier, B. (2021) *European Journal of Physics*, **42**, Article ID: 055404. https://doi.org/10.1088/1361-6404/ac0ecc

[9] Poveda, L.A., Grave de Peralta, L., Pittman, J. and Poirier, B. (2022) *Foundations of Physics*, **52**, 29. https://doi.org/10.1007/s10701-022-00541-5

[10] Jackson, J.D. (1975) Classical Electrodynamics. 2nd Edition, J. Wiley & Sons, New York.

[11] Grave de Peralta, L. and Farooq, H. (2021) *Journal of Modern Physics*, **12**, 1145-1159. https://doi.org/10.4236/jmp.2021.128068

[12] Grave de Peralta, L. (2020) *Journal of Modern Physics*, **11**, 788-802. https://doi.org/10.4236/jmp.2020.116051

[13] Grave de Peralta, L. (2020) *Scientific Reports*, **10**, Article No. 14925.

https://doi.org/10.1038/s41598-020-71505-w

[14] Grave de Peralta, L. (2022) *Foundations of Physics*. (under review)

[15] Goldstein, H., Poole, C. and Safko, J. (2000) Classical Mechanics. 3rd Edition, Addison & Wesley, Boston.

[16] Holland, P.R. (2004) The Quantum Theory of Motion: An Account of the Broglie-Bohm Causal Interpretation of Quantum Mechanics. Cambridge University Press, Cambridge.

[17] Jacobs, S., Olsson, M.G. and Suchyta III, C. (1986) *Physical Review D*, **33**, Article ID: 014003. https://doi.org/10.1103/PhysRevD.33.3338

[18] Lucha, W., Schöberl, F.F. and Gromes, D. (1991) *Physics Reports*, **200**, 127. https://doi.org/10.1016/0370-1573(91)90001-3

[19] Lucha, W. and Schöberl, F.F. (1995) *Physical Review A*, **51**, 4419. https://doi.org/10.1103/PhysRevA.51.4419

[20] Semay, C., Silvestre-Brac, B. and Narodetskii, I.M. (2004) *Physical Review D*, **69**, Article ID: 014003. https://doi.org/10.1103/PhysRevD.69.014003

Journal of Modern Physics, 2022, 13, 442-494
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Nonlinear Conformal Electromagnetism

J.-F. Pommaret

CERMICS, Ecole des Ponts ParisTech, France
Email: jean-francois.pommaret@wanadoo.fr

How to cite this paper: Pommaret, J.F. (2022) Nonlinear Conformal Electromagnetism. *Journal of Modern Physics*, **13**, 442-494.
https://doi.org/10.4236/jmp.2022.134031

Received: February 25, 2022
Accepted: April 19, 2022
Published: April 22, 2022

Abstract

In 1909 the brothers E. and F. Cosserat discovered a new nonlinear group theoretical approach to elasticity (EL), with the only experimental need to measure the EL constants. In a modern language, their idea has been to use the nonlinear Spencer sequence instead of the nonlinear Janet sequence for the Lie groupoid defining the group of rigid motions of space. Following H. Weyl, our purpose is to compute for the first time the nonlinear Spencer sequence for the Lie groupoid defining the conformal group of space-time in order to provide the mathematical foundations of electromagnetism (EM), with the only experimental need to measure the EM constant in vacuum. With a manifold of dimension n, the difficulty is to deal with the n nonlinear transformations that have been called "elations" by E. Cartan in 1922. Using the fact that dimension $n=4$ has very specific properties for the computation of the Spencer cohomology, we prove that there is thus no conceptual difference between the Cosserat EL field or induction equations and the Maxwell EM field or induction equations. As a byproduct, the well known field/matter couplings (piezzoelectricity, photoelasticity, streaming birefringence, …) can be described abstractly, with the only experimental need to measure the corresponding coupling constants. The main consequence of this paper is the need to revisit the mathematical foundations of gauge theory (GT) because we have proved that EM was depending on the conformal group and not on $U(1)$, with a shift by one step to the left in the physical interpretation of the differential sequence involved.

Keywords

Nonlinear Differential Sequences, Linear Differential Sequences, Lie Groupoids, Lie Algebroids, Conformal Group, Spencer Cohomology, Maxwell Equations, Cosserat Equations

1. Introduction

Let us start this paper with a personal but meaningful story that has oriented my

research during the last forty years or so, since the French "*Grandes Ecoles*" created their own research laboratories. Being a fresh permanent researcher of Ecole Nationale des Ponts et Chaussées in Paris, the author of this paper has been asked to become the scientific adviser of a young student in order to introduce him to research. As General Relativity was far too much difficult for somebody without any specific mathematical knowledge while remembering his own experience at the same age, he asked the student to collect about 50 books of Special Relativity and classify them along the way each writer was avoiding the use of the conformal group of space-time implied by the Michelson and Morley experiment, only caring about the Poincaré or Lorentz subgroups. After six months, the student (like any reader) arrived at the fact that most books were almost copying each other and could be nevertheless classified into three categories:

- 30 books, including the original 1905 paper ([1]) by Einstein, were at once, as a working assumption, deciding to restrict their study to a linear group reducing to the Galilée group when the speed of light was going to infinity. It must be noticed that people did believe that Einstein had not been influenced in 1905 by the Michelson and Morley experiment of 1887 till the discovery of hand written notes taken during lectures given by Einstein in Chicago (1921) and Kyoto (1922).
- 15 books were trying to "*prove*" that the conformal factor was indeed reduced to a constant equal to 1 when space-time was supposed to be homogeneous and isotropic.
- 5 books *only* were claiming that the conformal factor could eventually depend on the property of space-time, adding however that, if there was no surrounding electromagnetism or gravitation, the situation should be reduced to the preceding one but nothing was said otherwise.

The student was so disgusted by such a state of affair that he decided to give up on research and to become a normal civil engineer. As a byproduct, if group theory must be used, the underlying group of transformations of space-time *must* be related to the propagation of light *by itself* rather than by considering tricky signals between observers, thus *must* have to do with the biggest group of invariance of Maxwell Equations ([2] [3]). However, at the time we got the solution of this problem with the publication of ([4]) in 1988 (See [5] for recent results), a deep confusion was going on which is still not acknowledged though it can be explained in a few lines ([6]). Using standard notations of differential geometry, if the 2-form $F \in \wedge^2 T^*$ describing the EM field is satisfying the *first set* of Maxwell equations, it amounts to say that it is closed, that is killed by the exterior derivative $d : \wedge^2 T^* \to \wedge^3 T^*$. The EM field can be thus (locally) parametrized by the EM potential 1-form $A \in T^*$ with $dA = F$ where $d : T^* \to \wedge^2 T^*$ is again the exterior derivative, because $d^2 = d \circ d = 0$. Now, if E is a vector bundle over a manifold X of dimension n, then we may define its *adjoint* vector bundle $ad(E) = \wedge^n T^* \otimes E^*$ where E^* is obtained from E by inverting the

transition rules, like T^* is obtained from $T=T(X)$ and such a construction can be extended to linear partial differential operators between (sections of) vector bundles. When $n=4$, it follows that the *second set* of Maxwell equations for the EM induction is just described by $ad(d):\wedge^4T^*\otimes\wedge^2T\to\wedge^4T^*\otimes T$, *independently of any Minkowski constitutive relation between field and induction*. Using *Hodge duality* with respect to the volume form $dx=dx^1\wedge\cdots\wedge dx^4$, this operator is isomorphic to $d:\wedge^2T^*\to\wedge^3T^*$. It follows that *both the first set and second set of Maxwell equations are invariant by any diffeomorphism* and that the conformal group of space-time is the biggest group of transformations preserving the Minkowski constitutive relations *in vacuum* where the speed of light is truly c as a universal constant. It was thus natural to believe that the mathematical structure of electromagnetism and gravitation had only to do with such a group having:

$$4\ \text{translations}+6\ \text{rotations}+1\ \text{dilatation}+4\ \text{elations}=15\ \text{parameters}$$

the main difficulty being to deal with these later non-linear transformations. Of course, such a challenge could not be solved without the help of the non-linear theory of partial differential equations and Lie pseudogroups combined with homological algebra, that is before 1995 *at least* ([7]).

From a purely physical point of view, these new nonlinear methods have been introduced for the first time in 1909 by the brothers E. and F. Cosserat for studying the mathematical foundations of EL ([8]-[14]). We have presented their link with the *nonlinear Spencer differential sequences* existing in the formal theory of Lie pseudogroups at the end of our book "Differential Galois Theory" published in 1983 ([15]). Similarly, the conformal methods have been introduced by H. Weyl in 1918 for revisiting the mathematical foundations of EM ([3]). We have presented their link with the above approach through a *unique* differential sequence only depending on the structure of the *conformal group* in our book "Lie Pseudogroups and Mechanics" published in 1988 ([4]). However, the Cosserat brothers were only dealing with *translations* and *rotations* while Weyl was only dealing with *dilatation* and *elations*. Also, as an additional condition not fulfilled by the classical Einstein-Fokker-Nordström theory ([16]), if the conformal factor has to do with gravitation, it *must* be defined everywhere but at the central attractive mass as we already said.

From a purely mathematical point of view, the concept of a finite length *differential sequence*, now called *Janet sequence*, has been first described as a footnote by M. Janet in 1920 ([17]). Then, the work of D. C. Spencer in 1970 has been the first attempt to use the formal theory of systems of partial differential equations that he developed himself in order to study the formal theory of Lie pseudogroups ([18] [19] [20]). However, the *nonlinear Spencer sequences for Lie pseudogroups*, though never used in physics, largely supersede the "*Cartan structure equations*" introduced by E.Cartan in 1905 ([21] [22]) and are different from the "*Vessiot structure equations*" introduced by E. Vessiot in 1903 ([23]) or 1904 ([24]) for the same purpose but still not known today after more than a

century because they have never been acknowledged by Cartan himself or even by his successors.

The purpose of the present difficult paper is to apply these new methods for studying the common nonlinear conformal origin of electromagnetism *and* gravitation, *in a purely mathematical way*, by constructing explicitly the corresponding nonlinear Spencer sequence. All the physical consequences will be presented in another paper.

2. Groupoids and Algebroids

Let us now turn to the clever way proposed by Vessiot in 1903 ([23]) and 1904 ([24]). Our purpose is only to sketch the main results that we have obtained in many books ([4] [7] [13] [15], we do not know other references) and to illustrate them by a series of specific examples, asking the reader to imagine any link with what has been said. We break the study into 8 successive steps.

1) If $\mathcal{E}=X\times X$, we shall denote by $\Pi_q=\Pi_q(X,X)$ the open sub-fibered manifold of $J_q(X\times X)$ defined independently of the coordinate system by $det(y_i^k)\neq 0$ with *source projection* $\alpha_q:\Pi_q\to X:(x,y_q)\to(x)$ and *target projection* $\beta_q:\Pi_q\to X:(x,y_q)\to(y)$. We shall sometimes introduce a copy Y of X with local coordinates (y) in order to avoid any confusion between the source and the target manifolds. In order to construct another nonlinear sequence, we need a few basic definitions on *Lie groupoids* and *Lie algebroids* that will become substitutes for Lie groups and Lie algebras. The first idea is to use the chain rule for derivatives $j_q(g\circ f)=j_q(g)\circ j_q(f)$ whenever $f,g\in aut(X)$ can be composed and to replace both $j_q(f)$ and $j_q(g)$ respectively by f_q and g_q in order to obtain the new section $g_q\circ f_q$. This kind of "composition" law can be written in a symbolic way by introducing another copy Z of X with local coordinates (z) as follows:

$$\gamma_q:\Pi_q(Y,Z)\times_Y\Pi_q(X,Y)\to\Pi_q(X,Z):$$
$$\left(y,z,\frac{\partial z}{\partial y},\cdots\right),\left(x,y,\frac{\partial y}{\partial x},\cdots\right)\to\left(x,z,\frac{\partial z}{\partial y}\frac{\partial y}{\partial x},\cdots\right)$$

We may also define $j_q(f)^{-1}=j_q(f^{-1})$ and obtain similarly an "inversion" law.

DEFINITION 2.1: A fibered submanifold $\mathcal{R}_q\subset\Pi_q$ is called a *system of finite Lie equations* or a *Lie groupoid* of order q if we have an induced *source projection* $\alpha_q:\mathcal{R}_q\to X$, *target projection* $\beta_q:\mathcal{R}_q\to X$, *composition* $\gamma_q:\mathcal{R}_q\times_X\mathcal{R}_q\to\mathcal{R}_q$, *inversion* $\iota_q:\mathcal{R}_q\to\mathcal{R}_q$ and *identity* $j_q(id)=id_q:X\to\mathcal{R}_q$. In the sequel we shall only consider *transitive* Lie groupoids such that the map $(\alpha_q,\beta_q):\mathcal{R}_q\to X\times X$ is an epimorphism and we shall denote by $\mathcal{R}_q^0=id^{-1}(\mathcal{R}_q)$ the *isotropy Lie group bundle* of $\mathcal{R}_q$. Also, one can prove that the new system $\rho_r(\mathcal{R}_q)=\mathcal{R}_{q+r}$ obtained by differentiating r times all the defining equations of $\mathcal{R}_q$ is a Lie groupoid of order $q+r$.

Let us start with a Lie pseudogroup $\Gamma\subset aut(X)$ defined by a system

$\mathcal{R}_q \subset \Pi_q$ of order q. Roughly speaking, if $f, g \in \Gamma \Rightarrow g \circ f, f^{-1} \in \Gamma$ but such a definition is totally meaningless in actual practice as it cannot be checked most of the time. In all the sequel we shall suppose that the system is involutive ([4] [7] [13] [15] [25]) and that Γ is *transitive* that is $\forall x, y \in X, \exists f \in \Gamma, y = f(x)$ or, equivalently, the map $(\alpha_q, \beta_q): \mathcal{R}_q \to X \times X : (x, y_q) \to (x, y)$ is surjective.

2) The Lie algebra $\Theta \subset T$ of infinitesimal transformations is then obtained by linearization, setting $y = x + t\xi(x) + \cdots$ and passing to the limit $t \to 0$ in order to obtain the linear involutive system $R_q = id_q^{-1}(V(\mathcal{R}_q)) \subset J_q(T)$ by reciprocal image with $\Theta = \{\xi \in T \mid j_q(\xi) \in R_q\}$. We define the *isotropy Lie algebra bundle* $R_q^0 \subset J_q^0(T)$ by the short exact sequence $0 \to R_q^0 \to R_q \xrightarrow{\pi_0^q} T \to 0$.

3) Passing from source to target, we may *prolong* the vertical infinitesimal transformations $\eta = \eta^k(y)\frac{\partial}{\partial y^k}$ to the jet coordinates up to order q in order to obtain:

$$\eta^k(y)\frac{\partial}{\partial y^k} + \left(\frac{\partial \eta^k}{\partial y^r} y_i^r\right)\frac{\partial}{\partial y_i^k} + \left(\frac{\partial^2 \eta^k}{\partial y^r \partial y^s} y_i^r y_j^s + \frac{\partial \eta^k}{\partial y^r} y_{ij}^r\right)\frac{\partial}{\partial y_{ij}^k} + \cdots$$

where we have replaced $j_q(f)(x)$ by y_q, each component being the "formal" derivative of the previous one .

4) As $[\Theta, \Theta] \subset \Theta$, we may use the Frobenius theorem in order to find a generating fundamental set of *differential invariants* $\{\Phi^\tau(y_q)\}$ up to order q which are such that $\Phi^\tau(\overline{y}_q) = \Phi^\tau(y_q)$ by using the chain rule for derivatives whenever $\overline{y} = g(y) \in \Gamma$ acting now on Y. Specializing the Φ^τ at $id_q(x)$ we obtain the *Lie form* $\Phi^\tau(y_q) = \omega^\tau(x)$ of $\mathcal{R}_q$.

Of course, in actual practice *one must use sections of* R_q *instead of solutions* and we now prove why the use of the Spencer operator becomes crucial for such a purpose. Indeed, using the *algebraic bracket* $\{j_{q+1}(\xi), j_{q+1}(\eta)\} = j_q([\xi, \eta]), \forall \xi, \eta \in T$, we may obtain by bilinearity a *differential bracket* on $J_q(T)$ extending the bracket on T:

$$[\xi_q, \eta_q] = \{\xi_{q+1}, \eta_{q+1}\} + i(\xi) D\eta_{q+1} - i(\eta) D\xi_{q+1}, \forall \xi_q, \eta_q \in J_q(T)$$

which does not depend on the respective lifts ξ_{q+1} and η_{q+1} of ξ_q and η_q in $J_{q+1}(T)$. This bracket on sections satisfies the Jacobi identity and we set ([4] [7] [13] [25]):

DEFINITION 2.2: We say that a vector subbundle $R_q \subset J_q(T)$ is a *system of infinitesimal Lie equations* or a *Lie algebroid* if $[R_q, R_q] \subset R_q$, that is to say $\xi_q, \eta_q \in R_q \Rightarrow [\xi_q, \eta_q] \in R_q$. Such a definition can be tested by means of computer algebra. We shall also say that R_q is *transitive* if we have the short exact sequence $0 \to R_q^0 \to R_q \xrightarrow{\pi_0^q} T \to 0$. In that case, a *splitting* of this sequence, namely a map $\chi_q : T \to R_q$ such that $\pi_0^q \circ \chi_q = id_T$ or equivalently a section $\chi_q \in T^* \otimes R_q$ over $id_T \in T^* \otimes T$, is called a R_q-*connection* and its *curvature* $\kappa_q \in \wedge^2 T^* \otimes R_q^0$ is defined by $\kappa_q(\xi, \eta) = [\chi_q(\xi), \chi_q(\eta)] - \chi_q([\xi, \eta]), \forall \xi, \eta \in T$.

PROPOSITION 2.3: If $\left[R_q,R_q\right]\subset R_q$, then $\left[R_{q+r},R_{q+r}\right]\subset R_{q+r},\forall r\geq 0$.

Proof: When $r=1$, we have $\rho_1\left(R_q\right)=R_{q+1}=\left\{\xi_{q+1}\in J_{q+1}\left(T\right)\,|\,\xi_q\in R_q,D\xi_{q+1}\in T^*\otimes R_q\right\}$ and we just need to use the following formulas showing how D acts on the various brackets (See [7] and [25] for more details):

$$i\left(\zeta\right)D\left\{\xi_{q+1},\eta_{q+1}\right\}=\left\{i\left(\zeta\right)D\xi_{q+1},\eta_q\right\}+\left\{\xi_q,i\left(\zeta\right)D\eta_{q+1}\right\},\quad\forall\zeta\in T$$

$$\begin{aligned}i\left(\zeta\right)D\left[\xi_{q+1},\eta_{q+1}\right]&=\left[i\left(\zeta\right)D\xi_{q+1},\eta_q\right]+\left[\xi_q,i\left(\zeta\right)D\eta_{q+1}\right]\\&\quad+i\left(L\left(\eta_1\right)\zeta\right)D\xi_{q+1}-i\left(L\left(\xi_1\right)\zeta\right)D\eta_{q+1}\end{aligned}$$

because the right member of the second formula is a section of R_q whenever $\xi_{q+1},\eta_{q+1}\in R_{q+1}$. The first formula may be used when R_q is formally integrable. □

EXAMPLE 2.4: With $n=1,q=3,X=\mathbb{R}$ and evident notations, the components of $\left[\xi_3,\eta_3\right]$ at order zero, one, two and three are defined by the totally unusual successive formulas:

$$\left[\xi,\eta\right]=\xi\partial_x\eta-\eta\partial_x\xi$$

$$\left(\left[\xi_1,\eta_1\right]\right)_x=\xi\partial_x\eta_x-\eta\partial_x\xi_x$$

$$\left(\left[\xi_2,\eta_2\right]\right)_{xx}=\xi_x\eta_{xx}-\eta_x\xi_{xx}+\xi\partial_x\eta_{xx}-\eta\partial_x\xi_{xx}$$

$$\left(\left[\xi_3,\eta_3\right]\right)_{xxx}=2\xi_x\eta_{xxx}-2\eta_x\xi_{xxx}+\xi\partial_x\eta_{xxx}-\eta\partial_x\xi_{xxx}$$

For affine transformations, $\xi_{xx}=0,\eta_{xx}=0\Rightarrow\left(\left[\xi_2,\eta_2\right]\right)_{xx}=0$ and thus $\left[R_2,R_2\right]\subset R_2$.

For projective transformations, $\xi_{xxx}=0,\eta_{xxx}=0\Rightarrow\left(\left[\xi_3,\eta_3\right]\right)_{xxx}=0$ and thus $\left[R_3,R_3\right]\subset R_3$.

THEOREM 2.5: (*prolongation/projection* (PP) *procedure*) If an arbitrary system $R_q\subseteq J_q\left(E\right)$ is given, one can *effectively* find two integers $r,s\geq 0$ such that the system $R_{q+r}^{(s)}$ is formally integrable or even involutive.

COROLLARY 2.6: The bracket is compatible with the PP procedure:

$$\left[R_q,R_q\right]\subset R_q\quad\Rightarrow\quad\left[R_{q+r}^{(s)},R_{q+r}^{(s)}\right]\subset R_{q+r}^{(s)},\forall r,s\geq 0$$

EXAMPLE 2.7: With $n=m=2$ and $q=1$, let us consider the Lie pseudogroup $\Gamma\subset aut\left(X\right)$ of finite transformations $y=f\left(x\right)$ such that $y^2dy^1=x^2dx^1=\omega=\left(x^2,0\right)\in T^*$. Setting $y=x+t\xi\left(x\right)+\cdots$ and linearizing, we get the Lie operator $\mathcal{D}\xi=\mathcal{L}\left(\xi\right)\omega$ where $\mathcal{L}$ is the Lie derivative because it is well known that $\left[\mathcal{L}\left(\xi\right),\mathcal{L}\left(\eta\right)\right]=\mathcal{L}\left(\xi\right)\circ\mathcal{L}\left(\eta\right)-\mathcal{L}\left(\eta\right)\circ\mathcal{L}\left(\xi\right)=\mathcal{L}\left(\left[\xi,\eta\right]\right)$ in the operator sense. The system $R_1\subset J_1\left(T\right)$ of linear infinitesimal Lie equations is:

$$x^2\partial_1\xi^1+\xi^2=0,\quad\partial_2\xi^1=0$$

Replacing $j_1\left(\xi\right)$ by a section $\xi_1\in J_1\left(T\right)$, we have:

$$\xi_1^1=-\frac{1}{x^2}\xi^2,\quad\xi_2^1=0$$

Let us choose the two sections:

$$\xi_1 = \left\{\xi^1 = 0, \xi^2 = -x^2, \xi_1^1 = 1, \xi_2^1 = 0, \xi_1^2 = 0, \xi_2^2 = 0\right\} \in R_1$$

$$\eta_1 = \left\{\eta^1 = x^2, \eta^2 = 0, \eta_1^1 = 0, \eta_2^1 = -x^2, \eta_1^2 = 0, \eta_2^2 = 1\right\} \in R_1$$

We let the reader check that $[\xi_1, \eta_1] \in R_1$. However, we have the strict inclusion $R_1^{(1)} \subset R_1$ defined by the additional equation $\xi_1^1 + \xi_2^2 = 0$ and thus $\xi_1, \eta_1 \notin R_1^{(1)}$ though we have indeed $\left[R_1^{(1)}, R_1^{(1)}\right] \subset R_1^{(1)}$, a result not evident at all because the sections ξ_1 and η_1 have *nothing to do* with solutions. The reader may proceed in the same way with $x^2 dx^1 - x^1 dx^2$ and compare.

5) The main discovery of Vessiot, as early as in 1903 and thus fifty years in advance, has been to notice that the prolongation at order q of any horizontal vector field $\xi = \xi^i(x)\dfrac{\partial}{\partial x^i}$ commutes with the prolongation at order q of any vertical vector field $\eta = \eta^k(y)\dfrac{\partial}{\partial y^k}$, exchanging therefore the differential invariants. Keeping in mind the well known property of the Jacobian determinant while passing to the finite point of view, any (local) transformation $y = f(x)$ can be lifted to a (local) transformation of the differential invariants between themselves of the form $u \to \lambda\left(u, j_q(f)(x)\right)$ allowing to introduce a *natural bundle* $\mathcal{F}$ over X by patching changes of coordinates $\overline{x} = \varphi(x), \overline{u} = \lambda\left(u, j_q(\varphi)(x)\right)$. A section ω of $\mathcal{F}$ is called a *geometric object* or *structure* on X and transforms like $\overline{\omega}\left(f(x)\right) = \lambda\left(\omega(x), j_q(f)(x)\right)$ or simply $\overline{\omega} = j_q(f)(\omega)$. This is a way to generalize vectors and tensors ($q = 1$) or even connections ($q = 2$). As a byproduct we have $\Gamma = \left\{f \in aut(X) \,|\, \Phi_\omega\left(j_q(f)\right) = j_q(f)^{-1}(\omega) = \omega\right\}$ as a new way to write out the Lie form and we may say that Γ *preserves* ω. We also obtain $\mathcal{R}_q = \left\{f_q \in \Pi_q \,|\, f_q^{-1}(\omega) = \omega\right\}$. Coming back to the infinitesimal point of view and setting $f_t = exp(t\xi) \in aut(X), \forall \xi \in T$, we may define the *ordinary Lie derivative* with value in $F = \omega^{-1}\left(V(\mathcal{F})\right)$ by introducing the *vertical bundle* of $\mathcal{F}$ as a vector bundle over $\mathcal{F}$ and the formula:

$$\mathcal{D}\xi = \mathcal{L}(\xi)\omega = \frac{d}{dt} j_q(f_t)^{-1}(\omega)\bigg|_{t=0} \Rightarrow \Theta = \left\{\xi \in T \,|\, \mathcal{L}(\xi)\omega = 0\right\}$$

while we have $x \to x + t\xi(x) + \cdots \Rightarrow u^\tau \to u^\tau + t\partial_\mu \xi^k L_k^{\tau\mu}(u) + \cdots$ where $\mu = (\mu_1, \cdots, \mu_n)$ is a multi-index as a way to write down the system $R_q \subset J_q(T)$ of infinitesimal Lie equations in the *Medolaghi form*:

$$\Omega^\tau \equiv \left(\mathcal{L}(\xi)\omega\right)^\tau \equiv -L_k^{\tau\mu}\left(\omega(x)\right)\partial_\mu \xi^k + \xi^r \partial_r \omega^\tau(x) = 0$$

EXAMPLE 2.8: With $n = 1$, let us consider the Lie group of projective transformations $y = (ax + b)/(cx + d)$ as a lie pseudogroup. Differentiating three times in order to eliminate the parameters, we obtain the third order *Schwarzian* OD equation and its linearization over $y = x$:

$$\mathcal{R}_3 \subset \Pi_3, \quad \Psi \equiv \frac{y_{xxx}}{y_x} - \frac{3}{2}\left(\frac{y_{xx}}{y_x}\right)^2 = 0$$

$$R_3 \subset J_3(T), \quad \xi_{xxx} = 0$$

Accordingly, the prolongation $\#(\eta_3)$ of any $\eta_3 \in J_3(T(Y))$ over Y such that $\eta_{yyy} = 0$ becomes:

$$\eta(y)\frac{\partial}{\partial y} + \eta_y(y)\left(y_x\frac{\partial}{\partial y_x} + y_{xx}\frac{\partial}{\partial y_{xx}} + y_{xxx}\frac{\partial}{\partial y_{xxx}}\right)$$
$$+\eta_{yy}(y)\left((y_x)^2\frac{\partial}{\partial y_{xx}} + 3y_x y_{xx}\frac{\partial}{\partial y_{xxx}}\right)$$

It follows that Ψ is a generating third order differential invariant and R_3 is in Lie form.

Now, we have:

$$\overline{x} = \varphi(x) \Rightarrow y_x = y_{\overline{x}}\partial_x\varphi, y_{xx} = y_{\overline{xx}}(\partial_x\varphi)^2 + y_{\overline{x}}\partial_{xx}\varphi,$$
$$y_{xxx} = y_{\overline{xxx}}(\partial_x\varphi)^3 + 3y_{\overline{xx}}\partial_x\varphi\partial_{xx}\varphi + y_{\overline{x}}\partial_{xxx}\varphi$$

and the natural bundle $\mathcal{F}$ is thus defined by the transition rules:

$$\overline{x} = \varphi(x), \ u = \overline{u}(\partial_x\varphi)^2 + \left(\frac{\partial_{xxx}\varphi}{\partial_x\varphi} - \frac{3}{2}\left(\frac{\partial_{xx}\varphi}{\partial_x\varphi}\right)^2\right)$$

The general Lie form of $\mathcal{R}_3$ is:

$$\frac{y_{xxx}}{y_x} - \frac{3}{2}\left(\frac{y_{xx}}{y_x}\right)^2 + \omega(y)(y_x)^2 = \omega(x)$$

We obtain $R_3 \subset J_3(T)$ in Medolaghi form as follows:

$$\Omega \equiv \mathcal{L}(\xi)\omega \equiv \partial_{xxx}\xi + 2\omega(x)\partial_x\xi + \xi\partial_x\omega(x) = 0$$

Using a section $\xi_3 \in J_3(T)$, we finally get the formal Lie derivative:

$$\Omega \equiv L(\xi_3)\omega \equiv \xi_{xxx} + 2\omega(x)\xi_x + \xi\partial_x\omega(x) = 0$$

and let the reader ckeck directly that $\left[L(\xi_3), L(\eta_3)\right] = L\left(\left[\xi_3, \eta_3\right]\right)$, $\forall \xi_3, \eta_3 \in J_3(T)$, a result absolutely not evident at first sight.

6) By analogy with "special" and "general" relativity, we shall call the given section *special* and any other arbitrary section *general*. The problem is now to study the formal properties of the linear system just obtained with coefficients only depending on $j_1(\omega)$, exactly like L.P. Eisenhart did for $\mathcal{F} = S_2T^*$ when finding the constant Riemann curvature condition for a metric ω with $det(\omega) \neq 0$ ([7], Example 10, p 246 to 256, [26]). Indeed, if any expression involving ω and its derivatives is a scalar object, it must reduce to a constant because Γ is assumed to be transitive and thus cannot be defined by any zero order equation. Now one can prove that the CC for $\overline{\omega}$, thus for ω too, only depend on the Φ and take the quasi-linear symbolic form $v \equiv I(u_1) \equiv A(u)u_x + B(u) = 0$, allowing to define an affine subfibered manifold $\mathcal{B}_1 \subset J_1(\mathcal{F})$ over $\mathcal{F}$. Now, if one has two sections ω and $\overline{\omega}$ of $\mathcal{F}$, the *equivalence problem* is to look for $f \in aut(X)$ such that $j_q(f)^{-1}(\omega) = \overline{\omega}$. When the two sections satisfy the same CC, the problem is sometimes locally

possible (Lie groups of transformations, Darboux problem in analytical mechanics, …) but sometimes not ([25], p. 333).

7) Instead of the CC for the equivalence problem, let us look for the *integrability conditions* (IC) for the system of infinitesimal Lie equations and suppose that, for the given section, all the equations of order $q+r$ are obtained by differentiating r times only the equations of order q, then it was claimed by Vessiot ([23] with no proof, see [7], pp. 207-211) that such a property is held if and only if there is an equivariant section $c:\mathcal{F}\to\mathcal{F}_1:(x,u)\to(x,u,v=c(u))$ where $\mathcal{F}_1=J_1(\mathcal{F})/\mathcal{B}_1$ is a natural vector bundle over $\mathcal{F}$ with local coordinates (x,u,v). Moreover, any such equivariant section depends on a finite number of constants c called *structure constants* and the IC for the *Vessiot structure equations* $I(u_1)=c(u)$ are of a polynomial form $J(c)=0$.

EXAMPLE 2.9: Comig back to Example 2.7 first considered by Vessiot as early as in 1903 ([23]), the geometric object $\omega=(\alpha,\beta)\in T^*\otimes_X\wedge^2T^*$ must satisfy the Vessiot structure equation $d\alpha=c\beta$ with a single Vessiot structure constant $c=-1$ in the situation considered where $\alpha=x^2dx^1$ and $\beta=dx^1\wedge dx^2$ (See ([27]) for other examples and applications). As a byproduct, there is no conceptual difference between such a constant and the constant appearing in the constant Riemannian curvature condition of Eisenhart ([26]).

8) Finally, when Y is no longer a copy of X, a system $\mathcal{A}_q\subset J_q(X\times Y)$ is said to be an *automorphic system* for a Lie pseudogroup $\Gamma\subset aut(Y)$ if, whenever $y=f(x)$ and $\overline{y}=\overline{f}(x)$ are two solutions, then there exists one and only one transformation $\overline{y}=g(y)\in\Gamma$ such that $\overline{f}=g\circ f$. Explicit tests for checking such a property formally have been given in ([15]) and can be implemented on computer in the differential algebraic framework.

3. Nonlinear Sequences

Contrary to what happens in the study of Lie pseudogroups and in particular in the study of the algebraic ones that can be found in mathematical physics, *nonlinear operators do not in general admit* CC, unless they are defined by differential polynomials, as can be seen by considering the two following examples with $m=1,n=2,q=2$. With standard notations from differential algebra, if we are dealing with a ground differential field K, like $\mathbb{Q}$ in the next examples, we denote by $K\{y\}$ the ring (which is even an integral domain) of differential polynomials in y with coefficients in K and by $K\langle y\rangle=Q(K\{y\})$ the corresponding quotient field of differential rational functions in y. Then, if $u,v\in K\langle y\rangle$, we have the two towers $K\subset K\langle u\rangle\subset K\langle y\rangle$ and $K\subset K\langle v\rangle\subset K\langle y\rangle$ of extensions, thus the tower $K\subset K\langle u,v\rangle\subset K\langle y\rangle$. Accordingly, the differential extension $K\langle u,v\rangle/K$ is a finitely generated differential extension. If we consider u and v as new indeterminates, then $K\langle u\rangle$ and $K\langle v\rangle$ are both differential transcendental extensions of K and the kernel of the canonical differential morphism $K\{u\}\otimes_K K\{v\}\to K\langle y\rangle$ is a prime differential ideal in the differential integral domain $K\langle y\rangle\otimes_K K\langle v\rangle$, a way to describe by residue the smallest differential

field containing $K\langle u\rangle$ and $K\langle v\rangle$ in $K\langle y\rangle$. Of course, the true difficulty is to find out such a prime differential ideal.

EXAMPLE 3.1: First of all, let us consider the following nonlinear system in y with second member (u,v):

$$P\equiv y_{22}-\frac{1}{3}(y_{11})^3=u,\ Q\equiv y_{12}-\frac{1}{2}(y_{11})^2=v\Rightarrow y_{11}=\frac{u_1-v_2}{v_1}$$

The differential ideal $\mathfrak{a}$ generated by P and Q in $\mathbb{Q}\{y\}$ is prime because $d_2Q+d_1P-y_{11}d_1Q=0$ and thus $\mathbb{Q}\{y\}/\{P,Q\}\simeq\mathbb{Q}[y,y_1,y_2,y_{11},y_{111},\cdots]$ is an integral domain.

We may consider the following nonlinear involutive system with two equations:

$$\begin{cases}y_{22}-\frac{1}{3}(y_{11})^3=0 & \boxed{\begin{matrix}1 & 2\\ 1 & \bullet\end{matrix}}\\ y_{12}-\frac{1}{2}(y_{11})^2=0\end{cases}$$

We have also the linear inhomogeneous finite type second order system with three equations:

$$\begin{cases}y_{22}=u+\frac{1}{3}\left(\frac{u_1-v_2}{v_1}\right)^3 & \boxed{\begin{matrix}1 & 2\\ 1 & \bullet\\ 1 & \bullet\end{matrix}}\\ y_{12}=v+\frac{1}{2}\left(\frac{u_1-v_2}{v_1}\right)^2\\ y_{11}=\frac{u_1-v_2}{v_1}\end{cases}$$

Though we have *a priori* two CC, we let the reader prove, as a delicate exercise, that there is only the single nonlinear second order CC obtained from the bottom dot:

$$\boxed{d_2\left(\frac{u_1-v_2}{v_1}\right)-d_1\left(v+\frac{1}{2}\left(\frac{u_1-v_2}{v_1}\right)^2\right)=0}$$

EXAMPLE 3.2: On the contrary, if we consider the following new nonlinear system:

$$P\equiv y_{22}-\frac{1}{2}(y_{11})^2=u,\ Q\equiv y_{12}-y_{11}=v\Rightarrow(y_{11}-1)y_{111}=v_2+v_1-u_1=w$$

we obtain successively:

$$d_2Q+d_1Q-d_1P\equiv(y_{11}-1)y_{111}$$

$$y_{111}(d_{12}Q+d_{11}Q-d_{11}P)-y_{1111}(d_2Q+d_1Q-d_1P)=(y_{111})^3$$

The symbol at order 3 is thus not a vector bundle and no direct study as above can be used because the differential ideal generated by (P,Q) is not perfect as it contains $(y_{111})^3$ without containing y_{111} (See [15] and [28] for more de-

tails). The following nonlinear system is *not* involutive:

$$\begin{cases} y_{22} - \frac{1}{2}(y_{11})^2 = 0 \\ y_{12} - y_{11} = 0 \end{cases} \quad \boxed{\begin{matrix} 1 & 2 \\ 1 & \bullet \end{matrix}}$$

We have the following four generic nonlinear additional finite type third order equations:

$$\begin{cases} y_{222} - y_{11}\left(v_1 + \frac{w}{y_{11}-1}\right) = u_2 \\ y_{122} - y_{11}\frac{w}{y_{11}-1} = u_1 \\ y_{112} - \frac{w}{y_{11}-1} = v_1 \\ y_{111} - \frac{w}{y_{11}-1} = 0 \end{cases} \quad \boxed{\begin{matrix} 1 & 2 \\ 1 & \bullet \\ 1 & \bullet \\ 1 & \bullet \end{matrix}}$$

Though we have now *a priori* three CC and thus three additional equations because the system is not involutive, setting $y_{11} - 1 = z \Rightarrow y_{112} = z_2, y_{111} = z_1$, there is only the single additional nonlinear second order equation:

$$\boxed{v_{11}z^2 + (w_1 - w_2)z + v_1 w = 0}$$

Differentiating once and using the relation $zz_1 = w$, we get:

$$\boxed{v_{111}z^3 + (w_{11} - w_{12})z^2 + (v_1 w_1 + 3v_{11}w)z + (w_1 - w_2)w = 0}$$

a result leading to a tricky resultant providing a third order differential polynomial in (u, v).

However, the kernel of a linear operator $\mathcal{D}: E \to F$ is always taken with respect to the zero section of *F*, while it must be taken with respect to a prescribed section by a *double arrow* for a nonlinear operator. Keeping in mind the linear Janet sequence and the examples of Vessiot structure equations already presented, one obtains:

THEOREM 3.3: There exists a *nonlinear Janet sequence* associated with the Lie form of an involutive system of finite Lie equations:

$$0 \to \Gamma \to aut(X) \underset{\omega \circ \alpha}{\overset{\Phi \circ j_q}{\rightrightarrows}} \mathcal{F} \underset{0}{\overset{I \circ j_1}{\rightrightarrows}} \mathcal{F}_1$$

where the kernel of the first operator $f \to \Phi \circ j_q(f) = \Phi(j_q(f)) = j_q(f)^{-1}(\omega)$ is taken with respect to the section ω of $\mathcal{F}$ while the kernel of the second operator $\omega \to I(j_1(\omega)) \equiv A(\omega)\partial_x \omega + B(\omega)$ is taken with respect to the zero section of the vector bundle $\mathcal{F}_1$ over $\mathcal{F}$.

COROLLARY 3.4: By linearization at the identity, one obtains the involutive *Lie operator* $\mathcal{D}: T \to F_0: \xi \to \mathcal{L}(\xi)\omega$ with kernel $\Theta = \{\xi \in T \mid \mathcal{L}(\xi)\omega = 0\} \subset T$ satisfying $[\Theta, \Theta] \subset \Theta$ and the corresponding *linear Janet sequence*:

$$0 \to \Theta \to T \overset{\mathcal{D}}{\to} F_0 \overset{\mathcal{D}_1}{\to} F_1$$

where $F_0 = F = \omega^{-1}\left(V\left(\mathcal{F}\right)\right)$ and $F_1 = \omega^{-1}\left(\mathcal{F}_1\right)$.

Now we notice that T is a natural vector bundle of order 1 and $J_q\left(T\right)$ is thus a natural vector bundle of order $q+1$. Looking at the way a vector field and its derivatives are transformed under any $f \in aut\left(X\right)$ while replacing $j_q\left(f\right)$ by f_q, we obtain:

$$\eta^k\left(f\left(x\right)\right) = f_r^k\left(x\right)\xi^r\left(x\right) \Rightarrow \eta_u^k\left(f\left(x\right)\right) f_i^u\left(x\right) = f_r^k\left(x\right)\xi_i^r\left(x\right) + f_{ri}^k\left(x\right)\xi^r\left(x\right)$$

and so on, a result leading to:

LEMMA 3.5: $J_q\left(T\right)$ is *associated* with $\Pi_{q+1} = \Pi_{q+1}\left(X, X\right)$ that is we can obtain a new section $\eta_q = f_{q+1}\left(\xi_q\right)$ from any section $\xi_q \in J_q\left(T\right)$ and any section $f_{q+1} \in \Pi_{q+1}$ by the formula:

$$d_\mu \eta^k \equiv \eta_r^k f_\mu^r + \cdots = f_r^k \xi_\mu^r + \cdots + f_{\mu+1_r}^k \xi^r, \forall 0 \le \left|\mu\right| \le q$$

where the left member belongs to $V\left(\Pi_q\right)$. Similarly $R_q \subset J_q\left(T\right)$ is associated with $\mathcal{R}_{q+1} \subset \Pi_{q+1}$.

More generally, looking now for transformations "close" to the identity, that is setting $y = x + t\xi\left(x\right) + \cdots$ when $t \ll 1$ is a small constant parameter and passing to the limit $t \to 0$, we may linearize any (nonlinear) *system of finite Lie equations* in order to obtain a (linear) *system of infinitesimal Lie equations* $R_q \subset J_q\left(T\right)$ for vector fields. Such a system has the property that, if ξ, η are two solutions, then $\left[\xi, \eta\right]$ is also a solution. Accordingly, the set $\Theta \subset T$ of its solutions satisfies $\left[\Theta, \Theta\right] \subset \Theta$ and can therefore be considered as the Lie algebra of Γ.

More generally, the next definition will extend the *classical Lie derivative*:

$$\mathcal{L}\left(\xi\right)\omega = \left(i\left(\xi\right)d + di\left(\xi\right)\right)\omega = \frac{d}{dt} j_q\left(exp\, t\xi\right)^{-1}\left(\omega\right)\Big|_{t=0}.$$

DEFINITION 3.6: We say that a vector bundle F is *associated* with R_q if there exists a first order differential operator $L\left(\xi_q\right): F \to F$ called *formal Lie derivative* and such that:

1) $L\left(\xi_q + \eta_q\right) = L\left(\xi_q\right) + L\left(\eta_q\right) \quad \forall \xi_q, \eta_q \in R_q$.
2) $L\left(f\xi_q\right) = fL\left(\xi_q\right) \quad \forall \xi_q \in R_q, \forall f \in C^\infty\left(X\right)$.
3) $\left[L\left(\xi_q\right), L\left(\eta_q\right)\right] = L\left(\xi_q\right) \circ L\left(\eta_q\right) - L\left(\eta_q\right) \circ L\left(\xi_q\right) = L\left(\left[\xi_q, \eta_q\right]\right) \quad \forall \xi_q, \eta_q \in R_q$.
4) $L\left(\xi_q\right)\left(f\eta\right) = fL\left(\xi_q\right)\eta + \left(\xi f\right)\eta \quad \forall \xi_q \in R_q, \forall f \in C^\infty\left(X\right), \forall \eta \in F$.

LEMMA 3.7: If E and F are associated with R_q, we may set on $E \otimes F$:

$$L\left(\xi_q\right)\left(\eta \otimes \zeta\right) = L\left(\xi_q\right)\eta \otimes \zeta + \eta \otimes L\left(\xi_q\right)\zeta \quad \forall \xi_q \in R_q, \forall \eta \in E, \forall \zeta \in F$$

If $\Theta \subset T$ denotes the solutions of R_q, then we may set $\mathcal{L}\left(\xi\right) = L\left(j_q\left(\xi\right)\right), \forall \xi \in \Theta$ but no explicit computation can be done when Θ is infinite dimensional. However, we have:

PROPOSITION 3.8: $J_q\left(T\right)$ is associated with $J_{q+1}\left(T\right)$ if we define:

$$L\left(\xi_{q+1}\right)\eta_q = \left\{\xi_{q+1}, \eta_{q+1}\right\} + i\left(\xi\right)D\eta_{q+1} = \left[\xi_q, \eta_q\right] + i\left(\eta\right)D\xi_{q+1}$$

and thus R_q is associated with R_{q+1}.

Proof: It is easy to check the properties 1, 2, 4 and it only remains to prove property 3 as follows.

$$\begin{aligned}
&\left[L\left(\xi_{q+1}\right),L\left(\eta_{q+1}\right)\right]\zeta_q\\
&=L\left(\xi_{q+1}\right)\left(\left\{\eta_{q+1},\zeta_{q+1}\right\}+i\left(\eta\right)D\zeta_{q+1}\right)-L\left(\eta_{q+1}\right)\left(\left\{\xi_{q+1},\zeta_{q+1}\right\}+i\left(\xi\right)D\zeta_{q+1}\right)\\
&=\left\{\xi_{q+1},\left\{\eta_{q+2},\zeta_{q+2}\right\}\right\}-\left\{\eta_{q+1},\left\{\xi_{q+2},\zeta_{q+2}\right\}\right\}+\left\{\xi_{q+1},i\left(\eta\right)D\zeta_{q+2}\right\}\\
&\quad-\left\{\eta_{q+1},i\left(\xi\right)D\zeta_{q+2}\right\}+i\left(\xi\right)D\left\{\eta_{q+2},\zeta_{q+2}\right\}-i\left(\eta\right)D\left\{\xi_{q+2},\zeta_{q+2}\right\}\\
&\quad+i\left(\xi\right)D\left(i\left(\eta\right)D\zeta_{q+2}\right)-i\left(\eta\right)D\left(i\left(\xi\right)D\zeta_{q+2}\right)\\
&=\left\{\left\{\xi_{q+2},\eta_{q+2}\right\},\zeta_{q+1}\right\}+\left\{i\left(\xi\right)D\eta_{q+2},\zeta_{q+1}\right\}-\left\{i\left(\eta\right)D\xi_{q+2},\zeta_{q+1}\right\}+i\left(\left[\xi,\eta\right]\right)D\zeta_{q+1}\\
&=\left\{\left[\xi_{q+1},\eta_{q+1}\right],\zeta_{q+1}\right\}+i\left(\left[\xi,\eta\right]\right)D\zeta_{q+1}
\end{aligned}$$

by using successively the Jacobi identity for the algebraic bracket and the last proposition.

□

EXAMPLE 3.9: T and T^* both with any tensor bundle are associated with $J_1(T)$. For T we may define $L(\xi_1)\eta=[\xi,\eta]+i(\eta)D\xi_1=\{\xi_1,j_1(\eta)\}$. We have $\xi^r\partial_r\eta^k-\eta^s\partial_s\xi^k+\eta^s\left(\partial_s\xi^k-\xi^k_s\right)=-\eta^s\xi^k_s+\xi^r\partial_r\eta^k$ and the four properties of the formal Lie derivative can be checked directly. Of course, we find back $\mathcal{L}(\xi)\eta=[\xi,\eta],\forall\xi,\eta\in T$. We let the reader treat similarly the case of T^*.

PROPOSITION 3.10: There is a *first nonlinear Spencer sequence*:

$$0\to aut\left(X\right)\overset{j_{q+1}}{\to}\Pi_{q+1}\left(X,X\right)\overset{\bar{D}}{\to}T^*\otimes J_q\left(T\right)\overset{\bar{D}'}{\to}\wedge^2T^*\otimes J_{q-1}\left(T\right)$$

with $\bar{D}f_{q+1}\equiv f^{-1}_{q+1}\circ j_1\left(f_q\right)-id_{q+1}=\chi_q$

$\Rightarrow\bar{D}'\chi_q\left(\xi,\eta\right)\equiv D\chi_q\left(\xi,\eta\right)-\left\{\chi_q\left(\xi\right),\chi_q\left(\eta\right)\right\}=0$. Moreover, setting $\chi_0=A-id\in T^*\otimes T$, this sequence is locally exact if $det\left(A\right)\neq0$.

Proof: There is a canonical inclusion $\Pi_{q+1}\subset J_1\left(\Pi_q\right)$ defined by $y^k_{\mu,i}=y^k_{\mu+1_i}$ and the composition $f^{-1}_{q+1}\circ j_1\left(f_q\right)$ is a well defined section of $J_1\left(\Pi_q\right)$ over the section $f^{-1}_q\circ f_q=id_q$ of Π_q like id_{q+1}. The difference $\chi_q=f^{-1}_{q+1}\circ j_1\left(f_q\right)-id_{q+1}$ is thus a section of $T^*\otimes V\left(\Pi_q\right)$ over id_q and we have already noticed that $id_q^{-1}\left(V\left(\Pi_q\right)\right)=J_q\left(T\right)$. For $q=1$ we get with $g_1=f_1^{-1}$:

$$\chi^k_{,i}=g^k_l\partial_if^l-\delta^k_i=A^k_i-\delta^k_i,\quad\chi^k_{j,i}=g^k_l\left(\partial_if^l_j-A^r_if^l_{rj}\right)$$

We also obtain from Lemma 3.5 the useful formula $f^k_r\chi^r_{\mu,i}+\cdots+f^k_{\mu+1_r}\chi^r_{,i}=\partial_if^k_\mu-f^k_{\mu+1_i}$ allowing to determine χ_q inductively.

We refer to ([7], p 215-216) for the inductive proof of the local exactness, providing the only formulas that will be used later on and can be checked directly by the reader:

$$\partial_i\chi^k_{,j}-\partial_j\chi^k_{,i}-\chi^k_{i,j}+\chi^k_{j,i}-\left(\chi^r_{,i}\chi^k_{r,j}-\chi^r_{,j}\chi^k_{r,i}\right)=0\qquad(1)$$

$$\partial_i\chi^k_{l,j}-\partial_j\chi^k_{l,i}-\chi^k_{li,j}+\chi^k_{lj,i}-\left(\chi^r_{,i}\chi^k_{lr,j}+\chi^r_{l,i}\chi^k_{r,j}-\chi^r_{l,j}\chi^k_{r,i}-\chi^r_{,j}\chi^k_{lr,i}\right)=0\qquad(2)$$

$$\begin{aligned}&\partial_i\chi^k_{lr,j}-\partial_j\chi^k_{lr,i}-\chi^k_{lri,j}+\chi^k_{lrj,i}-\left(\chi^s_{,i}\chi^k_{lrs,j}+\chi^s_{r,i}\chi^k_{ls,j}+\chi^s_{l,i}\chi^k_{rs,j}\right.\\&\left.+\chi^s_{lr,i}\chi^k_{s,j}-\chi^s_{,j}\chi^k_{lrs,i}-\chi^s_{r,j}\chi^k_{ls,i}-\chi^s_{l,j}\chi^k_{rs,i}-\chi^s_{lr,j}\chi^k_{s,i}\right)=0\end{aligned}\tag{3}$$

There is no need for double-arrows in this framework as the kernels are taken with respect to the zero section of the vector bundles involved. We finally notice that the main difference with the gauge sequence is that *all the indices range from* 1 *to n* and that the condition $det(A)\neq 0$ amounts to $\Delta=det(\partial_i f^k)\neq 0$ because $det(f^k_i)\neq 0$ by assumption.

□

COROLLARY 3.11: There is a *restricted first nonlinear Spencer sequence*:

$$0\to\Gamma\xrightarrow{j_{q+1}}\mathcal{R}_{q+1}\xrightarrow{\bar{D}}T^*\otimes R_q\xrightarrow{\bar{D}'}\wedge^2T^*\otimes J_{q-1}(T)$$

DEFINITION 3.12: A *splitting* of the short exact sequence $0\to R^0_q\to R_q\xrightarrow{\pi^q_0}T\to 0$ is a map $\chi'_q:T\to R_q$ such that $\pi^q_0\circ\chi'_q=id_T$ or equivalently a section of $T^*\otimes R_q$ over $id_T\in T^*\otimes T$ and is called a R_q-*connection*. Its *curvature* $\kappa'_q\in\wedge^2T^*\otimes R^0_q$ is defined by $\kappa'_q(\xi,\eta)=\left[\chi'_q(\xi),\chi'_q(\eta)\right]-\chi'_q([\xi,\eta])$. We notice that $\chi'_q=-\chi_q$ is a connection with $\bar{D}'\chi'_q=\kappa'_q$ if and only if $A=0$. In particular $(\delta^k_i,-\gamma^k_{ij})$ is the only existing symmetric connection for the Killing system.

REMARK 3.13: Rewriting the previous local formulas with A instead of χ_0 we get:

$$\partial_iA^k_j-\partial_jA^k_i-A^r_i\chi^k_{r,j}+A^r_j\chi^k_{r,i}=0\tag{1*}$$

$$\partial_i\chi^k_{l,j}-\partial_j\chi^k_{l,i}-A^r_i\chi^k_{lr,j}+A^r_j\chi^k_{lr,i}-\chi^r_{l,i}\chi^k_{r,j}+\chi^r_{l,j}\chi^k_{r,i}=0\tag{2*}$$

$$\begin{aligned}&\partial_i\chi^k_{lr,j}-\partial_j\chi^k_{lr,i}-A^s_i\chi^k_{lrs,j}+A^s_j\chi^k_{lrs,i}\\&-\left(\chi^s_{r,i}\chi^k_{ls,j}+\chi^s_{l,i}\chi^k_{rs,j}+\chi^s_{lr,i}\chi^k_{s,j}-\chi^s_{r,j}\chi^k_{ls,i}-\chi^s_{l,j}\chi^k_{rs,i}-\chi^s_{lr,j}\chi^k_{s,i}\right)=0\end{aligned}\tag{3*}$$

When $q=1,g_2=0$ and though surprising it may look like, we find back *exactly* all the formulas presented by E. and F. Cosserat in ([10], p 123). Even more strikingly, *in the case of a Riemann structure*, *the last two terms disappear but the quadratic terms are left while*, *in the case of screw and complex structures*, *the quadratic terms disappear but the last two terms are left*. We finally notice that $\chi'_q=-\chi_q$ is a R_q-connection if and only if $A=0$, *a result contradicting the use of connections in physics*. However, when $A=0$, we have $\chi'_0(\xi)=\xi$ and thus:

$$\begin{aligned}\bar{D}'\chi_{q+1}&=\left(D\chi_{q+1}\right)(\xi,\eta)-\left(\left[\chi_q(\xi),\chi_q(\eta)\right]+i(\xi)D\left(\chi_{q+1}(\eta)\right)-i(\eta)D\left(\chi_{q+1}(\xi)\right)\right)\\&=-\left[\chi_q(\xi),\chi_q(\eta)\right]-\chi_q([\xi,\eta])\\&=-\left(\left[\chi'_q(\xi),\chi'_q(\eta)\right]-\chi'_q([\xi,\eta])\right)\\&=-\kappa'_q(\xi,\eta)\end{aligned}$$

does not depend on the lift of χ_q.

COROLLARY 3.14: When $det(A)\neq 0$ there is a *second nonlinear Spencer sequence* stabilized at order q:

$$0 \to aut(X) \to^{j_q} \Pi_q \to^{\bar{D}_1} C_1(T) \to^{\bar{D}_2} C_2(T)$$

where $\bar{D}_1$ and $\bar{D}_2$ are involutive and a *restricted second nonlinear Spencer sequence*:

$$0 \to \Gamma \to^{j_q} \mathcal{R}_q \to^{\bar{D}_1} C_1 \to^{\bar{D}_2} C_2$$

such that $\bar{D}_1$ and $\bar{D}_2$ are involutive whenever $\mathcal{R}_q$ is involutive.

Proof: With $|\mu| = q$ we have $\chi^k_{\mu,i} = -g^k_l A^r_i f^l_{\mu+1_r} + terms(order \le q)$. Setting $\chi^k_{\mu,i} = A^r_i \tau^k_{\mu,r}$, we obtain $\tau^k_{\mu,r} = -g^k_l f^l_{\mu+1_r} + terms(order \le q)$ and $\bar{D}: \Pi_{q+1} \to T^* \otimes J_q(T)$ restricts to $\bar{D}_1 : \Pi_q \to C_1(T)$.

Finally, setting $A^{-1} = B = id - \tau_0$, we obtain successively:

$$\partial_i \chi^k_{\mu,j} - \partial_j \chi^k_{\mu,i} + terms(\chi_q) - \left(A^r_i \chi^k_{\mu+1_r,j} - A^r_j \chi^k_{\mu+1_r,i}\right) = 0$$

$$B^i_r B^j_s \left(\partial_i \chi^k_{\mu,j} - \partial_j \chi^k_{\mu,i}\right) + terms(\chi_q) - \left(\tau^k_{\mu+1_r,s} - \tau^k_{\mu+1_s,r}\right) = 0$$

We obtain therefore $D\tau_{q+1} + terms(\tau_q) = 0$ and $\bar{D}': T^* \otimes J_q(T) \to \wedge^2 T^* \otimes J_{q-1}(T)$ restricts to $\bar{D}_2 : C_1(T) \to C_2(T)$.

In the case of Lie groups of transformations, the symbol of the involutive system R_q *must* be $g_q = 0$ providing an isomorphism $\mathcal{R}_{q+1} \simeq \mathcal{R}_q \Rightarrow R_{q+1} \simeq R_q$ and we have therefore $C_r = \wedge^r T^* \otimes R_q$ for $r = 1, \cdots, n$ like in the linear Spencer sequence.

□

REMARK 3.15: In the case of the (local) action of a Lie group G on X, we may consider the graph of this action, that is the morphism $X \times G \to X \times X : (x,a) \to (x, y = f(x,a))$. If q is large enough, then there is an isomorphism $X \times G \to \mathcal{R}_q \subset \Pi_q : (x,a) \to j_q(f)(x,a)$ obtained by eliminating the parameters and $C_r = \wedge^r T^* \otimes R_q$. If $\{\theta_\tau\}$ with $1 \le \tau \le dim(G)$ is a basis of infinitesimal generators of this action, there is a morphism of Lie algebroids over X, namely $X \times \mathcal{G} \to R_q : \lambda^\tau(x) \to \lambda^\tau(x) j_q(\theta_\tau)$ when q is large enough and the linear Spencer sequence $R_q \to^{D_1} T^* \otimes R_q \to^{D_2} \wedge^2 T^* \otimes R_q \to^{D_3} \cdots$ is locally exact because it is locally isomoprphic to the tensor product by $\mathcal{G}$ of the Poincaré sequence $\wedge^0 T^* \to^d \wedge^1 T^* \to^d \wedge^2 T^* \to^d \cdots$ where d is the exterior derivative ([7]).

We may also consider similarly $dy = dax = daa^{-1}y$ and $dx = dbb^{-1}dx = -a^{-1}dax$, depending on the choice of the independent variable among the *source* x or the *target* y.

Surprisingly, in the case of Lie pseudogroups or Lie groupoids, the situation is quite different. We recall the way to introduce a groupoid structure on $\Pi_{q,1} \subset J_1(\Pi_q)$ from the groupoid structure on Π_q when $\Delta = det(\partial_i f^k(x)) \ne 0$, that is how to define $j_1(h_q) = j_1(g_q \circ f_q) = j_1(g_q) \circ j_1(f_q)$. We get successively with $y = f(x)$:

$$h(x) = (g \circ f)(x) = g(f(x)) \Rightarrow \frac{\partial h^r}{\partial x^i} = \frac{\partial g^r}{\partial y^k}\frac{\partial f^k}{\partial x^i} \Rightarrow h^r_i(x) = g^r_k(f(x)) f^k_i(x)$$

$$\frac{\partial h_i^r}{\partial x^j}=\frac{\partial g_k^r}{\partial y^l}f_i^k\frac{\partial f^l}{\partial x^j}+g_k^r\frac{\partial f_i^k}{\partial x^j}$$

$$\Rightarrow h_{ij}^r(x)=g_{kl}^r\left(f(x)\right)f_i^k(x)f_j^l(x)+g_k^r\left(f(x)\right)f_{ij}^k(x)$$

$$\frac{\partial h_{ij}^r}{\partial x^s}=\frac{\partial g_{ki}^r}{\partial y^u}f_i^k f_j^l\frac{\partial f^u}{\partial x^s}+g_{kl}^r\left(\frac{\partial f_i^k}{\partial x^s}f_j^l+f_i^k\frac{\partial f_j^l}{\partial x^s}\right)+\frac{\partial g_k^r}{\partial y^u}\frac{\partial f^u}{\partial x^s}f_{ij}^k+g_k^r\frac{\partial f_{ij}^k}{\partial x^s}$$

$$\Rightarrow h_{ijs}^r=g_{klu}^r f_i^k f_j^l f_s^u+g_{kl}^r\left(f_{is}^k f_j^l+f_i^k f_{js}^l\right)+g_{ku}^r f_s^u f_{ij}^k+g_k^r f_{ijs}^k$$

and so on with more and more involved formulas.

Now, if we want to obtain objects over the *source x* according to the non-linear Spencer sequence, we have only two possibilities in actual practice, namely:

$$\chi_q=f_{q+1}^{-1}\circ j_1\left(f_q\right)-id_{q+1}\in T^*\otimes J_q(T)$$

$$\leftrightarrow\overline{\chi}_q=j_1\left(f_q\right)^{-1}\circ f_{q+1}-id_{q+1}\in T^*\otimes J_q(T)$$

As we have already considered the first, we have now only to study the second. In $J_1\left(\Pi_q\right)$, we have:

$$\chi_q+id_{q+1}=\left(A_r^k,\chi_{i,r}^k,\chi_{ij,r}^k,\cdots\right)\text{ and }\overline{\chi}_q+id_{q+1}=\left(\overline{A}_r^k,\overline{\chi}_{i,r}^k,\overline{\chi}_{ij,r}^k,\cdots\right)$$

over $(x,x,\delta,0,\cdots)$

LEMMA 3.16: $\overline{\chi}_q$ is a quasi-linear rational function of χ_q, $\forall q\geq 0$. With more details, when $q=0$, we have $\overline{\chi}_0=\overline{A}-id$ and $\chi_0=A-id$ with $\overline{A}=A^{-1}=B$ and when $q\geq 1$, we have $\overline{\chi}_q\circ A=-\chi_q$, that is to say $\overline{\chi}_q=-\tau_q$.

Proof: In the groupoid framework, we have:

$$\left(\overline{\chi}_q+id_{q+1}\right)\circ\left(\chi_q+id_{q+1}\right)=id_{q+1}\in J_1\left(\Pi_q\right)$$

Doing the substitutions:

$$\frac{\partial g^r}{\partial y^k}\to\overline{A}_k^r,\ \frac{\partial g_k^r}{\partial y^l}\to\overline{\chi}_{k,l}^r,\ \frac{\partial g_{kl}^r}{\partial y^u}\to\overline{\chi}_{kl,u}^r$$

$$\frac{\partial f^k}{\partial x^i}\to A_i^k,\ \frac{\partial f_i^k}{\partial x^j}\to\chi_{i,j}^k,\ \frac{\partial f_{ij}^k}{\partial x^s}\to\chi_{ij,s}^k$$

while using the fact that $f_i^k=\delta_i^k, f_{ij}^k=0,\cdots$ and $g_k^r=\delta_k^r, g_{kl}^r=0,\cdots$, we obtain at once:

$$\overline{A}_k^r A_i^k=\delta_i^r,\overline{\chi}_{k,l}^r A_j^l+\chi_{i,j}^k=0,\overline{\chi}_{ij,u}^r A_s^u+\chi_{ij,s}^r=0,\cdots$$

Proceeding by induction, we finally obtain:

$$\overline{\chi}_{\mu,r}^k A_s^r+\chi_{\mu,i}^k=0$$

that is to say $\overline{\chi}_{\mu,i}^k+\tau_{\mu,i}^k=0$ because $\Delta\neq 0\Rightarrow det(A)\neq 0$, thus $\overline{\chi}_q\circ A=-\chi_q$ or, equivalently, $\overline{\chi}_q=-\tau_q$.

□

REMARK 3.17: The passage from χ_q to τ_q is *exactly* the one done by E. and F. Cosserat in ([10], p 190), even though it is based on a *subtle misunderstanding* that we shall correct later on.

REMARK 3.18: According to the previous results, the "*field*" must be a sec-

tion of the natural bundle $\mathcal{F}$ of geometric objects if we use the nonlinear Janet sequence or a section of the first Spencer bundle C_1 if we use the nonlinear Spencer sequence. The aim of this paper is to prove that the second choice is by far more convenient for mathematical physics.

4. Variational Calculus

It remains to graft a variational procedure adapted to the previous results. Contrary to what happens in analytical mechanics or elasticity for example, *the main idea is to vary sections but not points*. Hence, we may introduce the variation $\delta f^k(x)=\eta^k(f(x))$ and set $\eta^k(f(x))=\xi^i\partial_i f^k(x)(x)$ along the "*vertical machinery*" but *notations like* $\delta x^i=\xi^i$ *or* $\delta y^k=\eta^k$ *have no meaning at all*.

As a major result first discovered in specific cases by the brothers Cosserat in 1909 and by Weyl in 1916, we shall prove and apply the following key result:

THE PROCEDURE ONLY DEPENDS ON THE LINEAR SPENCER OPERATOR AND ITS FORMAL ADJOINT.

In order to prove this result, if $f_{q+1},g_{q+1},h_{q+1}\in\Pi_{q+1}$ can be composed in such a way that $g'_{q+1}=g_{q+1}\circ f_{q+1}=f_{q+1}\circ h_{q+1}$, we get:

$$\begin{aligned}\bar{D}g'_{q+1}&=f_{q+1}^{-1}\circ g_{q+1}^{-1}\circ j_1(g_q)\circ j_1(f_q)-id_{q+1}=f_{q+1}^{-1}\circ\bar{D}g_{q+1}\circ j_1(f_q)+\bar{D}f_{q+1}\\&=h_{q+1}^{-1}\circ f_{q+1}^{-1}\circ j_1(f_q)\circ j_1(h_q)-id_{q+1}=h_{q+1}^{-1}\circ\bar{D}f_{q+1}\circ j_1(h_q)+\bar{D}h_{q+1}\end{aligned}$$

Using the local exactness of the first nonlinear Spencer sequence or ([25], p 219), we may state:

LEMMA 4.1: For any section $f_{q+1}\in\mathcal{R}_{q+1}$, the *finite gauge transformation*:

$$\chi_q\in T^*\otimes R_q\to\chi'_q=f_{q+1}^{-1}\circ\chi_q\circ j_1(f_q)+\bar{D}f_{q+1}\in T^*\otimes R_q$$

exchanges the solutions of the *field equations* $\bar{D}'\chi_q=0$.

Introducing the *formal Lie derivative* on $J_q(T)$ by the formulas:

$$L(\xi_{q+1})\eta_q=\{\xi_{q+1},\eta_{q+1}\}+i(\xi)D\eta_{q+1}=[\xi_q,\eta_q]+i(\eta)D\xi_{q+1}$$

$$\left(L(j_1(\xi_{q+1}))\chi_q\right)(\zeta)=L(\xi_{q+1})(\chi_q(\zeta))-\chi_q([\xi,\zeta])$$

LEMMA 4.2: Passing to the limit *over the source* with $h_{q+1}=id_{q+1}+t\xi_{q+1}+\cdots$ for $t\to 0$, we get an *infinitesimal gauge transformation* leading to the *infinitesimal variation*:

$$\delta\chi_q=D\xi_{q+1}+L(j_1(\xi_{q+1}))\chi_q \tag{3}$$

which *does not depend on the parametrization* of χ_q. Setting $\bar{\xi}_{q+1}=\xi_{q+1}+\chi_{q+1}(\xi)$, we get:

$$\delta\chi_q=D\bar{\xi}_{q+1}-\{\chi_{q+1},\bar{\xi}_{q+1}\} \tag{3*}$$

LEMMA 4.3: Passing to the limit *over the target* with $\chi_q=\bar{D}f_{q+1}$ and $g_{q+1}=id_{q+1}+t\eta_{q+1}+\cdots$, we get the other *infinitesimal variation* where $D\eta_{q+1}$ is *over the target*:

$$\delta\chi_q=f_{q+1}^{-1}\circ D\eta_{q+1}\circ j_1(f_q) \tag{4}$$

which *depends on the parametrization* of χ_q.

EXAMPLE 4.4: We obtain for $q=1$:

$$\delta\chi^k_{,i}=\left(\partial_i\xi^k-\xi^k_i\right)+\left(\xi^r\partial_r\chi^k_{,i}+\chi^k_{,r}\partial_i\xi^r-\chi^r_{,i}\xi^k_r\right)$$
$$=\left(\partial_i\overline{\xi}^k-\overline{\xi}^k_i\right)+\left(\chi^k_{r,i}\overline{\xi}^r-\chi^r_{,i}\overline{\xi}^k_r\right)$$
$$\delta\chi^k_{j,i}=\left(\partial_i\xi^k_j-\xi^k_{ij}\right)+\left(\xi^r\partial_r\chi^k_{j,i}+\chi^k_{j,r}\partial_i\xi^r+\chi^k_{r,i}\xi^r_j-\chi^r_{j,i}\xi^k_r-\chi^r_{,i}\xi^k_{jr}\right)$$
$$=\left(\partial_i\overline{\xi}^k_j-\overline{\xi}^k_{ij}\right)+\left(\chi^k_{rj,i}\overline{\xi}^r+\chi^k_{r,i}\overline{\xi}^r_j-\chi^r_{j,i}\overline{\xi}^k_r-\chi^r_{,i}\overline{\xi}^k_{jr}\right)$$

Introducing the inverse matrix $B=A^{-1}$, we obtain therefore equivalently:

$$\delta A^k_i=\xi^r\partial_r A^k_i+A^k_r\partial_i\xi^r-A^r_i\xi^k_r\Leftrightarrow\delta B^i_k=\xi^r\partial_r B^i_k-B^r_k\partial_r\xi^i+B^i_r\xi^r_k$$

both with:

$$\delta\chi^k_{j,i}=\left(\partial_i\xi^k_j-A^r_i\xi^k_{jr}\right)+\left(\xi^r\partial_r\chi^k_{j,i}+\chi^k_{j,r}\partial_i\xi^r+\chi^k_{r,i}\xi^r_j-\chi^r_{j,i}\xi^k_r\right)$$

For the Killing system $R_1\subset J_1(T)$ with $g_2=0$, these variations are *exactly* the ones that can be found in ([10], (50) + (49), p 124 with a printing mistake corrected on p 128) when replacing a 3 × 3 skew-symmetric matrix by the corresponding vector. *The last unavoidable Proposition is thus essential in order to bring back the nonlinear framework of finite elasticity to the linear framework of infinitesimal elasticity that only depends on the linear Spencer operator.*

For the conformal Killing system $\hat{R}_1\subset J_1(T)$ (see next section) we obtain:

$$\delta\chi^r_{r,i}=\left(\partial_i\xi^r_r-\xi^r_{ri}\right)+\left(\xi^r\partial_r\chi^s_{s,i}+\chi^s_{s,r}\partial_i\xi^r-\chi^s_{,i}\xi^r_{rs}\right)$$

but $\chi^r_{r,i}(x)dx^i$ is *far* from being a 1-form. However, $\left(\chi^k_{j,i}+\gamma^k_{js}\chi^s_{,i}\right)\in T^*\otimes T^*\otimes T$ and thus $\left(\alpha_i=\chi^r_{r,i}+\gamma^r_{rs}\chi^s_{,i}\right)\in T^*$ is a pure 1-form if we replace $\left(\chi^r_{r,i},\chi^r_{,i}\right)$ by $\left(\alpha_i,0\right)$. Hence, $\alpha(\zeta)$ is a scalar for any $\zeta\in T$ and we have $L(\xi_1)(\alpha(\zeta))-\alpha([\xi,\zeta])=\left(\alpha_r\partial_i\xi^r+\xi^r\partial_r\alpha_i\right)\zeta^i$. As we shall see in section *V.A*, we have $\left(L(\xi_2)\gamma\right)^k_{ij}=\xi^k_{ij}$ for any section $\xi_2\in J_2(T)$ and we obtain therefore successively:

$$\delta\alpha_i=\left(\partial_i\xi^r_r-\xi^r_{ri}\right)+\left(\alpha_r\partial_i\xi^r+\xi^r\partial_r\alpha_i\right)$$
$$\varphi_{ij}=\partial_i\alpha_j-\partial_j\alpha_i\Rightarrow\delta\varphi_{ij}=\left(\partial_j\xi^r_{ri}-\partial_i\xi^r_{rj}\right)+\left(\varphi_{rj}\partial_i\xi^r+\varphi_{ir}\partial_j\xi^r+\xi^r\partial_r\varphi_{ij}\right)$$

These are *exactly* the variations obtained by Weyl ([3], (76), p. 289) who was assuming implicitly $A=0$ when setting $\overline{\xi}^r_r=0\Leftrightarrow\xi^r_r=-\alpha_i\xi^i$ by introducing a connection. Accordingly, ξ^r_{ri} is the variation of the EM potential itself, that is the δA_i of engineers used in order to exhibit the Maxwell equations from a variational principle ([3], p. 26) but the introduction of the Spencer operator is new in this framework.

The explicit general formulas of the two lemma cannot be found somewhere else (The reader may compare them to the ones obtained in [19] by means of the so-called "diagonal" method that cannot be applied to the study of explicit examples). The following unusual difficult proposition generalizes well known variational techniques used in continuum mechanics and will be crucially used for applications:

PROPOSITION 4.5: The same variation is obtained whenever $\eta_q=f_{q+1}\left(\xi_q+\chi_q(\xi)\right)$ with $\chi_q=\overline{D}f_{q+1}$, a transformation only depending on

$j_1(f_q)$ and invertible if and only if $det(A)\neq 0$.

Proof: First of all, setting $\overline{\xi}_q=\xi_q+\chi_q(\xi)$, we get $\overline{\xi}=A(\xi)$ for $q=0$, a transformation which is invertible if and only if $det(A)\neq 0$. In the nonlinear framework, we have to keep in mind that there is no need to vary the object ω which is given but only the need to vary the section f_{q+1} as we already saw, using $\eta_q\in R_q(Y)$ *over the target* or $\xi_q\in R_q$ *over the source*. With $\eta_q=f_{q+1}(\xi_q)$, we obtain for example:

$$\begin{aligned}\delta f^k&=\eta^k=f_r^k\xi^r\\ \delta f_i^k&=\eta_u^k f_i^u=f_r^k\xi_i^r+f_{ri}^k\xi^r\\ \delta f_{ij}^k&=\eta_{uv}^k f_i^u f_j^v+\eta_u^k f_{ij}^u=f_r^k\xi_{ij}^r+f_{ri}^k\xi_j^r+f_{rj}^k\xi_i^r+f_{rij}^k\xi^r\end{aligned}$$

and so on. Introducing the formal derivatives d_i for $i=1,\cdots,n$, we have:

$$\delta f_\mu^k=\zeta_\mu^k(f_q,\eta_q)=d_\mu\eta^k=\eta_u^k f_\mu^u+\cdots=f_r^k\xi_\mu^r+\cdots+f_{\mu+1_r}^k\xi^r$$

We shall denote by $\#(\eta_q)=\zeta_\mu^k(y_q,\eta_q)\dfrac{\partial}{\partial y_\mu^k}\in V(\mathcal{R}_q)$ with $\zeta^k=\eta^k$ the corresponding vertical vector field, namely:

$$\begin{aligned}\#(\eta_q)=0\frac{\partial}{\partial x^i}+\eta^k(y)\frac{\partial}{\partial y^k}+(\eta_u^k(y)y_i^u)\frac{\partial}{\partial y_i^k}\\ +(\eta_{uv}^k(y)y_i^u y_j^v+\eta_u^k(y)y_{ij}^u)\frac{\partial}{\partial y_{ij}^k}+\cdots\end{aligned}$$

However, the standard prolongation of an infinitesimal change of source coordinates described by the horizontal vector field ξ, obtained by replacing all the derivatives of ξ by a section $\xi_q\in R_q$ over $\xi\in T$, is the vector field:

$$\begin{aligned}\flat(\xi_q)=\xi^i(x)\frac{\partial}{\partial x^i}+0\frac{\partial}{\partial y^k}-(y_r^k\xi_i^r(x))\frac{\partial}{\partial y_i^k}\\ -(y_r^k\xi_{ij}^r(x)+y_{rj}^k\xi_i^r(x)+y_{ri}^k\xi_j^r(x))\frac{\partial}{\partial y_{ij}^k}+\cdots\end{aligned}$$

It can be proved that $[\flat(\xi_q),\flat(\xi_q')]=\flat([\xi_q,\xi_q']),\forall\xi_q,\xi_q'\in R_q$ *over the source*, with a similar property for $\#(.)$ *over the target* ([25]). However, $\flat(\xi_q)$ *is not a vertical vector field and cannot therefore be compared to* $\#(\eta_q)$. The solution of this problem explains a strange comment made by Weyl in ([3], p 289 + (78), p 290) and which became a founding stone of classical gauge theory. Indeed, ξ_r^r is *not* a scalar because ξ_i^k is *not* a 2-tensor. However, when $A=0$, then $-\chi_q$ is a R_q-connection and $\overline{\xi}^r_r=\xi_r^r+\chi_{r,i}^r\xi^i$ is a true scalar that may be set equal to zero in order to obtain $\xi_r^r=-\chi_{r,i}^r\xi^i$, a fact explaining why the EM-potential is considered as a connection in quantum mechanics instead of using the second order jets ξ_{ri}^r of the conformal system, with a *shift by one step in the physical interpretation of the Spencer sequence* (See [4] for more historical details).

The main idea is to consider the vertical vector field $T(f_q)(\xi)-\flat(\xi_q)\in V(\mathcal{R}_q)$ whenever $y_q=f_q(x)$. Passing to the limit $t\to 0$ in the formula $g_q\circ f_q=f_q\circ h_q$, we first get $g\circ f=f\circ h\Rightarrow f(x)+t\eta(f(x))+\cdots=f(x+t\xi(x)+\cdots)$. Using the

chain rule for derivatives and substituting jets, we get successively:

$$\delta f^k(x) = \xi^r \partial_r f^k, \quad \delta f_i^k = \xi^r \partial_r f_i^k + f_r^k \xi_i^r,$$
$$\delta f_{ij}^k = \xi^r \partial_r f_{ij}^k + f_{rj}^k \xi_i^r + f_{ri}^k \xi_j^r + f_r^k \xi_{ij}^r$$

and so on, replacing $\xi^r f_{\mu+1_r}^k$ by $\xi^r \partial_r f_\mu^k$ in $\eta_q = f_{q+1}(\xi_q)$ in order to obtain:

$$\delta f_\mu^k = \eta_r^k f_\mu^r + \cdots = \xi^i \left(\partial_i f_\mu^k - f_{\mu+1_i}^k\right) + f_{\mu+1_r}^k \xi^r + \cdots + f_r^k \xi_\mu^r$$

where the right member only depends on $j_1(f_q)$ when $|\mu| = q$.

Finally, we may write the symbolic formula $f_{q+1}(\chi_q) = j_1(f_q) - f_{q+1} = Df_{q+1} \in T^* \otimes V(\mathcal{R}_q)$ in the explicit form:

$$f_r^k \chi_{\mu,i}^r + \cdots + f_{\mu+1_r}^k \chi_{,i}^r = \partial_i f_\mu^k - f_{\mu+1_i}^k$$

Substituting in the previous formula provides $\eta_q = f_{q+1}(\xi_q + \chi_q(\xi))$ and we just need to replace q by $q+1$ in order to achieve the proof.

Checking directly the proposition is not evident even when $q = 0$ as we have:

$$\left(\frac{\partial \eta^k}{\partial y^u} - \eta_u^k\right)\partial_i f^u = f_r^k \left[\left(\partial_i \overline{\xi}^r - \overline{\xi}_i^r\right) - \left(\chi_{,i}^s \overline{\xi}_s^r - \chi_{s,i}^r \overline{\xi}^s\right)\right]$$

but cannot be done by hand when $q \geq 1$.

□

For an arbitrary vector bundle E and involutive system $R_q \subseteq J_q(E)$, we may define the *r-prolongations* $\rho_r(R_q) = R_{q+r} = J_r(R_q) \cap J_{q+r}(E) \subset J_r(J_q(E))$ and their respective *symbols* $g_{q+r} = \rho_r(g_q)$ defined from $g_q \subseteq S_q T^* \otimes E$ where $S_q T^*$ is the vector bundle of q-symmetric covariant tensors. Using the Spencer δ-map, we now recall the definition of the *Spencer bundles*:

$$C_r = \wedge^r T^* \otimes R_q \Big/ \delta\left(\wedge^{r-1} T^* \otimes g_{q+1}\right)$$
$$\subseteq \wedge^r T^* \otimes J_q(E) \Big/ \delta\left(\wedge^{r-1} T^* \otimes S_{q+1}\right) T^* \otimes E = C_r(E)$$

and of the *Janet bundles*:

$$F_r = \wedge^r T^* \otimes J_q(E) \Big/ \left(\wedge^r T^* \otimes R_q + \delta \wedge^{r-1} T^* \otimes S_{q+1} T^* \otimes E\right)$$

When $D = \Phi \circ j_q$, we may obtain by induction on r the following *fundamental diagram I* relating the *second linear Spencer sequence* to the *linear Janet sequence* with epimorphisms $\Phi = \Phi_0, \cdots, \Phi_n$:

$$\begin{array}{ccccccccccccccc}
 & & & & 0 & & 0 & & 0 & & & & 0 & & \\
 & & & & \downarrow & & \downarrow & & \downarrow & & & & \downarrow & & \\
 & & & \overset{j_q}{\rightarrow} & C_0 & \overset{D_1}{\rightarrow} & C_1 & \overset{D_2}{\rightarrow} & C_2 & \overset{D_3}{\rightarrow} & \cdots & \overset{D_n}{\rightarrow} & C_n & \rightarrow & 0 \\
 & & & & \downarrow & & \downarrow & & \downarrow & & & & \downarrow & & \\
0 & \rightarrow & \Theta & \overset{j_q}{\rightarrow} & C_0(E) & \overset{D_1}{\rightarrow} & C_1(E) & \overset{D_2}{\rightarrow} & C_2(E) & \overset{D_3}{\rightarrow} & \cdots & \overset{D_n}{\rightarrow} & C_n(E) & \rightarrow & 0 \\
 & & \| & & \downarrow \Phi_0 & & \downarrow \Phi_1 & & \downarrow \Phi_2 & & & & \downarrow \Phi_n & & \\
0 \rightarrow & \Theta & \rightarrow & \overset{\mathcal{D}}{E} & \rightarrow & F_0 & \overset{\mathcal{D}_1}{\rightarrow} & F_1 & \overset{\mathcal{D}_2}{\rightarrow} & F_2 & \overset{\mathcal{D}_3}{\rightarrow} & \cdots \overset{\mathcal{D}_n}{\rightarrow} & F_n & \rightarrow & 0 \\
 & & & & & \downarrow & & \downarrow & & \downarrow & & & \downarrow & & \\
 & & & & & 0 & & 0 & & 0 & & & 0 & &
\end{array}$$

Chasing in the above diagram, the Spencer sequence is locally exact at C_1 if and only if the Janet sequence is locally exact at F_0 because the central sequence is locally exact (See [7] [13] [25] for more details). In the present situation, we shall always have $E=T$. *The situation is much more complicate in the nonlinear framework and we provide details for a later use*.

Let $\bar{\omega}$ be a section of $\mathcal{F}$ satisfying the same CC as ω, namely $I\left(j_1(\omega)\right)=0$. As $\mathcal{F}$ is a quotient of Π_q, we may find a section $f_q\in\Pi_q$ such that:

$$\Phi\circ f_q\equiv f_q^{-1}(\omega)=\bar{\omega}$$
$$\Rightarrow \rho_1(\Phi)\circ j_1(f_q)\equiv j_1\left(f_q^{-1}\right)\left(j_1(\omega)\right)=j_1\left(f_q^{-1}(\omega)\right)=j_1(\bar{\omega})$$

Similarly, as $\mathcal{F}$ is a natural bundle of order q, then $J_1(\mathcal{F})$ is a natural bundle of order $q+1$ and we can find a section $f_{q+1}\in\Pi_{q+1}$ such that:

$$\rho_1(\Phi)\circ f_{q+1}\equiv f_{q+1}^{-1}\left(j_1(\omega)\right)=j_1(\bar{\omega})$$

and we are facing two possible but quite different situations:

- Eliminating $\bar{\omega}$, we obtain:

$$j_1\left(f_q^{-1}\right)\left(j_1(\omega)\right)=f_{q+1}^{-1}\left(j_1(\omega)\right)$$
$$\Rightarrow\left(f_{q+1}\circ j_1\left(f_q^{-1}\right)\right)^{-1}\left(j_1(\omega)\right)-j_1(\omega)=L(\sigma_q)\omega=0$$

and thus $\sigma_q=\bar{D}f_{q+1}^{-1}\in T^*\otimes R_q=-f_{q+1}\circ\chi_q\circ j_1(f)^{-1}$ *over the target* if we set $\chi_q=\bar{D}f_{q+1}=f_{q+1}^{-1}\circ j_1(f_q)-id_{q+1}$ *over the source*, even if f_{q+1} may not be a section of $\mathcal{R}_{q+1}$. As σ_q is killed by $\bar{D}'$, we have related cocycles at $\mathcal{F}$ in the Janet sequence *over the source* with cocycles at $T^*\otimes R_q$ or C_1 *over the target*.

- Eliminating ω, we obtain successively:

$$\left(f_{q+1}^{-1}\circ j_1(f_q)\right)\left(j_1(\bar{\omega})\right)-j_1(\bar{\omega})$$
$$=-\left(f_{q+1}^{-1}\circ j_1(f_q)\right)\left[\left(f_{q+1}^{-1}\circ j_1(f_q)\right)^{-1}\left(j_1(\bar{\omega})\right)-j_1(\bar{\omega})\right]$$
$$=-\left(f_{q+1}^{-1}\circ j_1(f_q)\right)L(\chi_q)\bar{\omega}$$

where we have *over the source*:

$$L(\chi_q)\bar{\omega}=\left(\bar{\Omega}_i^\tau\equiv-L_k^{\tau\mu}\left(\bar{\omega}(x)\right)\chi_{\mu,i}^k+\chi_{,i}^r\partial_r\bar{\omega}^\tau(x)\right)\in T^*\otimes F_0$$

However, we know that F_0 is associated with R_q and is thus not affected by $f_{q+1}^{-1}\circ j_1(f_q)$ which projects onto $f_q^{-1}\circ f_q=id_q$. Hence, only T^* is affected by $f_1^{-1}\circ j_1(f)=A$ in a covariant way and we obtain therefore *over the source*:

$$\left(f_{q+1}^{-1}\circ j_1(f_q)\right)\left(j_1(\bar{\omega})\right)-j_1(\bar{\omega})=-BL(\chi_q)\bar{\omega}=-L(\tau_q)\bar{\omega}=0$$

where $B=A^{-1}$. It follows that $\chi_q\in T^*\otimes R_q(\bar{\omega})$ with $\bar{D}'\chi_q=0$ in the first non-linear Spencer sequence for $R_q(\bar{\omega})\subset J_q(T)$.

We invite the reader to follow all the formulas involved in these technical results on the next examples. Of course, whenever $\mathcal{R}_q$ is formally integrable and $f_{q+1}\in\mathcal{R}_{q+1}$ is a lift of $f_q\in\mathcal{R}_q$, then we have $\bar{\omega}=\omega$ and $\xi_q\in T^*\otimes R_q$ be-

cause $R_q(\omega) = R_q$.

EXAMPLE 4.6: In the case of Riemannian structures, we have $\mathcal{F} \in S_2T^*$ because we deal with a non-degenerate metric $\omega = (\omega_{ij}) \in S_2T^*$ with $det(\omega) \neq 0$ and may introduce $\omega^{-1} = (\omega^{ij}) \in S_2T$. We have by definition $\omega_{kl}(f(x))\, f_i^k(x)\, f_j^l(x) = \overline{\omega}_{ij}(x)$ that we shall simply write $\omega_{kl}(f)\, f_i^k f_j^l = \overline{\omega}_{ij}(x)$ and obtain therefore:

$$\omega_{kl}(f)\, f_j^l \partial_r f_i^k + \omega_{kl}(f)\, f_i^k \partial_r f_j^l + \frac{\partial \omega_{kl}}{\partial y^u}(f)\, f_i^k f_j^l \partial_r f^u - \partial_r \overline{\omega}_{ij}(x) = 0$$

Our purpose is now to compute the expression:

$$\omega_{kl}(f)\, f_j^l f_{ir}^k + \omega_{kl}(f)\, f_i^k f_{jr}^l + \frac{\partial \omega_{kl}}{\partial y^u}(f)\, f_i^k f_j^l f_r^u - \partial_r \overline{\omega}_{ij}(x) \neq 0$$

In order to eliminate the derivatives of ω over te target we may multiply the first equation by B and substract from the second while using the fact that $\omega_{kl}(f) = \overline{\omega}_{ij}(x) g_k^i g_l^j$ with $\chi_0 = A - id_T \Rightarrow \tau_0 = B\chi_0 = id_T - B$ in order to get:

$$-\left(\overline{\omega}_{sj}\tau_{i,r}^s + \overline{\omega}_{is}\tau_{j,r}^s + \tau_{,r}^s \partial_s \overline{\omega}_{ij}\right) = -\left(L(\tau_1)\overline{\omega}\right)_{ij,r}$$

These results can be extended at once to any tensorial geometric object but the conformal case needs more work and we let the reader treat it as an exercise. He will discover that the standard elimination of a conformal factor is not the best way to use in order to understand the conformal structure which has to do with a tensor density and no longer with a tensor.

In the non-linear case, the non-linear CC of the system $\mathcal{R}_q$ defined by $\Phi(y_q) = \overline{\omega}(x)$ only depend on the differential invariants and are exactly the ones satisfied by ω in the sense that they have the same Vessiot structure constants whenever $\mathcal{R}_q$ is formally integrable, in particular involutive as shown in Example 2.7. Accordingly, we can always find f_{q+1} over f_q. In the linear case, the procedure is similar but slightly simpler. Indeed, if $\mathcal{D}: T \to F_0$ is an involutive Lie operator, we may consider only the initial part of the *fundamental diagram I*:

$$\begin{array}{ccccccccccc}
 & & & & 0 & & 0 & & 0 \\
 & & & & \downarrow & & \downarrow & & \downarrow \\
0 & \to & \Theta & \overset{j_q}{\to} & C_0 & \overset{D_1}{\to} & C_1 & \overset{D_2}{\to} & C_2 \\
 & & & & \downarrow & & \downarrow & & \downarrow \\
0 & \to & T & \overset{j_q}{\to} & C_0(T) & \overset{D_1}{\to} & C_1(T) & \overset{D_2}{\to} & C_2(T) \\
 & & \| & & \downarrow \Phi_0 & & \downarrow \Phi_1 & & \\
0 \to & \Theta & \to & T & \overset{\mathcal{D}}{\to} & F_0 & \overset{\mathcal{D}_1}{\to} & F_1 & \\
 & & & & \downarrow & & \downarrow & & \\
 & & & & 0 & & 0 & &
\end{array}$$

$$\begin{array}{ccccccccc}
 & & & & 0 & & 0 & & \\
 & & & & \downarrow & & \downarrow & & \\
 & & & 0 & \to & g_{q+1} & \xrightarrow{-\delta} & \delta(g_{q+1}) & \to & 0 \\
 & & & & \downarrow & & \downarrow & & & \\
0 & \to & \Theta & \xrightarrow{j_{q+1}} & R_{q+1} & & \xrightarrow{D} & T^*\otimes R_q & & \\
 & & \| & & \downarrow & & \downarrow & & & \\
0 & \to & \Theta & \xrightarrow{j_q} & R_q & & \xrightarrow{D_1} & C_1 & & \\
 & & & & \downarrow & & \downarrow & & & \\
 & & & & 0 & & 0 & & &
\end{array}$$

and study the linear inhomogeneous involutive system $\mathcal{D}\xi=\Omega$ with $\Omega\in F_0$ and $\mathcal{D}_1\Omega=0$. If we pick up any lift $\xi_q\in C_0(T)=J_q(T)$ of Ω and chase, we notice that $X_1=D_1\xi_q\in C_1\subset C_1(T)$ is such that $D_2X_1=0$.

EXAMPLE 4.7: In the Example 2.7, using the involutive system $R_1'=R_1^{(1)}\subset R_1\subset J_1(T)$, we have $m=n=2,q=1$ and the fundamental diagram I:

$$\begin{array}{ccccccccccccc}
 & & & & & & 0 & & 0 & & 0 & & \\
 & & & & & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & \to & \Theta & \xrightarrow{j_1} & 3 & \xrightarrow{D_1} & 5 & \xrightarrow{D_2} & 2 & \to & 0 \\
 & & & & & & \downarrow & & \downarrow & & \| & & \\
 & & 0 & \to & 2 & \xrightarrow{j_1} & 6 & \xrightarrow{D_1} & 6 & \xrightarrow{D_2} & 2 & \to & 0 \\
 & & & & \| & & \downarrow\Phi_0 & & \downarrow\Phi_1 & & \downarrow & & \\
0 & \to & \Theta & \to & 2 & \xrightarrow{\mathcal{D}} & 3 & \xrightarrow{\mathcal{D}_1} & 1 & \to & 0 & & \\
 & & & & & & \downarrow & & \downarrow & & & & \\
 & & & & & & 0 & & 0 & & & &
\end{array}$$

with fiber dimensions:

$$\begin{array}{ccccccccc}
 & & & & 0 & & 0 & & \\
 & & & & \downarrow & & \downarrow & & \\
 & & & 0 & \to & 1 & \xrightarrow{-\delta} & 1 & \to & 0 \\
 & & & & \downarrow & & \downarrow & & & \\
0 & \to & \Theta & \xrightarrow{j_2} & 4 & & \xrightarrow{D} & 6 & & \\
 & & \| & & \downarrow & & \downarrow & & & \\
0 & \to & \Theta & \xrightarrow{j_1} & 3 & & \xrightarrow{D_1} & 5 & & \\
 & & & & \downarrow & & \downarrow & & & \\
 & & & & 0 & & 0 & & &
\end{array}$$

It is important to point out the importance of formal integrability and involution in this case. For this, let us start with a 1-form $\alpha=(\alpha_1,\alpha_2)$, denote its variation by $A=(A_1,A_2)$ and consider only the linear inhomogeneous system

$\mathcal{D}\xi = \mathcal{L}(\xi)\alpha = A$ with no CC for A. If the ground differential field is $K = \mathbb{Q}(x^1, x^2)$ with commuting derivations (d_1, d_2), let us choose $\alpha = x^2 dx^1 = (x^2, 0)$, $A = (x^2, x^1)$. As a lift $\xi_1 \in J_1(T)$ of A, we let the reader check that we may choose in K:

$$\xi^1 = 0, \xi^2 = 0, \xi^1_1 = 1, \xi^1_2 = \frac{x^1}{x^2}, \xi^2_1 = 0, \xi^2_2 = 0$$

Using one prolongation, we have:

$$d_1 A_1 \equiv x^2 \xi^1_{11} + \xi^2_1 = 0, d_2 A_1 \equiv x^2 \xi^1_{12} + \xi^1_1 + \xi^2_2 = 1,$$
$$d_1 A_2 \equiv x^2 \xi^1_{12} = 1, d_2 A_2 \equiv x^2 \xi^1_{22} + \xi^1_2 = 0$$

If $\beta = -d\alpha = dx^1 \wedge dx^2$, we may denote its variation by B and get at once $B = d_2 A_1 - d_1 A_2 \equiv \xi^1_1 + \xi^2_2 = 0$. Such a result is contradicting our initial choice $1 + 0 = 1$ and we cannot therefore find a lift ξ_2 of $j_1(A)$. Hence, we have to introduce the new geometric object $\omega = (\alpha, \beta)$ with $\Omega = (A, B)$ and CC $d\alpha + \beta = 0$ leading to $d_1 A_2 - d_2 A_1 + B = 0$ while using the previous diagrams. We can therefore lift $\Omega = (A, B)$ to $\xi_1 \in J_1(T)$ by choosing in K:

$$\xi^1 = 0, \xi^2 = 0, \xi^1_1 = 1, \xi^1_2 = \frac{x^1}{x^2}, \xi^2_1 = 0, \xi^2_2 = -1$$

However, we have now to add:

$$d_1 B \equiv \xi^1_{11} + \xi^2_{12} = 0, d_2 B \equiv \xi^1_{12} + \xi^2_{22} = 0$$

and lift $j_1(\Omega)$ to $\xi_2 \in J_2(T)$ over $\xi_1 \in J_1(T)$ by choosing in K:

$$\xi^1_{11} = 0, \xi^1_{12} = \frac{1}{x^2}, \xi^1_{22} = -\frac{x^1}{(x^2)^2}, \xi^2_{11} = 0, \xi^2_{12} = 0, \xi^2_{22} = -\frac{1}{x^2}$$

The image of the Spencer operator is $X_1 = D\xi_2 = j_1(\xi_1) - \xi_2$ that is to say:

$$X^1_{,1} = -1, X^1_{,2} = -\frac{x^1}{x^2}, X^2_{,1} = 0, X^2_{,2} = 1,$$

$$X^1_{1,1} = 0, X^1_{2,1} = 0, X^1_{1,2} = -\frac{1}{x^2}, X^1_{2,2} = 0, X^2_{1,1} = 0, X^2_{2,1} = 0, X^2_{1,2} = 0, X^2_{2,2} = \frac{1}{x^2}$$

and we check that $X_1 \in T^* \otimes R_1$, namely:

$$x^2 X^1_{1,i} + X^2_{,i} = 0, X^1_{2,i} = 0, X^1_{1,i} + X^2_{2,i} = 0, \forall i = 1, 2$$

a result which is not evident at first sight and has no meaning in any classical approach because we use *sections* and not *solutions*.

Now, if $f_{q+1}, f'_{q+1} \in \Pi_{q+1}$ are such that $f^{-1}_{q+1}(j_1(\omega)) = f'^{-1}_{q+1}(j_1(\omega)) = j_1(\bar{\omega})$, it follows that $(f'_{q+1} \circ f^{-1}_{q+1})(j_1(\omega)) = j_1(\omega) \Rightarrow \exists g_{q+1} \in \mathcal{R}_{q+1}$ such that $f'_{q+1} = g_{q+1} \circ f_{q+1}$ and the new $\sigma'_q = \bar{D} f'^{-1}_{q+1}$ differs from the initial $\sigma_q = \bar{D} f^{-1}_{q+1}$ by a gauge transformation.

Conversely, let $f_{q+1}, f'_{q+1} \in \Pi_{q+1}$ be such that $\sigma_q = \bar{D} f^{-1}_{q+1} = \bar{D} f'^{-1}_{q+1} = \sigma'_q$. It follows that $\bar{D}(f^{-1}_{q+1} \circ f'_{q+1}) = 0$ and one can find $g \in aut(X)$ such that $f'_{q+1} = f_{q+1} \circ j_{q+1}(g)$ providing $\bar{\omega}' = f'^{-1}_q(\omega) = (f_q \circ j_q(g))^{-1}(\omega) = j_q(g)^{-1}(f^{-1}_q(\omega)) = j_q(g)^{-1}(\bar{\omega})$.

PROPOSITION 4.8: Natural transformations of $\mathcal{F}$ *over the source* in the nonlinear Janet sequence correspond to gauge transformations of $T^*\otimes R_q$ or C_1 *over the target* in the nonlinear Spencer sequence. Similarly, the Lie derivative $\mathcal{D}\xi=\mathcal{L}(\xi)\omega\in F_0$ in the linear Janet sequence corresponds to the Spencer operator $D\xi_{q+1}\in T^*\otimes R_q$ or $D_1\xi_q\in C_1$ in the linear Spencer sequence.

With a slight abuse of language $\delta f=\eta\circ f\Leftrightarrow \delta f\circ f^{-1}=\eta\Leftrightarrow f^{-1}\circ\delta f=\xi$ when $\eta=T(f)(\xi)$ and we get $j_q(f)^{-1}(\omega)=\overline{\omega}\Rightarrow j_q(f+\delta f)^{-1}(\omega)=\overline{\omega}+\delta\overline{\omega}$ that is $j_q\left(f^{-1}\circ(f+\delta f)\right)^{-1}(\overline{\omega})=\overline{\omega}+\delta\overline{\omega}\Rightarrow\delta\overline{\omega}=\mathcal{L}(\xi)\overline{\omega}$ and
$j_q\left((f+\delta f)\circ f^{-1}\circ f\right)^{-1}(\omega)=j_q(f)^{-1}\left(j_q\left((f+\delta f)\circ f^{-1}\right)^{-1}(\omega)\right)$
$\Rightarrow\delta\overline{\omega}=j_q(f)^{-1}\left(\mathcal{L}(\eta)\omega\right)$.

Passing to the infinitesimal point of view, we obtain the following generalization of Remark 3.12 which is important for applications.

COROLLARY 4.9:
$\overline{\Omega}=\delta\overline{\omega}=L(\xi_q)\overline{\omega}=f_q^{-1}\left(L(\eta_q)\omega\right)\Rightarrow\delta\overline{\omega}=\mathcal{L}(\xi)\overline{\omega}=j_q(f)^{-1}\left(\mathcal{L}(\eta)\omega\right)$.

Recapitulating the results so far obtained concerning the links existing between the source and the target points of view, we may set in a symbolic way:

$$\delta f_q \overset{(f_q)}{\leftrightarrow}\eta_q \overset{(f_{q+1})}{\leftrightarrow}\overline{\xi}_q \overset{(\chi_q)}{\leftrightarrow}\xi_q$$

In order to help the reader maturing the corresponding nontrivial formulas, we compute explicitly the case $n=1,q=1,2$ and let the case n arbitrary left to the reader as a difficult exercise that cannot be achieved by hand when $q\geq 3$:

EXAMPLE 4.10: Using the previous formulas, we have $\delta f(x)=\eta(f(x))$, $\delta f_x(x)=\eta_y(f(x))f_x(x)$ and:

$$\eta_1=f_2(\overline{\xi}_1)\Rightarrow\left(\eta(f(x))=f_x(x)\overline{\xi}(x),\right.$$
$$\left.\eta_y(f(x))f_x(x)=f_x(x)\overline{\xi}_x(x)+f_{xx}(x)\overline{\xi}(x)\right)$$

The delicate point is that we have successively:

$$\chi_{,x}=\frac{\partial_x f}{f_x}-1=A-1,\quad \chi_{x,x}=\frac{1}{f_x}\left(\partial_x f_x-\frac{\partial_x f}{f_x}f_{xx}\right)$$

$$\overline{\xi}=\xi+\chi_{,x}(\xi)=\frac{\partial_x f}{f_x}\xi=A\xi,\quad \overline{\xi}_x=\xi_x+\chi_{x,x}\xi$$

$$\Rightarrow\quad \eta=\partial_x f\,\xi,\quad \eta_y=\xi_x+\frac{\partial_x f_x}{f_x}\xi$$

$$f_x\eta_{yy}=\xi_{xx}+\frac{f_{xx}}{f_x}\xi_x+\left(\frac{\partial_x f_{xx}}{f_x}-\frac{f_{xx}}{(f_x)^2}\partial_x f_x\right)\xi$$

When $z=g(y),y=f(x)\Rightarrow z=(g\circ f)(x)=h(x)$, we obtain therefore the simple groupoid composition formulas $h_x(x)=g_y(f(x))f_x(x)$ and thus:

$$\zeta=\partial_x h\xi=\partial_y g\eta=\partial_y g\partial_x f\xi,$$

$$\zeta_z=\eta_y+\frac{\partial_y g_y}{g_y}\eta=\xi_x+\left(\frac{\partial_y g_y}{g_y}\partial_x f+\frac{\partial_x f_x}{f_x}\right)\xi=\xi_x+\frac{\partial_x h_x}{h_x}\xi$$

Using indices in arbitrary dimension, we get successively:

$$\eta^k = f_r^k \overline{\xi}^r, \eta_u^k f_i^u = f_r^k \overline{\xi}_i^r + f_{ri}^k \overline{\xi}^r \eta^k$$

$$\Rightarrow \eta^k = \xi^r \partial_r f^k, \eta_u^k f_i^u = f_s^k \left(\xi_i^s + g_u^s \left(\partial_r f_i^u - A_r^t f_{ti}^u \right) \xi^r \right) + f_{ti}^k A_r^t \xi^r$$

$$\eta_u^k = g_u^i f_s^k \xi_i^s + \xi^r g_u^i \partial_r f_i^k \Rightarrow \eta_k^k = \xi_r^r + \xi^r g_u^i \partial_r f_i^u$$

As a very useful application, we obtain successively:

$$\Delta(x) = det\left(\partial_i f^k(x)\right) \Rightarrow \delta\Delta = \Delta \frac{\partial \eta^k}{\partial y^k} = \Delta \partial_r \xi^r + \xi^r \partial_r \Delta = \partial_r \left(\xi^r \Delta \right)$$

$$\delta det(A) = det(A) \left(\frac{\partial \eta^k}{\partial y^k} - \eta_k^k \right) = det(A) \left(\partial_r \xi^r - \xi_r^r \right) + \xi^r \partial_r det(A)$$

where sections of jet bundles are used in an essential way, and the important lemma:

LEMMA 4.11: When the transformation $y = f(x)$ is invertible with inverse $x = g(y)$, we have the *fundamental identity* over the source or over the target:

$$\frac{\partial}{\partial x^i} \left(\Delta(x) \frac{\partial g^i}{\partial y^k} \left(f(x) \right) \right) \equiv 0, \quad \forall x \in X$$

$$\Leftrightarrow \frac{\partial}{\partial y^k} \left(\frac{1}{\Delta(g(y))} \frac{\partial f^k}{\partial x^i} \left(g(y) \right) \right) \equiv 0, \quad \forall y \in Y$$

EXAMPLE 4.12: We proceed the same way for studying the links existing between $\chi_q = \overline{D} f_{q+1}$ *over the source*, $\chi_q^{-1} = \sigma_q = \overline{D} f_{q+1}^{-1}$ *over the target* and the nonlinear Spencer operator. First of all, we notice that:

$$\sigma_q = f_{q+1} \circ j_1 \left(f_q^{-1} \right) - id_{q+1} = f_{q+1} \circ \left(id_{q+1} - f_{q+1}^{-1} j_1 \left(f_q \right) \right) \circ j_1 \left(f_q \right)^{-1}$$

$$= - f_{q+1} \circ \chi_q \circ j_1 \left(f_q \right)^{-1}$$

and the components of σ_q thus factor through linear combinations of the components of χ_q. After tedious computations, we get successively when $m = n = 1$:

$$\chi_{,x} = \frac{\partial_x f}{f_x} - 1 = A - 1 = \frac{1}{f_x} \left(\partial_x f - f_x \right)$$

$$\chi_{x,x} = \frac{1}{f_x} \left(\partial_x f_x - \frac{\partial_x f}{f_x} f_{xx} \right) = \frac{1}{f_x} \left(\partial_x f_x - f_{xx} \right) - \frac{f_{xx}}{(f_x)^2} \left(\partial_x f - f_x \right)$$

$$\chi_{xx,x} = \frac{1}{f_x} \left(\partial_x f_{xx} - \frac{\partial_x f}{f_x} f_{xxx} \right) - 2 \frac{f_{xx}}{(f_x)^2} \left(\partial_x f_x - \frac{\partial_x f}{f_x} f_{xx} \right)$$

$$= \frac{1}{f_x} \left(\partial_x f_{xx} - f_{xxx} \right) - 2 \frac{f_{xx}}{(f_x)^2} \left(\partial_x f_x - f_{xx} \right) + \left(2 \frac{(f_{xx})^2}{(f_x)^3} - \frac{f_{xxx}}{(f_x)^2} \right) \left(\partial_x f - f_x \right)$$

These formulas agree with the successive constructive/inductive identities:

$$\begin{cases} \chi_{,x} f_x = \partial_x f - f_x \\ \chi_{x,x} f_x + \chi_{,x} f_{xx} = \partial_x f_x - f_{xx} \\ \chi_{xx,x} f_x + 2 \chi_{x,x} f_{xx} + \chi_{,x} f_{xxx} = \partial_x f_{xx} - f_{xxx} \end{cases}$$

showing that χ_q is linearly depending on Df_{q+1} and we finally get:

$$\begin{cases} \sigma_{,y} = -(\partial_x f - f_x)\dfrac{1}{\partial_x f} = \dfrac{f_x}{\partial_x f} - 1 = -f_x \chi_{,x} \dfrac{1}{\partial_x f} \\ \sigma_{y,y} = -\dfrac{1}{f_x}(\partial_x f_x - f_{xx})\dfrac{1}{\partial_x f} = -\left(\chi_{x,x} + \dfrac{f_{xx}}{f_x}\chi_{,x}\right)\dfrac{1}{\partial_x f} \\ \sigma_{yy,y} = -\left(\dfrac{1}{(f_x)^2}(\partial_x f_{xx} - f_{xxx}) - \dfrac{f_{xx}}{(f_x)^3}(\partial_x f_x - f_{xx})\right)\dfrac{1}{\partial_x f} \\ \qquad = -\left(\dfrac{1}{f_x}\chi_{xx,x} + \dfrac{f_{xx}}{(f_x)^2}\chi_{x,x} + \left(\dfrac{f_{xxx}}{(f_x)^2} - \dfrac{(f_{xx})^2}{(f_x)^3}\right)\chi_{,x}\right)\dfrac{1}{\partial_x f} \end{cases}$$

while using successively the relations $g_y f_x = 1$, $\partial_y g \partial_x f = 1$, $g_{yy}(f_x)^2 + g_y f_{xx} = 0$ and so on when $x = g(y)$ is the inverse of $y = f(x)$, in a coherent way with the action of f_3 on $J_2(T)$ which is described as follows:

$$\begin{cases} \eta = f_x \xi \\ \eta_y = \xi_x + \dfrac{f_{xx}}{f_x}\xi \\ \eta_{yy} = \dfrac{1}{f_x}\xi_{xx} + \dfrac{f_{xx}}{(f_x)^2}\xi_x + \left(\dfrac{f_{xxx}}{(f_x)^2} - \dfrac{(f_{xx})^2}{(f_x)^3}\right)\xi \end{cases}$$

Restricting these formulas to the affine case defined by $y_{xx} = 0 \Rightarrow \xi_{xx} = 0$, we have thus $y_{xx} = 0, y_{xxx} = 0 \Rightarrow f_{xx} = 0, f_{xxx} = 0$. It follows that $\eta = f_x \xi, \eta_y = \xi_x, \eta_{yy} = \dfrac{1}{f_x}\xi_{xx} = 0$ on one side and $\chi_{xx,x} = 0 \Leftrightarrow \sigma_{yy,y} = 0$ in a coherent way. It is finally important to notice that these results are not evident, even when $m = n = 1$, as soon as second order jets are involved.

We shall use *all* the preceding formulas in the next example showing that, contrary to what happens in elasticity theory where the source is usually identified with the Lagrange variables, in both the Vessiot/Janet and the Cartan/Spencer approaches, *the source must be identified with the Euler variables without any possible doubt*.

EXAMPLE 4.13: With $n = 1, q = 1, \mathcal{F} = T^*$ and the finite OD Lie equation $\omega(y) y_x = \omega(x)$ with $\omega \in T^*$ and corresponding Lie operator $\mathcal{D}\xi \equiv \mathcal{L}(\xi)\omega = \omega(x)\partial_x \xi + \xi \partial_x \omega(x)$ *over the source*, we have:

$$\omega(f(x))f_x(x) = \bar{\omega}(x), \quad \omega(f(x))f_{xx}(x) + \partial_y \omega(f(x))f_x^2(x) = \partial_x \bar{\omega}(x)$$

Differentiating once the first equation and substracting the second, we obtain therefore:

$$\omega \sigma_{y,y} + \sigma_{,y}\partial_y \omega \equiv -\omega(1/f_x)(\partial_x f_x - f_{xx})(1/\partial_x f) + ((f_x/\partial_x f) - 1)\partial_y \omega = 0$$

whenever $y = f(x)$. Finally, setting $\omega(f(x))\partial_x f(x) = \bar{\omega}(x)$, we get *over the target*:

$$\delta\bar{\omega} = \omega(f(x))\frac{\partial \eta}{\partial y}\partial_x f(x) + \partial_x f(x)\frac{\partial \omega}{\partial y}(f(x))\eta = \partial_x f \mathcal{L}(\eta)\omega$$

Differentiating $\eta = \xi\partial_x f$ in order to obtain $\dfrac{\partial\eta}{\partial y} = \partial_x\xi + \xi\left(\partial_{xx} f/\partial_x f\right)$, we get *over the source*:

$$\delta\overline{\omega} = \overline{\omega}\partial_x\xi + \xi\partial_x\overline{\omega} = \mathcal{L}(\xi)\overline{\omega}$$

We may summarize these results as follows:

$$\delta\overline{\omega} = \mathcal{L}(\xi)\overline{\omega} \overset{(j_1(f))}{\rightarrow} \delta\overline{\omega} = \partial_x f\,\mathcal{L}(\eta)\omega$$

We invite the reader to extend this result to an arbitrary dimension $n \geq 2$.

EXAMPLE 4.14: The case of an affine stucture needs more work with $n = m = 1, q = 2$. Indeed, let us consider the action of the affine Lie group of transformations $\overline{y} = ay + b$ with $a, b = cst$ acting on the target $y \in Y$ considered as a copy of the real line *X*. We obtain the prolongations up to order 2 of the 2 infinitesimal generators of the action:

$$a \rightarrow y\frac{\partial}{\partial y} + y_x\frac{\partial}{\partial y_x} + y_{xx}\frac{\partial}{\partial y_{xx}}, \quad b \rightarrow \frac{\partial}{\partial y} + 0\frac{\partial}{\partial y_x} + 0\frac{\partial}{\partial y_{xx}}$$

There cannot be any differential invariant of order 1 and the only generating one of order 2 can be $\Phi \equiv y_{xx}/y_x$. When $\overline{x} = \varphi(x)$ we get successively $y_x = y_{\overline{x}}\partial_x\varphi$, $y_{xx} = y_{\overline{xx}}\left(\partial_x\varphi\right)^2 + y_{\overline{x}}\partial_{xx}\varphi$ and Φ transforms like $u = \partial_x\varphi\,\overline{u} + \dfrac{\partial_{xx}\varphi}{\partial_x\varphi}$ a result providing the bundle of geometric objects $\mathcal{F}$ with local coordinates (x,u) and corresponding transition rules. For any section γ, we get the Vessiot general system $\mathcal{R}_2 \subset \Pi_2$ of second order finite Lie equations $\dfrac{y_{xx}}{y_x} + \gamma(y)y_x = \gamma(x)$ which is already in Lie form and relates the jet coordinates (x, y, y_x, y_{xx}) of order 2. The special section is $\gamma = 0$ and we may consider the automorphic system $\Phi \equiv \dfrac{y_{xx}}{y_x} = \overline{\gamma}(x)$ obtained by introducing *any second order section* $f_2(x) = \left(f(x), f_x(x), f_{xx}(x)\right)$, for example $f_2 = j_2(f)$ providing $\left(f(x), \partial_x f(x), \partial_{xx} f(x)\right)$. It is not at all evident, *even on such an elementary example*, to compute the variation $\overline{\Gamma} = \delta\overline{\gamma}$ induced by the previous formulas and to prove that, like any *field* quantity, it only depends on $\overline{\gamma}$ on the condition to use *only* source quantities. For this, setting $\dfrac{f_{xx}(x)}{f_x(x)} = \overline{\gamma}(x)$, varying and substituting, we obtain:

$$\overline{\Gamma} = \delta\overline{\gamma} = \frac{\delta f_{xx}}{f_x} - \frac{f_{xx}}{\left(f_x\right)^2}\delta f_x = f_x\eta_{yy} = \xi_{xx} + \overline{\gamma}\xi_x + \xi\partial_x\overline{\gamma}$$

Now, linearizing the preceding Lie equation over the identity transformation $y = x$, we get the Medolaghi equation:

$$L(\xi_2)\gamma \equiv \xi_{xx} + \gamma(x)\xi_x + \xi\partial_x\gamma(x) = 0, \forall\xi_2 \in R_2 \subset J_2(T)$$

and the striking formula $\overline{\Gamma} = \delta\overline{\gamma} = L(\xi_2)\overline{\gamma}$ over the source for an arbitrary $\xi_2 \in J_2(T)$. We finally point out the fact that, as we have just shown above and

contrary to what the brothers Cosserat had in mind, the first order operators involved in the nonlinear Spencer sequence have *strictly nothing to do* with the operators involved in the nonlinear Janet sequence whenever $q \geq 2$. For example, in the present situation, $\chi_{,x} = \frac{\partial_x f}{f_x} - 1$ has nothing to do with $\Phi \equiv \frac{f_{xx}}{f_x}$. Similarly, using the comment before example 4.7 in the linear framework, we have the first order Spencer operator $D_1 : (\xi, \xi_x) \to (\partial_x \xi - \xi_x, \partial_x \xi_x)$ on one side and the second order Lie operator $\mathcal{D} : \xi \to \partial_{xx} \xi$ on the other side.

The next delicate example proves nevertheless that target quantities may also be used.

EXAMPLE 4.15: In the last example, depending on the way we use $\overline{\gamma}(x)$ on the source or $\gamma(y)$ on the target, we may consider the two (very different) Medolaghi equations:

$$\xi_{xx} + \overline{\gamma}(x)\xi_x + \xi \partial_x \overline{\gamma}(x) = 0 \quad \leftrightarrow \quad \eta_{yy} + \gamma(y)\eta_y + \eta \partial_y \gamma(y) = 0$$

Now, starting from the single OD equation $\frac{f_{xx}}{f_x} = \overline{\gamma}(x)$ in sectional notations, we may successively *differentiate* and *prolongate* once in order to get:

$$\frac{\partial_x f_{xx}}{f_x} - \frac{f_{xx}}{(f_x)^2} \partial_x f_x = \partial_x \overline{\gamma}(x) \quad \leftrightarrow \quad \frac{f_{xxx}}{f_x} - \left(\frac{f_{xx}}{f_x} \right)^2 = \partial_x \overline{\gamma}(x)$$

Substracting the second from the first as a way to eliminate $\overline{\gamma}$, we obtain a linear relation involving only the components of the nonlinear Spencer operator in a coherent way with the theory of nonlinear systems, namely:

$$\frac{1}{f_x}(\partial_x f_{xx} - f_{xxx}) - \frac{f_{xx}}{(f_x)^2}(\partial_x f_x - f_{xx}) = 0$$

At first sight it does not seem possible to know whether we have a linear combination of the components of χ_2 or of the components of σ_2. However, if we come back to the original situation $f_q^{-1}(\omega) = \overline{\omega}$, we have eliminated $j_1(\overline{\gamma})$ over the source and we are thus only left with $j_1(\gamma)$ over the target. Hence it can only depend on σ_2 and we find indeed the striking relation:

$$-\frac{1}{f_x}\left[\frac{1}{f_x}(\partial_x f_{xx} - f_{xxx}) - \frac{f_{xx}}{(f_x)^2}(\partial_x f_x - f_{xx}) \right] \frac{1}{\partial_x f} = \sigma_{yy,y} = 0$$

provided by the simple second order Medolaghi equation $\gamma = 0 \Rightarrow \eta_{yy} = 0$ over the target. It is essential to notice that no classical technique can provide these results which are essentially depending on the nonlinear Spencer operator and are thus not known today.

EXAMPLE 4.16: The above methods can be applied to any explicit example. The reader may treat as an exercise the case of the pseudogroup of isometries of a non degenerate metric which can be found in any textbook of continuum mechanics or elasticity theory, though in a very different framework with methods only valid for tensors. With the previous notations, let $\omega \in S_2 T^*$ with $det(\omega) \neq 0$

and consider the following nonlinear system $\omega_{kl}\left(f\left(x\right)\right)\partial_i f^k\left(x\right)\partial_j f^l\left(x\right)=\overline{\omega}_{ij}\left(x\right)$ with $1\le i,j,k,l\le n$. One obtains therefore:

$$\delta\overline{\omega}_{ij}=\overline{\omega}_{rj}\partial_i\xi^r+\overline{\omega}_{ir}\partial_i\xi^u+\xi^r\partial_r\overline{\omega}_{ij}=\partial_i f^k\partial_j f^l\left(\omega_{ul}\frac{\partial\eta^u}{\partial y^k}+\omega_{ku}\frac{\partial\eta^u}{\partial y^l}+\eta^u\frac{\partial\omega_{kl}}{\partial y^u}\right)$$

and thus the same recapitulating formulas linking the source and target variations:

$$\delta\overline{\omega}=\mathcal{L}\left(\xi\right)\overline{\omega}\overset{\left(j_1(f)\right)}{\rightarrow}\delta\overline{\omega}=\partial_i f^k\partial_j f^l\left(\mathcal{L}\left(\eta\right)\omega\right)_{kl}$$

It is also difficult to compute or compare the variational formulas over the source and target in the nonlinear Spencer sequence, even when $m=n=1$ and $q=0,1$ ([29]).

EXAMPLE 4.17: Let us prove that the explicit computation of the gauge transformation is at the limit of what can be done with the hand, even when $m=n=1,q=1$. We have successively:

$$\chi_{,x}=\frac{\partial_x f}{f_x}-1,\ \chi_{x,x}=\frac{1}{f_x}\left(\partial_x f_x-\frac{\partial_x f}{f_x}f_{xx}\right)$$

$$f'\left(x\right)=g\left(f\left(x\right)\right)\Rightarrow f'_x=g_y f_x,\ f'_{xx}=g_{yy}\left(f_x\right)^2+g_y f_{xx}$$

and thus:

$$\chi'_{,x}=\frac{\partial_x f'}{f'_x}-1=\frac{\partial_y g\partial_x f}{g_y f_x}-1=\left(\chi_{,y}+1\right)\frac{\partial_x f}{f_x}-1=\frac{\partial_x f}{f_x}\chi_{,y}+\left(\frac{\partial_x f}{f_x}-1\right)$$

$$\begin{aligned}\chi'_{x,x}&=\frac{1}{f'_x}\left(\partial_x f'_x-\frac{\partial_x f'}{f'_x}f'_{xx}\right)\\&=\frac{1}{g_y f_x}\left(\partial_y g_y\left(\partial_x f\right)f_x+g_y\partial_x f_x\right)-\frac{\partial_y g\partial_x f}{g_y f_x}\left(g_{yy}\left(f_x\right)^2+g_y f_{xx}\right)\\&=\frac{1}{g_y}\partial_y g_y\partial_x f+\frac{\partial_x f_x}{f_x}-\frac{\partial_y g\partial_x f g_{yy}}{\left(g_y\right)^2}-\frac{\partial_y g\partial_x f f_{xx}}{g_y\left(f_x\right)^2}\\&=\left(\partial_x f\chi_{y,y}-\frac{\partial_x f f_{xx}}{\left(f_x\right)^2}\chi_{,y}\right)+\frac{1}{f_x}\left(\partial_x f_x-\frac{\partial_x f}{f_x}f_{xx}\right)\end{aligned}$$

Setting $f_2=id_2+t\xi_2+\cdots$ and passing to the limit when $t\to 0$, we finally obtain:

$$\begin{aligned}\delta\chi_{,x}&=\left(\partial_x\xi-\xi_x\right)+\left(\xi\partial_x\chi_{,x}+\chi_{,x}\partial_x\xi-\chi_{,x}\xi_x\right)\\\delta\chi_{x,x}&=\left(\partial_x\xi_x-\xi_{xx}\right)+\left(\xi\partial_x\chi_{x,x}+\chi_{x,x}\partial_x\xi-\chi_{,x}\xi_{xx}\right)\end{aligned}$$

If we use the standard euclidean metric $\omega=1\Rightarrow\gamma=0$, we may thus introduce the pure 1-form $\alpha=\chi_{x,x}+\gamma\chi_{,x}$. We should consider the defining formula $\chi'_1=f_2^{-1}\circ\chi_1\circ j_1\left(f_1\right)+\overline{D}f_2$ and have to introduce the additional term $f_2^{-1}\left(\gamma\right)\chi_{,x}$ which is only leading to the additional infinitesimal term $\left(L\left(\xi_2\right)\gamma\right)\chi_{,x}=\xi_{xx}\chi_{,x}$ because $\gamma=0$. We finally obtain:

$$\delta\alpha = \delta\chi_{x,x} + \xi_{xx}\chi_{,x} + \gamma\delta\chi_{,x} = \left(\partial_x\xi_x - \xi_{xx}\right) + \left(\xi\partial_x\alpha + \alpha\partial_x\xi\right)$$

and this result can be easily extended to an arbitrary dimension with the formula:

$$\alpha_i = \chi^r_{r,i} + \gamma^s_{sr}\chi^r_{,i} \Rightarrow \left(\delta\alpha\right)_i = \left(\partial_i\xi^r_r - \xi^r_{ri}\right) + \left(\xi^r\partial_r\alpha_i + \alpha_r\partial_i\xi^r\right)$$

Comparing this procedure with the one we have adopted in the previous exampes, we have:

$$\chi_{,x} = \frac{\partial_x f}{f_x} - 1 = A - 1 \Rightarrow \delta\chi_{,x} = \frac{\partial_x\delta f}{f_x} - \frac{\partial_x f}{\left(f_x\right)^2}\delta f_x = \frac{1}{f_x}\left(\frac{\partial\eta}{\partial y} - \eta_y\right)\partial_x f$$

However, taking into account the formulas $\eta = \xi\partial_x f$ and $\eta_y = \xi_x + \frac{\partial_x f_x}{f_x}\xi$, we also get:

$$\begin{aligned}\delta\chi_{,x} &= \frac{1}{f_x}\left(\partial_x\xi\partial_x f + \xi\partial_{xx} f\right) - \frac{\partial_x f}{\left(f_x\right)^2}\left(\xi_x f_x + \xi\partial_x f_x\right)\\ &= A\left(\partial_x\xi - \xi_x\right) + \xi\partial_x\chi_{,x}\\ &= \left(\partial_x\xi - \xi_x\right) + \left(\xi\partial_x\chi_{,x} + \chi_{,x}\partial_x\xi - \chi_{,x}\xi_x\right)\end{aligned}$$

Working over the target is more difficult. Indeed, we have successively (*care to the first step*):

$$\sigma_{,y} = \frac{f_x}{\partial_x f} - 1 \Rightarrow \delta\sigma_{,y} + \eta\frac{\partial\sigma_{,y}}{\partial y} = \frac{\delta f_x}{\partial_x f} - \frac{f_x}{\left(\partial_x f\right)^2}\partial_x\delta f = -\frac{f_x}{\left(\partial_x f\right)^2}\left(\frac{\partial\eta}{\partial y} - \eta_y\right)$$

$$\begin{aligned}\delta\sigma_{,y} &= -\left[\frac{f_x}{\partial_x f}\left(\frac{\partial\eta}{\partial y} - \eta_y\right) + \eta\frac{\partial\sigma_{,y}}{\partial y}\right]\\ &= -\left[\left(\frac{\partial\eta}{\partial y} - \eta_y\right) + \left(\eta\frac{\partial\sigma_{,y}}{\partial y} + \sigma_{,y}\frac{\partial\eta}{\partial y} - \sigma_{,y}\eta_y\right)\right]\end{aligned}$$

More generaly, we let the reader prove that the variation of σ_q over the target (respectively the source) is described by "*minus*" the same formula as the variation of χ_q over the source (respectively the target). In any case, the reader must not forget that the word "variation" just means that the section f_{q+1} is changed, *not* that the source is moved. Accordingly, getting in mind this example and for simplicity, we shall always prefer to work with vertical bundles over the source, closely following the purely mathematical definitions, contrary to Weyl ([3], §28, formulas (17) to (27), p 233-236). The reader must be now ready for comparing the variations of $\chi_{x,x}$ and $\sigma_{y,y}$.

In order to conclude this section, we provide without any proof two results and refer the reader to ([7]) for details.

PROPOSITION 4.18: Changing slightly the notation while setting $\sigma_{q-1} = \bar{D}'\chi_q$, we have:

$$\chi'_q = f^{-1}_{q+1}\circ\chi_q\circ j_1\left(f\right) + \bar{D}f_{q+1} \Rightarrow \sigma'_{q-1} = f^{-1}_q\circ\sigma_{q-1}\circ j_1\left(f\right)$$

where f^{-1}_q acts on $J_{q-1}\left(T\right)$ and $j_1\left(f\right)$ acts on $\wedge^2 T^*$. It follows that gauge transformations exchange the solutions of $\bar{D}'$ among themselves.

COROLLARY 4.19: Denoting by $\mathcal{C}(\ )$ the cyclic sum, we have the so-called *Bianchi identity*:

$$D\sigma_{q-1}(\xi,\eta,\zeta)+\mathcal{C}(\xi,\eta,\zeta)\{\sigma_{q-1}(\xi,\eta),\chi_{q-1}(\zeta)\}=0$$

5. Applications

Before studying in a specific way electromagnetism and gravitation, we shall come back to Example 4.10 and provide a technical result which, though looking like evident at first sight, is at the origin of a deep misunderstanding done by the brothers Cosserat and Weyl on the variational procedure used in the study of physical problems (Compare to [14]).

Setting $dx=dx^1\wedge\cdots\wedge dx^n$ for simplicity while using Lemma 4.11 and the fact that the standard Lie derivative is commuting with any diffeomorphism, we obtain at once:

$$y=f(x)\Rightarrow dy=det\left(\partial_i f^k(x)\right)dx=\Delta(x)dx$$

$$\eta=T(f)\xi\Rightarrow\mathcal{L}(\eta)dy=\mathcal{L}(\xi)\left(\Delta(x)dx\right)$$

$$\Rightarrow\delta\Delta=\Delta div_y(\eta)=\Delta div_x(\xi)+\xi^r\partial_r\Delta$$

The interest of such a presentation is to provide the right correspondence between the source/target and the Euler/Lagrange choices. Indeed, if we use the way followed by most authors up to now in continuum mechanics, we should have source = Lagrange, target = Euler, a result leading to the conservation of mass $dm=\rho dy=\rho_0 dx=dx$ when ρ_0 is the original initial mass per unit volume. We may set $\rho_0=1$ and obtain therefore $\rho(f(x))=1/\Delta(x)$, a choice leading to:

$$\delta\rho+\eta^k\frac{\partial\rho}{\partial y^k}=-\frac{1}{\Delta^2}\delta\Delta\Rightarrow\delta\rho=-\rho\frac{\partial\eta^k}{\partial y^k}-\eta^k\frac{\partial\rho}{\partial y^k}=-\rho\frac{\partial\xi^r}{\partial x^r}\Rightarrow\delta\rho=-\frac{\partial\left(\rho\eta^k\right)}{\partial y^k}$$

but the concept of "*variation*" is not mathematically well defined, though this result is coherent with the classical formulas that can be found for example in ([4] [9]) or ([3], (17) and (18) p 233, (20) to (21) p 234, (76) p 289, (78) p 290) where "*points are moved*".

On the contrary, if we adopt the *unusual* choice source = Euler, target = Lagrange, we should get $\rho(x)=\Delta(x)$, a choice leading to $\delta\rho=\delta\Delta$ and thus:

$$\delta\rho=\rho\frac{\partial\eta^k}{\partial y^k}=\rho\frac{\partial\xi^r}{\partial x^r}+\xi^r\partial_r\rho=\partial_r\left(\rho\xi^r\right)$$

which is the right choice agreeing, up to the sign, with classical formulas but with the important improvement that this result becomes a purely mathematical one, obtained from a well defined variational procedure involving only the so-called "*vertical*" machinery. This result fully explains why we had doubts about the sign involved in the variational formulas of ([4], p. 383) but without being able to correct them at that time. We may finally revisit Lemma 4.11 in

order to obtain the *fundamental identity over the source*:

$$\frac{\partial}{\partial x^i}\left(\Delta(x)\frac{\partial g^i}{\partial y^k}(f(x))\right)\equiv 0,\quad \forall x\in X$$

which becomes the conservation of mass when $n=4$ and $k=4$.

In addition, as many chases will be used through many diagrams in the sequel, we invite the reader not familiar with these technical tools to consult the books ([30] [31]) that we consider as the best references for learning about homological algebra. A more elementary approach can be found in ([32]) that has been used during many intensive courses on the applications of homological algebra to control theory. As for *differential homological algebra*, one of the most difficult tools existing in mathematics today, and its link with applications, we refer the reader to the various references provided in ([33]).

Finally, for the reader interested by a survey on more explicit applications, we particularly refer to ([2] [34] [35] [36]) for analytical mechanics and hydrodynamics, ([5] [37] [38]) for coupling phenomenas, ([36] [39] [40]) for the foundations of Gauge Theory, ([36] [41]) for the foundations of General Relativity.

A) POINCARE, WEYL AND CONFORMAL GROUPS

When constructing inductively the Janet and Spencer sequences for an involutive system $R_q\subset J_q(E)$, we have to use the following commutative and exact diagrams where we have set $F_0=J_q(E)/R_q$ and used a diagonal chase:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0\rightarrow & & \delta\left(\wedge^{r-1}T^*\otimes g_{q+1}\right) & \rightarrow & \wedge^r T^*\otimes R_q & \rightarrow & C_r & \rightarrow 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0\rightarrow & & \delta\left(\wedge^{r-1}T^*\otimes S_{q+1}T^*\otimes T\right) & \rightarrow & \wedge^r T^*\otimes J_q(E) & \rightarrow & C_r(E) & \rightarrow 0 \\
 & & \downarrow & & \downarrow & \searrow & \downarrow & & \\
0\rightarrow & \wedge^r T^*\otimes R_q+\delta\left(\wedge^{r-1}T^*\otimes S_{q+1}T^*\otimes E\right) & \rightarrow & \wedge^r T^*\otimes F_0 & \rightarrow & F_r & \rightarrow 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

It follows that the short exact sequences $0\rightarrow C_r\rightarrow C_r(E)\overset{\Phi_r}{\rightarrow}F_r\rightarrow 0$ are allowing to define the Janet and Spencer bundles inductively. If we consider *two* involutive systems $0\subset R_q\subset\hat{R}_q\subset J_q(E)$, it follows that the *kernels* of the induced canonical epimorphisms $F_r\rightarrow\hat{F}_r\rightarrow 0$ are isomorphic to the *cokernels* of the canonical monomorphisms $0\rightarrow C_r\rightarrow\hat{C}_r\subset C_r(E)$ and we may say that *Janet and Spencer play at see-saw* because we have the formula $dim(C_r)+dim(F_r)=dim(C_r(E))$.

When dealing with applications, we have set $E=T$ and considered systems of finite type Lie equations determined by Lie groups of transformations. Accordingly, we have obtained in particular $C_r=\wedge^r T^*\otimes R_2\subset\wedge^r T^*\otimes\hat{R}_2=\hat{C}_r\subset C_r(T)$ when comparing the classical and conformal Killing systems, but *these bundles have never been used in physics*. However, instead of the classical Killing sys-

tem $R_1 \subset J_1(T)$ defined by the infinitesimal first order PD Lie equations $\Omega \equiv \mathcal{L}(\xi)\omega = 0$ *and* its first prolongations $R_2 \subset J_2(T)$ defined by the infinitesimal additional second order PD Lie equations $\Gamma \equiv \mathcal{L}(\xi)\gamma = 0$ or the conformal Killing system $\hat{R}_2 \subset J_2(T)$ defined by $\Omega \equiv \mathcal{L}(\xi)\omega = 2A(x)\omega$ and $\Gamma \equiv \mathcal{L}(\xi)\gamma = \left(\delta_i^k A_j(x) + \delta_j^k A_i(x) - \omega_{ij}\omega^{ks}A_s(x)\right) \in S_2T^* \otimes T$ but we may also consider the formal Lie derivatives for geometric objects:

$$\Omega_{ij} \equiv \left(L(\xi_1)\omega\right)_{ij} \equiv \omega_{rj}\xi_i^r + \omega_{ir}\xi_j^r + \xi^r\partial_r\omega_{ij} = 0$$

$$\Gamma_{ij}^k \equiv \left(L(\xi_2)\gamma\right)_{ij}^k \equiv \xi_{ij}^k + \gamma_{rj}^k\xi_i^r + \gamma_{ir}^k\xi_j^r - \gamma_{ij}^r\xi_k^r + \xi^r\partial_r\gamma_{ij}^k = 0$$

We may now introduce the *intermediate differential system* $\tilde{R}_2 \subset J_2(T)$ defined by $\mathcal{L}(\xi)\omega = 2A(x)\omega$ and $\Gamma \equiv \mathcal{L}(\xi)\gamma = 0$, for the *Weyl group* obtained by adding the only dilatation with infinitesimal generator $x^i\partial_i$ to the Poincaré group. We have the relations $R_1 \subset \tilde{R}_1 = \hat{R}_1$ and the strict inclusions $R_2 \subset \tilde{R}_2 \subset \hat{R}_2$ when $R_2 = \rho_1(R_1)$, $\tilde{R}_2 = \rho_1(\tilde{R}_1)$, $\hat{R}_2 = \rho_1(\hat{R}_1)$ but we have to notice that we must have $\partial_i A - A_i = 0$ for the conformal system and thus $A_i = 0 \Rightarrow A = cst$ if we do want to deal again with an involutive second order system $\tilde{R}_2 \subset J_2(T)$. However, we must not forget that the comparison between the Spencer and the Janet sequences can only be done for involutive operators, that is we can indeed use the involutive systems $R_2 \subset \tilde{R}_2$ but we have to use $\hat{R}_3$ even if it is isomorphic to $\hat{R}_2$. Finally, as $\hat{g}_2 \simeq T^*$ and $\hat{g}_3 = 0, \forall n \geq 3$, the first Spencer operator $\hat{R}_2 \xrightarrow{D_1} T^* \otimes \hat{R}_2$ is induced by the usual Spencer operator $\hat{R}_3 \xrightarrow{D} T^* \otimes \hat{R}_2 : \left(0,0,\xi_{rj}^r, \xi_{rij}^r = 0\right) \to \left(0, \partial_i 0 - \xi_{ri}^r, \partial_i\xi_{rj}^r - 0\right)$ and thus projects by cokernel onto the induced operator $T^* \to T^* \otimes T^*$. Composing with δ, it projects therefore onto $T^* \xrightarrow{d} \wedge^2 T^* : A \to dA = F$ as in EM and so on by using the fact that D_1 *and d are both involutive* or the composite epimorphisms $\hat{C}_r \to \hat{C}_r/\tilde{C}_r \simeq \wedge^r T^* \otimes \left(\hat{R}_2/\tilde{R}_2\right) \simeq \wedge^r T^* \otimes \hat{g}_2 \simeq \wedge^r T^* \otimes T^* \xrightarrow{\delta} \wedge^{r+1} T^*$. The main result we have obtained is thus to be able to increase the order and dimension of the underlying jet bundles and groups as we have ([29]):

$$POINCARE \quad GROUP \subset WEYL \quad GROUP \subset CONFORMAL \quad GROUP$$

that is $10 < 11 < 15$ when $n = 4$ and our aim is now to prove that *the mathematical structures of electromagnetism and gravitation only depend on the second order jets*.

With more details, the *Killing system* $R_2 \subset J_2(T)$ is defined by the infinitesimal Lie equations in *Medolaghi form* with the well known *Levi-Civita isomorphism* $(\omega, \gamma) \simeq j_1(\omega)$ for geometric objects:

$$\begin{cases} \Omega_{ij} \equiv \omega_{rj}\xi_i^r + \omega_{ir}\xi_j^r + \xi^r\partial_r\omega_{ij} = 0 \\ \Gamma_{ij}^k \equiv \gamma_{rj}^k\xi_i^r + \gamma_{ir}^k\xi_j^r - \gamma_{ij}^r\xi_r^k + \xi^r\partial_r\gamma_{ij}^k = 0 \end{cases}$$

We notice that $R_2(\bar{\omega}) = R_2(\omega) \Leftrightarrow \bar{\omega} = a\omega, a = cst, \bar{\gamma} = \gamma$ and refer the reader to ([27]) for more details about the link between this result and the deformation theory of algebraic structures. We also notice that R_1 is formally integrable and

thus R_2 is involutive if and only if ω has constant Riemannian curvature along the results of L. P. Eisenhart ([26]). The only structure constant c appearing in the corresponding Vessiot structure equations is such that $\overline{c} = c/a$ and the normalizer of R_1 is R_1 if and only if $c \neq 0$. Otherwise R_1 is of codimension 1 in its normalizer $\tilde{R}_1$ as we shall see below by adding the only dilatation. In all what follows, ω is assumed to be flat with $c = 0$ and vanishing Weyl tensor.

The *Weyl system* $\tilde{R}_2 \subset J_2(T)$ is defined by the infinitesimal Lie equations:

$$\begin{cases} \omega_{rj}\xi_i^r + \omega_{ir}\xi_j^r + \xi^r\partial_r\omega_{ij} = 2A(x)\omega_{ij} \\ \xi_{ij}^k + \gamma_{rj}^k\xi_i^r + \gamma_{ri}^k\xi_j^r - \gamma_{ij}^r\xi_r^k + \xi^r\partial_r\gamma_{ij}^k = 0 \end{cases}$$

and is involutive if and only if $\partial_i A = 0 \Rightarrow A = cst$. Introducing for convenience the *metric density* $\hat{\omega}_{ij} = \omega_{ij} \Big/ \left(\left|det(\omega)\right|\right)^{\frac{1}{n}}$, we obtain the *Medolaghi form* for $(\hat{\omega}, \gamma)$ with $\left|det(\hat{\omega})\right| = 1$:

$$\begin{cases} \hat{\Omega}_{ij} \equiv \hat{\omega}_{rj}\xi_i^r + \hat{\omega}_{ir}\xi_j^r - \dfrac{2}{n}\hat{\omega}_{ij}\xi_r^r + \xi^r\partial_r\hat{\omega}_{ij} = 0 \\ \Gamma_{ij}^k \equiv \xi_{ij}^k + \gamma_{rj}^k\xi_i^r + \gamma_{ri}^k\xi_j^r - \gamma_{ij}^r\xi_r^k + \xi^r\partial_r\gamma_{ij}^k = 0 \end{cases}$$

Finally, the *conformal system* $\hat{R}_2 \subset J_2(T)$ is defined by the following infinitesimal Lie equations:

$$\begin{cases} \omega_{rj}\xi_i^r + \omega_{ir}\xi_j^r + \xi^r\partial_r\omega_{ij} = 2A(x)\omega_{ij} \\ \xi_{ij}^k + \gamma_{rj}^k\xi_i^r + \gamma_{ri}^k\xi_j^r - \gamma_{ij}^r\xi_r^k + \xi^r\partial_r\gamma_{ij}^k = \delta_i^k A_j(x) + \delta_j^k A_i(x) - \omega_{ij}\omega^{kr}A_r(x) \end{cases}$$

and is involutive if and only if $\partial_i A - A_i = 0$ or, equivalently, if ω has vanishing Weyl tensor.

However, introducing again the *metric density* $\hat{\omega}$ while substituting, we obtain after prolongation and division by $\left(\left|det(\omega)\right|\right)^{\frac{1}{n}}$ the second order system $\hat{R}_2 \subset J_2(T)$ in *Medolaghi form* and the Levi-Civita isomorphim $(\omega, \gamma) \simeq j_1(\omega)$ *restricts* to an isomorphism $(\hat{\omega}, \hat{\gamma}) \simeq j_1(\hat{\omega})$ if we set:

$$\hat{\gamma}_{ij}^k = \gamma_{ij}^k - \frac{1}{n}\left(\delta_i^k\gamma_{rj}^r + \delta_j^k\gamma_{ri}^r - \omega_{ij}\omega^{ks}\gamma_{rs}^r\right) \Rightarrow \hat{\gamma}_{ri}^r = 0 \left(tr(\hat{\gamma}) = 0\right)$$

$$\begin{cases} \hat{\Omega}_{ij} \equiv \hat{\omega}_{rj}\xi_i^r + \hat{\omega}_{ir}\xi_j^r - \dfrac{2}{n}\hat{\omega}_{ij}\xi_r^r + \xi^r\partial_r\hat{\omega}_{ij} = 0 \Rightarrow \omega^{ij}\overline{\Omega}^{ij} = 0 \\ \hat{\Gamma}_{ij}^k \equiv \xi_{ij}^k - \dfrac{1}{n}\left(\delta_i^k\xi_{rj}^r + \delta_j^k\xi_{ri}^r - \hat{\omega}_{ij}\hat{\omega}^{kr}\xi_{rs}^s\right) + \hat{\gamma}_{rj}^k\xi_i^r + \hat{\gamma}_{ri}^k\xi_j^r - \hat{\gamma}_{ij}^r\xi_r^k + \xi^r\partial_r\hat{\gamma}_{ij}^k = 0 \Rightarrow \hat{\Gamma}_{ri}^r = 0 \end{cases}$$

Contracting the first equations by $\hat{\omega}^{ij}$ we notice that ξ_r^r is *no longer vanishing* while, contracting in k and j the second equations, we now notice that ξ_{ri}^r is *no longer vanishing*. It is also essential to notice that the symbols $\hat{g}_1$ and $\hat{g}_2$ only depend on ω and not on any conformal factor.

The following Proposition does not seem to be known:

PROPOSITION 5.A.1: $(id, -\hat{\gamma})$ is the only symmetric $\hat{R}_1$-connection with vanishing trace.

Proof: Using a direct substitution, we have to study:

$$-\hat{\omega}_{ir}\hat{\gamma}^r_{jt}-\hat{\omega}_{rj}\hat{\gamma}^r_{it}+\frac{2}{n}\hat{\omega}_{ij}\hat{\gamma}^r_{rt}+\partial_t\hat{\omega}_{ij}$$

Multiplying by $\left(\left|det(\omega)\right|\right)^{\frac{1}{n}}$, we have to study:

$$-\omega_{ir}\hat{\gamma}^r_{jt}-\omega_{rj}\hat{\gamma}^r_{it}+\frac{2}{n}\omega_{ij}\hat{\gamma}^r_{rt}+\left(\left|det(\omega)\right|\right)^{\frac{1}{n}}\partial_t\hat{\omega}_{ij}$$

or equivalently:

$$-\omega_{ir}\hat{\gamma}^r_{jt}-\omega_{rj}\hat{\gamma}^r_{it}+\frac{2}{n}\omega_{ij}\hat{\gamma}^r_{rt}+\partial_t\omega_{ij}-\frac{1}{n}\omega_{ij}\left(\left|det(\omega)\right|\right)^{-1}\partial_t\left(\left|det(\omega)\right|\right)$$

that is to say:

$$-\omega_{ir}\hat{\gamma}^r_{jt}-\omega_{rj}\hat{\gamma}^r_{it}+\partial_t\omega_{ij}-\frac{2}{n}\omega_{ij}\gamma^s_{st}$$

Now, we have:

$$-\omega_{ir}\left(\gamma^r_{jt}-\frac{1}{n}\left(\delta^r_j\gamma^s_{st}+\delta^r_t\gamma^s_{sj}-\omega_{jt}\omega^{ru}\gamma^s_{su}\right)\right)=-\omega_{ir}\gamma^r_{jt}+\frac{1}{n}\omega_{ij}\gamma^s_{st}+\frac{1}{n}\omega_{it}\gamma^s_{sj}-\frac{1}{n}\omega_{jt}\gamma^s_{si}$$

Finally, taking into account that $(id,-\gamma)$ is a R_1-connection, we have:

$$-\omega_{ir}\gamma^r_{jt}-\omega_{rj}\gamma^r_{it}+\partial_t\omega_{ij}=0$$

Hence, collecting all the remaining terms, we are left with $\frac{2}{n}\omega_{ij}\gamma^s_{st}-\frac{2}{n}\omega_{ij}\gamma^s_{st}=0$.

As for the unicity, it comes from a chase in the commutative and exact diagram:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & \hat{g}_2 & \overset{\delta}{\to} & T^*\otimes\hat{g}_1 & \overset{\delta}{\to} & \wedge^2T^*\otimes T & \to & 0 \\
 & & \downarrow & & \downarrow & & \| & & \\
0 & \to & S_2T^*\otimes T & \overset{\delta}{\to} & T^*\otimes T^*\otimes T & \overset{\delta}{\to} & \wedge^2T^*\otimes T & \to & 0 \\
 & & & & & & \downarrow & & \\
 & & & & & & 0 & &
\end{array}$$

obtained by counting the respective dimensions with $dim(\hat{g}_1)=\left(n(n-1)/2\right)+1=\left(n^2-n+2\right)/2$ and $dim(\hat{g}_2)=n$ while checking that $-n+n\left(n^2-n+2\right)-n^2(n-1)/2=0$. The lower sequence splits because the short exact δ-sequence $0\to S_2T^*\overset{\delta}{\to}T^*\otimes T^*\overset{\delta}{\to}\wedge^2T^*\to 0$ splits and the upper sequence also splits because we have a composite monomorphism $\wedge^2T^*\otimes T\simeq T^*\otimes g_1\to T^*\otimes\hat{g}_1$.

□

COROLLARY 5.A.2: The R_1-connection $(id,-\gamma)$ is also a $\hat{R}_1$-connection.

Proof: This result first follows from the fact that $(id,-\gamma)\in T^*\otimes R_1$ is over $id\in T^*\otimes T$ and $R_1\subset\hat{R}_1$. However, we may also check such a property directly. Indeed, mutiplying $-\hat{\omega}_{rj}\gamma^r_{it}-\hat{\omega}_{ir}\gamma^r_{rt}+\frac{2}{n}\hat{\omega}_{ij}\gamma^r_{rt}+\partial_t\hat{\omega}_{ij}$ by $\left(\left|det(\omega)\right|\right)^{\frac{1}{n}}$ as in the

last Proposition, we obtain:

$$-\omega_{rj}\gamma^r_{it}-\omega_{ir}\gamma^r_{jt}+\frac{2}{n}\omega_{ij}\gamma^r_{rt}+\partial_t\hat{\omega}_{ij}=-\omega_{rj}\gamma^r_{it}-\omega_{rj}\gamma^r_{jt}+\partial_t\omega_{ij}=0$$

because $(id,-\gamma)$ is a R_1-connection.

□

REMARK 5.A.3: If one is using $(id,-\gamma)$, then $\left(L(\xi_2)\gamma\right)^k_{ij}\xi^k_{ij}$ when $\gamma=0$ locally and we have $(\delta\alpha)_i=\left(\partial_i\xi^r_r-\xi^r_{ri}\right)+\left(\alpha_r\partial_i\xi^r+\xi^r\partial_r\alpha-i\right)$ as the simplest variation. However, we have $f_2^{-1}(\gamma)=\overline{\gamma}\neq\gamma$ and we cannot thus split the Spencer operator over the target by means of a pull-back. Nevertheless, if one is using $(id,-\hat{\gamma})$, then $L(\xi_2)\hat{\gamma}=0$ when $\xi_2\in\hat{R}_2$ and the variation $(\delta\alpha)_i$ contains an additional term $\xi^s_{sr}\chi^r_{,i}$ but $f_2^{-1}(\hat{\gamma})=\hat{\gamma}$ and one can split the Spencer operator over the source and over the target with the help of $\hat{\gamma}$ but we have to point out that $\gamma=0\Rightarrow\hat{\gamma}=0$ *locally*.

□

We let the reader exhibit similarly the finite *Lie forms* of the previous equations that will be presented when needed. We have to distinguish the strict inclusions $\Gamma\subset\tilde{\Gamma}\subset\hat{\Gamma}\subset aut(X)$ with:

- The Lie pseudogroup $\Gamma\subset aut(X)$ of isometries which is preserving the metric $\omega\in S_2T^*$ with $det(\omega)\neq 0$ *and thus also* γ.
- The Lie pseudogroup $\tilde{\Gamma}$ which is preserving $\hat{\omega}$ *and* γ.
- The Lie pseudogroup $\hat{\Gamma}$ of conformal isometries which is preserving $\hat{\omega}$ *and thus also* $\hat{\gamma}$ with:

$$\begin{aligned}&g^k_l(x)\left(f^l_{ij}(x)+\gamma^l_{rs}\left(f(x)\right)f^r_i(x)f^s_j(x)\right)\\&=\overline{\gamma}^k_{ij}(x)=\gamma^k_{ij}(x)+\delta^k_i a_j(x)+\delta^k_i a_j(x)-\omega_{ij}(x)\omega^{kr}(x)a_r(x)\end{aligned}$$

where $a_i(x)dx^i\in T^*$ and thus $\overline{\gamma}-\gamma\in\hat{g}_2\subset S_2T^*\otimes T^*\otimes T$.

B) ELECTROMAGNETISM

The key idea, still never acknowledged, is that, even if we shall prove that *electromagnetism only depends on the elations of the conformal group* which are clearly non-linear transformations, we shall see that *electromagnetism has* "*by chance*" *a purely linear behaviour*.

Indeed, setting as we already did $\chi_0=A-id$ and defining $\chi^k_{lr,j}=A^s_j\tau^k_{lr,s}$, we may rewrite the defining equation of the second non-linear Spencer operator $\overline{D}'$ in the form:

$$\begin{cases}\partial_i A^k_j-\partial_j A^k_i=A^r_i\chi^k_{r,j}-A^r_j\chi^k_{r,i}=A^r_iA^s_j\left(\tau^k_{r,s}-\tau^k_{s,r}\right)\\\partial_i\chi^k_{l,j}-\partial_j\chi^k_{l,i}-\chi^r_{l,i}\chi^k_{r,j}+\chi^r_{l,j}\chi^k_{r,i}=A^r_i\chi^k_{lr,j}-A^r_j\chi^k_{lr,i}=A^r_iA^s_j\left(\tau^k_{lr,s}-\tau^k_{ls,r}\right)\end{cases}$$

Hence, contracting in *k* and *l*, *the quadratic terms in* χ *disappear* and we get:

$$\partial_i\chi^r_{r,j}-\partial_j\chi^r_{r,i}=A^r_iA^s_j\left(\tau^k_{kr,s}-\tau^k_{ks,r}\right)$$

By analogy with EM it should be tempting to introduce $\alpha_i=\chi^r_{r,i}$ and denote

by φ_{ij} the right member of the last formula but the relation $\partial_i\alpha_j-\partial_j\alpha_i=\varphi_{ij}$ thus obtained has no intrinsic meaning because α is far from being a 1-form while φ is far from being a 2-form.

REMARK 5.B.1: The *target* "*y*" could be called "*hidden variable*" as it is just used in order to construct objects over the *source* "*x*". As a byproduct, the changes of local coordinates are of the form $\overline{x}=\varphi(x),\overline{y}=\psi(y)$ but the second one does not appear through the implicit summations over the target because the first order transition rules are:

$$\overline{y}^l_j\frac{\partial\varphi^j}{\partial x^i}(x)=\frac{\partial\psi^l}{\partial y^k}(y)\,y^k_i\Rightarrow\overline{f}^l_j\left(\varphi(x)\right)\frac{\partial\varphi^j}{\partial x^i}(x)=\frac{\partial\psi^l}{\partial y^k}\left(f(x)\right)f^k_i(x)$$

It follows therefore that $A\in T^*\otimes T$ indeed and is thus a well defined object over the source.

LEMMA 5.B.2: The short exact δ-sequence $0\to S_2T^*\overset{\delta}{\to}T^*\otimes T^*\overset{\delta}{\to}\wedge^2T^*\to 0$ admits a canonical splitting, that is a splitting coherent with the tensor nature of the vector bundles involved.

Proof: The splitting of the above sequence is obtained by setting $\left(\tau_{i,j}\right)\in T^*\otimes T^*\to\left(\frac{1}{2}\left(\tau_{i,j}+\tau_{j,i}\right)\right)\in S_2T^*$ in such a way that

$\left(\tau_{i,j}=\tau_{j,i}=\tau_{ij}\right)\in S_2T^*\Rightarrow\frac{1}{2}\left(\tau_{ij}+\tau_{ji}\right)=\tau_{ij}$.

Similarly, $\left(\varphi_{ij}=-\varphi_{ji}\right)\in\wedge^2T^*\to\left(\frac{1}{2}\varphi_{ij}\right)\in T^*\otimes T^*$ and

$\left(\frac{1}{2}\varphi_{ij}-\frac{1}{2}\varphi_{ji}\right)=\left(\varphi_{ij}\right)\in\wedge^2T^*$.

□

We shall revisit the previous results by showing that, *in fact*, all the maps and splittings existing for the Killing operator are coming from maps and splittings existing for the conformal Killing operator, *though surprising it may look like*. As these results are based on a systematic use of the Spencer δ-map, they are neither known nor acknowledged.

We now recall the commutative diagrams allowing to define the (analogue) of the first Janet bundle and their dimensions when $n=4$:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & g_3 & \to & S_3T^*\otimes T & \to & S_2T^*\otimes F_0 & \to & F_1 & \to 0\\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & T^*\otimes g_2 & \to & T^*\otimes S_2T^*\otimes T & \to & T^*\otimes T^*\otimes F_0 & \to & 0 & \\
 & & \downarrow & & \downarrow & & \downarrow & & & \\
0 & \to & \wedge^2T^*\otimes g_1 & \to & \underline{\wedge^2T^*\otimes T^*\otimes T} & \to & \wedge^2T^*\otimes F_0 & \to & 0 & \\
 & & \downarrow & & \downarrow & & \downarrow & & & \\
0 & \to & \wedge^3T^*\otimes T & = & \wedge^3T^*\otimes T & \to & 0 & & & \\
 & & \downarrow & & \downarrow & & & & & \\
 & & 0 & & 0 & & & & &
\end{array}$$

$$\begin{array}{ccccccccc} & & & & 0 & & 0 & & \\ & & & & \downarrow & & \downarrow & & \\ & & 0 & \to & 80 & \to & 100 & \to & 20 & \to 0 \\ & & & & \downarrow & & \downarrow & & \\ & & 0 & \to & 160 & \to & 160 & \to & 0 & \\ & & & & \downarrow & & \downarrow & & & \\ & 0 \to & 36 & \to & 96 & \to & 60 & \to & 0 & \\ & & \downarrow & & \downarrow & & \downarrow & & & \\ & 0 \to & 16 & = & 16 & \to & 0 & & & \\ & & \downarrow & & \downarrow & & & & & \\ & & 0 & & 0 & & & & & \end{array}$$

PROPOSITION 5.B.3: Recalling that we have $F_1 = H^2(g_1) = Z^2(g_1)$ in the Killing case and $\hat{F}_1 = H^2(\hat{g}_1) \neq Z^2(\hat{g}_1)$ in the conformal Killing case, we have the unusual commutative diagram:

$$\begin{array}{ccccccccc} & & 0 & & 0 & & 0 & & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ 0 \to & & Z^2(g_1) & \to & Z^2(\hat{g}_1) & \subset & Z^2(T^* \otimes T) & \to & S_2T^* \\ & & \downarrow & & \downarrow & & \downarrow & & \downarrow \delta \\ 0 \to & & \wedge^2 T^* \otimes g_1 & \to & \wedge^2 T^* \otimes \hat{g}_1 & \subset & \wedge^2 T^* \otimes T^* \otimes T & \to & T^* \otimes T^* \\ & & \downarrow \delta & & \downarrow \delta & & \downarrow \delta & & \downarrow \delta \\ 0 \to & & \wedge^3 T^* \otimes T & = & \wedge^3 T^* \otimes T & = & \wedge^3 T^* \otimes T & \to & \wedge^2 T^* \\ & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ & & 0 & & 0 & & 0 & & 0 \end{array}$$

Proof: First of all, we must point out that the surjectivity of the bottom δ in the central column is well known from the exactness of the δ-sequence for S_3T^* and thus also after tensoring by T. However, the surjectivity of the bottom δ in the left column is *not evident at all* as it comes from a delicate circular chase in the preceding diagram, using the fact that the Riemann and Weyl operators are second order operators. Then, setting $\varphi_{ij} = \rho^r_{r,ij} = -\varphi_{ji}$ and $\rho_{ij} = \rho^r_{i,rj} \neq \rho_{ji}$, we may define the right central horizontal map by $\rho^k_{l,ij} \to \rho_{ij} - \frac{1}{2}\varphi_{ij}$ and the right bottom horizontal map by $\omega \otimes \xi \to -i(\xi)\omega$ by introducing the interior product $i(\,)$. We obtain at once:

$$-\left(\rho^r_{r,ij} + \rho^r_{i,jr} + \rho^r_{j,ri}\right) = -\varphi_{ij} + \rho_{ij} - \rho_{ji} = \left(\rho_{ij} - \frac{1}{2}\varphi_{ij}\right) - \left(\rho_{ji} - \frac{1}{2}\varphi_{ji}\right)$$

and the right bottom diagram is commutative, clearly inducing the upper map. If we restrict to the Killing symbol, then $\varphi_{ij} = 0$ and we obtain $\rho_{ij} - \rho_{ji} = 0 \Rightarrow (\rho_{ij} = \rho_{ji}) \in S_2T^*$, that is the classical contraction allowing to obtain the Ricci tensor from the Riemann tensor but *there is no way to go backwards with a canonical lift*. A similar comment may be done for the conformal Killing symbol and the $\frac{1}{2}$ coefficient.

□

Using the previous diagram allowing to define both $F_1 = H^2(g_1) = Z^2(g_1)$ if we use ω or $\hat{F}_1 = H^2(\hat{g}_1) = Z^2(\hat{g}_1)/\delta(T^* \otimes \hat{g}_2)$ if we use $\hat{\omega}$ while taking into account that $dim(\hat{g}_1/g_1) = 1$ and $\hat{g}_2 \simeq T^*$, we obtain the crucial theorem *which is in fact only depending on* ω:

THEOREM 5.B.4: We have the commutative and exact "*fundamental diagram II*":

$$
\begin{array}{ccccccccc}
 & & & & & & 0 & & \\
 & & & & & & \downarrow & & \\
 & & & & 0 & & S_2T^* & & \\
 & & & & \downarrow & & \downarrow & & \\
 & & & 0 & \to & Z^2(g_1) & \to & H^2(g_1) & \to 0 \\
 & & & \downarrow & & \downarrow & & \downarrow & \\
 & & 0 & \to & T^* \otimes \hat{g}_2 & \xrightarrow{\delta} & Z^2(\hat{g}_1) & \to & H^2(\hat{g}_1) & \to 0 \\
 & & & & \downarrow & & \downarrow & & \downarrow & \\
0 & \to & S_2T^* & \xrightarrow{\delta} & T^* \otimes T^* & \xrightarrow{\delta} & \wedge^2 T^* & \to & 0 & \\
 & & & & \downarrow & & \downarrow & & & \\
 & & & & 0 & & 0 & & &
\end{array}
$$

The following theorem will provide *all* the classical formulas of both Riemannian and conformal geometry in a totally unusual framework *not depending on any conformal factor*:

THEOREM 5.B.5: All the short exact sequences of the preceding diagram split in a canonical way, that is in a way compatible with the underlying tensorial properties of the vector bundles involved. With more details:

$$
\begin{aligned}
T^* \otimes T^* \simeq S_2T^* \oplus \wedge^2 T^* &\Rightarrow Z^2(\hat{g}_1) \simeq Z^2(g_1) + \delta(T^* \otimes \hat{g}_2) \simeq Z^2(g_1) \oplus \wedge^2 T^* \\
&\Rightarrow H^2(g_1) \simeq H^2(\hat{g}_1) \oplus S_2T^* \\
&\Rightarrow F_1 \simeq \hat{F}_1 \oplus S_2T^*
\end{aligned}
$$

Proof: First of all, we recall that:

$$
\begin{aligned}
g_1 &= \left\{ \xi_i^k \in T^* \otimes T \mid \omega_{rj}\xi_i^r + \omega_{ir}\xi_j^r = 0 \right\} \\
&\subset \hat{g}_1 = \left\{ \xi_i^k \in T^* \otimes T \mid \omega_{rj}\xi_i^r + \omega_{ir}\xi_j^r - \frac{2}{n}\omega_{ij}\xi_r^r = 0 \right\}
\end{aligned}
$$

$$
\Rightarrow \quad 0 = g_2 \subset \hat{g}_2 = \left\{ \xi_{ij}^k \in S_2T^* \otimes T \mid n\xi_{ij}^k = \delta_i^k \xi_{rj}^r + \delta_j^k \xi_{ri}^r - \omega_{ij}\omega^{ks}\xi_{rs}^r \right\}
$$

Now, if $\left(\tau_{li,j}^k\right) \in T^* \otimes \hat{g}_2$, then we have:

$$
n\tau_{li,j}^k = \delta_l^k \tau_{ri,j}^r + \delta_i^k \tau_{rl,j}^r - \omega_{li}\omega^{ks}\tau_{rs,j}^r
$$

and we may set $\tau_{ri,j}^r = \tau_{i,j} \neq \tau_{j,i}$ with $\left(\tau_{i,j}\right) \in T^* \otimes T$ and such a formula does not depend on any conformal factor. Taking into account Proposition 4.B.5, we have:

$$
\delta\left(\tau_{li,j}^k\right) = \left(\tau_{li,j}^k - \tau_{lj,i}^k\right) = \left(\rho_{l,ij}^k\right) \in B^2(\hat{g}_1) \subset Z^2(\hat{g}_1)
$$

with:

$$Z^2(\hat{g}_1)=\left\{\left(\rho^k_{l,ij}\right)\in\wedge^2T^*\otimes\hat{g}_1 \mid \delta\left(\rho^k_{l,ij}\right)=0\right\}\Rightarrow\varphi_{ij}=\rho^r_{r,ij}\neq 0$$

$$\delta\left(\rho^k_{l,ji}\right)=\left(\mathcal{C}_{(l,i,j)}\rho^k_{l,ij}=\rho^k_{l,ij}+\rho^k_{i,jl}+\rho^k_{j,li}\right)\in\wedge^3T^*\otimes T$$

- The splitting of the central vertical column is obtained from a lift of the epimorphism $Z^2(\hat{g}_1)\to\wedge^2T^*\to 0$ which is obtained by lifting $\left(\varphi_{ij}\right)\in\wedge^2T^*$ to $\left(\frac{1}{2}\varphi_{ij}\right)\in T^*\otimes T^*$, setting $\tau^r_{ri,j}=\frac{1}{2}\varphi_{ij}$ and applying δ to obtain $\left(\tau^r_{ri,j}-\tau^r_{rj,i}=\frac{1}{2}\varphi_{ij}-\frac{1}{2}\varphi_{ji}=\varphi_{ij}\right)\in B^2(\hat{g}_1)\subset Z^2(\hat{g}_1)$.
- Now, let us define $\left(\rho_{i,j}=\rho^r_{i,rj}\neq\rho_{j,i}\right)\in T^*\otimes T^*$. Hence, elements of $Z^2(g_1)$ are such that:

$$\varphi_{ij}=\rho^r_{r,ij}=0,\ \varphi_{ij}-\rho_{i,j}+\rho_{j,i}=0\Rightarrow\left(\rho_{ij}=\rho_{i,j}=\rho_{j,i}=\rho_{ji}\right)\in S_2T^*$$

while elements of $Z^2(\hat{g}_1)$ are such that:

$$\left(\rho^r_{r,ij}=\varphi_{ij}=\rho_{i,j}-\rho_{j,i}=\tau_{i,j}-\tau_{j,i}\neq 0\right)\in\wedge^2T^*$$

Accordingly, $\left(\rho_{i,j}-\frac{1}{2}\varphi_{ij}=\rho_{j,i}-\frac{1}{2}\varphi_{ji}\right)\in S_2T^*$. More generally, we may consider $\rho^k_{l,ij}-\left(\tau^k_{li,j}-\tau^k_{lj,i}\right)$ with $\tau^r_{ri,j}=\frac{1}{2}\varphi_{ij}$. Such an element is killed by δ and thus belongs to $Z^2(\hat{g}_1)$ because each member of the difference is killed by δ. However, we have $\rho^r_{r,ij}-\left(\tau^r_{ri,j}-\tau^r_{rj,i}\right)=\varphi_{ij}-\varphi_{ij}=0$ and the element does belong indeed to $Z^2(g_1)$, providing a lift $Z^2(\hat{g}_1)\to Z^2(g_1)\to 0$.

- Of course, *the most important result is to split the right column*. As this will be the hard step, we first need to describe the monomorphism $0\to S_2T^*\to H^2(g_1)$ which is in fact produced by a north-east diagonal snake type chase. Let us choose $\left(\tau_{ij}=\tau_{i,j}=\tau_{j,i}=\tau_{ji}\right)\in S_2T^*\subset T^*\otimes T^*$. Then, we may find $\left(\tau^k_{li,j}\right)\in T^*\otimes\hat{g}_2$ by deciding that $\tau^r_{ri,j}=\tau_{i,j}=\tau_{j,i}=\tau^r_{rj,i}$ in $Z^2(\hat{g}_1)$ and apply δ in order to get $\rho^k_{l,ij}=\tau^k_{li,j}-\tau^k_{k,lj,i}$ such that $\rho^r_{r,ij}=\varphi_{ij}=0$ and thus $\left(\rho^k_{l,ij}\right)\in Z^2(g_1)=H^2(g)_1$. We obtain:

$$\begin{aligned}n\rho^k_{l,ij}&=\delta^k_l\tau^r_{ri,j}-\delta^k_l\tau^r_{rj,i}+\delta^k_i\tau^r_{rl,j}-\delta^k_j\tau^r_{rli}-\omega^{ks}\left(\omega_{li}\tau^r_{rs,j}-\omega_{lj}\tau^r_{rs,i}\right)\\&=\left(\delta^k_i\tau_{lj}-\delta^k_j\tau_{li}\right)-\omega^{ks}\left(\omega_{li}\tau_{sj}-\omega_{lj}\tau_{si}\right)\end{aligned}$$

Contracting in k and i while setting simply $tr(\tau)=\omega^{ij}\tau_{ij}, tr(\rho)=\omega^{ij}\rho_{ij}$, we get:

$$\begin{aligned}&n\rho_{ij}=n\tau_{ij}-\tau_{ij}-\tau_{ij}+\omega_{ij}tr(\tau)=(n-2)\tau_{ij}+\omega_{ij}tr(\tau)=n\rho_{ji}\\&\Rightarrow ntr(\rho)=2(n-1)tr(\tau)\end{aligned}$$

Substituting, we finally obtain $\tau_{ij}=\frac{n}{n-2}\rho_{ij}-\frac{n}{2(n-1)(n-2)}\omega_{ij}tr(\rho)$ and thus the tricky formula:

$$\begin{aligned}\rho^k_{l,ij}=&\frac{1}{n-2}\left(\left(\delta^k_i\rho_{lj}-\delta^k_j\rho_{li}\right)-\omega^{ks}\left(\omega_{li}\rho_{sj}-\omega_{lj}\rho_{si}\right)\right)\\&-\frac{1}{(n-1)(n-2)}\left(\delta^k_i\omega_{lj}-\delta^k_j\omega_{li}\right)tr(\rho)\end{aligned}$$

Contracting in k and i, we check that $\rho_{ij}=\rho_{ij}$ indeed, obtaining therefore the desired canonical lift $H^2(g_1)\to S_2T^*\to 0:\rho^k_{i,lj}\to\rho^r_{i,rj}=\rho_{ij}$. Finally, using again Proposition 3.4, the epimorphism $H^2(g_1)\to H^2(\hat{g}_1)\to 0$ is just described by the formula:

$$\sigma^k_{l,ij}=\rho^k_{l,ij}-\frac{1}{n-2}\Big(\big(\delta^k_i\rho_{lj}-\delta^k_j\rho_{li}\big)-\omega^{ks}\big(\omega_{li}\rho_{sj}-\omega_{lj}\rho_{si}\big)\Big)$$
$$+\frac{1}{(n-1)(n-2)}\big(\delta^k_i\omega_{ij}-\delta^k_j\omega_{li}\big)tr(\rho)$$

which is just the way to define the Weyl tensor. We notice that $\sigma^r_{r,ij}=\rho^r_{r,ij}=0$ and $\sigma^r_{i,rj}=0$ by using indices or a circular chase showing that $Z^2(\hat{g}_1)=Z^2(g_1)+\delta\big(T^*\otimes\hat{g}_2\big)$. This purely algebraic result only depends on the metric ω and does not depend on any conformal factor. In actual practice, the lift $H^2(g_1)\to S_2T^*$ is described by $\rho^k_{l,ij}\to\rho^r_{i,rj}=\rho_{ij}=\rho_{ji}$ but it is not evident at all that the lift $H^2(\hat{g}_1)\to H^2(g_1)$ is described by the strict inclusion $\sigma^k_{l,ij}\to\rho^k_{l,ij}=\sigma^k_{l,ij}$ providing a short exact sequence as in Proposition 3.4 because $\rho_{ij}=\rho^r_{i,rj}=\sigma^r_{i,rj}=0$ by composition.

□

PROPOSITION 5.B.6: We have the following commutative and exact diagram made by splitting sequences according to a circular chase in the right upper commutative square:

$$\begin{array}{ccccccccc}
 & & & & 0 & & 0 & & \\
 & & & & \downarrow & & \downarrow & & \\
 & & & 0 & \to & \hat{g}_2 & \to & T^* & \to & 0 \\
 & & & \downarrow & & \downarrow & & \| & & \\
0 & \to & \tilde{R}_2 & \to & \hat{R}_2 & \to & T^* & \to & 0 & \\
 & & \downarrow & & \downarrow & & \downarrow & & & \\
0 & \to & \tilde{R}_1 & = & \hat{R}_1 & \to & 0 & & & \\
 & & \downarrow & & \downarrow & & & & & \\
 & & 0 & & 0 & & & & &
\end{array}$$

This diagram is thus leading to the short exact sequence:

$$0\to T^*\otimes\tilde{R}_2\to T^*\otimes\hat{R}_2\to T^*\otimes T^*\to 0$$

with a canonical splitting $T^*\otimes T^*\simeq S_2T^*\oplus\wedge^2T^*$.

Proof: According to the definition of the Christoffel symbols γ for the metric ω, we have:

$$2\omega_{rk}\gamma^k_{ij}=\partial_i\omega_{rj}+\partial_j\omega_{ri}-\partial_r\omega_{ij}\Leftrightarrow\omega_{kj}\gamma^k_{ir}+\omega_{ik}\gamma^k_{jr}-\partial_r\omega_{ij}=0$$

It follows that $-\gamma$ (*care*) is the unique symmetric R_1-connection, that is a map $T\to R_1$ considered as an element of $T^*\otimes R_1$ projecting onto $id_T\in T^*\otimes T$. Accordingly, any $\chi_1\in T^*\otimes J_1(T)$ provides $\big(\chi^k_{j,i}+\gamma^k_{jr}\chi^r_{,i}\big)\in T^*\otimes T^*\otimes T$ and thus a true 1-form $\big(\alpha_i=\chi^r_{r,i}+\gamma^r_{r,s}\chi^s_{,i}\big)\in T^*$. However, such an approach cannot be extended to higher orders and we prefer to consider half of the morphism defining the Killing operator, namely the morphism $J_1(T)\to S_2T^*:\xi_1\to\frac{1}{2}L(\xi_1)\omega$,

tensor it by T^* and contract it by ω^{-1} in order to get:

$$\frac{1}{2}\omega^{st}\left(\omega_{rt}\chi^r_{s,i}+\omega_{sr}\chi^r_{t,i}+\chi^r_{,i}\partial_r\omega_{st}\right)=\chi^r_{r,i}+\frac{1}{2}\chi^r_{,i}\omega^{st}\partial_r\omega_{st}=\alpha_i$$

where we notice that:

$$2\gamma^r_{ri}=\omega^{st}\partial_r\omega_{st}=\left(1/det(\omega)\right)\partial_i det(\omega)\Rightarrow\partial_i\gamma^r_{rj}-\partial_j\gamma^r_{r,i}=0$$

Similarly, there is a well defined map $J_2(T)\to S_2T^*\otimes T:\xi_2\to L(\xi_2)\gamma$ that can be tensored by T^* and restricted to $T^*\otimes\hat{R}_2$ in order to obtain a well defined map $T^*\otimes\hat{R}_2\to T^*\otimes S_2T^*\otimes T$ that can be contracted to $T^*\otimes T^*$ according to the following local formulas:

$$\beta^k_{lr,s}=\tau^k_{lr,s}+\gamma^k_{ur}\tau^u_{l,s}+\gamma^k_{lu}\tau^u_{r,s}-\gamma^u_{lr}\tau^k_{u,s}+\tau^u_{,s}\partial_u\gamma^k_{lr}$$
$$\beta^k_{kr,s}=\tau^k_{kr,s}+\gamma^k_{ku}\tau^u_{r,s}+\tau^u_{,s}\partial_u\gamma^k_{kr}$$

We can "*twist*" by A and apply $\delta:T^*\otimes S_2T^*\otimes T\to\wedge^2T^*\otimes T^*\otimes T$ that can be contracted to $\wedge^2T^*$ according to the following local formulas:

$$\varphi^k_{l,ij}=A^r_iA^s_j\left(\beta^k_{lr,s}-\beta^k_{ls,r}\right)\Rightarrow\varphi_{ij}=\varphi^r_{r,ij}=A^r_iA^s_j\left(\beta^k_{kr,s}-\beta^k_{ks,r}\right)$$

□

As $\varphi\in\wedge^2T^*$ though $\chi_2\in T^*\otimes\hat{R}_2$, we obtain the following crucial theorem ([4] [8]):

THEOREM 5.B.7: The non-linear Spencer sequence for the conformal group of transformations projects onto a part of the Poincaré sequence for the exterior derivative according to the following commutative and locally exact diagram:

$$\begin{array}{ccccccccc}
0\to & \hat{\Gamma} & \xrightarrow{j_2} & \hat{\mathcal{R}}_2 & \xrightarrow{\bar{D}_1} & T^*\otimes\hat{R}_2 & \xrightarrow{\bar{D}_2} & \wedge^2T^*\otimes\hat{R}_2 \\
 & & & \downarrow & \swarrow & \downarrow & & \downarrow \\
 & & & T^* & \xrightarrow{d} & \wedge^2T^* & \xrightarrow{d} & \wedge^3T^* \\
 & & & \alpha & & d\alpha=\varphi & & d\varphi=0
\end{array}$$

Accordingly, *this purely mathematical result contradicts classical gauge theory*.

Proof: Substituting the previous results in the last formula, we obtain successively:

$$\begin{aligned}
\varphi_{ij}&=A^r_iA^s_j\left(\tau^k_{kr,s}-\tau^k_{ks,r}\right)+\gamma^k_{ku}A^r_iA^s_j\left(\tau^u_{r,s}-\tau^u_{s,r}\right)+A^r_iA^s_j\left(\tau^u_{,s}\partial_u\gamma^k_{kr}-\tau^u_{,r}\partial_u\gamma^k_{ks}\right)\\
&=\left(\partial_i\chi^r_{r,j}-\partial_j\chi^r_{r,i}\right)+\gamma^r_{ru}\left(\partial_iA^u_j-\partial_jA^u_i\right)+\left(\delta^r_i+\chi^r_{,i}\right)\chi^s_{,j}\partial_r\gamma^k_{ks}-\left(\delta^s_j+\chi^s_{,j}\right)\chi^r_{,i}\partial_s\gamma^k_{kr}\\
&=\left(\partial_i\chi^r_{r,j}-\partial_j\chi^r_{r,i}\right)+\gamma^r_{rs}\left(\partial_i\chi^s_{,j}-\partial_j\chi^s_{,i}\right)+\left(\chi^s_{,j}\partial_i\gamma^r_{rs}-\chi^s_{,i}\partial_j\gamma^r_{rs}\right)\\
&=\partial_i\alpha_j-\partial_j\alpha_i
\end{aligned}$$

because $\partial_i\gamma^r_{rj}-\partial_j\gamma^r_{ri}=0$. It follows that $d\alpha=\varphi\in\wedge^2T^*$ and thus $d\varphi=0$, that is $\partial_i\varphi_{jk}+\partial_j\varphi_{ki}+\partial_k\varphi_{ij}=0$, has an intrinsic meaning in $\wedge^3T^*$. It is important to notice that the corresponding EM Lagrangian is defined on sections of $\hat{C}_1$ killed by $\bar{D}_2$ but *not* on $\hat{C}_2$, *contrary to gauge theory*. Finally, the south west arrow in the left square is the composition:

$$f_2 \in \hat{\mathcal{R}}_2 \xrightarrow{\bar{D}_1} \chi_2 \in T^* \otimes \hat{R}_2 \xrightarrow{\pi_1^2} \chi_1 \in T^* \otimes \hat{R}_1 \xrightarrow{(\gamma)} \alpha \in T^*$$

Accordingly, though α is a potential for φ, it can also be considered as a part of the *field* but the important fact is that the first set of (*linear*) Maxwell equations $d\varphi = 0$ is induced by the (*nonlinear*) operator $\bar{D}_2$. The linearized framework will explain this point.

One of the most important but difficult result of this paper will be the following direct proof of the existence of the right square in the previous diagram.

Supposing for simplicity that ω is a (locally) constant metric (in fact the Minkowski metric!) and thus $\gamma = 0$. When we are considering the conformal group of space-time, it first follows that the jets of order three vanish and the Formula (3*) can be now written:

$$\partial_i \chi^k_{lr,j} - \partial_j \chi^k_{lr,i} - \left(\chi^s_{r,i} \chi^k_{ls,j} + \chi^s_{l,i} \chi^k_{rs,j} + \chi^s_{lr,i} \chi^k_{s,j} - \chi^s_{r,j} \chi^k_{ls,i} - \chi^s_{l,j} \chi^k_{rs,i} - \chi^s_{lr,j} \chi^k_{s,i} \right) = 0$$

Contracting in $k = l = u$ and replacing *r* by *t*, we obtain the simple formula:

$$\partial_i \chi^u_{ut,j} - \partial_j \chi^u_{ut,i} - \chi^s_{t,i} \chi^u_{us,j} + \chi^s_{t,j} \chi^u_{us,i} = 0$$

Multiplyig by A^t_k the *two last terms* and replacing χ by τ, we get *for these terms only*:

$$A^r_i A^s_j A^t_k \left(\tau^v_{t,s} \tau^u_{uv,r} - \tau^v_{t,r} \tau^u_{uv,s} \right)$$

Now, denoting by $\mathcal{C}(i,j,k)$ the cyclic sum on the permutation $(i,j,k) \to (j,k,i) \to (k,i,j)$ and proceeding in this way on the last formula, we obtain easily:

$$\mathcal{C}(i,j,k) A^r_i A^s_j A^t_k \left(\tau^v_{t,s} - \tau^v_{s,t} \right) \tau^u_{uv,r}$$

or, equivalently:

$$A^r_i A^s_j A^t_k \mathcal{C}(r,s,t) \left(\tau^v_{t,s} - \tau^v_{s,t} \right) \tau^u_{uv,r} = A^r_i A^s_j A^t_k \mathcal{C}(r,s,t) \left(\tau^v_{s,r} - \tau^v_{r,s} \right) \tau^u_{uv,t}$$

Let us now similarly consider only the *two first terms*. After multiplication by A^t_k and integration by part, we get for the first:

$$A^t_k \left(\partial_i \left(A^s_j \tau^u_{ut,s} \right) \right) = \partial_i \left(A^s_j A^t_k \tau^u_{ut,s} \right) - A^s_j \tau^u_{ut,s} \partial_i A^t_k$$

Applying the same procedure to the second term and considering the sum $\mathcal{C}(i,j,k)$ while rearranging the six terms of the summation two by two, we obtain:

$$\mathcal{C}(i,j,k) \left(\partial_k \left(A^t_j A^s_i \tau^u_{ut,s} - A^t_i A^r_j \tau^u_{ut,r} \right) + A^t_k \tau^u_{ur,t} \partial_j A^r_i - A^t_k \tau^u_{ut,r} \partial_i A^r_j \right)$$

Exchanging the dumb indices between themselves, we finally obtain:

$$\mathcal{C}(i,j,k) \left(\partial_k \left(A^r_i A^s_j \left(\tau^u_{us,r} - \tau^u_{ur,s} \right) \right) + A^t_k \tau^u_{ur,t} \left(\partial_i A^r_j - \partial_j A^r_i \right) \right)$$

that is to say, taking into account the Equations (1*) and changing the signs:

$$\mathcal{C}(i,j,k) \left(\partial_k \varphi_{ij} \right) - \mathcal{C}(i,j,k) \left(A^r_i A^s_j A^t_k \left(\tau^v_{r,s} - \tau^v_{s,r} \right) \tau^u_{uv,t} \right)$$

or, equivalently:

$$\mathcal{C}(i,j,k)\left(\partial_k\varphi_{ij}\right)-A_i^rA_j^sA_k^t\mathcal{C}(r,s,t)\left(\tau_{r,s}^v-\tau_{s,r}^v\right)\tau_{uv,t}^u$$

Collecting all the results, we are only left with $\mathcal{C}(i,j,k)\left(\partial_k\varphi_{ij}\right)=0$ as we wished.

□

COROLLARY 5.B.8: The linear Spencer sequence for the conformal group of transformations projects onto a part of the Poincaré sequence for the exterior derivative according to the following commutative and locally exact diagram:

$$\begin{array}{ccccccccc}
0\rightarrow & \hat{\Theta} & \xrightarrow{j_2} & \hat{R}_2 & \xrightarrow{D_1} & T^*\otimes\hat{R}_2 & \xrightarrow{D_2} & \wedge^2T^*\otimes\hat{R}_2 \\
 & & & \downarrow & \swarrow & \downarrow & & \downarrow \\
 & & & T^* & \xrightarrow{d} & \wedge^2T^* & \xrightarrow{d} & \wedge^3T^* \\
 & & & A & & dA=F & & dF=0
\end{array}$$

Accordingly, *this purely mathematical result also contradicts classical gauge theory* because it proves that EM only depends on the structure of the conformal group of space-time but not on $U(1)$.

Proof: Considering ω and γ as geometric objects, we obtain at once the formulas:

$$\overline{\omega}_{ij}=\mathrm{e}^{2a(x)}\omega_{ij}\quad\Rightarrow\quad\overline{\gamma}_{ri}^r=\gamma_{ri}^r+\partial_i a$$

Though looking like the key Formula (69) in ([3], p 286), this transformation is quite different because the sign is not coherent and the second object has nothing to do with a 1-form. Moreover, if we use $n=2$ and set $\mathcal{L}(\xi)\omega=2A\omega$ for the standard euclidean metric, we should have $(\partial_{11}+\partial_{22})A=0$, contrary to the assumption that A is arbitrary which is *only* agreeing with the following jet formulas improving the ones already provided in ([29] [36] [40]) in order to point out the systematic use of the Spencer operator:

$$L(\xi_1)\omega=2A\omega\quad\Rightarrow\quad\left(\xi_r^r+\gamma_{ri}^r\xi^i\right)=nA,\quad\left(L(\xi_2)\gamma\right)_{ri}^r=nA_i,\quad\forall\xi_2\in\hat{R}_2$$

Now, if we make a change of coordinates $\overline{x}=\varphi(x)$ on a function $a\in\wedge^0T^*$, we get:

$$\overline{a}\left(\varphi(x)\right)=a(x)\quad\Rightarrow\quad\frac{\partial\overline{a}}{\partial\overline{x}^j}\frac{\partial\varphi^j}{\partial x^i}=\frac{\partial a}{\partial x^i}$$

We obtain therefore an isomorphism $J_1\left(\wedge^0T^*\right)\simeq\wedge^0T^*\times_X T^*$, a result leading to the following commutative diagram:

$$\begin{array}{ccccccccc}
0\rightarrow & R_2 & \rightarrow & \hat{R}_2 & \rightarrow & J_1\left(\wedge^0T^*\right) & \rightarrow 0 \\
 & \downarrow D & & \downarrow D & & \downarrow D & \\
0\rightarrow & T^*\otimes R_1 & \rightarrow & T^*\otimes\hat{R}_1 & \rightarrow & T^* & \rightarrow 0
\end{array}$$

where the rows are exact by counting the dimensions. The operator $D:(A,A_i)\rightarrow(\partial_iA-A_i)$ on the right is induced by the central Spencer operator, a result that could not have been even imagined by Weyl and followers. This result provides a good transition towards the conformal origin of electromagnetism.

As the nonlinear aspect has been already presented, we restrict our study to the linear framework. A first problem to solve is to construct vector bundles from the components of the image of D_1. Using the corresponding capital letter for denoting the linearization, let us introduce:

$$\left(B^k_{l,i} = X^k_{l,i} + \gamma^k_{ls} X^s_{,i}\right) \in T^* \otimes T^* \otimes T \Rightarrow \left(B^r_{r,i} = B_i\right) \in T^*$$

$$\left(B^k_{lj,i} = X^k_{lj,i} + \gamma^k_{sj} X^s_{l,i} + \gamma^k_{ls} X^s_{j,i} - \gamma^s_{lj} X^k_{s,i} + X^r_{,i} \partial_r \gamma^k_{lj}\right) \in T^* \otimes S_2 T^* \otimes T$$

$$\Rightarrow \left(B^r_{ri,j} - B^r_{rj,i} = F_{ij}\right) \in \wedge^2 T^*$$

We obtain from the relations $\partial_i \gamma^r_{rj} = \partial_j \gamma^r_{ri}$ and the previous results:

$$\begin{aligned} F_{ij} &= B^r_{ri,j} - B^r_{rj,i} = X^r_{ri,j} - X^r_{rj,i} + \gamma^r_{rs} X^s_{i,j} - \gamma^r_{rs} X^s_{j,i} + X^r_{,j} \partial_r \gamma^s_{si} - X^r_{,i} \partial_r \gamma^s_{sj} \\ &= \partial_i X^r_{r,j} - \partial_j X^r_{r,i} + \gamma^r_{rs} \left(X^s_{i,j} - X^s_{j,i}\right) + X^r_{,j} \partial_i \gamma^s_{sr} - X^r_{,i} \partial_j \gamma^s_{sr} \\ &= \partial_i \left(X^r_{r,j} + \gamma^r_{rs} X^s_{,j}\right) - \partial_j \left(X^r_{r,i} + \gamma^r_{rs} X^s_{s,i}\right) \\ &= \partial_i B_j - \partial_j B_i \end{aligned}$$

Now, using the contracted formula $\xi^r_{ri} + \gamma^r_{rs} \xi^s_i + \xi^s \partial_s \gamma^r_{ri} = n A_i$ from section *A*, we obtain:

$$\begin{aligned} B_i &= \left(\partial_i \xi^r_r - \xi^r_{ri}\right) + \gamma^r_{rs} \left(\partial_i \xi^s - \xi^s_i\right) \\ &= \partial_i \xi^r_r + \gamma^r_{rs} \partial_i \xi^s + \xi^s \partial_s \gamma^r_{ri} - n A_i \\ &= \partial_i \left(\xi^r_r + \gamma^r_{rs} \xi^s\right) - n A_i \\ &= n \left(\partial_i A - A_i\right) \end{aligned}$$

and we finally get $F_{ij} = n\left(\partial_j A_i - \partial_i A_j\right)$ *which is no longer depending on A*, a result fully solving the dream of Weyl. Of course, when $n = 4$ and ω is the Minkowski metric, then we have $\gamma = 0$ in actual practice and the previous formulas become particularly simple.

It follows that $dB = F \Leftrightarrow -ndA = F$ in $\wedge^2 T^*$ and thus $dF = 0$, that is $\partial_i F_{jk} + \partial_j F_{ki} + \partial_k F_{ij} = 0$, has an intrinsic meaning in $\wedge^3 T^*$. It is finally important to notice that the usual EM Lagrangian is defined on sections of $\hat{C}_1$ killed by D_2 but *not* on $\hat{C}_2$. Finally, the south west arrow in the left square is the composition:

$$\xi_2 \in \hat{R}_2 \xrightarrow{D_1} X_2 \in T^* \otimes \hat{R}_2 \xrightarrow{\pi^2_1} X_1 \in T^* \otimes \hat{R}_1 \xrightarrow{(\gamma)} \left(B_i\right) \in T^* \Leftrightarrow \xi_2 \in \hat{R}_2 \rightarrow \left(nA_i\right) \in T^*$$

Accordingly, though *A* and *B* are potentials for *F*, then *B* can also be considered as a part of the *field* but the important fact is that the first set of (*linear*) Maxwell equations $dF = 0$ is induced by the (*linear*) operator D_2 because we are only dealing with involutive and thus formally integrable operators, a *fact* justifying the commutativity of the square on the left of the diagram.

□

REMARK 5.B.9: Taking the determinant of each term of the non-linear second order PD equations defining $\hat{\Gamma}$, we obtain successively:

$$det(\omega)\left(det\left(f^k_i(x)\right)\right)^2 = e^{2na(x)} det(\omega) \Rightarrow det\left(f^k_i(x)\right) = e^{na(x)}$$

in such a way that we may define $b\left(f\left(x\right)\right)=a\left(x\right)\Leftrightarrow b\left(y\right)=a\left(g\left(y\right)\right)$ and set $\Theta\left(y\right)=\mathrm{e}^{-b(y)}>0$ over the target when caring only about the connected component $0\to 1\to\infty$ of the dilatation group. The problem is thus to change at the same time the numerical value of the section and /or the nature of the geometric object cosifered, passing therefore from a (metric) tensor to a (metric) tensor density, exactly what also happens with the contact structure when it was necessary to pass from a 1-form to a 1-form density ([4] [7]). In a more specific way, the idea has been to consider successively the two non-linear systems of finite defining Lie equations:

$$\omega_{kl}\left(y\right)y_i^k y_j^l=\omega_{ij}\left(x\right)\to\hat{\omega}_{kl}y_i^k y_j^l\left(det\left(y_i^k\right)\right)^{\frac{-2}{n}}=\hat{\omega}_{ij}\left(x\right)$$

Now, with $\gamma=0$ we have $\chi_{r,i}^r=g_k^s\left(\partial_i f_s^k-A_i^r f_{rs}^k\right)$ and:

$$g_k^s\partial_i f_s^k=\left(1/det\left(f_i^k\right)\right)\partial_i det\left(f_i^k\right)=n\partial_i a,\ g_k^s f_{rs}^k=na_r\left(x\right)$$

Finally, we have the jet compositions and contractions:

$$g_k^r f_i^k=\delta_i^r\Rightarrow g_k^r f_{ij}^k=-g_{kl}^r f_i^k f_j^l$$

$$\Rightarrow na_i\left(x\right)=g_k^s f_{is}^k=-f_i^k f_r^l g_{kl}^r=-n f_i^k\left(x\right)b_k\left(f\left(x\right)\right)$$

It follows that $\alpha_i=n\left(\partial_i a\left(x\right)-A_i^r a_r\left(x\right)\right)$ but we may also set $a_i\left(x\right)=f_i^k\left(x\right)b_k\left(f\left(x\right)\right)$ in order to obtain $\alpha_i=n\left(\frac{\partial b}{\partial y^k}-b_k\right)\partial_i f^k$ as a way to pass from source to target. We have:

PROPOSITION 5.B.10: EM does not depend on the choice between source and target.

Proof: Replacing the groupoid by its inverse in each formula, we may introduce:

$$\alpha=\alpha_i\left(x\right)dx^i,\ \alpha_i=n\left(\partial_i a-A_i^r a_r\right)\Leftrightarrow\beta=\beta_k\left(y\right)dy^k,\ \beta_k=n\left(\frac{\partial b}{\partial y^k}-b_k\right)$$

and compare:

$$x\overset{a}{\to}\left(\alpha,\varphi\right)\Leftrightarrow y\overset{b}{\to}\left(\beta,\psi\right)$$

while setting $\psi_{kl}=\frac{\partial\beta_l}{\partial y^k}-\frac{\partial\beta_k}{\partial y^l}$. We have successively:

$$\begin{aligned}\varphi_{ij}&=\partial_i\alpha_j-\partial_j\alpha_i=-n\left(\partial_i\left(A_j^s a_s\right)-\partial_j\left(A_i^r a_r\right)\right)\\&=-n\left(\partial_i\left(b_l\partial_j f^l\right)-\partial_j\left(b_k\partial_i f^k\right)\right)\\&=-n\left(\frac{\partial b_l}{\partial y^k}-\frac{\partial b_k}{\partial y^l}\right)\partial_i f^k\partial_j f^l\\&=\left(\frac{\partial\beta_l}{\partial y^k}-\frac{\partial\beta_k}{\partial y^l}\right)\partial_i f^k\partial_j f^l\\&=\psi_{kl}\partial_i f^k\partial_j f^l\end{aligned}$$

and we notice that φ *does not depend any longer on a* while ψ *does not de-*

pend any longer on b. Accordingly, we have the equivalences:

$$NO\ EM \Leftrightarrow \varphi = 0 \Leftrightarrow \psi = 0 \Leftrightarrow \frac{\partial b_l}{\partial y^k} - \frac{\partial b_k}{\partial y^l} = 0 \Leftrightarrow \partial_i \left(A_j^r a_r \right) - \partial_j \left(A_i^r a_r \right) = 0$$

□

REMARK 5.B.11: If we use *only* the conformal group, we must use the metric density $\hat{\omega}$ instead of the metric ω. However, if we can define $\hat{\omega}$ from ω by setting $\hat{\omega}_{ij} = \omega_{ij} \Big/ \left(\left| det(\omega) \right| \right)^{\frac{1}{n}}$, we cannot recover ω from $\hat{\omega}$. The way to escape from such a situation is to notice that:

$$\omega \to \mathrm{e}^{2a(x)} \omega \Rightarrow \gamma_{ij}^k \to \gamma_{ij}^k + \delta_i^k \partial_j a(x) + \delta_j^k \partial_i a(x) - \omega_{ij} \omega^{kr} \partial_r a(x)$$
$$\Rightarrow \gamma_{ri}^r \to \gamma_{ri}^r + n \partial_i a(x)$$

a result showing that the conformal symbols $\hat{g}_1$ and $\hat{g}_2$ do not depend on any conformal factor.

REMARK 5.B.12: In fact, our purpose is quite different now though it is also based on the combined use of group theory and the Spencer operator. The idea is to notice that the brothers are *always* dealing with the same group of rigid motions because the lines, surfaces or media they consider are all supposed to be in the same 3-dimensional background/surrounding space which is acted on by the group of rigid motions, namely a group with 6 parameters (3 *translations* + 3 *rotations*). In 1909 it should have been *strictly impossible* for the two brothers to extend their approach to bigger groups, in particular to include the only additional *dilatation* of the Weyl group that will provide the virial theorem and, *a fortiori*, the *elations* of the conformal group considered later on by H. Weyl ([29]). In order to explain the reason for using Lie equations, we provide the explicit form of the n finite elations and their infinitesimal counterpart with $1 \le r, s, t \le n$:

$$\boxed{y = \frac{x - x^2 b}{1 - 2(bx) + b^2 x^2} \Rightarrow \theta_s = -\frac{1}{2} x^2 \partial_s + \omega_{st} x^t x^r \partial_r \Rightarrow \partial_r \theta_s^r = n \omega_{st} x^t, \left[\theta_s, \theta_t \right] = 0}$$

where the underlying metric is used for the scalar products x^2, bx, b^2 involved. It is easy to check that $\xi_2 \in S_2 T^* \otimes T$ defined by $\xi_{ij}^k(x) = \lambda^s(x) \partial_{ij} \theta_s^k(x)$ belongs to $\hat{g}_2$ with $A_i = \omega_{si} \lambda^s$. In view of these local formulas, we understand how important it is to use "*equations*" rather than "*solutions*".

REMARK 5.B.13: Setting $\sigma_{q-1} = \bar{D}' \chi_q \in \wedge^2 T^* \otimes J_{q-1}(T)$, we let the reader prove, as an exercise, that the following so-called *Bianchi identities* hold ([7], p 221):

$$D \sigma_{q-1}(\xi, \eta, \zeta) + \mathcal{C}(\xi, \eta, \zeta) \left\{ \sigma_{q-1}(\xi, \eta), \chi_{q-1}(\zeta) \right\} = 0, \ \forall \xi, \eta, \zeta \in T$$

In the nonlinear conformal framework, it follows that the first set of Maxwell equations has only to do with $\bar{D}'$ in the nonlinear Spencer sequence and thus nothing to do with the Bianchi identities, contrary to what happens with $U(1)$ in classical gauge theory. Similarly, in the linear conformal framework, the first set of Maxwell equations has only to do with D_2 and thus nothing to do with

D_3 in the linear Spencer sequence. Indeed, the EM potential A is a section of $\hat{C}_0$ while the EM field F is a section of $\hat{C}_1$ killed by D_2. This "*shift by one step to the left*" is the most important result of this section and could not be even imagined with any other approach.

6. Conclusions

This paper is part of the achievement of a lifetime research work on the common conformal origin of electromagnetism and gravitation. Roughly speaking, the Cosserat brothers have only been dealing with the 3 translations and 3 rotations of the group of rigid motions of space with 6 parameters while Weyl has only been dealing with the dilatation and the 4 elations of the conformal group of space-time having now $4+6+1+4=15$ parameters ([29]). Among the most striking results obtained from this conformal extension, we successively notice:

- The generating nonlinear *first order* (care) compatibility conditions (CC) for the Cosserat fields are *exactly* described by the *first order* nonlinear second Spencer operator $\bar{D}_2$. Accordingly, there is no conceptual difference between these nonlinear CC and the first set $d:\wedge^2T^*\to\wedge^3T^*:F\to dF=0$ of Maxwell equations where d is the exterior derivative, which is parametrized by $d:T^*\to\wedge^2T^*:A\to dA=F$. However, the classical CC of elasticity are described by the *nonlinear second order* (care) *Riemann* operator existing in the nonlinear Janet sequence but this different canonical nonlinear differential sequence could not explain the existence of field-matter couplings like piezzoelectricity, photoelasticity or even streaming birefringence ([5] [37]). On the contrary, in the conformal approach, it is essential to notice that *the elastic and electromagnetic fields are both specific sections of* $\hat{C}_1=T^*\otimes\hat{R}_2$ *killed by* $\bar{D}_2$ and parametrized by $\bar{D}_1$. They can thus be coupled in a natural way but *cannot be associated to the concept of curvature* described by $\hat{C}_2$. Meanwhile, we insist on the fact that the phenomenological laws of these quoted couplings have been discovered by… Maxwell himself. This *shift by one step to the left*, even in the nonlinear framework, can be considered as the main novelty of this paper.
- The linear Cosserat equations are *exactly* described by the formal adjoint $ad(D_1)$ of the *linear* first Spencer operator $D_1:\hat{C}_0\to\hat{C}_1$ which is a first order operator ([38]). Accordingly, there is no conceptual difference between these equations and the second set $ad(d)$ of Maxwell equations where $d:T^*\to\wedge^2T^*$. This result explains why the *Cosserat* equations are quite different from the *Cauchy* equations which are described by the formal adjoint of the *Killing* operator in the Janet sequence used in classical elasticity, that is $Cauchy=ad(Killing)$ in the language of operators. It follows that the elastic and electromagnetic inductions are both specific sections of $\wedge^4T^*\otimes\hat{C}_1^*\simeq\wedge^3T^*\otimes\hat{R}_2^*$, *independently of any constitutive relation*. The specific use of the 1-dimensional *dilatation subgroup* allows to understand the mathematical origin of *thermoelectricity* and the so-called *virial theorem* through the trace of the Cauchy tensor ([36] [40]).

- Combining the two previous comments, respectively related to "*geometry*" and to "*physics*" according to H. Poincaré ([34]), there is no conceptual difference between the elastic constitutive constants of elasticity and the magnetic constant μ or rather $1/\mu$ of electromagnetism in the case of homogeneous isotropic materials on one side (*space*) or between the mass per unit volume and the dielectric constant ε on the other side (*time*), a result confirmed by the speeds of the various elastic or electromagnetic existing waves ([5] [37]). In general one has $\varepsilon\mu c^2 = n^2$ where n is the index of refraction but in vacuum we have $\varepsilon_0\mu_0 c^2 = 1$ and we have thus only one electromagnetic constant involved in the corresponding Minkowski constitutive law of vacuum ([2]).
- As for gravitation and the possibility to exhibit a conformal factor defined everywhere but at the origin, we may simply say that we needed 25 years in order to correct the result we already obtained in 1994 ([7] [41]). Such a possibility highly depends on the new mathematical tools involved in the construction of the Janet or Spencer nonlinear differential sequences for the conformal group of space-time because, *in this case*, the Spencer δ-cohomology has very specific properties for the dimension $n = 4$ *only*. This will be the subject of a forthcoming companion paper (arXiv: 2007.01710).

We end this paper with the French proverb "AUTRES TEMPS, AUTRES MOEURS" adapted from the famous Latin sentence "AUT TEMPORA, AUT MORES" as we do believe that a modern scientific translation could be "NEW MATHEMATICS, NEW PHYSICS".

Conflicts of Interest

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

References

[1] Einstein, A. (1905) *Annalen der Physik*, **17**, 891-921. https://doi.org/10.1002/andp.19053221004

[2] Ougarov, V. (1969) Théorie de la Relativité Restreinte. MIR, Moscow.

[3] Weyl, H. (1918, 1958) Space, Time, Matter. Springer, Berlin.

[4] Pommaret, J.-F. (1988) Lie Pseudogroups and Mechanics. Gordon and Breach, New York.

[5] Pommaret, J.-F. (2019) *Journal of Modern Physics*, **10**, 1566-1595. https://doi.org/10.4236/jmp.2019.1013104

[6] Gutt, S. (1983) Invariance of Maxwell's Equations. In: Cahen, M., Lemaire, L. and Vanhecke, L., Eds., *Differential Geometry and Mathematical Physics. Mathematical Physics Studies* (*A Supplementary Series to Letters in Mathematical Physics*), Vol. 3, Springer, Dordrecht, 27-29. https://doi.org/10.1007/978-94-009-7022-9_3 https://link.springer.com/chapter/10.1007/978-94-009-7022-9_3

[7] Pommaret, J.-F. (1994) Partial Differential Equations and Group Theory. Wolters Kluwer, Alphen aan den Rijn. https://doi.org/10.1007/978-94-017-2539-2

[8] Appell, P. (1909) Traité de Mécanique Rationnelle. Gauthier-Villars, Paris.

[9] O'Chwolson, O. (1914) Traité de Physique (See III, 2, 537 + III, 3, 994 + V, 209). Hermann, Paris.

[10] Cosserat, E. and Cosserat, F. (1909) Théorie des Corps Déformables. Hermann, Paris.

[11] Koenig, G. (1897) Leçons de Cinématique (The Note "Sur la Cinématique d'un Milieu Continu" by E. Cosserat and F. Cosserat Has Rarely Been Quoted). Hermann, Paris, 391-417.

[12] Pommaret, J.-F. (1997) *Annales des Ponts et Chaussées*, **82**, 59-66. (Translation by D.H. Delphenich)

[13] Pommaret, J.-F. (2001) Partial Differential Control Theory. Kluwer, Dordrecht. https://doi.org/10.1007/978-94-010-0854-9

[14] Teodorescu, P.P. (1975) Dynamics of Linear Elastic Bodies. Abacus Press, Tunbridge Wells. (Editura Academiei, Bucuresti, Romania)

[15] Pommaret, J.-F. (1983) Differential Galois Theory. Gordon and Breach, New York.

[16] Einstein, A. and Fokker, A.D. (1914) *Annalen der Physik*, **44**, 321-328. https://doi.org/10.1002/andp.19143491009

[17] Janet, M. (1920) *Journal de mathématiques pures et appliquées*, **8**, 65-151.

[18] Goldschmidt, H. (1972) *Journal of Differential Geometry*, **6**, 357-373 and **7**, 67-95. https://doi.org/10.4310/jdg/1214430498

[19] Kumpera, A. and Spencer, D.C. (1972) Lie Equations. Annals of Mathematics Studies No. 73, Princeton University Press, Princeton. https://doi.org/10.1515/9781400881734

[20] Spencer, D.C. (1965) *Bulletin of the American Mathematical Society*, **75**, 1-114.

[21] Cartan, E. (1904) *Annales Scientifiques de l'École Normale Supérieure*, **21**, 153-206, **22** (1905) 219-308. https://doi.org/10.24033/asens.538

[22] Cartan, E. (1923) *Annales de la Société Polonaise de Mathématique*, **2**, 171-221.

[23] Vessiot, E. (1903) *Annales Scientifiques de l'École Normale Supérieure*, **20**, 411-451. https://doi.org/10.24033/asens.529

[24] Vessiot, E. (1904) *Annales Scientifiques de l'École Normale Supérieure*, **21**, 9-85. https://doi.org/10.24033/asens.534

[25] Pommaret, J.-F. (1978) Systems of Partial Differential Equations and Lie Pseudogroups. Gordon and Breach, New York. Russian Translation: MIR, Moscow (1983).

[26] Eisenhart, L.P. (1926) Riemannian Geometry. Princeton University Press, Princeton.

[27] Pommaret, J.-F. (2016) Deformation Theory of Algebraic and Geometric Structures. Lambert Academic Publisher (LAP), Saarbrucken. https://arxiv.org/abs/1207.1964

[28] Pommaret, J.-F. (2021) *Advances in Pure Mathematics*, **11**, 835-882. https://doi.org/10.4236/apm.2021.1111056

[29] Pommaret, J.-F. (2021) *Journal of Modern Physics*, **12**, 1822-1842. https://doi.org/10.4236/jmp.2021.1213106

[30] Northcott, D.G. (1966) An Introduction to Homological Algebra. Cambridge University Press, Cambridge.

[31] Rotman, J.J. (1979) An Introduction to Homological Algebra. Academic Press, Cambridge.

[32] Pommaret, J.-F. (2005) Algebraic Analysis of Control Systems Defined by Partial

Differential Equations. In: *Advanced Topics in Control Systems Theory*, Lecture Notes in Control and Information Sciences 311, Springer, Berlin, Chapter 5, 155-223. https://doi.org/10.1007/11334774_5

[33] Pommaret, J.-F. (2015) *Multidimensional Systems and Signal Processing*, **26**, 405-437. https://doi.org/10.1007/s11045-013-0265-0

[34] Poincaré, H. (1901) *Comptes Rendus de l'Académie des Sciences Paris*, **132**, 369-371.

[35] Pommaret, J.-F. (2012) Spencer Operator and Applications: From Continuum Mechanics to Mathematical Physics. In: Gan, Y., Ed., *Continuum Mechanics-Progress in Fundamentals and Engineering Applications*, InTech, Rijeka, 1-32. https://doi.org/10.5772/35607

[36] Pommaret, J.-F. (2018) New Mathematical Methods for Physics. Mathematical Physics Books, Nova Science Publishers, New York, 150 p.

[37] Pommaret, J.-F. (2001) *Acta Mechanica*, **149**, 23-39. https://doi.org/10.1007/BF01261661

[38] Pommaret, J.-F. (2010) *Acta Mechanica*, **215**, 43-55. https://doi.org/10.1007/s00707-010-0292-y

[39] Pommaret, J.-F. (2014) *Journal of Modern Physics*, **5**, 157-170. https://doi.org/10.4236/jmp.2014.55026

[40] Pommaret, J.-F. (2015) From Thermodynamics to Gauge Theory: The Virial Theorem Revisited. In: *Gauge Theories and Differential Geometry*, NOVA Science Publisher, Hauppauge, 1-46.

[41] Pommaret, J.-F. (2013) *Journal of Modern Physics*, **4**, 223-239. https://doi.org/10.4236/jmp.2013.48A022

List of the Main Notations

$\mu=(\mu_1,\cdots,\mu_n)$ multi-index, $|\mu|=\mu_1+\cdots+\mu_n$,
$\mu+1_i=(\mu_1,\cdots,\mu_{i-1},\mu_i+1,\mu_{i+1},\cdots,\mu_n)$.

$\Pi_q=\Pi_q(X,X)\subset J_q(X\times X)$ Lie groupoid of order q over X.

$(x,y_q)=(x,y_\mu^k)$ local coordinates with $0\le|\mu|\le q$ and such that $det(y_i^k)\neq 0$.

$\alpha_q:\Pi_q\to X:(x,y_q)\to x$ source projection, $\beta_q:\Pi_q\to X=Y:(x,y_q)\to y$ target projection.

$f:X\to X\times X:(x)\to(x,f^k(x))$ section of $X\times X$ $(id:(x)\to(x,x))$.

$f_q:X\to\Pi_q:(x)\to(x,f_\mu^k(x))$ section of Π_q with $j_q(f):X\to\Pi_q:(x)\to(x,\partial_\mu f^k(x))$.

$E\to X$ vector bundle over X with section ξ, $J_q(E)$ q-jet bundle of E with section ξ_q over ξ.

$\pi_q^{q+r}:J_{q+r}(E)\to J_q(E):\xi_{q+r}\to\xi_q$ canonical projection.

$D\xi_{q+1}=j_1(\xi_q)-\xi_{q+1}:(\partial_i\xi^k(x)-\xi_i^k(x),\partial_i\xi_j^k(x)-\xi_{ij}^k(x),\cdots)\in T^*\otimes J_q(E)$ linear Spencer operator.

$d:\wedge^rT^*\to\wedge^{r+1}T^*:\alpha\to d\alpha$ exterior derivative with $d\circ d=0$.

$D(\alpha\otimes\xi_{q+1})=d\alpha\otimes\xi_q+(-1)^r\alpha\wedge D\xi_{q+1}\in\wedge^{r+1}T^*\otimes J_q(E),\forall\alpha\in\wedge^rT^*$ extension of D.

$\{\xi_{q+1},\eta_{q+1}\}=j_q([\xi,\eta]),\forall\xi,\eta\in T$ algebraic bracket.

$[\xi_q,\eta_q]=\{\xi_{q+1},\eta_{q+1}\}+i(\xi)D\eta_{q+1}-i(\eta)D\xi_{q+1},\forall\xi_q,\eta_q\in J_q(T)$ differential bracket.

$R_q\subset J_q(T)$ with $[R_q,R_q]\subset R_q$ system of infinitesimal Lie equations or Lie algebroid.

$0\to R_q^0\to R_q\xrightarrow{\pi_0^q}T\to 0$ exact $\Rightarrow$ transitive algebroid.

$\chi_q'\in T^*\otimes R_q$ over $id_T\in T^*\otimes T$ is called a R_q-connection.

$\kappa_q'(\xi,\eta)=[\chi_q'(\xi),\chi_q'(\eta)]-\chi_q'([\xi,\eta])\in R_q^0$ is called the curvature of χ_q'.

$\rho_1(R_q)=J_1(R_q)\cap J_{q+1}(T)\subset J_1(J_q(T))$ first prolongation of R_q.

$\rho_1(R_q)=\{\xi_{q+1}\in J_{q+1}(T)\,|\,\xi_q\in R_q,D\xi_{q+1}\in T^*\otimes R_q\}$ alternative definition.

$R_{q+r}^{(s)}=\pi_{q+r}^{q+r+s}(R_{q+r+s})\subseteq R_{q+r}$ prolongation/projection (PP) procedure.

$L(\xi_{q+1})\eta_q=\{\xi_{q+1},\eta_{q+1}\}+i(\xi)D\eta_{q+1}=[\xi_q,\eta_q]+i(\eta)D\xi_{q+1}$ formal Lie derivative.

$\bar{D}f_{q+1}=f_{q+1}^{-1}\circ j_1(f_q)\in T^*\otimes J_q(T)$ first nonlinear Spencer operator.

$\bar{D}'\chi_q(\xi,\eta)=D\chi_q(\xi,\eta)-\{\chi_q(\xi),\chi_q\eta\}\in J_{q-1}(T)$ second nonlinear Spencer operator.

$\delta:\wedge^rT^*\otimes S_{q+1}T^*\otimes E\to\wedge^{r+1}T^*\otimes S_qT^*\otimes E:\omega\to(\delta\omega)_\mu^k=dx^i\wedge\omega_{\mu+1_i}^k$ Spencer δ-map.

$F_r=\wedge^rT^*\otimes J_q(E)\big/\left(\wedge^rT^*\otimes R_q+\delta(\wedge^{r-1}T^*\otimes S_{q+1}T^*\otimes E)\right)$ Janet bundles.

$C_r=\wedge^rT^*\otimes R_q\big/\delta(\wedge^{r-1}T^*\otimes g_{q+1})$ Spencer bundles.

Journal of Modern Physics, 2022, 13, 495-508
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Coulomb Interaction between Electrons and a New Concept of Atom

Marina V. Krasinkova

Ioffe Institute, Politekhnicheskaya 26, St. Petersburg, Russia
Email: marinak@mail.ioffe.ru

How to cite this paper: Krasinkova, M.V. (2022) Coulomb Interaction between Electrons and a New Concept of Atom. *Journal of Modern Physics*, **13**, 495-508.
https://doi.org/10.4236/jmp.2022.134032

Received: March 23, 2022
Accepted: April 22, 2022
Published: April 25, 2022

Abstract

It is shown that the approximation of a strong Coulomb interaction between electrons results in a new model of the atom with a spatial quantization of electrons accompanied by their quantization in energy. This model implies that electrons rotate in circular orbits centered outside the atomic nucleus and only orbit axes pass through it. The Coulomb interaction between electrons leads to a spherically symmetric distribution of their orbits on the surfaces of equipotential spheres of a spherically symmetric electrostatic field of the nucleus. The distribution is similar to "inscribing" electron orbits into faces of regular nucleus-centered polyhedra so each polyhedron corresponds to a certain electron state (s, p, d, f), and a certain set of polyhedra corresponds to a certain period of the Mendeleev Table. It is shown that a spherically symmetric distribution of electron orbits gives rise to the formation of electron pairs in which electron orbits with a common axis are located symmetrically with respect to the nucleus and the orbital magnetic moments of the electrons are oppositely directed. The physical meaning of the electron spin concept becomes clear. The spin turns out to be related to the orbital magnetic moment of an electron and reflects the fact that two electrons of a pair rotate in opposite directions relative to their common axis. So the spin is one of characteristics of the electron state in the atom associated with electron rotation in the orbit centered outside the nucleus. The atomic model gives an insight into the periodicity of changes in the atomic properties with increasing nuclear charge and the reasons for an electron double energy quantization associated with different states and periods. The model shows that the atomic structure and properties can be explained by using concepts of classical mechanics and classical electrodynamics which regard the electron as a particle.

Keywords

Correlated Electron State, Electron Shell Structure, Electron Energy Quantization, Electron Pairing, Electron Spin, Magnetic Dipole-Dipole

Interaction, Closed System

1. Introduction

The studies of the π electron state in a flat carbon monolayer (graphene) [1] included some assumptions which were helpful for the understanding of this carbon modification. The assumptions concerned, first, the role of Coulomb interaction between neighboring electrons, which resulted in an alternating distribution of electrons on both sides of the carbon frame and the formation of two-dimensional electron crystals from them. Second, it was supposed that the formation of three resonating π bonds by one electron resulted from a p_z electron rotation in the orbit in the plane parallel to the carbon frame plane. The orbit was in the limits of the $2p_z$ state. It is important that the center of the p_z electron orbits in this case was not the carbon atom nucleus, only the rotation axis passed through the nucleus. Third, to compensate for the orbital magnetic momenta of electrons, opposite directions of p_z electron rotation (clockwise and counterclockwise) on opposite sides of the carbon frame were assumed. It turned out that the difference in the directions of electron rotation could replace the difference in the spin directions of the electrons forming π bonds. This result was unexpected and interesting because a physical understanding of the spin concept is still lacking. The spin is interpreted either as a purely quantum phenomenon that has no analogue in classical mechanics or as some inherent magnetic moment of the electron itself, *i.e.*, as its property.

Since all of the above assumptions, which allowed explanation of the origin of graphene specific features concerned the behavior of electrons in the 2 p_z state of carbon atoms, it was supposed that they could be useful for the consideration of the electron state in the atom as well. First of all, the influence of Coulomb interaction between electrons on their spatial distribution and also the relationship between the spin and the orbital magnetic moment of electron attracted attention.

This paper is devoted to the consideration of the atom in the approximation of a strong Coulomb interaction between electrons.

At first, a few words on the history of this research subject should be said. The development of quantum mechanics in the early 20th century sparked a serious scientific debate in the scientific community because it meant a change in the scientific approach and the transition from the consideration of real phenomena within the framework of classical mechanics to the consideration of real phenomena probabilities, *i.e.*, moving from a deterministic approach to a non-materialized one. As a result of these disputes, the view prevailed that quantum mechanics completely solves the atomic problem. However, questions such as the reason for the periodic changes in the atomic properties with an increase in the nuclear charge; the atomic ability to form directional bonds and the me-

chanism of electron transition from the wave state into the particle one remain unresolved. Therefore, attempts to explain atomic properties within the framework of classical mechanics continued and continue [2] [3] [4] [5].

2. Coulomb Interaction between Electrons and Spherically Symmetric Distribution of Electron Orbits

Let us consider the influence of Coulomb interaction between electrons on their state in the atom by using the analogy with its influence on p_z electrons in graphene [1]. Let us assume, first, that electrons in the atom rotate in circular orbits centered outside the nucleus, and only the rotation axis passes through the nucleus. Second, all electron orbits in the same state are located on one equipotential surface of the nucleus electrostatic field. In the case of the atom, this surface is an equipotential sphere. In the presence of Coulomb interaction between electrons their orbits on such a sphere must lie at equal distances from each other. To keep this distance constant as electrons rotate, the electron rotation must be synchronous, with the same velocity and in the same direction relative to the sphere surface. In other words, the electron state on the equipotential surface must be strongly correlated.

It is possible to achieve a uniform distribution of electron orbits on a spherical surface by "inscribing" a regular polyhedron into the sphere. In this case the polyhedron center must be in the atomic nucleus and the number of its faces must be equal to the number of electrons in a given state. In this case each electron orbit will be on a separate face of the polyhedron and each electron state will have its own polyhedron. The discussion below will show that the number of regular polyhedra really specifies the number of electron states in the atom, and the number of faces of regular polyhedra specifies the number of electrons in each state.

Five regular polyhedra are known. However, only those the opposite faces of which are parallel to each other were found to be useful for the electron distribution in the atom. These are a hexahedron, an octahedron, a dodecahedron, and an icosahedron. It is natural to assume that a hexahedron corresponds to the s and p states, a dodecahedron corresponds to the d state and an icosahedron corresponds to the f state.

It turned out that difficulties in the description of the spatial distribution of electron orbits by using polyhedra arise primarily with electrons in the s state. There are only two of them, and the smallest number of faces in regular polyhedra with the face parallelism is six (hexahedron). Therefore, s electrons can fill their hexahedron only partially, occupying only two faces out of six (**Figure 1**).

The atomic structure which consists of polyhedra and also possible variants of polyhedron coalescence when the nuclear charge increases are shown schematically in **Figure 2**. It shows, as an example, completely filled electron shells of noble gases. The sequence of polyhedra for other atoms is the same, only the outer polyhedron remains incompletely filled.

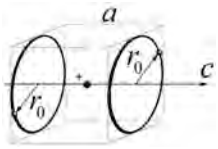


Figure 1. Two 1s electron orbits with radius r_o inscribed into two opposite faces of hexahedron with edge a (He atom).

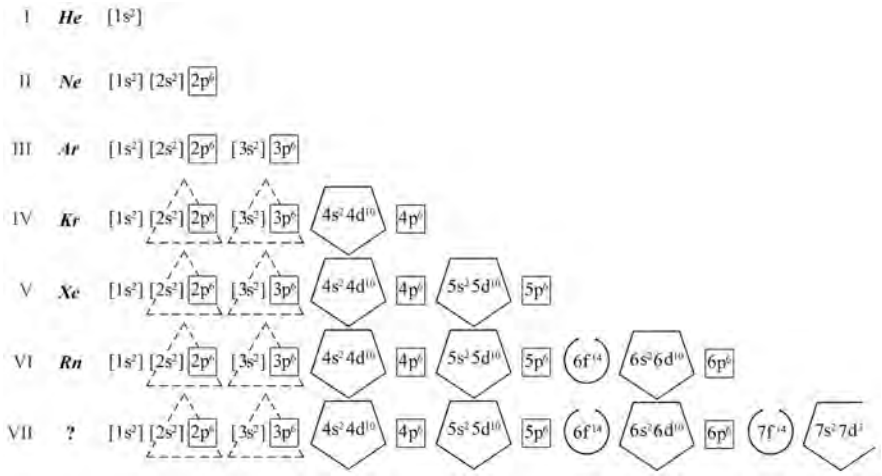


Figure 2. Electron shells of noble gases shown by a sequence of polyhedra: incompletely filled hexahedron of s electrons (square brackets), completely filled hexahedron of p electrons (a square), octahedron (a regular triangle), dodecahedron (a regular pentagon), icosahedron (a circle). Incompletely filled icosahedra are shown by open circles. 7s + 7d dodecahedron is still incomplete. The octahedra formed by coalescence are shown by the dashed line. Replacement of the 3d state by 4d in the fourth period will be discussed in Section 4.

Figure 2 shows that the electron shell of each noble gas atom consists of a certain sequence of alternating polyhedra, each of which corresponds to a definite distribution of electron orbits in the spherically symmetric electrostatic nuclear field. Such a distribution can be regarded as a spatial quantization of electron orbits in the atomic shell.

Since each electron state in the atomic model corresponds to a definite polyhedron, a limited number of states is associated with a limited number of regular polyhedra.

3. On the Fine Structure of Electron Energy Levels

Before proceeding to a detailed discussion of the consequences of spatial quantization of electron orbits, some words should be said about regular polyhedra. It is known [6] that three spheres are associated with each regular polyhedron: a sphere circumscribed around the polyhedron and passing through all its vertices; a sphere inscribed into the polyhedron and passing through the midpoints of its faces; and a median sphere passing through the midpoints of its edges. There are certain relations between the polyhedron edge length and the radii of these three spheres, so if you know the edge length or the radius of one of the spheres, you

can find all the others.

The median sphere is of greatest interest for the consideration of the electron state in the atom because it is the equipotential sphere on the surface of which electron orbits are located and which determines the electron energy. The median sphere also gives important information about the electron orbits themselves. So, the polyhedron face which cuts off a spherical segment from the median sphere determines the orbit plane, its orientation relative to the atomic nucleus, and its maximum possible radius (Figure 3). In regular polyhedra a perpendicular from the polyhedron center to the face passes through the face center. This means that the electron rotation axis must pass through the atomic nucleus because the polyhedron is centered by the nucleus.

It can be seen from Figure 3 that the electron orbit radius $r_o = \rho\sin\vartheta$, *i.e.* it is determined by the radius of the median sphere ρ (which is also the electron radius-vector relative to the nucleus) and the angle ϑ between this radius and the rotation axis. The distance between the orbit center and the nucleus will be defined by the radius r of the sphere that can be inscribed into the same hexahedron.

Angle ϑ for a hexahedron is 45 degrees. For other polyhedra it is smaller and decreases when the number of faces increases. It follows that the largest radius is for the orbit inscribed into the hexahedron face, *i.e.* for the electrons of the s and p states.

Since each electron state corresponds to a specific polyhedron, and the median spheres of these polyhedra are different, it becomes clear that the electron energies in different states of the same period of the Mendeleev Table must also be different. This means that the electron distribution over different polyhedra leads to a quantization of electrons in energy. It is possible that the difference in energies in such a quantization is insignificant as compared with the difference in electron energies between neighboring periods of the Mendeleev Table, but the difference must exist. In other words, each energy level corresponding to any

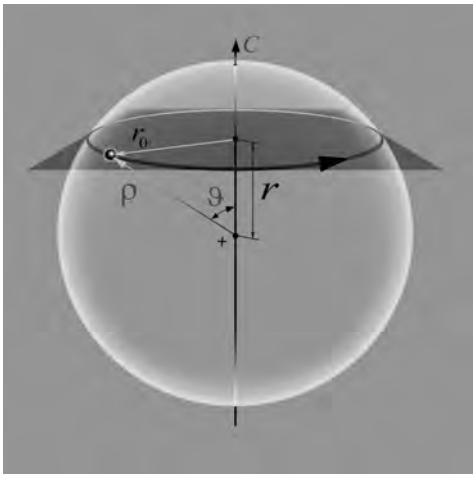


Figure 3. An electron in orbit with radius r_o lying on the surface of the median sphere with radius ρ. The orbit is inscribed into a hexahedron face.

period of the Mendeleev Table must have sublevels corresponding to different electron states. Perhaps this is the reason for the fine structure of atomic spectra observed experimentally [7].

4. Specific Features of Polyhedron Filling, Electric Field Changes with Increasing Nuclear Charge and Quantization of Electrons in Energy Associated with Periods of the Mendeleev Table

A detailed consideration of the sequence of filling of electron states (polyhedra) presented in the general form in **Figure 2** allowed us to identify a number of features in the formation of an atomic electron shell.

First, there is a certain sequence of polyhedra filling which is seen in **Figure 2**. However, it is important that the set of polyhedra specific for each period of the Mendeleev Table is not just a set, but a set of interconnected polyhedra, because starting from the hexahedron of s electrons each subsequent polyhedron is inscribed into the median sphere of the previous one. As a result, each set of polyhedra is a sequence of polyhedra consistently inscribed into each other. This way of formation of the set brings the median spheres, on which the electron orbits of different states are located, closer. Thus, the difference in energies between different states in each set is reduced.

Second, it seems that a weak Coulomb interaction between two s electrons makes their state unstable. Therefore, the s electrons either combine with p electrons under certain conditions to form an octahedron (e.g.: 2s and 3s electrons) or help d electrons to fill their dodecahedron (4s, 5s, 6s, 7s electrons). The only exception is the 1s state, which is closest to the nucleus, and perhaps it remains unchanged in all atoms for this reason. However, despite its instability, the s state is important for the atomic structure formation, because it is the first in each period of the Mendeleev Table and its equipotential sphere determines the energy of different electron states in the period. (A dodecahedron and then the other polyhedra are inscribed consistently into the median sphere of its hexahedron). In addition, the hexahedron of s electrons which becomes empty after the transition of s electrons into the d state is later filled with p electron orbits. Therefore, the hexahedron of p electrons is located at a larger distance from the nucleus and encloses each period of the Mendeleev Table.

Third, as long as any state, *i.e.*, the polyhedron, is filled, the energy of the electrons filling it remains the same for all electrons despite the increase in the nuclear charge during this process. This electron energy equality is apparently maintained by compensation of a stronger electron attraction to the nucleus by an increasing Coulomb repulsion from the increasing number of electrons in the polyhedron which they fill. This is possible only under conditions of a spherically symmetric arrangement of electrons, when each electron is repelled from all other electrons in the same state.

Fourth, a periodic appearance of noble gas atoms at the end of each period of the Mendeleev Table is accompanied by a sharp change in the electric field

strength at the atomic surface. The inertness of noble gases means the absence of electric field at their surfaces, *i.e.*, a complete shielding of the nuclear charge by the electron shell (by a set of polyhedra ending with a hexahedron of p-electrons). As the nuclear charge increases further, the electric field at the atomic surface appears again, but its strength is no longer determined by the point nuclear charge. It is determined by the charge of a charged sphere with the radius equal to the radius of the noble gas atom preceding the given period of the Mendeleev Table. The nucleus remains a point charge only for the electrons of periods I and II. Therefore, for other periods the field strength at the atomic surface is determined by not only the nuclear charge, but also by the charged sphere radius: $E = K \cdot Ze/(R_a + x)^2$, where K is a known $1/4\pi\varepsilon_o$ constant, Ze is the nuclear charge, R_a is the noble gas atom radius in the preceding period of the Mendeleev Table, and x is the distance to the charged sphere surface.

Finally, in addition to a certain sequence of filling, the filling of each next polyhedron begins only after a preceding polyhedron is filled. Deviations from this rule are observed only in atoms with f electrons. Perhaps that is why these atoms have a high chemical activity, which is also true for gas Rn, which occupies the position of a noble gas in the Mendeleev Table.

In Section 3 we discussed the energy quantization of electrons associated with the distribution of electron orbits over different polyhedra. A small difference in the energies of different states was assumed. The sequence of polyhedra filling considered above confirmed that the difference in energies of the states should really be low, since the polyhedra are consistently inscribed into the median sphere of the previous polyhedron. Let us call this quantization a small-scale quantization in comparison with the quantization associated with different periods, the nature of which becomes obvious from the consideration of the electron shell structure of the atom, which consists of sets of polyhedra.

Indeed, the Mendeleev Table shows that each new period begins after the nuclear charge changes by a strictly defined number of units: 8, 8, 18, 18, 32, 32. This means that the filling of each new set of polyhedra begins at the field strength quite different from the previous one. It has been shown above that the field strength depends on the nuclear charge and on the charged sphere radius. Since both the charged sphere radius and the field strength at which each new set of polyhedra begins to be filled are discrete, there must be discrete energies which the electrons can have. It is obvious that energy quantization of electrons associated with different periods of the Mendeleev Table (different sets of polyhedra) takes place already with a much larger difference in energies than quantization associated with different electron states in each period. This is a large-scale quantization.

Thus, the Coulomb interaction between electrons, which distributes electrons in a certain manner in the space around the nucleus (**Figure 4**), leads to two different scales of their energy quantization, *i.e.*, quantization associated with state and period.

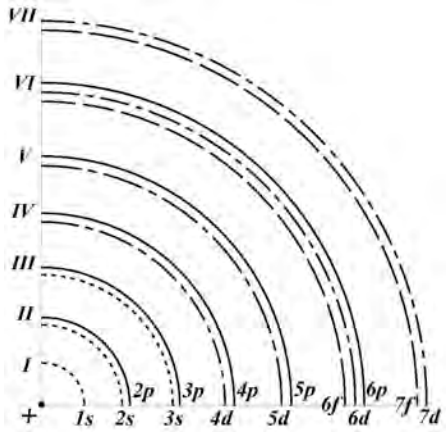


Figure 4. Sequence of equipotential spheres on which electron orbits are in an atomic shell (after the s electron transition into d states). Atomic shell compression when the nuclear charge is increased is ignored.

Note, that 3d electrons in Figure 4 are absent because the model does not imply electron transitions between the polyhedron sets with different energies in the cases when the electrons are not affected by an external factor (such as an external strong electric field, for instance). Therefore, the designation 3d for d electrons in period IV of the Mendeleev Table is not entirely correct. They, apparently, should be 4d electrons. This also applies to d electrons of other periods and also to f electrons.

5. Specific Features of the Electron State in a Circular Orbit Centered Outside a Nucleus

Let us consider, as an example, the forces acting on an electron in a circular orbit in the spherically symmetric field of the nucleus by using a He atom in which, as already noted, two electron orbits are on opposite sides of the hexahedron (Figure 1) Figure 5.

One can see from Figure 5 that two forces act on the electron I: the force of attraction to the nucleus $\boldsymbol{F}_a$ and the force of repulsion between electrons (I and II) $\boldsymbol{F}_r$. The total force $\boldsymbol{F}_\Sigma$ equal to $\boldsymbol{F}_\Sigma = \boldsymbol{F}_a - \boldsymbol{F}_r$ is directed at angle ϑ to the electron rotation axis.

Let us decompose $\boldsymbol{F}_\Sigma$ into two components, *i.e.*, $\boldsymbol{F}_\perp$ directed perpendicularly to the orbit plane and $\boldsymbol{F}_c$ lying in the orbit plane and directed towards its center. The momentum of force $\boldsymbol{T}$ created by force $\boldsymbol{F}_\perp$ and defined as the vector product $\boldsymbol{T} = [\boldsymbol{r}_o\ \boldsymbol{F}_\perp]$, where $\boldsymbol{r}_o$ is the electron radius-vector relative to the orbit center, must be perpendicular to $\boldsymbol{r}_o$ and $\boldsymbol{F}_\perp$ and directed tangentially to the orbit in the direction from the reader. This direction of the force momentum $\boldsymbol{T}$ determines the direction of the electron linear velocity $\boldsymbol{U}$ in the orbit and also the direction of the electron angular velocity $\boldsymbol{\omega}$ directed always along the rotation axis and away from the nucleus in this case. Normally, the presence of a force momentum $\boldsymbol{T}$ entails a change in angular velocity $\boldsymbol{\omega}$ and hence impulse moment $\boldsymbol{L}$ [8]. However, if a stationary rotation axis along which the angular velocity is directed is

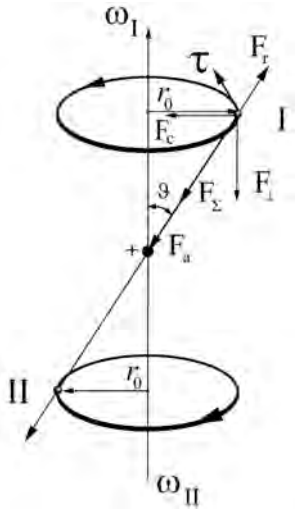


Figure 5. Forces acting on an electron in a circular orbit in the nuclear field in the He atom and resolution of the total force into two components.

chosen as the origin (and this case is of interest to us) the projection of force moment $\boldsymbol{T}$ on this axis is zero, since $\boldsymbol{T}$ lies in the orbit plane perpendicular to the axis. This means that the force moment $\boldsymbol{T}$ does not change the angular velocity $\boldsymbol{\omega}$ in this case. Thus, $\boldsymbol{\omega}$ = Const and $\boldsymbol{L}$ = Const.

As known, this is the case when the force acting on the particle perpendicular to the trajectory of its motion does not do the work [8].

The second component $\boldsymbol{F}_c$ of the total force $\boldsymbol{F}_\Sigma$ acting on the electron is the centripetal force, which provides a centripetal acceleration $\boldsymbol{a}_c$, depending on the angular velocity and electron radius-vector relative to the orbit center, *i.e.* $\boldsymbol{a}_c = -\omega^2 \boldsymbol{r}_o$. This centripetal acceleration changes only the direction of electron linear velocity in the orbit (or the position of the radius-vector $\boldsymbol{r}_o$ on the orbital plane).

Thus, the total force $\boldsymbol{F}_\Sigma$ acting on the electron determines its rotation in a circular orbit in a certain direction with a constant angular velocity $\boldsymbol{\omega}$, linear velocity $\boldsymbol{U}$ equal to the vector product $\boldsymbol{U} = [\boldsymbol{r}_o \boldsymbol{\omega}]$, centripetal acceleration $\boldsymbol{a}_c = -\omega^2 \boldsymbol{r}_o$ and constant kinetic energy $E_k = 1/2\ m r_o^2 \omega^2$.

Note that the Coulomb repulsion force in the He atom is determined by the interaction between two electrons alone. In the states where there are more electrons and where each electron under the condition of the spherically symmetric distribution is repelled simultaneously from all electrons in the same state, the contribution of the repulsive force to the total force is greater. Therefore, the Coulomb interaction between electrons in the atom can be considered to be strong.

Note also that it is the Coulomb repulsion force between electrons that ensures a stable position of the electron in the circular orbit centered outside a nucleus. The lack of such a force in hydrogen, apparently, explains the instability of its electron in the circular orbit and the impossibility of existence of hydrogen in the atomic state. It is known that hydrogen exists only in the form of a molecule H_2 or a positively charged ion in compounds with other elements. The dis-

sociation constant of the hydrogen molecule is only 2.56 * 10^{-34} at 300 K and 1.22 * 10^{-3} at 2000 K, dissociation energy is large 436 kJ/mole [9].Therefore, consideration of the atom structure should begin with the He atom rather than hydrogen.

6. Electron Pairs and Physical Meaning of the Spin Concept

A spherically symmetric distribution of electron orbits on faces of regular polyhedra which are characterized by opposite face parallelism leads to the electron pair formation in any state, *i.e.*, electron pairing. The electrons of each pair are arranged symmetrically with respect to the nucleus. This means that the electrons of the pair have a common rotation axis passing through the nucleus, but they are on opposite sides of the nucleus (**Figure 5**). Thus, the nucleus is the nodal point of any electron state in the atom. The electrons in such pairs have opposite directions of impulse momenta and angular velocities, and also of orbital magnetic momenta. This is not surprising, since the vectors of impulse moment and orbital magnetic moment are not true vectors. They are only pseudo-vectors which reverse their direction under mirror reflection [8]. Thus, the feature manifested itself in conditions of a spherically symmetric distribution of electrons in the atom.

Opposite directions of the electron impulse and orbital magnetic momenta in a pair mean that these electrons rotate in opposite directions relative to their common axis, while they rotate in the same direction relative to the surface of the equipotential sphere on which their orbits are located. This direction (relative to the surface) is the counterclockwise one always and is governed by the direction of the force moment acting on the electron in the orbit. In an atom, therefore, the impulse momenta of all the electrons are directed away from the nucleus, and their orbital magnetic momenta are directed towards the nucleus.

The fact that electrons in a pair have opposite directions of the orbital magnetic moment changes the concept of electron spin [10] [11] [12]. The spin turns out to be related to the orbital magnetic moment and to its direction. It essentially reflects the fact that electrons of each pair having the same axis passing through the nucleus are rotating in opposite directions relative to their axis. Since only two directions (clockwise and counterclockwise) are possible in the orbital plane, the spin has only two directions (spin up, spin down). Thus, spin is neither a property peculiar to the electron itself nor a purely quantum phenomenon. It is only one of the characteristics of electron state in the nuclear spherically symmetric field.

It follows from the above that we can still speak about the electron spin only when the electron is in one of the atomic states. Outside the atom, the concept of spin loses its meaning and only the orbital magnetic moment of the rotating electron remains.

It should be noted in this connection that the O. Stern and W. Gerlach experiment [13] which was explained by using the spin concept was successful only

because the investigation was carried out with individual Ag atoms containing one unpaired electron, *i.e.*, one electron of the pair. In this case the electrons remained in the atomic 5s state and could rotate clockwise in one atom and counterclockwise in another (relative to their rotation axis in each atom and relative to the direction of the outer magnetic field which deflected the Ag atoms).

7. Magnetic Dipole-Dipole Interaction in the Atom

Let us consider another kind of interaction between electrons in the atom connected with the electron rotation in circular orbits. In Section 6 we mentioned the electron orbital magnetic moment which depends, as known, on the current created by the electron rotation and the orbital area covered. Such a rotating electron can be regarded as a small magnetic dipole. This means that a magnetic dipole-dipole interaction must exist between neighboring electrons in the atom. Since the dipole-dipole interaction strength depends on the cosine of the angle between the dipole directions, it is maximal when the neighboring dipole directions are parallel (repulsion) or antiparallel (attraction) to each other, and is zero when their directions are perpendicular.

It follows from the geometry of polyhedra that the angle between neighboring dipole directions in the atom is 2ϑ, where ϑ is the angle between the electron radius-vector $\boldsymbol{\rho}$ relative to the nucleus and orbital axis (**Figure 3**). Since both the angle ϑ and radius r_0 are different for electrons of different states, the dipole-dipole interaction strength will also be different. For example, if the angle between the directions of neighboring dipoles for p electrons is 90˚, there will be no magnetic dipole-dipole interaction between them. This fact distinguishes the p state from all other electron states in the atom, since the magnetic dipole-dipole interaction between neighboring electrons in the s, d and f states must exist.

The lower the angle ϑ or the shorter the distance between the dipoles, the stronger the magnetic interaction.

Since all orbital magnetic momenta (magnetic dipoles) in the atom are directed towards the nucleus, the dipole-dipole interaction between them is repulsive, *i.e.*, it acts on the electrons just as the Coulomb repulsion does.

Despite the fact that the magnetic dipole-dipole interaction between electrons is much weaker than the Coulomb interaction, it may affect the atomic properties and, above all, its chemical activity. This is evidenced by the absence of chemical activity of noble gases where the magnetic dipole-dipole interaction between neighboring dipoles is absent.

Since each electron in the atom that creates its magnetic field is also influenced by magnetic fields of other electrons of the same state, we may conclude that the electron in the atom is in a spherically symmetric electromagnetic field which, like the electric field, varies periodically along the atomic radius. Therefore, the resulting force acting on the electron in an atom and including both electric and magnetic components should be regarded as the Lorentz force.

8. Noble Gas Atoms as Examples of Closed Systems

As it is known, atoms of noble gases differ from other atoms in that their outer electron shell is a completely filled p shell. When all states are completely filled, all p electrons are paired. In the case of a spherically symmetric distribution of electron orbits this means that the vector sums of angular momenta of all electrons, as well as their orbital magnetic momenta, are zero. The sum of all charges is also zero. In addition, there is no magnetic dipole-dipole interaction between neighboring electrons in the outer shell. At the same time there are strong interactions in the atomic system itself. However, they do not manifest themselves outside the system, remaining inside it. Such systems are typically referred to as closed systems which interact neither with other external systems nor among themselves [14]. Therefore, the atoms in question exist in the gaseous phase and as single-atomic gases.

It should be emphasized once more that the specific feature of the atomic structure is a periodic repetition (with increasing nuclear charge and number of electrons) of the atomic state as a closed system, accompanied by periodic changes in the electric and magnetic fields in the atomic shell and periodic changes in the atomic properties. All this is due to a periodic spatial electron distribution in the electron shell (**Figure 4**) or, speaking more generally, due to a limited number of regular polyhedra.

9. Conclusions

It has been shown that the electron shell of the atom is spatially structured and an important role in this is played by the Coulomb interaction between electrons. First, it ensures a stable rotation of electrons in the orbits centered outside the nucleus and, second, it distributes these orbits on the equipotential surfaces in a spherically symmetrical manner. This distribution is similar to "inscribing" the orbits into faces of regular polyhedra centered by the atomic nucleus. In other words, in each state the electrons are distributed so that planes of electron orbits form a regular polyhedron by crossing each other. Therefore, the number of possible electron states in an atom corresponds to the number of regular polyhedra, excluding the tetrahedron, and the number of electrons in each state is equal to the number of faces of the polyhedron corresponding to the given state, excluding the s and f states.

It has been shown that the description of the electron orbit distribution using regular polyhedra permits determination of the equipotential surfaces on which the orbits are located, the orientation of their planes and rotation axes relative to the nucleus, and also their maximum possible radius in each state.

It has been shown that a spherically symmetric distribution of electron orbits is also associated with specific features of electron states in the atom, in particular, the formation of electron pair having a common rotation axis passing through the nucleus, but with oppositely directed impulse and orbital magnetic momenta. This makes the concept of spin questionable. The spin appears to be

related to the electron orbital magnetic moment or rather to the direction of the orbital rotation of the electrons in pair relative to its rotation axis passing through the nucleus. Thus, the spin seems to be one of characteristics of the electron state in the spherically symmetric field of the atom rather than a property inherent in the electron itself or a purely quantum phenomenon.

It has also been shown that the spherically symmetric distribution of electron orbits, similar to the distribution of faces in regular polyhedra, gives an insight into basic properties of the atomic system as a whole, *i.e.*, a periodic change in the properties of atoms with increasing nuclear charge, a periodic formation of closed atomic systems and also the origin of the double quantization of electrons in energy.

To summarize, the electron shell structure of the atom and its properties are determined by the Coulomb interaction between electrons under conditions of a spherical symmetry of the nuclear field. Under these conditions, electrons behave as particles obeying the laws of classical mechanics and classical electrodynamics.

Conflicts of Interest

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

References

[1] Krasinkova, M.V. (2016) *MPM e-Journal*, **29**, 93-100.

[2] Smulsky, J. (2014) *Open Access Library Journal*, **1**, 1-23. https://doi.org/10.4236/oalib.1100773

[3] Vlasov, A.D. (1993) Classical Trend in Quantum Mechanics. Moscow Radiotechnical Institute RAN, Moscow.

[4] Gryzinski, M. (1987) *International Journal of Theoretical Physics*, **26**, 967-980. https://doi.org/10.1007/BF00670821

[5] Gryzinski, M. (1965) *Physical Review Journals Archive*, **138**, 336-358. https://doi.org/10.1103/PhysRev.138.A336

[6] Wenninger, M. (1971) Polyhedron Models. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511569746

[7] Yelyashevich, M.A. (1962) Atomic and Molecular Spectroscopy. Physmathgiz, Moscow.

[8] Gianconi, D.C. (1995) Physics for Scientists and Engineers with Modern Physics. 2nd Edition, Prentice Hall College Div.

[9] Chemical Encyclopedia (1991) Nauka, Soviet Encyclopedia, Moscow.

[10] Uhlenbeck, G.E. and Goudsmit S. (1925) *Die Naturwissenschaften*, **13**, 953-954. https://doi.org/10.1007/BF01558878

[11] Thomas, L.H. (1926) *Nature*, **117**, 514. https://doi.org/10.1038/117514a0

[12] Dirac, P.A.M. (1928) The Quantum Theory of the Electron. *Proceedings of the Royal Society of London. Series A*, **117**, 610-624. https://doi.org/10.1098/rspa.1928.0023

[13] Gerlach W. and Stern, O. (1922) *Zeitschrift für Physik*, **9**, 349-352. https://doi.org/10.1007/BF01326983

[14] Landau L.D. and Lifshitz, E.M. (1993) Mechanics. 3rd Edition, Nauka, Moscow.

Journal of Modern Physics, 2022, 13, 509-517
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

The Dispersion Effect in Photonic Crystal Fibers

Mingxiang Gui[1], Jing Huang[2]

[1]School of Information Science and Engineering, Hunan University, Changsha, China
[2]Physics Department, South China University of Technology, Guangzhou, China
Email: huanggesheng@tom.com

How to cite this paper: Gui, M.X. and Huang, J. (2022) The Dispersion Effect in Photonic Crystal Fibers. *Journal of Modern Physics*, **13**, 509-517.
https://doi.org/10.4236/jmp.2022.134033

Received: March 3, 2022
Accepted: April 23, 2022
Published: April 26, 2022

Abstract

Based on the Harmonitor theory of field in PCFs in which the second order differential to the transmission distance is included, and by the Darboux solution, the dispersion effect on the field is re-discussed. The results can be used in the dispersion parameter design of photonics crystal fibers. The field will expand and split faster than that in common fibers so that (as the description of Harmonitor theory) the higher-order dispersion should be taken into account. The high-order dispersion can also induce pulse compression while its pre-order dispersion values are zeroes. The dispersion coefficient changes with distance.

Keywords

Harmonitor Theory, Darboux Solution, Dispersion

1. Introduction

With the development of nanophotonic technologies, it is available to engineer waveguides with the complex dispersion profiles. This is not adequately described by a simple second-order term in the expansion of the wave number with frequency [1]. Similarly, the nonlinear Schrodinger equation cannot describe fast modulations adequately because it is based on a Taylor expansion in the frequency domain and only several dispersion items are taken into account [2]. Therefore, the influence of the modulation of optical fiber dispersion on soliton propagation has attracted a lot of attentions recently.

More dispersion items have been included [2]. In [3], authors studied the effects of the entire modal dispersion curve and the frequency dependence of the nonlinear coefficients on the formation of modulation instabilities by using the harmonic analysis. New regions of modulation for the dispersion-flattened fibers were found and characterized by this approach [4]. In the strongly dispersion fi-

ber, the chirped Gaussian pulse, as in a linear superposition of the Hermite-Gaussian harmonics and as the zeroth harmonic, presented a periodically varying modulation width [5]. The multipeak modulation instability spectrum was observed in the dispersion oscillating optical fibers [6].

In the fiber with a periodic dispersion, by the variation approach, the stochastic decay of pulses was predicted [7]. Also, the finite-energy Airy waves may endlessly retain their shapes and follow the pre-engineered temporal trajectories [8].

With an appropriately designed dispersion profile, the chirped pulses can be retrieved through the chirp reversal [9] [10]. In this case, of the nonlinear Schrödinger equation with designed group velocity dispersion, variable nonlinearity, and gain or loss, had been demonstrated that the chirp reversal is crucial for the pulse reproduction [9]. A wave suffered the strong modulation of dispersion and then it resulted in a significant chirp development. The periodic modulation of dispersion led to the radiation damping [11] [12], the existence of vibrating solitons [13] and splitting of solitons [14].

Pulse evolution is highly sensitive to the input pulse profile [15]. It was found that light pulses with Gaussian input profile produced less radiation in the fiber system than that hyperbolic-secant or raised-cosine pulses could do.

In this paper, based on our established theory on the field in PCFs[16] which contained both the second order differential to the transmission distance ($\frac{\partial^2 A}{\partial z^2}$) and the higher-order dispersion items, we will re-discuss the dispersion effect on pulse. These discussions include the various pulses with profiles of Gauss, chirped Gauss, super-Gauss and hyperbolic functions (the common features of pulse profile). It also presents a Darboux state description for the field in a dispersion medium [17].

2. Theory on the Dispersion Effect in PCFs

In Micro-fibers (the PCFs in **Figure 1**), the transmission equation of field should include the second-order differential to transmission distance. It is [16]:

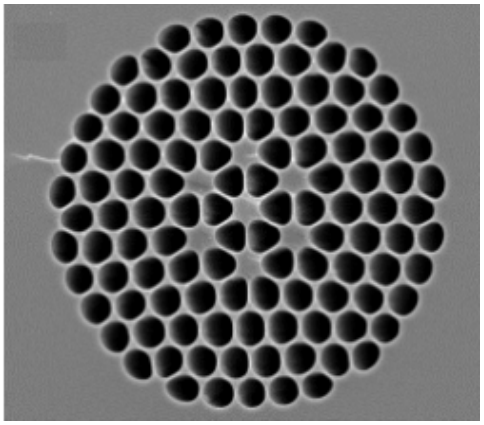


Figure 1. Scanning electron microscope image of the cross-section of the PCFs [18].

$$\frac{\partial^2 A}{\partial z^2} + 2i\beta\frac{\partial A}{\partial z} = -\left(\Delta\beta^2 - 2i\beta\Delta\beta\right)A \tag{1}$$

Generally, for the discussion of field in fibers, people adopt the nonlinear Schordinger equation which ignores the item $\frac{\partial^2 A}{\partial z^2}$ because of the suppose of the slow-varying envelope. But in PCFs, the envelope of field varies very fast. Thus, the field equation should be (1).

And the dispersion item β should include many orders of its Taylor expansion or be the entire β coefficient curve.

Then only taking the dispersion effect into account ($\Delta\beta = 0$), we get:

$$\frac{\partial^2 A}{\partial z^2} + 2i\beta\frac{\partial A}{\partial z} = 0 \tag{2}$$

The field is:

$$\begin{aligned} A(z,\omega) = A(0,\omega)\exp\left[-2iz\left(\beta_0 + \beta_1\omega + \beta_2/2\,\omega^2 + \cdots\right)\right] \\ \Big/\left[-2i\left(\beta_0 + \beta_1\omega + \beta_2/2\,\omega^2 + \cdots\right)\right] \end{aligned} \tag{3}$$

From the view of the quantum theory, in the phase-amplitude representation, the system's Hamiltonian operator is [17]:

$$\hat{H}_S = \omega a'^{+} a' - i\beta\tanh(\beta\beta) \tag{4}$$

In [18], we have derived the (4). In the Darboux space, the lax pair is:

$$\beta^2 = \beta_0^2 + 2\frac{\partial^2(\ln A)}{\partial z^2} \tag{5}$$

$$A = \frac{\partial A(\beta^2)}{\partial z} - \frac{\partial A(\beta_0)/\partial z}{A(\beta_0)}A(\beta) \tag{6}$$

The simulation results based on (6) are the same as those on (3).

3. Validation

In this section, we will compare the simulations of this theory with those of [19] [20] [21] [22] to prove the correction of our theory. The simulation procedure is shown in **Figure 2**.

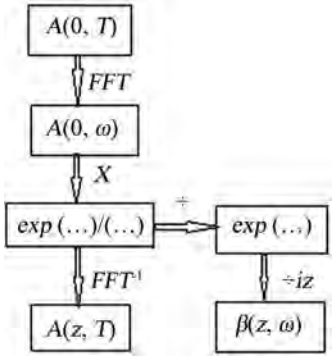


Figure 2. The simulation procedure.

For the Gauss pulse, there is:

$$A(0,T)=\exp\left[-(1+C)T^2/2/T_0^2\right] \tag{7}$$

$$A(0,\omega)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}A(0,T)\mathrm{e}^{-i\omega T}\mathrm{d}T \tag{8}$$

$$\begin{aligned}A(z,T)=\int_{-\infty}^{+\infty}\mathrm{d}\omega\Big\{&A(0,\omega)\exp\left[-2iz\left(\beta_0+\beta_1\omega+\beta_2/2\omega^2+\cdots\right)\right]\\&\Big/\left[-2zi\left(\beta_0+\beta_1\omega+\beta_2/2\omega^2+\cdots\right)\right]\exp\left[i\omega T\right]\Big\}\end{aligned} \tag{9}$$

We simulate the Gauss pulse evolution in **Figure 3** to prove the corrections of (3) and (6).

Although the form of (3) is completely different from the equations in [19] [20], the pulse evolution is the same in the following points:

1) The time axis values minus $z\beta_1$. It is $T=t-z\beta_1$. The whole pulse envelope is moved at the velocity of $1/\beta_1$;

2) The second order dispersion β_2 causes the pulse broaden;

3) The compressed chirped pulse (when $C=0.5$ and $Z=0.1L_D$, L_D is the dispersion length [23]) is shown in subfigure (c).

And differences as follow:

1) The second order dispersion β_2 causes the symmetrical fluctuations at the pulse edges (high frequency brims);

2) When the chirped pulse is compressed, as subfigure (c), the edge fluctuations turn severer;

3) The item $2zi\left(\beta_0+\beta_1\omega+\beta_2/2\omega^2+\cdots\right)$ in phase will cause the fluctuations in the envelopes of pulse.

These differences occur in our simulation, are not apparently shown in [19] [20].

Figure 4 is the demonstration of hyperbolic Secant pulse,

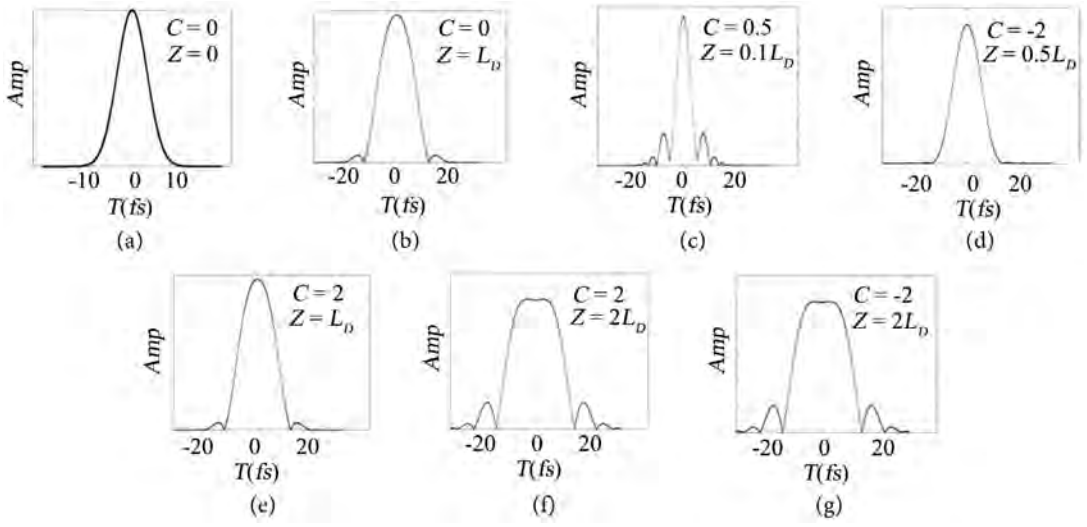


Figure 3. The evolution of Gauss pulse, only including the zero-, first- and the second-order dispersion items. $P_P=1(\mathrm{mW})$, $L=70(\mathrm{cm})$, $T_0=300(\mathrm{fs})$, $\gamma=0.1(\mathrm{W/m})$, $\beta_2=-4.22(\mathrm{ps^2/km})$, $\beta_3=9.9\times10^{-2}(\mathrm{ps^3/km})$, $\lambda=780(\mathrm{nm})$.

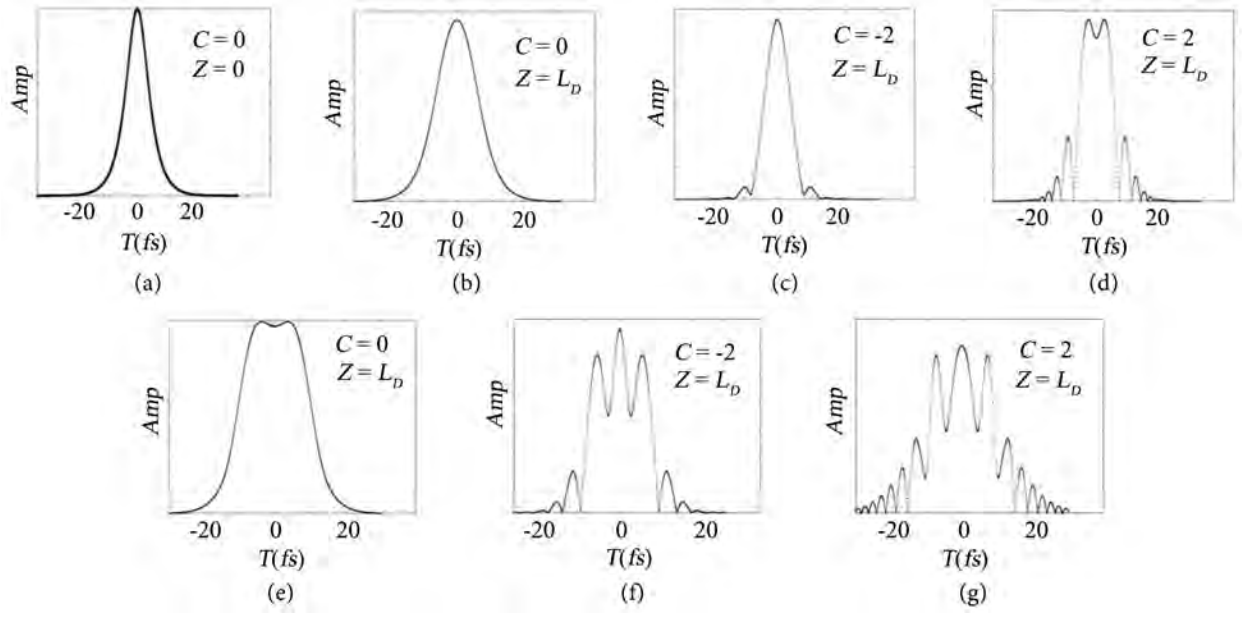


Figure 4. The evolution of the hyperbolic Secant pulse. Parameters are the same as **Figure 3**. (a) The initial pulse; (b)-(d) The results of [19], the fields in common fibers; and (e)-(f) The results of (3), the fields in PCFs.

$$A(0,T)=\operatorname{sech}\left(\frac{T}{T_0}\right)\exp\left(-\frac{iCT^2}{2T_0}\right).$$

By a transmission, this pulse is broadened due to the second-order dispersion according to the Formula (3). The fluctuations are more evident than those in **Figure 3**, and even the pulse splittings occur. The simulations based on our theory exhibit more severe pulse expandings and splittings (e)-(g) than those in [19] [21]. This is consistent with the experiment result that in PCFs [2], and we have observed strong super-continuum spectrum which is not induced in the common fibers.

The same results are obtained from **Figure 5** for the super-Gauss pulse. $A(0,T)=\exp\left[-\frac{1+iC}{2}\left(\frac{T}{T}\right)^{2m}\right]$. There are fluctuations in the pulse edges and the pulses will split with increase of the distance [19] [22]. These phenomena are more evident than those occurred in a Gauss-pulse.

4. Expanding Results by the Darboux State Description

4.1. The Compression of Pulse

We can get a very interesting result that from (3). The Harmonitor is reproduced at some situations and at some frequency components. Actually, it is very difficult to observe this state (reproduced pulse) for the pulse expands very fast. This can be realized by limiting the second-order dispersion (regarded as 0), while the third-order dispersion is taken into account. Then the symmetry is broken. We simulate this case in **Figure 6**. In the figure, the forward of "C" pulse is compressed and the backward keep unchanged.

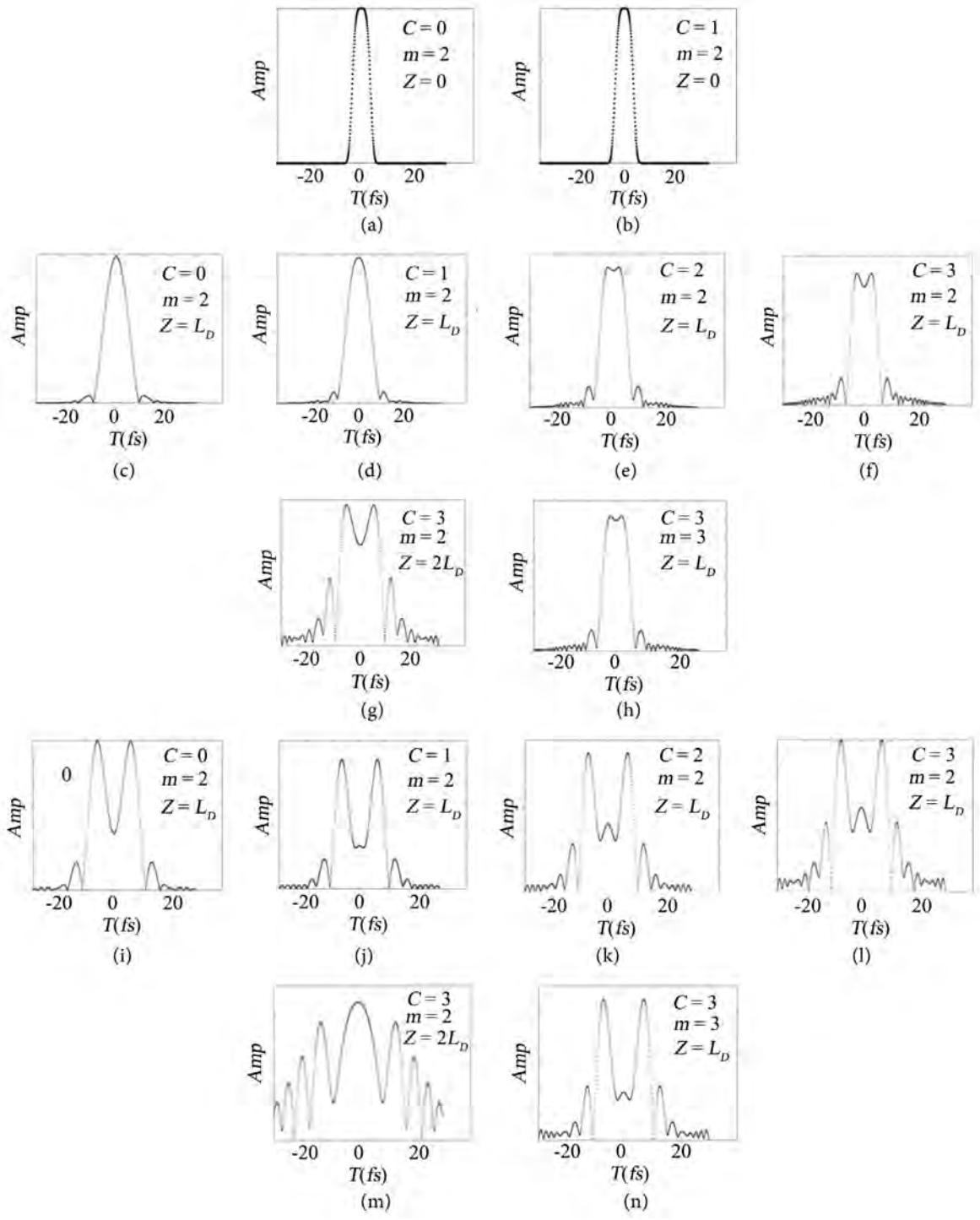


Figure 5. The evolution of super-Gauss pulse. (a) and (b) The initial pulses; (c)-(h) The results of [19] [22], the fields in common fibers; and (i)-(n) The results of (3), the fields in PCFs. Parameters are the same as **Figure 3**.

4.2. The Dispersion Coefficient $\beta(\omega)$

In **Figure 7**, we plot the entire dispersion coefficient. It complete differs from the Fourier serious results [19]. At the different places, it has different values, and the discrepancies are apparent. These exhibit another property of PCFs [4].

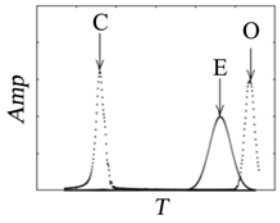


Figure 6. The pulse compression by the β_3. "O", "C" and "E" refer to the origin pulse, the compressed pulse, and the expanded pulse respectively.

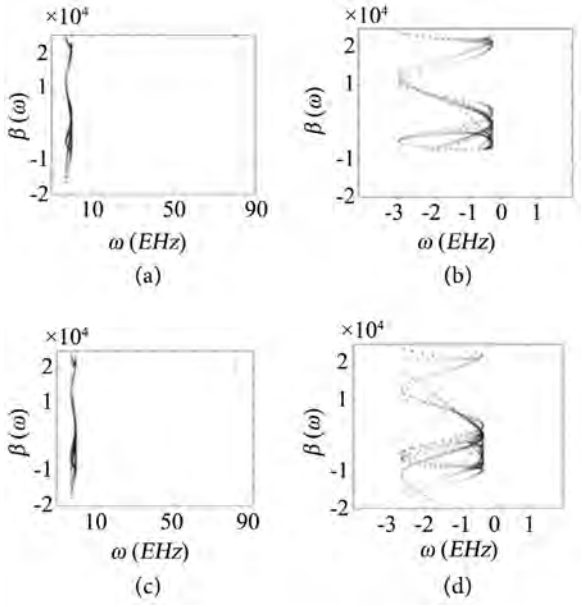


Figure 7. The dispersion curve (a) and (b) at $z = 0.2L_D$; and (c) and (d) at $z = 2L_D$. *EHz*: 10^{15} (*Hz*).

4.3. The Impacts of Birefringence and Confinement Loss in PCF Structure

If there are asymmetric and twist in the PCFs, these will result in the birefringence. The frequency component is much and will lead to more kinds of split and more induced waves [24] [25] [26].

The impact of confinement loss can be discussed by:

$$\frac{\partial^2 A}{\partial z^2} + 2i\beta\frac{\partial A}{\partial z} = \left(\frac{\alpha_{cl}}{\beta_0}\right)^2 A - 2\beta\alpha_{cl}/\beta_0 A \tag{10}$$

The first item of right side brings a gain or a loss on the field. The second item results in a phase shift along the transmission. The confinement loss value ($<10^{-7}$ dB/km compared the fiber loss 10^{-3} dB/km) is very small, so both the impacts on amplitude and phase are enough small to be ignored.

5. Conclusions

By resolving the transmission equation of field in PCFs, the dispersion effect is re-discussed. Although the second order differential to the transmission distance

($\partial^2 A/\partial z^2$) is taken into account, the first-order (β_1) and second-order dispersion (β_2) effects have the same influences as those of the nonlinear Schordinger equation, which cause the envelope of pulse move and pulse broadened, respectively. Additionally, the second-order dispersion also brings the fluctuations in the pulse edges, and with the increase of distance, even results in the pulse splittings.

But by our calculation, the third-order dispersion (β_3) will also induce pulse compression while the second-order dispersion is assumed as zero. The dispersion coefficient $\beta(\omega)$ alters with the distance.

Conflicts of Interest

The authors declare no conflict of interests.

References

[1] Kruglov, V.I. and Harvey, J.D. (2018) *Physical Review A*, **98**, Article ID: 063811. https://doi.org/10.1103/PhysRevA.98.063811

[2] Berkovsky, A.N., Kozlov, S.A. and Shpolyanskiy, Y.A. (2005) *Physical Review A*, **72**, Article ID: 043821. https://doi.org/10.1103/PhysRevA.72.043821

[3] Yu, M., McKinstrie, C.J. and Agrawal, G.P. (1998) *Physical Review E*, **52**, 1072-1080. https://doi.org/10.1103/PhysRevE.52.1072

[4] Wang, W.C., Wang, N. and Jia, H.Z. (2021) *Optik*, **241**, Article ID: 166935. https://doi.org/10.1016/j.ijleo.2021.166935

[5] Lakoba, T. and Kaup, D. 1998) *Physical Review E*, **58**, 6728. https://doi.org/10.1103/PhysRevE.58.6728

[6] Droques, M., Kudlinski, A., Bouwmans, G., Martinelli, G. and Mussot, A. (2013) *Physical Review A*, **87**, Article ID: 013813. https://doi.org/10.1103/PhysRevA.87.013813

[7] Abdullaev, F.K. and Caputo, J.G. (1998) *Physical Review E*, **58**, 6637-6648. https://doi.org/10.1103/PhysRevE.58.6637

[8] Wang, S.F., Fan, D.F., Bai, X.K. and Zeng, X.L. (2014) *Physical Review A*, **89**, Article ID: 023802. https://doi.org/10.1103/PhysRevA.89.039901

[9] Atre, R. and Panigrahi, P.K. (2007) *Physical Review A*, **76**, Article ID: 043838. https://doi.org/10.1103/PhysRevA.76.043838

[10] Kumar, S. and Hasegawa, A. (1997) *Optics Letters*, **22**, 372. https://doi.org/10.1364/OL.22.000372

[11] Gordon, J.P. (1992) *Journal of the Optical Society of America B*, **9**, 91. https://doi.org/10.1364/JOSAB.9.000091

[12] Abdullaev, F.Kh., Caputo, J.G. and Flytzanis, N. (1994) *Physical Review E*, **50**, 1552. https://doi.org/10.1103/PhysRevE.50.1552

[13] Malomed, B.A., Parker, D.F. and Smyth, N. (1993) *Physical Review E*, **48**, 1418-1425. https://doi.org/10.1103/PhysRevE.48.1418

[14] Mak, W.C.K., Malomed, B.A. and Chu, P.L. (2004) *Journal of Modern Optics*, **51**, 2141-2158. https://doi.org/10.1080/09500340408232519

[15] Ngabireng, C.M., *et al.* (2005) *Physical Review E*, **72**, Article ID: 036613. https://doi.org/10.1103/PhysRevE.72.036613

[16] Huang, J. (2016) *Journal of Nano-Photonics*, **10**, Article ID: 016004. https://doi.org/10.1117/1.JNP.10.016004

[17] Gui, M.X. and Huang, J. (2018) *IEEE Photonics Journal*, **10**, Article ID: 2400408.

[18] Luo, Z.Z., Yu, H.H., Zheng, Y., Zheng, Z., Li, Y.J. and Jiang, X. (2020) *Journal of Lightwave Technology*, **38**, 4560-4571. https://doi.org/10.1109/JLT.2020.2989926

[19] Agrawal, G.P. (2007) Nonlinear Fiber Optics. 4th Edition, Academic, San Diego.

[20] Nair, A.A., Boopathi, C.S., Jayaraju, M. and Mani Rajan, M.S. (2019) *Optik*, **179**, 718-725. https://doi.org/10.1016/j.ijleo.2018.11.021

[21] Shaheen, S., Gris-Sánchez, I. and Gasulla, I. (2020) *Journal of Lightwave Technology*, **38**, 6237-6246. https://doi.org/10.1109/JLT.2020.3011548

[22] Suresh, M.I., Hammer, J., Joly, N.Y., Russell, P.St.J. and Tani, F. (2021) Extension of Supercontinuum via Tapered Single-Ring PCF. *Conference on Lasers and Electro-Optics Europe & European Quantum Electronics Conference* (*CLEO/Europe-EQEC*), Virtual, 21-25 June 2021, Paper cd_5_6.

[23] Huang, J., Gui, M.X. and Man, W.Q. (2020) *Journal of Nano-Photonics*, **14**, Article ID: 026010.

[24] Zhang, J.-X. (2022) *Optik*, **251**, Article ID: 168425. https://doi.org/10.1016/j.ijleo.2021.168425

[25] Wong, G.K.L., Roth, P., Frosz, M.H. and Philip, St.J. (2020) Russell, Cross-Phase Modulational Instability of Vortex Modes in a Twisted Three-Core Photonic Crystal Fibre. *Conference on Lasers and Electro-Optics/Pacific Rim* (*CLEOPR*), Virtual, 2-6 August 2020, C8B_3.

[26] Guo, X., Han, L.Y., Liu, F. and Li, S.T. (2020) *Optik*, **218**, Article ID: 164796. https://doi.org/10.1016/j.ijleo.2020.164796

Journal of Modern Physics, 2022, 13, 518-573
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Experimental Verification of the Theory of Non-Force (Information) Interaction

Iurii Teslia

Department of Information Technology of Design, Cherkasy State Technological University, Cherkasy, Ukraine
Email: dr.teslia@gmail.com

How to cite this paper: Teslia, I. (2022) Experimental Verification of the Theory of Non-Force (Information) Interaction. *Journal of Modern Physics*, **13**, 518-573.
https://doi.org/10.4236/jmp.2022.134034

Received: March 10, 2022
Accepted: April 26, 2022
Published: April 29, 2022

Abstract

The work is devoted to the demonstration of the possibility of applying the formulas of information handling obtained in the theory of non-force interaction for the natural language processing. These formulas were obtained in computer experiments in modelling the movement and interaction of material objects by changing the amount of information that triggers this movement. The hypothesis, objective and tasks of the experimental research were defined. The methods and software tools were developed to conduct the experiments. To compare different results of the simulation of the processes in a human brain during speech production, there was a range of methods proposed to calculate the estimate of sequence of fragments of natural language texts including the methods based on linear approximation. The experiments confirmed that the formulas of information handling obtained in the theory of non-force interaction reflect the processes of language formation. It is shown that the offered approach can successfully be used to create systems of reactive artificial intelligence machines. Experimental and, presented in this work, practical results constitute that the non-force (informational) interaction formulae are generally valid.

Keywords

Non-Force Interaction, Non-Force Interaction Method, Computer Linguistics, Mechanical Movement, Special Relativity

1. Introduction

Information is the basis of a human life. Can it be that it is the source of the Universe's existence? Can the Universe be digital, computable? Today there is a range of theories that declare this variant of world formation, particularly, the computer simulation hypothesis of Nick Bostrom [1], or the research of the laws

of physics as a cellular automaton by Stephen Wolfram and Nobel laureate Gerard't Hooft [2] [3]. But, unfortunately, these, as well as other similar theories, have not been proven experimentally yet.

One of these theories is the non-force interaction theory (NFIT). It is based on a computer model in which the source of movement is information that triggers the movement [4] [5] [6], not only at social, biological or technical level of the matter existence, but also the mechanical movement of any material objects. If information is the basis of the Universe's existence, it should definitely manifest itself in mechanical movement as movement is a general form of the existence of matter. The NFIT is built on the information-probabilistic interpretation of movement. There are formulae received that correspond to Newton's classical and Einstein's relativistic mechanics with one addition that the source of movement is the information. Therefore, during interaction informational contents changes first, which then leads to changes in movement (speaking about mechanical movement, it leads to changes in the direction and speed of movement).

The research purpose of the work is to confirm the hypothesis of the theory of non-force interaction, which claims that all interactions in nature are based on informational reasons and are described by the same formulas. For this purpose, it is offered to verify the correspondence of the formulae received in the theory of non-force interaction to informational processes of speech production in a human brain by the means of computer experiments. The main question the answer to which is given in the article through experimental research is that probabilities of sequence of letter combinations in the texts in different languages can be received from the formulae of non-force interaction, which in their turn were received from information-probabilistic interpretation of mechanical movement.

If we can confirm it, it will mean that, when forming natural language texts, the processes of interaction in a human brain correspond to the formulae received in the theory of non-force interaction. This, therefore, will prove the universality of these formulae, and therefore, the validity of the results received from the theory of non-force interaction. Moreover, the main aspect is that this will significantly prove the hypothesis of the theory that all interactions in the Nature are caused by informational reasons.

2. Primary Research Materials

2.1. General Information about Non-Force Interaction Theory

The theory of non-force (information) interaction gives information-probabilistic interpretation of the laws that describe mechanical movement and immediate interaction of material objects. To do this, one more parameter was introduced into the non-force interaction theory: information that triggers the movement. On the basis of this, the formulae of transformation of information contents of the objects during their interaction were received from the known laws of physics. These formulae were obtained based on the following logical conclusions:

1) The theory of non-force interaction is based on the thesis that each object has a memory that contains information about previous interactions and that motivates it to move. Geometrically, memory is represented by the areas of determining the direction of movement of the object (Figure 1). The greater the difference in the amount of information that triggers the movement of two objects, the greater the relative velocity of the objects.

2) In the non-force interaction theoryphysical fields [7] provide information that triggers the movement. Field is not physical, but informational category. It is important what information the field can transfer rather than how it is done. **Physical field is the carrier of information and not the source of forceful action.**

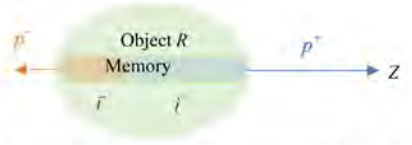
3) The movement of objects is determined not by physical laws, but by their "ability to react correctly" to the existence of other objects and which was formed during 13.8 billion years. In almost 13.8 billion years the matter learned to react to other objects exactly as it does when it changes the direction and velocity of movement during interaction. Physics explains the movementby the constant physical laws of the Nature. However, in the non-force interaction theory the movementis information.

4) We provide information to other people through speech. This information changes (complements) their available information. Such changed information triggers changes in the "movement" (actions) of people. Now, let us use the method of analogies. Physical fields "provide" information to the objects about the existence of other objects. The changed amount of information that triggers the movement, triggers the changes in mechanical movement (direction and velocity).

5) During the interaction of material objects, their amount of information that triggers the movement changes first, which then leads to the changes in their movement.

6) From the known "physical laws" there can be received the laws of change of the amount of information that triggers the movement, during interaction.

7) In the non-force interaction theory, the movement of the object is interpreted by shifts in possible directions (one quantum of space per one quantum of time). The probability of the shift is determined by the amount of information that triggers the movement [4] [5]. For a one-dimension movement in direction Z (Figure 1) [5].



i^+ - the size of the area to determine the movement in the direction Z;
i^- - the size of the area to determine the movement in direction opposite to Z;
p^+ - probability of a shift in direction Z;
p^- - probability of a shift in direction opposite to Z.

Figure 1. Information-probabilistic model of mechanical movement.

8) On **Figure 1**, the formulae that are proposed in the non-force interaction theory [4] are stated to calculate the probability of a shift.

$$p^+ = \frac{i^+}{i^- + i^+};\ p^- = \frac{i^-}{i^- + i^+}. \tag{1}$$

They are based on the geometrical representation of the ranges of definition of the object movement, that are in the object's memory (**Figure 1**).

9) In the NFIT, the difference in the size of the ranges is called the amount of information that triggers the movement and the addition is the awareness:

$$\begin{aligned} &d = i^+ - i^- \text{ - the amount of information that triggers the movement;} \\ &i = i^+ + i^- \text{ - awareness.} \end{aligned} \tag{2}$$

10) The amount of information that triggers the movement ≡ information about movement direction collected in interactions = non-force (information) influence on the object as to the choice of this exact movement direction.

11) Awareness ≡ information about movement direction collected in interactions = non-force (information) influence on the object as to possible movement directions.

12) Expected drift velocity (or just velocity) of an object (see **Figure 1**) [5]

$$\boldsymbol{V} = \left(p^+ - p^-\right)\cdot \boldsymbol{c} = \left(2\cdot p^+ - 1\right)\cdot \boldsymbol{c}, \tag{3}$$

where V is expected drift velocity of an object; c is the speed of light in vacuum.

13) The drift velocity of one object regarding the drift velocity of another object is the relative movement velocity.

14) Relative velocity from the formula of relativistic addition of velocities [7] in the information-probabilistic interpretation of mechanical movement looks as follows (using Formula (3)) [5]:

$$\Delta \boldsymbol{V} = \frac{\boldsymbol{V}_2 - \boldsymbol{V}_1}{1 - \frac{\boldsymbol{V}_1 \cdot \boldsymbol{V}_2}{\boldsymbol{c}^2}} \to \frac{p_2 \cdot \left(1 - p_1\right) - p_1 \cdot \left(1 - p_2\right)}{p_2 \cdot \left(1 - p_1\right) + p_1 \cdot \left(1 - p_2\right)} \cdot \boldsymbol{c}, \tag{4}$$

where $\Delta \boldsymbol{V}$ is relative velocity of some object R_2 in relation to the object R_1; p_2 is probability of the shift of object R_2 in the direction in relation to which the difference in velocities is measured; p_1 is probability of the shift of object R_1 in the direction in relation to which the difference in velocities is measured; c is the speed of light in vacuum.

However, from (3) it follows:

$$\Delta \boldsymbol{V} = \left(2\cdot \Delta p - 1\right)\cdot \boldsymbol{c}, \tag{5}$$

where Δp is probability of the shift of object R_2 in relation to object R_1 in the direction, in relation to which the difference in velocities is measured.

Then, the probability of the shift of object R_2 which would be "measured" by the observer who is in object R_1 (shift in the direction "away from the observer") from (4) and (5) is [1]

$$\Delta p = \frac{p_2 \cdot \left(1 - p_1\right)}{p_2 \cdot \left(1 - p_1\right) + p_1 \cdot \left(1 - p_2\right)}. \tag{6}$$

15) In the non-force interaction theory [4], it is shown that the implementation of mechanical movement according to Formula (6) is possible only under the condition that the ratio of quantity of actual shifts of all objects one in relation to another will be equal to the ratio of the sizes of definition rangesof motion (see **Figure 1**). However, it is only possible in the situation:

$$i^{+} \cdot i^{-} = const.$$

Based on the commonly used representation of formulae in relativistic mechanics [7], it is convenient to accept that [4]:

$$i^{+} \cdot i^{-} = 0.25. \tag{7}$$

Then from (1), (2) and (7) the connection between the probability of shift in direction (p), the amount of information that triggers the movement (d) and awareness as to the directions of shift (i) in one-dimensional virtual space is defined by the following formulae [1]:

$$i = \frac{1}{2\sqrt{p \cdot (1-p)}}; \tag{8}$$

$$i = \sqrt{d^2 + 1}; \tag{9}$$

$$d = \pm\sqrt{i^2 - 1}; \tag{10}$$

$$d = \begin{cases} 0.5 \cdot \sqrt{\dfrac{p}{1-p} + \dfrac{1-p}{p} - 2}, & p \geq 0.5 \\ -0.5 \cdot \sqrt{\dfrac{p}{1-p} + \dfrac{1-p}{p} - 2}, & p < 0.5 \end{cases} \tag{11}$$

$$p = 0.5 + \frac{d}{2i}. \tag{12}$$

2.2. Velocity and Momentum in Non-Force Interaction Theory

In the core of the non-force interaction theory, there is a computer model of mechanical movement in which the shift of objects one in relation to another is explained by the difference in the amount of information that triggers the movement, as it is shown in Sub-section 2.1. In essence, the non-force interaction theory interprets mechanical movement of material objects "in a new way". Physical categories, such as velocity and momentum, are interpreted through the information that triggers the movement. Accordingly, physical formulae that include these values are represented in a new way, through the amount of information that triggers the movement. Thus, such their interpretation creates possibility to find analogies and check for conformity with informational interaction of people. Let us look at the following interpretation:

1) Representation of physical formulae through the amount of information that triggers the movement:

a) Applying in (12) Formula (3), new representation of velocity formula:

$$\boldsymbol{V} = \frac{d}{i} \cdot \boldsymbol{c}. \tag{13}$$

b) Relativistic mass [7]

$$m = \frac{m_0}{\sqrt{1-\frac{\boldsymbol{V}^2}{\boldsymbol{c}^2}}}, \tag{14}$$

where m_0 is the mass in the state of rest.

Applying in (14) Formula (13), information-probabilistic interpretation of relativistic mass is obtained [1].

$$m = \frac{m_0}{\sqrt{1-\frac{\boldsymbol{V}^2}{\boldsymbol{c}^2}}} = \frac{m_0}{\sqrt{1-\frac{d^2\cdot\boldsymbol{c}^2}{i^2\cdot\boldsymbol{c}^2}}} = \frac{m_0\cdot i}{\sqrt{i^2-d^2}} = m_0\cdot i. \tag{15}$$

c) The law of conservation of momentum (quantity of movement). It characterizes mechanical interaction of material objects. It determines that in a closed system the total momentum of all bodies is preserved [8]:

$$\sum_j \boldsymbol{P}_j = const, \tag{16}$$

where $\boldsymbol{P}_i$ is momentum.

$$\boldsymbol{P}_j = m_j\boldsymbol{V}_j, \tag{17}$$

where m_j is mass; $\boldsymbol{V}_j$ is velocity.

Applying in (17) Formulae (13) and (15), it is obtained

$$\boldsymbol{P}_j = m_{0j}\cdot i\cdot\frac{d_j}{i}\boldsymbol{c} = m_{0j}\cdot d_j\cdot\boldsymbol{c},$$

where m_{0j} is rest mass; d_j is the amount of information that triggers the movement.

Hence, information-probabilistic interpretation of the law of conservation of momentum can be stated [4]:

$$\sum_j \boldsymbol{P}_j = \sum_j m_{0j}\cdot d_j\cdot\boldsymbol{c} = const.$$

Out of the fact that in a closed system the total rest mass does not change, the "law of conservation of the amount of information that triggers movement" arises [4]:

$$\sum_j d_j = const. \tag{18}$$

d) The difference in velocity is a consequence of the difference in the amount of information that triggers the movement. In the work [4] it is shown that if (13)

$$\boldsymbol{V}_1 = \frac{d_1}{i_1}\cdot\boldsymbol{c};$$

$$\boldsymbol{V}_2 = \frac{d_2}{i_2}\cdot\boldsymbol{c};$$

$$\Delta\boldsymbol{V}_1 = \frac{\Delta d_1}{\Delta i_1}\cdot\boldsymbol{c},$$

where V_1 is the drift velocity of object R_1; V_2 is the drift velocity of object R_2; d_2 is the amount of information that triggers the movementof object R_2; d_1 is the amount of information that triggers the movementof object R_1; i_1 is awareness of object R_1, then from Formulas (2) - (4) it is possible to obtain [4]

$$\Delta d = d_2 i_1 - d_1 i_2, \tag{19}$$

where Δd is the difference in determination of objects R_2 and R_1 in relation to the movement in direction Z.

Having d_1 and Δd known, it is possible to determine d_2. See this from (9)

$$i_1 = \sqrt{d_1^2 + 1};$$

$$i_2 = \sqrt{d_2^2 + 1};$$

$$\Delta i = \sqrt{\Delta d^2 + 1},$$

where Δi is the difference in the amount of information that triggers the movement of objects R_1 and R_2.

Then, from (9) it follows

$$d_2 = d_1 \Delta i + i_1 \Delta d. \tag{20}$$

2) The change in the amount of information that triggers the movement with direct (contact) interaction of objects. When examining this question, the mass of interacting objects will not be looked at, as the change of velocity during direct (contact) interaction depends on the mass and is considered in the calculations provided. Additionally, there is considered a closed system of three objects the mass of which is constant.

a) Assuming there is object R …

- … moving in direction Z with velocity V_R (13)

$$\boldsymbol{V}_R = \frac{d_R}{i_R}\boldsymbol{c},$$

where V_R is the drift velocity of object R in direction Z until collision; d_R is the amount of information that triggers the movement of object R in direction Z until collision; i_R is awareness of object R about movement until collision;

- … will collide with object X and its velocity will become V_{RX}. The change of the amount of information that triggers the movement will correspond to the change of velocity. As it follows from Formula (19)

$$\Delta d_{RX} = d_{RX} i_R - d_R i_{RX},$$

where Δd_{RX} is the change of the amount of information that triggers the movement of object R in direction Z after collision with object X; i_{RX} is awareness of object R about movement after collision with object X; d_{RX} —the amount of information that triggers the movementof object R in direction Z after collision with object X;

- … will collide with object Y and its velocity will become V_{RY}. The change of determination will correspond to the change of velocity. As it follows from Formula (19)

$$\Delta d_{RY} = d_{RY} i_R - d_R i_{RY},$$

where Δd_{RY} is the change of the amount of information that triggers the movement of object R in direction Z after collision with object Y; i_{RY} is awareness of object R about movement after collision with object Y; d_{RY} —the amount of information that triggers the movementof object R in direction Z after collision with object Y;

- … will collide with both objects X and Y and its velocity will become V_{RXY}. From the Formula (19) there follows a new amount of information that triggers the movement

$$\Delta d_{RXY} = d_{RXY} i_R - d_R i_{RXY}, \tag{21}$$

where Δd_{RXY} —is the change of the amount of information that triggers the movementof object R in direction Z after collision with objects X and Y; i_{RXY} — is awareness of object R about movement after collision with object X and Y; d_{RXY} —the amount of information that triggers the movementof object R in direction Z after collision with objects X and Y.

b) Out of information-probabilistic interpretation of the law of conservation of momentum (18) it follows that if two objects interact, the total value of the change of momentum of these objects should equal 0. Only in this case the total momentum does not change. Hence, it can be written

$$\begin{aligned} \Delta d_{RX} + \Delta d_{XR} &= 0; \\ \Delta d_{RY} + \Delta d_{YR} &= 0, \end{aligned} \tag{22}$$

where Δd_{XR} is the change of the amount of information that triggers the movement of object X in direction Z after colliding with object R; d_{YR} is the change of the amount of information that triggers the movementof object Y in direction Z after colliding with object R.

Then, for the case of collision of object R with objects X and Y; taking into account (18), it can be written as

$$\Delta d_{RXY} + \Delta d_{XR} + \Delta d_{YR} = 0.$$

Using (22), it can be obtained:

$$\Delta d_{RXY} = -\Delta d_{XR} - \Delta d_{YR} = \Delta d_{RX} + \Delta d_{RY}. \tag{23}$$

From (20) it follows

$$d_{RXY} = d_R \Delta i_{RXY} + i_R \Delta d_{RXY}, \tag{24}$$

where Δi_{XRY} is the change of awareness of object R about the movement after collision with objects X and Y.

From (9) it follows

$$i_{XRY} = \sqrt{d_{XRY}^2 + 1},$$

where i_{XRY} is awareness of object R about the movement after collision with objects X and Y.

c) New probability of shift of object R in direction Z (12)

$$p_{XRY} = 0.5 + \frac{d_{XRY}}{2 \cdot i_{XRY}}, \tag{25}$$

where p_{XRY} is the probability of shift of object R in direction Z after colliding with objects X and Y.

If the presented model of transformation of probabilistic movement is really implemented in mechanical movement, its formulae may depict some general-laws of our Nature. What if the same formulae are used to transform the information in a human brain? What if they are generally valid? It is the answer to this question that should be found in experiments. To do this, let us first simplify the information-probabilistic interpretation of mechanical movement to the form which makes it possible to solve other problems connected to the existence of on-force interaction.

2.3. Non-Force Interaction Method

The non-force interaction method (NFIM) allows to move from the operation "the amount of information that triggers the movement (mechanical)" to the operation of the amount of information that triggers the reaction (manifestation) (in social, biological or "technical" form of motion of matter).

In the subject field, there are objects that react to non-force influence. There are objects and processes that influence this reaction. The reaction itself can influence the reaction of objects. For instance, the result of a football match is influenced by referees, players, the field condition, audience, etc. However, the result of previous matches also influences the result (through the psychology of players).

The steps of the method:

1) The calculation of the amount of information that triggers the reaction of objects of subject field when non-force influences are absent

$$\forall R_i \in R, \exists r_{ij} \in R_j,$$

where r_{ij} is the possible reactions of object R_i; R_i is the object; R is the multitude of objects of subject field.

$$0 < p\left(r_{ij}\right) < 1, \sum\nolimits_j p\left(r_{ij}\right) = 1,$$

where $p\left(r_{ij}\right)$ is the unconditional probability of reaction r_{ij}.

Out of (11) it follows:

$$d\left(r_{ij}\right) = \begin{cases} 0.5 \cdot \sqrt{\dfrac{p\left(r_{ij}\right)}{1-p\left(r_{ij}\right)} + \dfrac{1-p\left(r_{ij}\right)}{p\left(r_{ij}\right)} - 2}, & p\left(r_{ij}\right) \geq 0.5 \\ -0.5 \cdot \sqrt{\dfrac{p\left(r_{ij}\right)}{1-p\left(r_{ij}\right)} + \dfrac{1-p\left(r_{ij}\right)}{p\left(r_{ij}\right)} - 2}, & p\left(r_{ij}\right) < 0.5 \end{cases}$$

where $d\left(r_{ij}\right)$ is the amount of information that triggers the reaction r_{ij}.

Out of (8) it follows:

$$i\left(r_{ij}\right)=\frac{1}{2\sqrt{p\left(r_{ij}\right)\cdot\left(1-p\left(r_{ij}\right)\right)}},$$

where $i\left(r_{ij}\right)$ is the awareness about the reaction r_{ij}.

2) The calculation of determination and awareness of reactions of objects of subject field to which the non-force influences are applied

$$\forall w_k \in W, \exists R_i \in R, r_{ij} \in R_i : p\left(r_{ij}/w_k\right) \neq p\left(r_{ij}\right),$$

where w_k are influences; $p\left(r_{ij}/w_k\right)$ is conditional probability of reaction r_{ij} if to the object R_i influence w_k is applied.

Out of (11) it follows:

$$\forall w_k \in W, \forall R_i \in R, \forall r_{ij} \in R_i :$$

$$d\left(r_{ij}/w_k\right)=\begin{cases}0.5\cdot\sqrt{\dfrac{p\left(r_{ij}/w_k\right)}{1-p(r_{ij}/w_k)}+\dfrac{1-p\left(r_{ij}/w_k\right)}{p\left(r_{ij}/w_k\right)}-2}, & p\left(r_{ij}/w_k\right)\geq 0.5\\ -0.5\cdot\sqrt{\dfrac{p\left(r_{ij}/w_k\right)}{1-p\left(r_{ij}/w_k\right)}+\dfrac{1-p\left(r_{ij}/w_k\right)}{p\left(r_{ij}/w_k\right)}-2}, & p\left(r_{ij}/w_k\right)<0.5\end{cases}$$

where $d\left(r_{ij}/w_k\right)$ is the determination of the reaction r_{ij} under the conditional that to the object R_i the influence w_k is applied.

Out of (8) it follows:

$$\forall w_k \in W, \forall R_i \in R, \forall r_{ij} \in R_i : i\left(r_{ij}/w_k\right)=\frac{1}{2\sqrt{p\left(r_{ij}/w_k\right)\cdot\left(1-p\left(r_{ij}/w_k\right)\right)}},$$

where $i\left(r_{ij}/w_k\right)$ is the awareness of the object R_i about the reaction r_{ij} under the condition that to it the influence w_k is applied.

3) The calculation of total non-force influence to the objects from (19) and (23)

$$\forall w_k \in W, \forall R_i \in R, \forall r_{ij} \in R_i : \Delta d\left(r_{ij}/W\right)=\sum_{w_k} d\left(r_{ij}/w_k\right)\cdot i\left(r_{ij}\right)-d\left(r_{ij}\right)\cdot i\left(r_{ij}/w_k\right),$$

where $\Delta d\left(r_{ij}/W\right)$ is the value of non-force influence on the reaction r_{ij} of object R_i that equals to the change in the amount of information that triggers the reaction.

Additional awareness that equals to non-force influences (9)

$$\forall w_k \in W, \forall R_i \in R, \forall r_{ij} \in R_i : \Delta i\left(r_{ij}/W\right)=\sqrt{\left(\Delta d\left(r_{ij}/W\right)\right)^2+1},$$

where $\Delta i\left(r_{ij}/W\right)$ is additional awareness about non-force influence on the reaction r_{ij} of object R_i that corresponds to additional non-force influence.

4) The calculation of a new (after influences) amount of information that triggers the reaction of objects of subject field (24)

$$\forall R_i \in R, \forall r_{ij} \in R_i : d\left(r_{ij}/W\right)=\Delta d\left(r_{ij}/W\right)\cdot i\left(r_{ij}\right)+d\left(r_{ij}\right)\cdot\Delta i\left(r_{ij}/W\right), \quad (26)$$

where $d\left(r_{ij}/W\right)$ is a new (after all influence) amount of information that trig-

gers the reaction r_{ij} of object R_i.

Additional awareness that corresponds to new amount of information that triggers the reaction (9)

$$\forall R_i \in R, \forall r_{ij} \in R_i : i\left(r_{ij}/W\right) = \sqrt{\left(d\left(r_{ij}/W\right)\right)^2 + 1},$$

where $i\left(r_{ij}/W\right)$ is new (after all influences) awareness about the reaction r_{ij} of object R_i.

5) The estimate of combined conditional probability of reactions of objects of subject field (12)

$$\forall R_i \in R, \forall r_{ij} \in R_i : \hat{p}\left(r_{ij}/W\right) = 0.5 + \frac{d\left(r_{ij}/W\right)}{2 \cdot i\left(r_{ij}/W\right)},$$

where $\hat{p}\left(r_{ij}/W\right)$ is the estimate of combined conditional probability of reaction r_{ij} after all influences.

The initial data of the method are unconditional and individual conditional probabilities of reactions of objects to non-force influences. The resulting data are the estimate of combined conditional probability of reactions. Therefore, the question arises. Why the estimate and not the probability itself? Considering that out of information-probabilistic interpretation of mechanical movement it follows that the result of calculation is the probability of the shift of objects itself. However, in this case the scheme of calculations for any stochastic subject fields is proposed. Thus, these fields can be characterized by such connection between objects and processes when their combinations can have synergetic effect.

Hence, it refers to the estimate of combined conditional probability of reactions and not to the probability itself.

3. Related Works

Wernicke's and Broca's areas are responsible for the speech processes in a human brain. Wernicke's area, which is also called Wernicke's speech area, is one of the two parts of cerebral cortex connected to speech, the other is Broca's area. Wernicke's area takes part in processing written and spoken language in contrast to Broca's area, which is responsible for speech production [9].

Wernicke's area supports an important component of speech production which is called phonological search where phonemes subject to articulation and their time order are represented mentally. This process is necessary for all the tasks of speech production including repetition, search of words (for instance, during spontaneous speaking or naming) and reading aloud. Repetition of language means entering the system of phonological search through the system of auditory comprehension of phonemes. Similar mechanism supports reading aloud, except the fact that the entrance to the system of phonological search is through the system of visual perception of letters in ventral occipitotemporal area. The production of communicative speech includes the step of phonological search, during which the concept depicting what the speaker wants to say is re-

ceived. Then the search of words is made through the reflection of these meanings of words in phonological representations. Thus, in contrast to repetition or reading, entrance to the system of phonological search in these tasks is made through internal semantic system (the meaning of word). This semantic processing network is widely-spread in the cortexes of associations of higher level in temporal, parietal and frontal lobes. The processing of language provides the reflection of the sequences of phonemes (they are perceived in the system of perception of auditory phoneme) in the meanings on words (represented in the semantic system) [10].

Today the researches of the speech production processes are mainly dedicated to the determination of neurobiological features of information interaction in a human brain. For instance, in the work "The Brain Basis of Language Processing: From Structure to Function" [11] the structural and functional neural system that is the basis of sentence processing and how this process evolves with time when the sentence is perceived is described. The authors conduct the overview of short outline of timeline of different subprocesses that constitute the process of sentence processing. Then, the general networking function of speech, which is in the brain cortex, and its neuroanatomical architecture are determined. On the basis of this there were described different processes that happen during processing, such as acoustic and phonological analysis as well as syntax and semantic processes. Also, authors [12] [13] argued that speech production includes sensory systems in posterior upper temporal lobe of the left hemisphere and that the interface between perception and movement systems are supported by sensorimotor circuit for the actions of voice routes that is really similar to sensorimotor systems found in the parietal part of primates, and that verbal short-term memory can be understood as integral feature of this sensorimotor circle.

From the position of simulating of speech production processes in a human brain with the aid of mathematical apparatus and computer technologies, nowadays most of the researches are aimed at speech synthesis [14]. The technology of speech synthesis has become the centre of researches in the sphere of intelligent computer. For almost 50 years, the studying of speech synthesis has had considerable development as an interdisciplinary approach [15]. The example of speech synthesis research is the work [16] where there are described concepts of automatic formation of personal digital template of voice to solve the following issues on its basis: computer speech synthesis, continuous speech recognition and identification of a person by voice (three main branches of language technologies). As well as that there is an article [17] where the authors implement and compare the models of identification of spoken language on the basis of deep learning. The authors also use two most modern and popular methods of speech recognition, such as Wav2Vec and SpecAugment, in their classifiers and verify if they are also used in the sphere of speech identification. Among the models that are implemented by authors the classifier of the deep data networks on the basis of X-vector is given the highest rank F1 - 0.91 where target set is composed of five languages. SpecAugment data augmentation

method, as it happens, improves the accuracy of classification when it is applied to input mel-spectrograms of CRNN architecture. Even though they receive the accuracy of classification lower than some other methods, Wav2Vec linguistic representations also provide rather promising results.

The researches provided give the answer to the questions how the brain works when it forms or processes the speech. However, despite the considerable number of researches dedicated to the problems of speech production and processing in a human brain, in the process of analysis of modern scientific works not detected were the works that could give the answer to a more important question: why the brain works in this way in the first place? What information laws are in the basis of brain development and structures? How do they function in the process of speech production and processing? If the underlying hypothesis of the NFIT is correct, then in both mechanical movement and the process in a human brain the same laws of non-force (information) interaction should be manifested. Thus, the formulae received from the information-probabilistic interpretation of the mechanical movement should also reflect the processes that are present in a human brain including the ones during speech production and processing. Moreover, in the studied works the correspondence of the existing formulae to the information processes in a human brain connected to speech was not verified, hence the hypothesis of the present research is neither confirmed nor refuted, therefore it requires experimental verification.

4. Hypothesis and Objective of Experimental Research

The hypothesis of research: The hypothetical model of non-force interaction created out of the laws of mechanical movement corresponds to the informational processes in a human brain, connected to speech. This is the indication of the laws' general validity.

Scientific objective of the experiment: is to prove the general validity of the non-force interaction formulae through demonstrating the applicability of the formulae received from information-probabilistic interpretation of mechanical movement in relation to the processes of speech production in a human brain.

The practical objective of the experiment: is to demonstrate the effectiveness of the method of non-force interaction in relation to the processing of natural language texts, which allows to create scientific-practical tools in the form of methods and algorithms of creation of the artificial intelligence systems that can develop reflexes to non-force influences in the functional environment.

Language is the manifestation of informational processed in a human brain. The idea of experiments is to show through the language that informational processes in a human brain run according to the formulae received in the non-force interaction theory. In fact, the objective of the experiments is to demonstrate the fact that the formulae obtained out of the physical laws "work" at the level of human intellectual activity as well. Particularly, in language. In the experiments it will be checked if it is possible to evaluate combined conditional

probability by individual conditional probabilities and unconditional probability of text fragment sequence. Mathematically p (O/XY) needs to be evaluated by p (O), p (O/X), p (O/Y). As it is provided in Subsection 2.3 applied to natural language texts.

Such formulation of the problem has also practical value, and in essence it reflects the mechanism of reacting to events (reflexes) of living creatures. In the process of life, each living creature learns how to react to fluences so that ensure its own life-sustaining activity in the best possible way. There is a rule in the core of the reflexes, which says that if X happened, action A is needed. Whereas if Y happened, action B is needed. What is both X and Y happened? In this case, the ability to evaluate combined conditional probability by individual ones can be useful.

If numerical values of speech creation correspond to the formulae received in the theory of non-force interaction, it will mean that intellectual apparatus of a human also uses them when producing speech (and maybe not only speech). In its turn, it means that they most probably are generally valid for the interaction processes in the Nature. Therefore, they can be used to create artificial intelligence systems as well. Is this true? Let us first conduct the experiments, and then look for the answer to this question.

5. Methodology of Experimental Research

All information on experiments is in the database [18].

There are natural language texts in different languages selected. They are the works of William Shakespeare [19] ("English" base), the epic novel of Lev Tolstoy "War and Peace) [20] [21] [22] [23] ("Russian" base), the works of Friedrich Nietzsche, Johann Wolfgang Goethe and Franz Kafka [24] [25] [26] ("German" base) and the works of Ukrainian poets [27] [28] [[29] ("Ukrainian" base).

1) In every sentence, all symbols are rejected except the letters of the corresponding alphabet (the fragments of the such texts are provided in **Figure 2**).

"Russian" Base 2165328 letters	*"English" Base 3784551 letters*	*"German" Base 1062279 letters*	*"Ukrainian" Base 531926 letters*
сейчасприбавилаонаопя тьуспокоиваясьнынчеум енядваоченьинтересные человекакстативиконтм ортемаронвродствесмон морансичрезрогановодн аизлучшихфамилийфра нцииэтоодинизхороших эмигрантовизнастоящих ипотомаббатмориовызн аетеэтотглубокийумонб ылпринятгосударемвыз наете	fromfairestcreatureswedesirein creasethattherebybeautysrosem ightneverdiebutastheripershoul dbytimedeceasehistenderheirm ightbearhismemorybutthoucon tractedtothineownbrighteyesfe edstthylightsflamewithselfsubs tantialfuelmakingafaminewher eabundanceliesthyselfthyfoetot hysweetselftoocruelthouthatart nowtheworldsfreshornamentan donlyheraldtothegaudyspringw ithinthineownbudburiestthycon tentandtenderchurlmakstwastei nniggardingpitytheworldorelse thisgluttonbetoeattheworldsdu ebythegraveandthee	gleicheineraltenhalbv erklungnensagekom mtersteliebundfreund schaftmitheraufdersc hmerzwirdneueswied erholtdieklagedesleb enslabyrinthischirren lauf	огожсиротинакогозап итаєіхтоїйрозкажеіхт отеєзнаєдемилийноч уєчивтемномугаючив бистрімдунаюконяна повачиможездругоюд ругуюкохаєїїчорнобр ивуужезабува

Figure 2. The fragments of texts prepared for experiments.

2) Each sentence is divided into consecutive letter combinations (fragments) with the length of L letters ($L = 1$, $L = 2$).

3) The statistical probability of the order of each fragment is calculated, as well as the probability of appearance of a fragment between two other fragments. For the text: $a_1, a_2, \cdots, a_{i-1}, a_i, a_{i+1}, \cdots, a_n$, the probability

$$p(a_i) = \frac{n(a_i)}{N_1},$$

where $n(a_i)$ is the number of times fragment a_i appears between other fragments in the text sentences; N_1 is the total number of fragment combinations present in the sentences of a text; $p(a_i)$ is the statistic unconditional probability of the sequence of text fragment a_i.

$$p(a_i / a_{i-1}a_{i+1}) = \frac{n(a_{i-1}, a_i, a_{i+1})}{n(a_{i-1}, a_{i+1})},$$

where $n(a_{i-1}, a_i, a_{i+1})$ is the number of times a fragment combination a_{i-1}, a_i, a_{i+1} appears in the sentences of a text; $n(a_{i-1}, a_{i+1})$ is the number of times fragment combination $a_{i-1}, \cdots, a_{i+1}$ appears in the sentences of a text; $p(a_i / a_{i-1}a_{i+1})$ is the statistical unconditional probability of the sequence of text fragment a_i if the previous text fragment is a_{i-1}, and the following is a_{i+1}.

4) The statistical probability of sequence of each fragment after another fragment is calculated:

$$p(a_i / a_{i-1}) = \frac{n(a_{i-1}, a_i)}{n(a_{i-1})},$$

where $n(a_{i-1}, a_i)$ is the number of times fragment combination $a_{i-1}, a_i, \cdots$ appears in the sentences of a text; $n(a_{i-1})$ is the number of times a fragment $a_{i-1}, \cdots$ appears in the sentences of a text; $p(a_i / a_{i-1})$ is the statistic conditional probability of sequence of the text fragment a_i if the previous text fragment is a_{i-1}.

5) The statistical probability of sequence of each fragment before another fragment is calculated:

$$p(a_i / a_{i+1}) = \frac{n(a_i, a_{i+1})}{n(a_{i+1})},$$

where $n(a_i, a_{i+1})$ is the number of times the combination of fragments $\cdots, a_i, a_{i+1}$ appears in the sentences of a text; $n(a_{i+1})$ is the number of times fragment $\cdots, a_{i+1}$ appears in the sentences of a text; $p(a_i / a_{i+1})$ is the statistical conditional probability of sequence of text fragment a_i if the next fragment is a_{i+1}.

6) B All the combinations of fragments for which $(a_{i-1}, a_i) > 0$ or $n(a_i, a_{i+1}) > 0$ (even if $n(a_{i-1}, a_i, a_{i+1}) = 0$) are recorded in the database.

7) Methods M_s of determination of combined conditional probability are chosen. The value received by such methods will be called estimate of sequence of the text fragments.

$$p_s(a_i / a_{i-1}a_{i+1}),$$

where $p_s\left(a_i/a_{i-1}a_{i+1}\right)$ is the estimate of combined conditional probability of sequence of text fragment a_i by method M$_s$; s the number of a method.

8) The estimate of combined conditional probability is calculated by using the chosen methods M$_s$.

9) The chosen methods are assessed by the deviation of the estimates of combined conditional probability of sequence of text fragment a_i by method M$_s$ from the conditional probabilities themselves.

10) The results are analysed. If the results are received by the method of non-force interaction are found to be better, it can be stated that the objective of the experiments is achieved.

6. Experimental Methods

For each combination of fragments, the combined conditional probability estimate is calculated by 9 different methods:

1) By unconditional probability:

$$p_1\left(a_i/a_{i-1}a_{i+1}\right)=p\left(a_i\right),$$

where $p_1\left(a_i/a_{i-1}a_{i+1}\right)$ is the estimate of combined conditional probability of sequence of text fragment a_i by its unconditional probability.

2) By conditional probability of the previous fragment:

$$p_2\left(a_i/a_{i-1}a_{i+1}\right)=p\left(a_i/a_{i-1}\right),$$

where $p_2\left(a_i/a_{i-1}a_{i+1}\right)$ is the estimate of combined conditional probability of sequence of text fragment a_i by conditional probability of its appearance if the previous text fragment is a_{i-1}.

3) By conditional probability of the following fragment:

$$p_3\left(a_i/a_{i-1}a_{i+1}\right)=p\left(a_i/a_{i+1}\right),$$

where $p_3\left(a_i/a_{i-1}a_{i+1}\right)$ is the estimate of combined conditional probability of sequence of text fragment a_i by conditional probability of its appearance if the following text fragment is a_{i+1}.

4) By the average value of individual conditional probabilities:

$$p_4\left(a_i/a_{i-1}a_{i+1}\right)=\frac{p\left(a_i/a_{i-1}\right)+p\left(a_i/a_{i+1}\right)}{2},$$

where $p_4\left(a_i/a_{i-1}a_{i+1}\right)$ is the estimate of combined conditional probability of sequence of text fragment a_i by the average value of conditional probabilities of its appearance if the previous text fragment is a_{i-1} and the following text fragment is a_{i+1}.

5) By weighted sum of individual conditional probabilities (the approximation by linear function using the method of least squares) (hereinafter, Method 1)

$$p_5\left(a_i/a_{i-1}a_{i+1}\right)=\alpha_1\cdot p\left(a_i/a_{i-1}\right)+\alpha_2\cdot p\left(a_i/a_{i+1}\right)+\alpha_3, \tag{27}$$

under the condition that

$$\sum_{a_i,a_{i-1},a_{i+1}}\left(p_5\left(a_i/a_{i-1}a_{i+1}\right)-p\left(a_i/a_{i-1}a_{i+1}\right)\right)^2\to\min, \tag{28}$$

where $p_5\left(a_i/a_{i-1}a_{i+1}\right)$ is the estimate of combined conditional probability of sequence of text fragment a_i by linear approximation of conditional probabilities of its appearance: if the previous text fragment is a_{i-1}; the following text fragment is a_{i+1}; $\alpha_1,\alpha_2,\alpha_3$ are approximation coefficients.

6) Taking into account the values of individual conditional probabilities (hereinafter, Method 2). Also, the approximation by linear function with minimizing of deviation is done:

$$\begin{aligned} p_6\left(a_i/a_{i-1}a_{i+1}\right) = \beta_1 \cdot \max\left(p\left(a_i/a_{i-1}\right), p\left(a_i/a_{i+1}\right)\right) \\ + \beta_2 \cdot \min\left(p\left(a_i/a_{i-1}\right), p\left(a_i/a_{i+1}\right)\right) + \beta_3, \end{aligned} \tag{29}$$

under the condition that

$$\sum_{a_i,a_{i-1},a_{i+1}} \left(p_6\left(a_i/a_{i-1}a_{i+1}\right) - p\left(a_i/a_{i-1}a_{i+1}\right)\right)^2 \to \min, \tag{30}$$

where $p_6\left(a_i/a_{i-1}a_{i+1}\right)$ is the estimate of conditional probability of sequence of text fragment a_i comparing to the values of individual conditional probabilities: if the previous text fragment is a_{i-1}; the following text fragment is a_{i+1}; β_1,β_2,β_3 are approximation coefficients.

7) Taking into account the deviation of the second conditional probability from unconditional (hereinafter, Method 3):

$$p_7\left(a_i/a_{i-1}a_{i+1}\right) = \gamma_1 \cdot p\left(a_i/a_{i-1}\right) + \gamma_2 \cdot \left(p\left(a_i/a_{i+1}\right) - p\left(a_i\right)\right) + \gamma_3, \tag{31}$$

under the condition that

$$\sum_{a_i,a_{i-1},a_{i+1}} \left(p_7\left(a_i/a_{i-1}a_{i+1}\right) - p\left(a_i/a_{i-1}a_{i+1}\right)\right)^2 \to \min, \tag{32}$$

where $p_7\left(a_i/a_{i-1}a_{i+1}\right)$ is the estimate of conditional probability of sequence of text fragment a_i by linear approximation of deviation of the second conditional probability from unconditional: if the previous text fragment is a_{i-1}; the following text fragment is a_{i+1}; γ_1,γ_2 are approximation coefficients.

8) Taking into account the deviation of smaller conditional probability from unconditional (hereinafter Method 4). Let us assume that $p_i^{\max}$ is maximum. Accordingly, $p_i^{\min}$ is the smallest of these two values. Then,

$$p_8\left(a_i/a_{i-1}a_{i+1}\right) = \delta_1 \cdot p_i^{\max} + \delta_2 \cdot \left(p_i^{\min} - p\left(a_i\right)\right) + \delta_3, \tag{33}$$

under the condition that

$$\sum_{a_i,a_{i-1},a_{i+1}} \left(p_8\left(a_i/a_{i-1}a_{i+1}\right) - p\left(a_i/a_{i-1}a_{i+1}\right)\right)^2 \to \min, \tag{34}$$

where $p_8\left(a_i/a_{i-1}a_{i+1}\right)$ is the estimate of conditional probability of text fragment a_i sequence by linear approximation of deviation of a small conditional probability from an unconditional one: if the previous text fragment is a_{i-1}; the following text fragment is a_{i+1}; δ_1,δ_2 are approximation coefficients.

It is possible to invent the multitude of heuristic methods of calculating the estimate of combined conditional probability, not only the ones presented in pp. 5-8. The theory of non-force interaction started from the search of such heuris-

tics that could give the best result when processing natural language texts (the experiments were conducted with the aim to create the system of natural speech access to databases in the basis of which was not time-consuming process of syntax, morphological, semantic machine analysis of a text, but rather a reaction to the fragments of texts [30]. Thus, when there was chosen heuristics which could provide the best result, the search of the answer to the question: "why it is like this?" has begun. The solution was found through information-probabilistic interpretation of mechanical movement [4] [5].

It is this solution that was put in the basis of the non-force interaction theory.

9) Using the **method of non-force interaction**, proposed in the theory of non-force interaction (see Subsection 2.3). By "reaction" we mean "fragment of text". Regarding the language, in the method the following steps of calculation of non-force influences on the process of text formation and determination of probabilities of sequence of its fragments are implemented:

a) The calculation of the influence size. Let us determine the amount of information that triggers a piece of text (11)

$$0 < p(a_i) < 1;\ 0 < p(a_i/a_{i-1}) < 1;\ 0 < p(a_i/a_{i+1}) < 1:$$

$$d(a_i) = \begin{cases} 0.5 \cdot \sqrt{\dfrac{p(a_i)}{1-p(a_i)} + \dfrac{1-p(a_i)}{p(a_i)} - 2}, & p(a_i) \geq 0.5 \\ -0.5 \cdot \sqrt{\dfrac{p(a_i)}{1-p(a_i)} + \dfrac{1-p(a_i)}{p(a_i)} - 2}, & p(a_i) < 0.5 \end{cases}$$

$$d(a_i/a_{i-1}) = \begin{cases} 0.5 \cdot \sqrt{\dfrac{p(a_i/a_{i-1})}{1-p(a_i/a_{i-1})} + \dfrac{1-p(a_i/a_{i-1})}{p(a_i/a_{i-1})} - 2}, & p(a_i/a_{i-1}) \geq 0.5 \\ -0.5 \cdot \sqrt{\dfrac{p(a_i/a_{i-1})}{1-p(a_i/a_{i-1})} + \dfrac{1-p(a_i/a_{i-1})}{p(a_i/a_{i-1})} - 2}, & p(a_i/a_{i-1}) < 0.5 \end{cases}$$

$$d(a_i/a_{i+1}) = \begin{cases} 0.5 \cdot \sqrt{\dfrac{p(a_i/a_{i+1})}{1-p(a_i/a_{i+1})} + \dfrac{1-p(a_i/a_{i+1})}{p(a_i/a_{i+1})} - 2}, & p(a_i/a_{i+1}) \geq 0.5 \\ -0.5 \cdot \sqrt{\dfrac{p(a_i/a_{i+1})}{1-p(a_i/a_{i+1})} + \dfrac{1-p(a_i/a_{i+1})}{p(a_i/a_{i+1})} - 2}, & p(a_i/a_{i+1}) < 0.5 \end{cases}$$

where $d(a_i)$ is the amount of information that triggers a piece of text a_i; $d(a_i/a_{i-1})$ is the amount of information that triggers a piece of text a_i after fragment a_{i-1}; $d(a_i/a_{i+1})$ is the amount of information that triggers a piece of text a_i before fragment a_{i+1}.

b) Let us calculate the difference in the determination (amount of information) by applying Formula (9) in (19)

$$\Delta d(a_i/a_{i-1}) = d(a_i/a_{i-1}) \cdot \sqrt{d(a_i)^2 + 1} - d(a_i) \cdot \sqrt{d(a_i/a_{i-1})^2 + 1};$$

$$\Delta d(a_i/a_{i+1}) = d(a_i/a_{i+1}) \cdot \sqrt{d(a_i)^2 + 1} - d(a_i) \cdot \sqrt{d(a_i/a_{i+1})^2 + 1},$$

where $\Delta d(a_i/a_{i-1})$ is the size of influence which appears in the process of

production of fragment a_{i-1} on the probability of fragment a_i as the following fragment (non-force influence of a human brain on the presence of fragment a_i when creating fragment a_{i-1}); $\Delta d\left(a_i/a_{i+1}\right)$ is the size of influence which appears in the process of production of fragment a_{i+1} on the probability of production of fragment a_i as the previous fragment (the non-force influence of a human brain on the presence of fragment a_i when creating fragment a_{i+1}).

c) The non-force influence on the appearance of text fragment a_i (23):

$$\Delta d\left(a_i/a_{i-1}a_{i+1}\right)=\Delta d\left(a_i/a_{i-1}\right)+\Delta d\left(a_i/a_{i+1}\right),$$

where $\Delta d\left(a_i/a_{i-1}a_{i+1}\right)$ is a combined additional influence that appears in the process of production of fragments a_{i-1} and a_{i+1} and which influences the possibility of fragment a_i being formed next.

d) New determination of the amount of information that triggers a piece of text a_i (from Formulae (9) and (26)):

$$d\left(a_i/a_{i-1}a_{i+1}\right)=d\left(a_i\right)\cdot\sqrt{\left(\Delta d\left(a_i/a_{i-1}a_{i+1}\right)\right)^2+1}+\Delta d\left(a_i/a_{i-1}a_{i+1}\right)\cdot\sqrt{\left(d\left(a_i\right)\right)^2+1},$$

where $d\left(a_i/a_{i-1}a_{i+1}\right)$ is the amount of information that triggers a piece of text a_i taking into account non-force influence.

e) The estimate of probability of fragment a_i presence (after both influences) (12):

$$p_9\left(a_i/a_{i-1}a_{i+1}\right)=0.5+\frac{d\left(a_i/a_{i-1}a_{i+1}\right)}{2\cdot\sqrt{\left(d\left(a_i/a_{i-1}a_{i+1}\right)\right)^2+1}},\tag{35}$$

where $p_9\left(a_i/a_{i-1}a_{i+1}\right)$ is the estimate of conditional probability of text fragment a_i sequence, through non-force influence on the possibility of fragment a_i production, which appears in the process of producing fragments a_{i-1} and a_{i+1}.

7. Methods' Effectiveness Evaluation Criteria

The effectiveness of the methods is to be evaluated by the following criteria:

1) Standard deviation of the estimate of combined conditional probability, received by method M_s from statistical combined conditional probability of text fragment sequence:

$$\sigma_s=\sqrt{\frac{\sum_{i=1}^{N_s}\left(p_s\left(a_i/a_{i-1}a_{i+1}\right)-p\left(a_i/a_{i-1}a_{i+1}\right)\right)^2}{N_s}},\tag{36}$$

where N_s is quantity of text fragments about which the estimate of combined conditional probability is calculated by method M_s; $p_s\left(a_i/a_{i-1}a_{i+1}\right)$ is estimate of combined conditional probability of text fragment sequence calculated by method M_s; $p\left(a_i/a_{i-1}a_{i+1}\right)$ is combined conditional probability of text fragment sequence; σ_s is standard deviation of estimate of combined conditional probability of text fragment sequence received by method M_s from statistical combined conditional probabilities.

The advantage of non-force interaction method over the other will be calculated by dividing the difference between deviation of NFIM and the smallest deviation of other methods by deviation of NFIM and multiplying by 100%:

$$Y_\sigma = \frac{\sigma_9 - \max\limits_{1\le i\le 8}\sigma_i}{\sigma_9}\cdot 100\%, \tag{37}$$

where Y_σ is the advantage of non-force interaction method in standard deviation of estimate of combined conditional probability of text fragment sequence.

2) The percentage of correctly predicted text fragments:

$$\mu_s = \frac{\sum_{a_{i-1}a_{i+1}}\left(n\left(a_i/a_{i-1}a_{i+1}\right)\left|\max_{a_i}\left[p_s\left(a_i/a_{i-1}a_{i+1}\right)\right]\right.\right)}{\sum_{a_{i-1}a_{i+1}} n\left(a_{i-1}a_{i+1}\right)}\cdot 100\%, \tag{38}$$

where μ_m is percentage of correct predictions of text fragment sequence made by M$_s$; $p_s\left(a_i/a_{i-1}a_{i+1}\right)$ is estimate of combined conditional probability of text fragment a_i sequence made by method M$_s$; $n\left(a_i/a_{i-1}a_{i+1}\right)$ is the quantity of letter combinations where the combination of fragments $a_{i-1}a_ia_{i+1}$ is present; $n\left(a_{i-1}a_{i+1}\right)$ is the quantity of letter combinations where the combination of fragments a_{i-1} < any fragment > a_{i+1} is present.

The advantage of non-force interaction method over the others will be calculated by diving the difference between the percentage (quantity) of correctly predicted text fragments, made by NFIM and the biggest percentage (quantity) of text fragments predicted correctly by other methods by percentage (quantity) of text fragments predicted correctly by NFIM and multiplying by 100%:

$$Y_\mu = \frac{\mu_9 - \max\limits_{1\le i\le 8}\mu_i}{\mu_9}\cdot 100\%, \tag{39}$$

where Y_μ is the advantage of non-force interaction method in predicting text fragments appearance.

3) Standard deviation of the rank of estimate of combined conditional probability received by method M_s from the range of nonzero statistical combined conditional probability of text fragment sequence:

$$V_s = \sqrt{\frac{\sum_{i=1}^{N_s}\left(u_s\left(a_i/a_{i-1}a_{i+1}\right) - u\left(a_i/a_{i-1}a_{i+1}\right)\right)^2}{N_s}}, \tag{40}$$

under the condition that

$$p\left(a_i/a_{i-1}a_{i+1}\right) > 0,$$

where V_s is standard deviation of the rank of estimate of combined conditional probability of text fragment sequence from the rank of statistical conditional probability in method M$_s$; $u_s\left(a_i/a_{i-1}a_{i+1}\right)$ is rank of estimate of combined conditional probability of text fragment sequence received by method M$_s$; $p\left(a_i/a_{i-1}a_{i+1}\right)$ is combined conditional probability of text fragment sequence; $u\left(a_i/a_{i-1}a_{i+1}\right)$ is rank of combined conditional probability of text fragment sequence.

As a rank, an ordinate number in the ordered range of probabilities of text fragment sequence is understood. For the sequence:

$$p\left(a_{i_1}/a_{i-1}a_{i+1}\right) > p\left(a_{i_2}/a_{i-1}a_{i+1}\right) > \cdots > p\left(a_{i_f}/a_{i-1}a_{i+1}\right) > \cdots > p\left(a_{i_k}/a_{i-1}a_{i+1}\right),$$

the rank of conditional probabilities will be the following:

$$u\left(a_{i_1}/a_{i-1}a_{i+1}\right) = 1;$$

$$u\left(a_{i_2}/a_{i-1}a_{i+1}\right) = 2;$$

$$\vdots$$

$$u\left(a_{i_f}/a_{i-1}a_{i+1}\right) = f;$$

$$\vdots$$

$$u\left(a_{i_k}/a_{i-1}a_{i+1}\right) = k.$$

For the sequence received by method M_s:

$$p_s\left(a_{j_1}/a_{i-1}a_{i+1}\right) > p_s\left(a_{j_2}/a_{i-1}a_{i+1}\right) > \cdots > p_s\left(a_{j_r}/a_{i-1}a_{i+1}\right) > \cdots > p_s\left(a_{j_k}/a_{i-1}a_{i+1}\right),$$

the rank of estimates of combined conditional probabilities will be the following:

$$u_s\left(a_{j_1}/a_{i-1}a_{i+1}\right) = 1;$$

$$u_s\left(a_{j_2}/a_{i-1}a_{i+1}\right) = 2;$$

$$\vdots$$

$$u_s\left(a_{j_r}/a_{i-1}a_{i+1}\right) = r;$$

$$\vdots$$

$$u_s\left(a_{j_k}/a_{i-1}a_{i+1}\right) = k.$$

This criterion is chosen because the estimate of combined conditional probability received by any method and the probability itself do not coincide. This is caused, in the first place, by the synergetic effect. That is by one more direction of evaluation was chosen **the ranking of probability estimates and probabilities themselves of text fragments.** How each method allows to rank fragments correctly according to the frequency of their appearance in a natural text. As well as the criterion of prediction, this one is really important and often used in artificial intelligence systems.

The advantage of non-force interaction method over the others will be calculated by diving the difference between the deviation of NFIM and the smallest deviation of other methods by the deviation of NFIM and multiplying by 100%:

$$Y_V = \frac{V_9 - \max\limits_{1\le i\le 8} V_i}{V_9} \cdot 100\%, \tag{41}$$

where Y_V is the advantage of non-force interaction method in standard deviation of the rank of estimate of combined conditional probability from the rank of combined conditional probability of text fragment sequence.

Methods provided in pp. 5-8 can form the estimate of combined conditional probability which is more than 1. To eliminate this, the author used the normalizing, which is division by the biggest value of result. However, the experiments showed that in this case the deviation of combined conditional probability estimates from the conditional probability of text fragment sequence itself has increased. In order not to be "blamed" in lowering the results of approximation methods, below there are non-normalized values of approximation methods are compared to non-force interaction method.

As to the criteria of the best prediction and the deviation in ranks, normalizing does not influence the result. Since for predicting the highest estimate is chosen (and normalizing does not change this choice), and ranking decides on the estimate sequence (normalizing does not change it either).

8. Software Tools

To implement the methodology stated in Section 5, there was developed a program "Experiments on NFIT" in VBA Access. It contains the interpreter of the texts (implements pp. 1-2 of Section 5), statistics formation module (pp. 3-5 of Section 5), module to calculate estimates of combined conditional probability by different methods (pp. 6-8 of Section 5), module of results analysis and display (pp. 9-10 of Section 5).

Time spent on conduction of experiments connected with the application of provided methodology to four natural language texts on a desktop computer (PC) is approximately 200 hours. Generalized algorithm of calculation of estimates of combined conditional probabilities of text fragment sequence includes the following points:

1) Transformation of the text into the set of 1- and 2-letter sequenced fragments.

2) Calculation of statistical characteristics of a text (unconditional and conditional probabilities of text fragment sequence).

3) Setting of the parameters of calculation (random or input text, with or without reference to unconditional probability of current text fragment sequence, with division of the text into educating and check samples, merger of texts, etc.)

4) Calculation of coefficients in approximation equations.

5) Calculation of estimates of combined conditional probability by all methods.

6) Evaluation of results and formation of reports.

7) Display of results.

To make the algorithm of calculation more understandable, let us consider the example of calculation of estimates of combined conditional probability by different methods.

9. The Example of Calculation (Fragment of "English" Base)

Let us consider three two-letter fragments: $a_{i-1} =$ "ab"; $a_i =$ "le"; $a_{i+1} =$ "in".

$$p(\text{“le”}) = 0.006458;$$

$$p(\text{“le”}/\text{“ab”}) = 0.233333;$$

$$p(\text{“le”}/\text{“in”}) = 0.002379;$$

$$p(\text{“le”}/\text{“ab”“in”}) - ?$$

For each combination of fragments, the calculation of estimate of combined conditional probability is made by the methods provided above. The results are summarized in Table 1. What conclusions can be made:

1) The best method for this combination of fragments if the one based on the results received in the non-force interaction theory.

2) None of the methods gives exact value of combined conditional probability. If the intellectual apparatus of a human had also developed for 13.8 billion years, the regularity in formation of text fragments would exactly correspond to the formulae interpreted from the laws of mechanical movement?!

It is a joke, of course. The peculiarity of human intellectual apparatus is the implementation of synergistic features of a language. The combination of fragments, by contents, is considerably bigger than their total sum! However, the fact that the non-force interaction theory provides the biggest approximation even in this example requires a more serious confirmation, as it is wrong to judge by one fragment. Within the experiment, there were processed the statistics by all fragments in the selected works of writing [19]-[29]. Let is consider the results of the conducted experiments.

10. Results of Experiments

Having set input texts (Section 5) for the developed program (Section 8) and calculated unconditional and conditional probabilities of appearance of fragments (letter combinations) of texts of one- and two-letter length, there were formed statistical tables, in accordance to the chosen criteria (Section 7).

Unconditional statistical probabilities for 1-letter fragments are shown in Tables 2-5. For 2-letter fragments such tables contain 941 records (“Russian” base), 609 (“English base”), 956 (“Ukrainian” base), 709 (“German” base).

Let us consider the results received.

Experiment 1. Calculation of approximation coefficients.

For each text, there are determined approximation coefficients that correspond to conditions (28), (30), (32), and (34). Approximation coefficients are stated in Table 6. As it can be seen from Table 6, these coefficients are different enough for the texts in different languages. This means that it is impossible to find such linear approximation which will be optimal for different texts in different languages.

Experiment 2. Determination of standard deviation of combined conditional probability estimate from statistical combined conditional probability of text fragment sequence.

Evaluation criterion: standard deviation (36).

Table 1. Results of estimating of combined conditional probability of appearance of fragment "le" with neighbouring fragments "ab" and "in" in "English" base made by different methods.

№	Name of method	Formula	Value	Difference from actual value 0.192308	Difference rank
1.	By unconditional probability	$p(\text{“le”})$	0.006458	−0.18585	8
2.	By conditional probability of a previous fragment	$p(\text{“le”}/\text{“ab”})$	0.233333	0.041025	5
3.	By conditional probability of a following fragment	$p(\text{“le”}/\text{“in”})$	0.002379	−0.189929	9
4.	By mean probability	$\frac{p(\text{“le”}/\text{“ab”})+p(\text{“le”}/\text{“in”})}{2}$	0.117856	−0.07445	7
5.	Method 1. Coefficients: $\alpha_1=0.664$; $\alpha_2=0.714$; $\alpha_3=-0.001$.	$\alpha_1\cdot p(\text{“le”}/\text{“ab”})+\alpha_2\cdot p(\text{“le”}/\text{“in”})+\alpha_3$	0.155632	−0.03668	4
6.	Method 2. Coefficients: $\beta_1=0.484$; $\beta_2=3.562$; $\beta_3=-0.002$.	$\beta_1\cdot p(\text{“le”}/\text{“ab”})+\beta_2\cdot p(\text{“le”}/\text{“in”})+\beta_3$	0.119407	−0.0729	6
7.	Method 3. Coefficient: $\gamma_1=0.707$; $\gamma_2=0.743$; $\gamma_3=0.001$.	$\gamma_1\cdot p(\text{“le”}/\text{“ab”})+\gamma_2\cdot\left(p(\text{“le”}/\text{“in”})-p(\text{le})\right)+\gamma_3$	0.167734	−0.02457	3
8.	Method 4. Coefficient: $\delta_1=0.682$; $\delta_2=1.990$; $\delta_3=0.004$.	$\delta_1\cdot p(\text{“le”}/\text{“ab”})+\delta_2\cdot\left(p(\text{“le”}/\text{“in”})-p(\text{le})\right)+\delta_3$	0.167867	−0.02444	2
9.	**Non-force interaction method** a) Calculation of influence size $d(\text{“le”})$ $\Delta d(\text{“le”}/\text{“ab”})$ $\Delta d(\text{“le”}/\text{“in”})$		−6.161437 3.348290 −0.522639		
	b) Combined additional influence $\Delta d(\text{“le”}/\text{“ab”“in”})$	$\Delta d(\text{“le”}/\text{“ab”})+\Delta d(\text{“le”}/\text{“in”})$	2.825651	**−0.01171**	**1**
	c) New quantity of information about the presence of fragment "le" $d(\text{“le”}/\text{“ab”“in”})$	$d(\text{“le”})\cdot\sqrt{\left(\Delta d(\text{“le”}/\text{“ab”“in”})\right)^2+1}+\Delta d(\text{“le”}/\text{“ab”“in”})\cdot\sqrt{\left(d(\text{“le”})\right)^2+1}$	−0.830304		
	d) Estimate of probability of fragment a_i presence (after both influences).	$0.5+\frac{d(\text{“le”}/\text{“ab”“in”})}{2\cdot\sqrt{\left(d(\text{“le”}/\text{“ab”“in”})\right)^2+1}}$	**0.180596**		

Table 2. Frequency of appearance of Russian alphabet letters in "Russian" base.

Fragment	Frequency	Probability
а	175,827	0.083292
б	37,245	0.017644
в	98,741	0.046775
г	42,105	0.019946
д	65,559	0.031056
е	176,940	0.083819
ж	23,074	0.010931
з	37,060	0.017556
и	145,085	0.068729
й	23,616	0.011187
к	71,762	0.033995
л	108,142	0.051229
м	63,055	0.029870
н	138,196	0.065466
о	248,121	0.117539
п	56,211	0.026628
р	95,332	0.045160
с	114,635	0.054304
т	125,962	0.059670
у	57,720	0.027343
ф	4385	0.002077
х	18,300	0.008669
ц	7694	0.003645
ш	20,128	0.009535
щ	6390	0.003027
ъ	923	0.000437
ы	41,442	0.019632
ь	41,546	0.019681
э	6409	0.003036
ю	12,849	0.006087
я	46,515	0.022035

Table 3. Frequency of Latin letter appearance in "English" base.

Fragment	Frequency	Probability
a	284,554	0.07714
b	58,963	0.015984

Continued

c	84,602	0.022935
d	143,045	0.038778
e	433,237	0.117446
f	77,559	0.021025
g	65,513	0.01776
h	233,002	0.063164
i	250,734	0.067971
j	4347	0.001178
k	32,997	0.008945
l	165,938	0.044984
m	107,653	0.029184
n	238,352	0.064615
o	311,075	0.084329
p	54,977	0.014904
q	3090	0.000838
r	232,522	0.063034
s	238,711	0.064712
t	319,496	0.086612
u	127,585	0.034587
v	36,926	0.01001
w	86,719	0.023509
x	5223	0.001416
y	90,381	0.024501
z	1616	0.000438

Table 4. Frequency of appearance of German alphabet letters in "German" base.

Fragment	Frequency	Probability
a	49,004	0.050259
ä	4965	0.005092
b	17,785	0.018241
c	36,940	0.037886
d	44,264	0.045398
e	149,403	0.15323
f	17,020	0.017456
g	30,778	0.031566

Continued

h	61,621	0.063199
i	76,184	0.078135
j	2546	0.002611
k	11,361	0.011652
l	39,694	0.040711
m	27,415	0.028117
n	92,783	0.09516
o	26,590	0.027271
ö	3229	0.003312
p	7638	0.007834
q	283	0.00029
r	65,295	0.066968
s	64,256	0.065902
ß	3486	0.003575
t	60,144	0.061685
u	37,803	0.038771
ü	6380	0.006543
v	7429	0.007619

Table 5. Frequency of appearance of Ukrainian alphabet letters in "Ukrainian" base.

Fragment	Frequency	Probability
а	35,710	0.09545
б	7862	0.02101
в	16,595	0.04436
г	6644	0.01776
д	12,866	0.03439
е	20,693	0.05531
є	2333	0.00624
ж	3885	0.01038
з	7913	0.02115
и	25,255	0.06750
і	15,511	0.04146
ї	1914	0.00512

Continued

й	6100	0.01630
к	12,404	0.03315
л	17,528	0.04685
м	12,277	0.03281
н	20,282	0.05421
о	36,527	0.09763
п	9957	0.02661
р	15,599	0.04169
с	16,822	0.04496
т	21,300	0.05693
у	12,561	0.03357
ф	120	0.00032
х	4285	0.01145
ц	1831	0.00489
ч	5345	0.01429
ш	2862	0.00765
щ	1806	0.00483
ь	6855	0.01832
ю	4348	0.01162
я	8146	0.02177

Processing of texts by the program "Experiments by NFIT" allowed to receive the integral values of deviation of calculated combined conditional probability estimates (36) (Table 7). As it can be seen from Table 7, non-force interaction method gives smaller deviation of probability estimate from its statistical value than other methods in all the texts except "Russian" base (one-letter fragments).

Having applied Fisher distribution to values stated in Table 7, it is received a confirmation of statistical hypotheses about exceedance of variance of other methods over the variance of non-force interaction method [31] (excluding Russian text with singular length of fragments) with significance level of $\alpha = 0.01$ (Table 8). In Table 8, for comparison it is considered the deviation variance of combined conditional probability estimate from combined conditional probability itself received by NFIM and the smallest deviation variance of combined conditional probability estimate from combined conditional probability itself received by other methods. The confirmation of these hypotheses indicates that for other methods (which have a bigger deviation with the same degrees of freedom) they will be confirmed too.

Table 6. Coefficients in approximation methods that provide the minimum standard deviation of combined conditional probability estimate from combined conditional probability of text fragment sequence (fulfilment of conditions of Formulas (28), (30), (32) and (34)).

Base	L	Method 1			Method 2			Method 3			Method 4		
		α_1	α_2	α_3	β_1	β_2	β_3	χ_1	χ_2	χ_3	δ_1	δ_2	δ_3
"Russian"	1	0.613	0.588	−0.006	0.292	1.592	−0.005	0.741	0.669	0.008	0.663	1.085	0.018
	2	0.663	0.673	−0.001	0.472	4.293	−0.002	0.692	0.689	0.000	0.644	2.367	0.004
"German"	1	0.785	0.527	−0.011	0.529	1.130	−0.011	0.915	0.684	0.003	0.815	1.006	0.010
	2	0.723	0.708	−0.002	0.496	3.017	−0.002	0.777	0.760	0.000	0.720	1.948	0.003
"Ukrainian"	1	0.720	0.618	−0.011	0.374	1.419	−0.007	0.857	0.786	0.004	0.746	1.226	0.016
	2	0.782	0.813	−0.002	0.498	4.516	−0.002	0.829	0.851	0.000	0.741	3.074	0.004
"English"	1	0.769	0.600	−0.014	0.521	1.186	−0.015	0.914	0.715	0.003	0.809	1.040	0.012
	2	0.664	0.714	−0.001	0.484	3.562	−0.002	0.707	0.743	0.001	0.682	1.990	0.004

Table 7. Standard deviation of combined conditional probability estimates from statistical combined conditional probability of text fragment sequence.

Base	L	By unconditional probability	By 1st fragment	By 2nd fragment	By mean probability	Method 1	Method 2	Method 3	Method 4	**NFIM**	Advantage of NFIM (%) (37)
"Russian"	1	0.09457	0.08879	0.09018	0.08228	0.08170	**0.07842**	0.08182	0.08117	0.08090	−3.05
	2	0.04313	0.04236	0.04227	0.04121	0.04107	0.04028	0.04106	0.04073	**0.03951**	1.96
"German"	1	0.10891	0.09320	0.10388	0.08973	0.08714	0.08690	0.08559	0.08596	**0.08142**	5.12
	2	0.05479	0.05296	0.05322	0.05148	0.05110	0.05019	0.05102	0.05058	**0.04839**	3.71
"Ukrainian"	1	0.08289	0.07339	0.07588	0.06849	0.06694	0.06486	0.06544	0.06455	**0.06204**	4.06
	2	0.04803	0.04731	0.04715	0.04662	0.04637	0.04561	0.04634	0.04589	**0.04491**	1.56
"English"	1	0.10037	0.08645	0.09298	0.08129	0.07845	0.07769	0.07768	0.07787	**0.07466**	4.05
	2	0.04393	0.04292	0.04265	0.04149	0.04127	0.04049	0.04124	0.04092	**0.03948**	2.55

Note: The best results are highlighted.

Table 8. Verification of hypothesis about exceedance of the smallest standard deviation received by different methods over the standard deviation received by non-force interaction method using Fisher criterion.

Base	L	$\min_{i\neq 9}\sigma_i$	σ_9	$\min_{i\neq 9}\sigma_i^2$	σ_9^2	$F_{n;n}=\frac{\min_{i\neq 9}\sigma_i^2}{\sigma_9^2}$	n	$F_{0,01;n;n}$	Confirmed hypothesis
"Russian"	1	**0.07842**	0.0809	0.00615	0.006545	1.064249*	32,758	1.02604	$\min_{i\neq 9}\sigma_i^2 < \sigma_9^2$
	2	0.04028	**0.03951**	0.001622	0.001561	1.039357	65,238,532	1.00058	$\min_{i\neq 9}\sigma_i^2 > \sigma_9^2$
"German"	1	0.08596	**0.08142**	0.007389	0.006629	1.11463	23,349	1.03092	$\min_{i\neq 9}\sigma_i^2 > \sigma_9^2$
	2	0.05019	**0.04839**	0.002519	0.002342	1.075779	21,284,121	1.00101	$\min_{i\neq 9}\sigma_i^2 > \sigma_9^2$

Continued

"Ukrainian"	1	0.06455	**0.06204**	0.004167	0.003849	1.082552	32,178	1.02628	$\min_{i\neq 9}\sigma_i^2 > \sigma_9^2$
	2	0.04561	**0.04491**	0.00208	0.002017	1.031416	42,253,549	1.00072	$\min_{i\neq 9}\sigma_i^2 > \sigma_9^2$
"English"	1	0.07768	**0.07466**	0.006034	0.005574	1.082536	16,783	1.03657	$\min_{i\neq 9}\sigma_i^2 > \sigma_9^2$
	2	0.04049	**0.03948**	0.001639	0.001559	1.05182	40,791,720	1.00073	$\min_{i\neq 9}\sigma_i^2 > \sigma_9^2$

* - there is a hypothesis considered about exceedance of standard deviation received by non-force interaction method over the smallest standard deviation received by other methods.

However, there is one exception, which is the Russian text. As it is shown in the results of additionally conducted experiments with other texts in Russian ("Anna Karenina" by Lev Tolstoy and "Master and Margarita" by Mikhail Bulgakov), this regularity is preserved. There was a separate research conducted, which showed that the main reason is four words that happen often and contain letter which almost do not appear in other combinations. Thus, in the word "всё" the probability of letter "с" being the predecessor of "ё" is 0.962389. The letter combination "сё" has significant synergetic effect as the probability of appearance of letter "с" equals 0.05327, whereas for letter "ё" it is 0.0004175. Therefore, for all letter combinations "асё", "бсё", …, "ясё" the combined conditional probability estimate p_9 ("с"/"аё"), p_9 ("с"/"бё"), …, p_9 ("с"/"яё") is calculated by deviation 0.962389 from 0.05327 (because letter "ё" provides the probability almost equal to 1 of the fact that there will be letter combination with letter "с" in it). However, in reality in the text, letter combinations "асё", "бсё", …, "ясё" (except "всё") did not appear (probability equals 0). The same case is for words "эт[о, а, у]". In these words, letter "э" appears with probability 0.90176.

Synergetic effect and restriction of choices in letter combinations for the stated words gives the difference in standard deviation which equals 0.00302. This exceeds common difference in deviations (0.00248, see Table 7, columns "Method 2" and "NFIM"). Without words "всё" and "это" the values of deviations are completely different (Table 9).

The verification showed that in other languages there are no such frequently used words that form total high probability of appearance of certain letters.

Experiment 3. Calculation of standard deviation of combined conditional probability estimate from statistical combined conditional probability of sequence of fragments of one part of texts using the approximation coefficients of the other texts.

Evaluation criterion: standard deviation (36).

As it was shown in experiment 1, there is a problem with using the methods which are based on approximation. Coefficients that give the biggest approximation of linear equation to the values of combined conditional probability in one text are not optimal for other texts. For example, coefficients of method 2 (see Table 7) for one-letter fragments of the Russian texts, which provide result even better than NFIM, were not the best for other texts. There were the experiments

Table 9. Standard deviation of combined conditional probability estimates from statistical combined conditional probability of text fragment sequence in "Russian" base.

Base	L	By unconditional probability	By 1st fragment	By 2nd fragment	By mean probability	Method 1	Method 2	Method 3	Method 4	**NFIM**	Advantage of NFIM (%) (37)
All words	1	0.09457	0.08879	0.09018	0.08228	0.08170	**0.07842**	0.08182	0.08117	0.08090	−3.05
Without words "всё" and "это"	1	0.095	0.08781	0.08957	0.08204	0.08112	0.07844	0.08079	0.08021	**0.0779**	0.7

Note: the best results are highlighted.

conducted which showed the increase of standard deviation if substituting coefficients of one part of texts with coefficients from the other texts (Table 10).

Taking into account the results provided in Table 10, it is possible to state once again that method which is based on the model of non-force interaction is universal (is suitable for different texts without changes in coefficients), while methods based on approximation can work effectively only with those texts for which the formula is developed (for example, only this piece of writing or, probably, works of only one writer).

Comment to experiments 2 and 3. *It has already been said that with set initial data (unconditional and certain conditional probabilities) it is impossible to calculate combined conditional probability. However, it can be evaluated. As it is seen from Tables 7-10, the biggest approximation to combined conditional probability is provided by non-force interaction method. Moreover, in methods that are based on approximation the calculation of coefficients was made from the position of minimum deviation from combined conditional probability (Formulas (28), (30), (32) and (34)). Nevertheless, non-force interaction method is still the best.*

Undoubtedly, any method, including the stated one, does not allow to calculate combined conditional probability precisely. The deviation of non-force interaction method is explained by the structure of natural language where a word is more than just the sum of letters, but also contents, because letters do not have contents, and a word has.

This is a synergetic effect of natural language. Earlier (in comments to Table 9) it was explained how words "всё" and "это" influence the statistics. Indeed, if one letter from the text is taken, let it be "с", it speaks about nothing. There are a lot of possible words with this letter. For two letters "ст" the probability of appearance of any word containing letters "ст" is higher than probability of appearance of a word with letter "т" or "с". Let us add one more letter "ста". It is even higher. If we add letter "н", for this combination of letters (a word "стан" – "condition" in Ukrainian), the probability is almost 1. Therefore, for more "informative" letter combinations the probability tends to 1, and for non-informative it tends to 0. This creates a deviation when using any other method. However,

Table 10. Standard deviation of estimates from probability of text fragment sequence by non-force interaction method and approximation methods with different coefficients.

Base	Base from which coefficients are taken	L	Approximation methods				**Non-Force Interaction Method**
			Method 1	Method 2	Method 3	Method 4	
"Russian"	"Russian"	1	0.08170	**0.07842**	0.08182	0.08117	0.08090
		2	0.04107	0.04028	0.04106	0.04073	**0.03951**
	"German"	1	0.08226	**0.07937**	0.08243	0.08181	0.08090
		2	0.04109	0.04039	0.04109	0.04078	**0.03951**
"German"	"German"	1	0.08714	0.08690	0.08559	0.08596	**0.08142**
		2	0.05110	0.05019	0.05102	0.05058	**0.04839**
	"English"	1	0.08727	0.08698	0.08561	0.08597	**0.08142**
		2	0.05112	0.05024	0.05104	0.05059	**0.04839**
"Ukrainian"	"Ukrainian"	1	0.06694	0.06486	0.06544	0.06455	**0.06204**
		2	0.04637	0.04561	0.04634	0.04589	**0.04491**
	"Russian"	1	0.06723	0.06500	0.06595	0.06489	**0.06204**
		2	0.04642	0.04562	0.04640	0.04596	**0.04491**
"English"	"English"	1	0.07845	0.07769	0.07768	0.07787	**0.07466**
		2	0.04127	0.04049	0.04124	0.04092	**0.03948**
	"Ukrainian"	1	0.07854	0.07831	0.07787	0.07825	**0.07466**
		2	0.04134	0.04060	0.04131	0.04120	**0.03948**

Note: the best results are highlighted.

the fact that the non-force interaction method gives the smallest deviation from statistical combined conditional probability confirms that in the mechanisms of speech production (human intellectual apparatus) the processes of interaction can be based on the algorithms which result from the information-probabilistic interpretation of mechanical movement.

It is no wonder, probably, that in the basis of different processes of interaction (on micro- and macro-levels of matter existence) there are same laws. It would be strange if this did not happen. Thus, it can be used in many situations, including the development of artificial intelligence systems.

One of the tasks connected to such systems is the task of predicting. To create a reliable forecast in stochastic subject fields, it is not really crucial which is the probability of occurrence of all events, the most important is which event has the highest probability. There were conducted experimental researches that allow to determine the most probable text fragment by the highest estimate of combined conditional probability using different methods. Let us consider the results.

Experiment 4. Forecasting of text fragment sequence

Evaluation criterion: the percentage of correctly predicted text fragments (38)

The forecast is based on the choice of the biggest combined conditional probability estimate (38). As a result of processing of the texts results stated in Table 11 were obtained.

As it can be seen in Table 11, the advantage of the non-force interaction method is rather substantial. Among the methods which are based on linear approximation the best is the one that uses the difference between the smallest conditional probability and unconditional probability (33). Moreover, it is more effective than the approximation methods that are oriented at conditional probabilities only. However, it is the deviation of conditional probability from unconditional one that interprets relative velocity of objects and is in the basis of the Formula (19). This, once again, confirms that in a human brain the value of non-force (information) action is represented by the deviation of conditional probability of certain reaction (in experiments it was language) from unconditional one. Returning to NFIM, it gives the best result even without aiming at the selection of the biggest conditional probability (which is dictated with the formulas 29, 33).

Experiment 5. Ranking of estimates of combined conditional probability of text fragment sequence.

Criterion of evaluation: deviation in ranks (40)

Within experiments, there was made ranking of probability estimates with comparison with the ranks of the probabilities themselves. The deviation of ranks of combined conditional probability estimates received by different methods from the rank of statistical combined conditional probability was evaluated. The deviation in ranks by Formula (40) for different methods is provided in Table 12.

The received distribution of ranks states even bigger advantage of non-force interaction method in comparison with others provided in this paper. With the

Table 11. Forecasting of text fragment sequence by different methods.

Base	L	By unconditional probability (μ_1)	By 1st fragment (μ_2)	By 2nd fragment (μ_3)	By mean probability (μ_4)	Method 1 (μ_5)	Method 2 (μ_6)	Method 3 (μ_7)	Method 4 (μ_8)	NFIM (μ_9)	Advantage of NFIM (%) (39)
"Russian"	1	11.54	19.37	19.17	22.92	22.88	25.61	25.22	26.19	**27.98**	6.40
	2	1.77	12.58	12.88	20.06	20.04	26.45	20.15	25.08	**28.02**	5.60
"German"	1	16.05	26.97	26.64	34.27	33.25	35.80	35.22	36.76	**38.44**	4.37
	2	3.63	15.83	13.89	22.90	22.89	26.28	23.43	26.62	**30.25**	12.00
"Ukrainian"	1	9.21	17.30	17.72	21.77	21.42	24.28	23.36	25.21	**25.83**	2.40
	2	1.34	9.12	9.63	15.38	15.38	19.93	15.44	19.94	**22.48**	11.30
"English"	1	11.79	22.35	21.93	28.11	27.86	29.21	28.66	29.25	**30.39**	3.75
	2	3.19	12.61	12.36	19.44	19.33	23.35	19.89	23.06	**25.61**	8.82

Note: the best results are highlighted.

help of this method, it is possible to determined more exactly which of the variations of forecast can be counted upon and which cannot.

Experiment 6. The graph of deviation of combined conditional probability estimate received by the non-force interaction method from statistical combined conditional probability of text fragment sequence.

The deviations of values of combined conditional probability received by the non-force interaction method from statistical combined conditional probability of text fragment sequence in "English" base which were received in the process of experiments are provided in Table 13. According to these data, the graph was created (Figure 3)

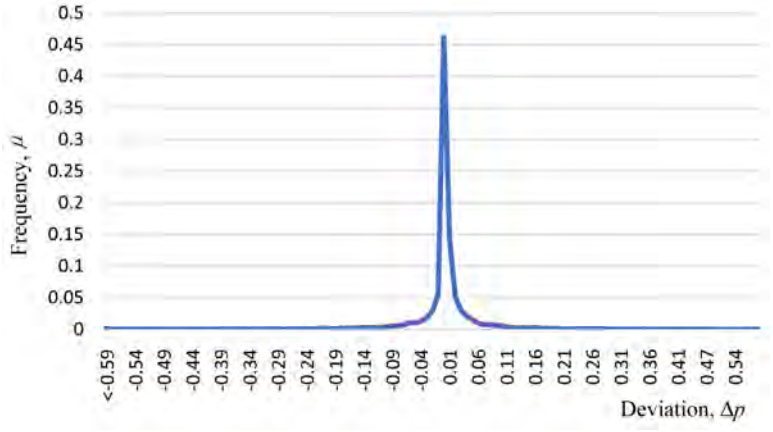


Figure 3. Distribution of deviation of combined conditional probability estimate received by non-force interaction method from its statistical value in "English" base (one-letter fragments).

Table 12. Standard deviation of ranks of combined conditional probability estimates and the rank of conditional probability of text fragment sequence.

Base	L	% of correct prediction of text fragment sequence									Advantage of NFIM (%) (41)
		By unconditional probability	By 1st fragment	By 2nd fragment	By mean probability	Method 1	Method 2	Method 3	Method 4	**Non-force interaction method**	
"Russian"	1	37.74	25.71	24.19	19.62	19.65	16.38	17.28	19.18	**11.48**	42.62
	2	21.88	14.63	14.59	11.80	11.81	11.27	10.89	11.17	**8.98**	21.30
"German"	1	24.10	18.11	13.11	12.47	12.99	11.61	10.85	10.70	**7.60**	40.85
	2	20.02	14.17	12.11	11.18	11.20	10.15	10.11	9.78	**7.52**	30.03
"Ukrainian"	1	41.62	27.53	25.18	23.81	23.91	18.28	19.76	17.09	**10.07**	69.78
	2	3.27	2.07	2.15	1.84	1.84	1.74	1.73	1.62	**1.50**	8.03
"English"	1	30.99	19.32	18.52	14.78	14.87	13.40	12.43	13.02	**8.74**	42.21
	2	63.06	42.61	41.05	32.79	32.79	31.63	29.96	31.07	**24.62**	21.69

Note: the best results are highlighted.

Table 13. The deviation of combined conditional probability estimate received by non-force interaction method ("English" base) from statistical combined conditional probabilities.

Deviation	Quantity
≤−0.60	32
−0.59	1
−0.57	2
−0.56	3
−0.55	4
−0.54	1
−0.53	1
−0.52	2
−0.51	1
−0.5	5
−0.49	5
−0.48	3
−0.47	3
−0.46	6
−0.45	1
−0.44	2
−0.43	5
−0.42	6
−0.41	8
−0.4	6
−0.39	6
−0.38	12
−0.37	9
−0.36	6
−0.35	10
−0.34	6
−0.33	14
−0.32	8
−0.31	13
−0.3	12
−0.29	10
−0.28	5

Continued

−0.27	13
−0.26	15
−0.25	16
−0.24	19
−0.23	12
−0.22	21
−0.21	29
−0.2	20
−0.19	18
−0.18	24
−0.17	30
−0.16	29
−0.15	35
−0.14	38
−0.13	43
−0.12	43
−0.11	46
−0.1	52
−0.09	75
−0.08	77
−0.07	98
−0.06	151
−0.05	141
−0.04	177
−0.03	250
−0.02	403
−0.01	762
0	6471
0.01	2015
0.02	724
0.03	427
0.04	298
0.05	212
0.06	135
0.07	99

Continued

0.08	105
0.09	89
0.1	78
0.11	52
0.12	36
0.13	38
0.14	39
0.15	27
0.16	39
0.17	25
0.18	26
0.19	26
0.2	19
0.21	15
0.22	13
0.23	12
0.24	11
0.25	17
0.26	14
0.27	7
0.28	8
0.29	5
0.3	9
0.31	5
0.32	4
0.33	3
0.34	2
0.35	9
0.36	2
0.37	4
0.38	3
0.39	4
0.4	5
0.41	2
0.42	5

Continued

0.43	3
0.45	1
0.46	2
0.47	1
0.48	2
0.49	1
0.5	1
0.51	1
0.54	1
0.55	1
0.56	1
0.59	1
≥0.60	5

As it can be seen from the graph, the received combined conditional probability estimates differ slightly from combined conditional probabilities themselves and form the peak of distribution in point "0" (deviation is absent).

To confirm the received results, there were conducted additional experiments which did not characterise the accuracy of determination of text structure, but rather show the versatility of methods developed in the non-force interaction theory.

11. Additional Experimental Researches

As it can be seen from the tables stated in Section 10, the best result is provided by the method based on the non-force interaction theory, even though, from the point of view of probabilistic approach, the approximation-based methods should be better, as coefficients for equations are chosen in such a way that they should give the smallest deviation square, thus the biggest approximation to the probabilities themselves.

Thus, a question arises. What if non-force interaction method does not operate the information part of the text production, but rather uses only the deviations of probabilities as a characteristics of fragment sequence appearance in a text? In this way, the information-probabilistic interpretation of mechanical movement is not reflected in a human brain during text production? In order to verify or deny this, the following experiment was conducted.

Experiment 7. Research of random texts with even distribution of the letters of alphabet.

Evaluation criterion: all stated in Section 7.

There were texts generated, the sizes of which were equal to those used in ex-

periments, but they had equal distribution of letter sequence. Of course, when increasing the size of text indefinitely there would be received the following value of probabilities:

$$p(a_i/a_{i-1}a_{i+1}) \approx p(a_i/a_{i-1}) \approx p(a_i/a_{i+1}) \approx p(a_i).$$

Therefore, all methods used in this paper would give the same result!

When the size of the text is limited, it is highly likely that there will be deviations of values of conditional probabilities from unconditional ones. In this case, the results of predicting of text fragment sequence and standard deviation of probability estimates made by different methods will be different.

In total, for each language, there were generated 40 random texts equal by the size to the texts of bases used. Standard deviations of combined conditional probability estimates from combined conditional probabilities of text fragment sequence of a random text are shown in Table 14. Mean values of the forecast of text fragments made by different methods are shown in Table 15. The deviations in ranks are in Table 16. In contrast to the experiments with real texts, here only one-letter fragments were used, as even distribution is the same for both one- and two-letter fragments. Moreover, for one-letter pieces, each fragment appears more often. Therefore, the statistics is better.

Table 14. Standard deviation of estimates from fragment sequence probabilities for the texts with even distribution of the letters of alphabet ($\sigma \cdot 10^2$).

Text by base size	By unconditional probability	By 1st fragment	By 2nd fragment	By mean probability	Method 1	Method 2	Method 3	Method 4	**NFIM**	Advantage of NFIM (%) (37)
"Russian"	0.323559	0.318828	0.318831	0.316438	0.314176	0.314174	**0.314028**	**0.314026**	0.314028	−0.000695
"German"	0.461893	0.454350	0.454352	0.450529	0.446929	0.446920	0.446682	**0.446675**	0.446677	−0.000373
"Ukrainian"	0.686833	0.676246	0.676257	0.670897	0.665845	0.665828	0.665501	0.665487	0.665485	0.000275
"English"	0.195578	0.192050	0.192052	0.190263	0.188596	0.188591	**0.188458**	**0.188453**	0.188459	−0.003413

Note: highlighted are the results which are better than those received by the non-force interaction method.

Table 15. Average number of correctly determined text fragments by 40 random generation of text with even distribution of the letters of alphabet.

Text by base size	By unconditional probability	By 1st fragment	By 2nd fragment	By mean probability	Method 1	Method 2	Method 3	Method 4	**NFIM**	Advantage of NFIM (%) (39)
"Russian"	78999.1	81436.3	81426.3	82665.6	82664.8	82659.4	**82691.9**	**82691.1**	82688.9	−0.003659
"German"	42545.5	44295.3	44271.9	45212.1	45213.0	45213.4	45228.9	**45236.5**	45235.5	−0.002155
"Ukrainian"	20153.7	21463.4	21477.5	22175.2	22176.6	**22184.0**	22170.3	22172.4	22177.1	−0.030888
"English"	183095.9	185953.9	185933.2	187219.2	187214.9	187241.0	187277.4	187267.9	187,278.7	0.000674

Note: highlighted are the results which are better than those received by the non-force interaction method.

Table 16. Deviation in ranks of estimates and probabilities of fragment sequence for the texts with even distribution of the letters of alphabet.

Text by example of base	% correct prediction of text fragment sequence									Advantage of NFIM (%)
	By unconditional probability	By 1st fragment	By 2nd fragment	By mean probability	Method 1	Method 2	Method 3	Method 4	**NFIM**	
"Russian"	164.760	139.896	139.819	130.470	130.486	130.485	130.215	**130.206**	130.210	−0.00769
"German"	134.129	112.559	112.505	104.600	104.623	104.611	104.378	**104.364**	104.373	−0.01375
"Ukrainian"	146.245	121.950	121.923	114.341	114.372	114.365	**114.105**	**114.098**	114.106	−0.01675
"English"	104.389	88.326	88.238	81.159	81.169	81.158	**80.918**	**80.913**	80.923	−0.03079

Note: highlighted are the results which are better than those received by the non-force interaction method.

As it can be seen from Tables 14-16, method based on the non-force interaction theory is inferior by almost all points comparing to the methods which use the approximation of probabilities. It means that it is the content part of the text, the processes of interaction that form what the method "processes", and not just probabilities. However, for "random" text, it is not better than the methods of approximation, as it should be according to classical mathematical principles.

In Table 14, it can be seen that the more the size of the text is, the smaller is standard deviation. This complies with the idea that for "indefinite" sizes of random texts all probability estimates will be equal to probabilities themselves.

The non-force interaction method is the best only for small sizes of texts ("Ukrainian" base) (see Table 14). However, even in this case, the advantage is minimum (0.000275%). Such exception only confirms the rule that the non-force interaction method works best with the objects that include the idea (meaning). This, in its turn, once more proves the important role of the non-force (informational) interactions when producing a text. Moreover, it also proves that processes of text production in a human brain correspond to the non-force interaction model which was received from information-probabilistic interpretation of mechanical movement and is the basis of the non-force interaction theory.

The NFIM was worse than other methods at ranking probability estimates (see Table 16). Comparing with Table 12, it is possible to reiterate that the non-force interaction method is effective only for "sensible" (intelligent) rather than random texts. This characterizes the level of influence of mechanics of natural text creating in a human brain (including forming of words) which is recognized by the non-force interaction method. Therefore, this means it is the contextual part of language that it operates with. Thus, this part of language functions according to the laws which are depicted in the non-force interaction method, that is movement and direct (contact) interaction in the information-probabilistic interpretation. Hence, what if it is not an interpretation but rather a reality?

Furthermore, what can happen if all texts are combined into one? There was an experiment with the text conducted which includes all four bases: Russian, German, Ukrainian and English.

Experiment 8. The research of a combined text

Criterion of evaluation: all stated in Section 7.

There were the rates of effectiveness developed for the combined text: standard deviation from combined conditional probability (Table 17) and quantity of correctly predicted text fragments (Table 18). Apart from that, the deviation of probability estimate rank from combined conditional probability rank was evaluated for the combined text (Table 19).

What can be seen from Tables 17-19? All rates of the NFIM are much higher than for other methods! Even if compared with linear approximation. By the way, the author tried using linear approximation, that is deviation square of conditional probability from an unconditional one. Of course, it does not eliminate the possibility of that someone will discover such an equation which will

Table 17. Standard deviation of estimates from probabilities of text fragment sequence of combined text.

L	By unconditional probability	By 1st fragment	By 2nd fragment	By mean probability	Method 1	Method 2	Method 3	Method 4	**NFIM**	Advantage of NFIM (%) (37)
1	0.09561	0.08543	0.08991	0.08024	0.07874	0.07693	0.07800	0.07765	**0.07508**	2.46
2	0.04628	0.04535	0.04527	0.04423	0.04404	0.04324	0.04401	0.04365	**0.04230**	2.22

Note: the best results are highlighted.

Table 18. Quantity of correctly predicted fragments of combined text.

L	By unconditional probability	By 1st fragment	By 2nd fragment	By mean probability	Method 1	Method 2	Method 3	Method 4	**NFIM**	Fragment total	Advantage of NFIM (%) (39)
1	894,574	1,617,312	1,595,614	2,001,922	1,992,145	2,136,515	2,083,939	2,165,393	**2,260,531**	7,456,824	4.21
2	201,116	957,000	946,025	1,479,064	1,476,830	1,826,359	1,498,931	1,794,563	**1,991,002**	7,456,816	8.27

Note: the best results are highlighted.

Table 19. Deviations in estimate ranks from ranks of fragment sequence probability of combined text.

L	By unconditional probability	By 1st fragment	By 2nd fragment	By mean probability	Method 1	Method 2	Method 3	Method 4	**NFIM**	Advantage of NFIM (%) (41)
1	35.10	23.62	21.33	18.65	18.79	15.80	15.93	15.79	**9.79**	61.27
2	34.80	23.54	22.64	18.36	18.36	17.60	16.80	17.53	**13.72**	22.47

Note: the best results are highlighted.

give bigger approximation to combined conditional probability than the NFIM for some text. As in "Russian" base, where there are one-letter fragments (see Table 7). However, whether this equation will fit other texts it unknown. According to the results stated in Table 10, it is highly unlikely.

As it can be seen from the provided tables (see Table 11 and Table 18), the approximation method which operates with deviations of conditional probability from an unconditional one (33) is not bad for forecasting, which Method 4. Therefore, the question arises. What if this difference is replaced with the non-force action value received in NFIT (which is also based on the difference between unconditional and conditional probability), how will the forecast change?

Experiment 9. Comparison of the effectiveness of the methods which are based on the deviation of conditional probability from an unconditional one and the value of non-force influence.

Criterion of evaluation: the percentage of correctly predicted text fragments (38).

Since method 4 is quite effective for predicting text fragment sequence, it is proposed to replace the difference between conditional and unconditional probability with the value of non-force influence (in the basis of which there is also a deviation of conditional probability from an unconditional one) (19), and the deviation of estimate from probability is replaced by achievement of the maximum results in predicting of the texts:

$$p_8\left(a_i/a_{i-1}a_{i+1}\right)=\delta_1\cdot p_i^{\max}+\delta_2\cdot\left(p_i^{\min}-p\left(a_i\right)\right)+\delta_3.$$

Let us apply Formula (9) in Formula (19) and "improve" Formula (33), replacing the difference in probabilities with non-force influence value

$$p_{10}\left(a_i/a_{i-1}a_{i+1}\right)=\delta_1\cdot p_i^{\max}+\delta_2\cdot\left(d_i^{\min}\cdot\sqrt{d\left(a_i\right)^2+1}-d\left(a_i\right)\cdot\sqrt{\left(d_i^{\min}\right)^2+1}\right)+\delta_3, \tag{42}$$

under the condition that

$$\mu_8\left(\delta_1,\delta_2\right)\cdot\sum_{a_{i-1}a_{i+1}}n\left(a_{i-1}a_{i+1}\right)\to\max; \tag{43}$$

$$\mu_{10}\left(\delta_1,\delta_2\right)\cdot\sum_{a_{i-1}a_{i+1}}n\left(a_{i-1}a_{i+1}\right)\to\max, \tag{44}$$

where $\mu_8\left(\delta_1,\delta_2\right)$ is the percentage of correct predictions of text fragment sequence by method 4; $\mu_{10}\left(\delta_1,\delta_2\right)$ is the percentage of correct predictions of text fragment sequence by method 5; $p_{10}\left(a_i/a_{i-1}a_{i+1}\right)$ is the estimate of conditional probability of sequence of text fragment a_i by linear approximation of non-force influence of the fragment with the smallest conditional probability: if the previous text fragment is a_{i-1}; if the following text fragment is a_{i+1} (Method 5); δ_1,δ_2 are approximation coefficients; $d_i^{\min}$ is the amount of information that triggers the reaction, which is calculated by Formula (11) with argument $p_i^{\min}$; $d\left(a_i\right)$ is the amount of information that triggers the reaction, which is calculated by Formula (11) with argument $p\left(a_i\right)$.

The results of predicting made by such combined method (method 5) are provided in Table 20. The result once again confirms higher descriptiveness of non-force interaction method regarding calculation of combined conditional probability, for example, the actions of artificial intelligence systems, and its higher "usefulness" in solving problems of predicting.

Experiment 10. Using instead of unconditional probability the conditional one, which excludes the cases when previous and subsequent fragments are the current ones.

Criterion of evaluation: all stated in Section 7.

To calculate the values of influence, it would be necessary to use conditional probability which would not include those cases when previous fragment was a_{i-1} and the following was a_{i+1}. Then the non-force value itself would exactly reflect the appearance of these fragments. However, the research has shown

$$p(a_i) \approx p\left(a_i / \overline{a_{i-1}}\right);$$

$$p(a_i) \approx p\left(a_i / \overline{a_{i+1}}\right),$$

where $p\left(a_i / \overline{a_{i-1}}\right)$ is statistical conditional probability of the sequence of fragment a_i when the previous fragment was not a_{i-1}; $p\left(a_i / \overline{a_{i+1}}\right)$ is statistical conditional probability of the sequence of fragment a_i when the following fragment is not a_{i+1}.

This was verified in the selected texts. The results are provided in Table 21.

Table 20. Comparison of approximation methods oriented at predicting of combined text fragments.

L	Quantity of correctly determined text fragments	
	Method 4	Method 5
1	2,242,440	**2,259,673**
2	1,985,547	**2,000,457**

Note: the best results are highlighted.

Table 21. Deviation of unconditional probabilities from conditional probabilities with exclusion of reviewed fragments.

Base	L	Deviation of probability		% deviation	
		Mean	Maximum	Mean	Maximum
"Russian"	1	0.000746	0.003368	2.460	11.115
	2	0.000002	0.000022	0.147	2.083
"German"	1	0.000887	0.004520	2.662	13.560
	2	0.000003	0.000073	0.195	5.257
"Ukrainian"	1	0.000830	0.002823	2.655	9.034
	2	0.000002	0.000016	0.145	1.500
"English"	1	0.001045	0.003581	2.717	9.310
	2	0.000003	0.000064	0.199	3.923

Nevertheless, in order to avoid doubts regarding the objectivity of the received results, there were conducted experiments using provided conditional probabilities $p\left(a_i/\overline{a_{i-1}}\right)$, $p\left(a_i/\overline{a_{i+1}}\right)$ instead of unconditional $p\left(a_i\right)$. The results are stated in Tables 22-24.

As it can be seen in Tables 22-24, the non-force interaction method remains the best of all others.

Experiment 11. Consideration of the combinations of fragments in which both conditional probabilities of appearance of a text fragment are not equal to zero ($p\left(a_i/a_{i-1}\right)>0$, $p\left(a_i/a_{i+1}\right)>0$).

Table 22. Standard deviation of combined conditional probability estimates from statistical combined conditional probability of text fragment sequence using conditional probabilities $p\left(a_i/\overline{a_{i-1}}\right), p\left(a_i/\overline{a_{i+1}}\right)$ instead of unconditional probabilities.

Base	L	By unconditional probability	By 1st fragment	By 2nd fragment	By mean probability	Method 1	Method 2	Method 3	Method 4	**NFIM**	Advantage of NFIM (%)
"Russian"	1	0.09545	0.08880	0.09019	0.08228	0.08170	**0.07843**	0.08176	0.08120	0.08107	−3.26
	2	0.04314	0.04236	0.04227	0.04121	0.04107	0.04028	0.04106	0.04074	**0.03952**	1.94
"German"	1	0.11050	0.09320	0.10388	0.08973	0.08715	0.08690	0.08565	0.08617	**0.08233**	4.03
	2	0.05481	0.05296	0.05323	0.05148	0.05110	0.05019	0.05102	0.05058	**0.04838**	3.73
"Ukrainian"	1	0.08394	0.07339	0.07588	0.06849	0.06694	0.06486	0.06543	0.06469	**0.06223**	3.95
	2	0.04804	0.04731	0.04715	0.04662	0.04637	0.04561	0.04634	0.04589	**0.04490**	1.58
"English"	1	0.10177	0.08646	0.09298	0.08130	0.07845	0.07769	0.07775	0.07808	**0.07504**	3.53
	2	0.04394	0.04292	0.04265	0.04149	0.04127	0.04049	0.04124	0.04093	**0.03948**	2.55

Note: the best results are highlighted.

Table 23. Predicting of text fragment sequence by different methods using conditional probabilities $p\left(a_i/\overline{a_{i-1}}\right), p\left(a_i/\overline{a_{i+1}}\right)$ instead of unconditional probabilities.

Base	L	By unconditional probability (μ_1)	By 1st fragment (μ_2)	By 2nd fragment (μ_3)	By mean probability (μ_4)	Method 1 (μ_5)	Method 2 (μ_6)	Method 3 (μ_7)	Method 4 (μ_8)	**NFIM (μ_9)**	Advantage of NFIM (%) (39)
"Russian"	1	11.54	19.37	19.17	22.92	22.88	25.61	25.36	26.17	**27.66**	5.39
	2	1.77	12.58	12.88	20.06	20.04	26.45	20.16	25.04	**27.84**	4.99
"German"	1	16.05	26.97	26.64	34.27	33.25	35.80	35.28	36.65	**37.39**	1.98
	2	3.31	15.83	13.89	22.90	22.89	26.28	23.45	26.59	**30.00**	11.37
"Ukrainian"	1	6.62	17.30	17.72	21.77	21.42	24.28	23.79	25.15	**25.60**	1.76
	2	0.74	9.12	9.63	15.38	15.38	19.93	15.45	19.90	**22.40**	11.03
"English"	1	11.79	22.35	21.93	28.11	27.86	29.21	28.64	29.26	**29.86**	2.01
	2	3.19	12.61	12.36	19.44	19.33	23.35	19.89	23.01	**25.43**	8.18

Note: the best results are highlighted.

Table 24. Standard deviation of ranks of combined conditional probability estimates from ranks of statistical combined conditional probabilities of text fragment sequence using conditional probabilities $p\left(a_i/\overline{a_{i-1}}\right), p\left(a_i/\overline{a_{i+1}}\right)$.

Base	L	% of correct prediction of text fragment sequence									Advantage of NFIM (%) (41)
		By unconditional probability	By 1st fragment	By 2nd fragment	By mean probability	Method 1	Method 2	Method 3	Method 4	**Non-force interaction method**	
"Russian"	1	40.32	25.71	24.19	19.62	19.65	16.38	17.44	19.52	**11.43**	43.33
	2	22.72	14.63	14.59	11.80	11.81	11.27	10.90	11.17	**9.05**	20.38
"German"	1	26.12	18.11	13.11	12.47	12.99	11.61	10.90	10.94	**7.64**	42.61
	2	21.07	14.17	12.11	11.18	11.20	10.15	10.12	9.84	**7.61**	29.22
"Ukrainian"	1	44.41	27.53	25.18	23.81	23.91	18.28	19.63	17.30	**10.14**	70.61
	2	3.37	2.07	2.15	1.84	1.84	1.74	1.73	1.62	**1.50**	7.78
"English"	1	33.74	19.32	18.52	14.78	14.87	13.40	12.61	13.28	**8.91**	41.42
	2	65.33	42.61	41.05	32.79	32.79	31.63	29.99	31.10	**24.75**	21.15

Note: the best results are highlighted.

Criterion of evaluation: all stated in Section 7.

One more experiment was connected to the decision accepted in the methodology of experimental research. It is to consider combinations of fragments for which at least one probability is not equal to zero (pp.6 of Section 5): $n\left(a_{i-1}, a_i\right) > 0$, or $n\left(a_i, a_{i+1}\right) > 0$. Since, under the condition that $n\left(a_{i-1}, a_i\right) = 0$ or $n\left(a_i, a_{i+1}\right) = 0$, combined conditional probability estimate calculated by the non-force interaction method will also be equal to zero, this does not create additional advantage to it comparing with others.

However, from the point of view of the development of artificial intelligence systems this approach is correct. As it is not improbable that when increasing the statistics, zero probability will become not equal to zero. Hence, these cases should be considered, too.

Results are provided in Tables 25-27.

As it can be seen from Tables 25-27, the change of methodology does not violate previous result, when the non-force interaction method is the most effective (according to the criteria stated in Section 7) of all others.

For artificial intelligence systems, it is highly important to have consistency of received results under different conditions. Usually, to verify any methods, samplings are divided into training and control ones. Probabilities received in training samplings and coefficients of approximation equations are applied to the control sampling. Within experimental research, the results of which are provided in this article, there also was such an experiment.

Experiment 12. Research of the text divided into training and control samplings.

Criterion of evaluation: all stated in Section 7.

To have objective results, especially those concerning approximation methods, all the texts were divided randomly into two samplings, each having 50% the size of the text: training and control. By the training sampling, approximation coefficients and unconditional probabilities of text fragment appearance were calculated. In the control sampling, according to the criteria described in Section 7, the effectiveness of provided methods was verified. The results are stated in Tables 28-30.

As it is seen from Tables 28-30, there has nothing changed. The NFIM is the best and its advantage over the other methods in some cases even increased (see Table 7, Table 11, Table 12).

Table 25. Standard deviation of combined conditional probability estimates from statistical combined conditional probability of text fragment sequence under the condition: $p(a_i/a_{i-1})>0$ and $p(a_i/a_{i+1})>0$.

Base	L	By unconditional probability	By 1st fragment	By 2nd fragment	By mean probability	Method 1	Method 2	Method 3	Method 4	**NFIM**	Advantage of NFIM (%) (37)
"Russian"	1	0.10384	0.09712	0.09882	0.09044	0.08966	**0.08653**	0.08974	0.08894	0.08943	−3.24
	2	0.08976	0.08699	0.08672	0.08522	0.08404	0.08312	0.08395	0.08301	**0.08237**	0.78
"German"	1	0.12585	0.10803	0.11927	0.10399	0.10074	0.10063	0.09885	0.09900	**0.09507**	3.98
	2	0.12943	0.12273	0.12361	0.12064	0.11757	0.11694	0.11714	0.11610	**0.11469**	1.23
"Ukrainian"	1	0.08782	0.07757	0.08030	0.07262	0.07083	0.06888	0.06924	0.06826	**0.06601**	3.40
	2	0.10761	0.10501	0.10463	0.10401	0.10212	0.10141	0.10194	0.10088	**0.10076**	0.12
"English"	1	0.10856	0.09373	0.10035	0.08808	0.08480	0.08405	0.08407	0.08413	**0.08111**	3.61
	2	0.08147	0.07850	0.07817	0.07652	0.07534	0.07451	0.07521	0.07446	**0.07341**	1.43

Note: the best results are highlighted.

Table 26. Predicting of text fragment sequence by different methods under the condition: $p(a_i/a_{i-1})>0$ and $p(a_i/a_{i+1})>0$.

Base	L	By unconditional probability (μ_1)	By 1st fragment (μ_2)	By 2nd fragment (μ_3)	By mean probability (μ_4)	Method 1 (μ_5)	Method 2 (μ_6)	Method 3 (μ_7)	Method 4 (μ_8)	**NFIM (μ_9)**	Advantage of NFIM (%) (39)
"Russian"	1	11.56	19.42	19.17	22.95	22.98	25.50	25.27	26.39	**27.98**	5.68
	2	2.66	15.51	15.49	21.72	21.69	26.46	21.85	27.11	**28.02**	3.25
"German"	1	16.19	27.02	26.85	34.32	33.32	35.70	35.98	36.79	**38.44**	4.29
	2	5.42	19.08	17.73	24.86	24.86	26.49	25.52	28.40	**30.25**	6.12
"Ukrainian"	1	9.42	17.41	17.73	21.78	21.50	24.15	23.40	25.21	**25.83**	2.40
	2	2.47	12.56	13.04	17.95	17.94	20.62	18.18	21.86	**22.48**	2.76
"English"	1	11.89	22.36	21.95	28.11	27.75	29.07	28.63	29.31	**30.39**	3.55
	2	4.18	14.23	13.82	20.26	20.20	23.15	20.75	24.22	**25.61**	5.43

Note: the best results are highlighted.

Table 27. Standard deviation of ranks of combined conditional probability estimates from the ranks of statistical combined conditional probabilities of text fragment sequence under the condition: $p(a_i/a_{i-1})>0$ and $p(a_i/a_{i+1})>0$.

Base	L	% of correct prediction of text fragment sequence									Advantage of NFIM (%)
		By unconditional probability	By 1st fragment	By 2nd fragment	By mean probability	Method 1	Method 2	Method 3	Method 4	**NFIM**	
"Russian"	1	37.74	25.71	24.19	19.62	19.66	16.51	17.26	19.75	**11.48**	43.77
	2	21.88	14.63	14.59	11.80	11.81	11.22	10.90	11.48	**8.98**	21.34
"German"	1	24.10	18.11	13.11	12.47	12.89	11.84	10.72	10.86	**7.60**	41.09
	2	20.02	14.17	12.11	11.18	11.20	10.48	10.10	9.84	**7.52**	30.90
"Ukrainian"	1	41.62	27.53	25.18	23.81	23.91	18.62	19.76	17.05	**10.07**	69.32
	2	3.27	2.07	2.15	1.84	1.84	1.75	1.73	1.63	**1.50**	8.71
"English"	1	30.99	19.32	18.52	14.78	14.88	13.48	12.42	13.27	**8.74**	42.06
	2	63.06	42.61	41.05	32.79	32.79	31.43	29.96	31.91	**24.62**	21.66

Note: the best results are highlighted.

Table 28. Standard deviation of combined conditional probability estimates from statistical combined conditional probability of text fragment sequence in control text sampling.

Base	L	By unconditional probability	By 1st fragment	By 2nd fragment	By mean probability	Method 1	Method 2	Method 3	Method 4	**NFIM**	Advantage of NFIM (%) (37)
"Russian"	1	0.09627	0.09058	0.09147	0.08388	0.08189	**0.07887**	0.08181	0.08116	0.08217	−4.01
	2	0.04718	0.04629	0.04623	0.04518	0.04478	0.04387	0.04477	0.04435	**0.04324**	1.45
"German"	1	0.11047	0.09504	0.10505	0.09139	0.09011	0.08998	0.08863	0.08908	**0.08278**	7.06
	2	0.05956	0.05748	0.05777	0.05601	0.05537	0.05436	0.05526	0.05473	**0.05256**	3.42
"Ukrainian"	1	0.08554	0.07608	0.07845	0.07129	0.06995	0.06790	0.06857	0.06772	**0.06470**	4.66
	2	0.05486	0.05398	0.05385	0.05325	0.05256	0.05155	0.05252	0.05187	**0.05116**	0.78
"English"	1	0.10079	0.08688	0.09318	0.08161	0.07838	0.07742	0.07758	0.07769	**0.07487**	3.41
	2	0.04686	0.04571	0.04552	0.04437	0.04407	0.04327	0.04403	0.04368	**0.04222**	2.47

Note: the best results are highlighted.

Table 29. Predicting of fragment sequence of control text sampling made by different methods.

Base	L	By unconditional probability (μ_1)	By 1st fragment (μ_2)	By 2nd fragment (μ_3)	By mean probability (μ_4)	Method 1 (μ_5)	Method 2 (μ_6)	Method 3 (μ_7)	Method 4 (μ_8)	**NFIM (μ_9)**	Advantage of NFIM (%) (39)
"Russian"	1	11.57	19.37	19.19	23.11	23.20	25.54	25.37	26.24	**28.05**	−4.01
	2	1.79	12.61	12.94	20.16	20.16	26.59	20.24	25.55	**28.44**	1.45

Continued

"German"	1	16.00	26.92	26.64	34.28	33.56	35.73	35.24	36.72	**38.44**	7.06
	2	3.64	15.87	13.98	22.87	22.85	26.30	23.40	26.83	**30.53**	3.42
"Ukrainian"	1	9.24	17.39	17.72	21.66	21.65	24.22	23.75	25.47	**25.98**	4.66
	2	1.38	9.22	9.82	15.82	15.81	20.91	15.91	21.25	**23.67**	0.78
"English"	1	11.79	22.34	21.90	28.13	27.66	29.22	28.64	29.25	**30.41**	3.41
	2	3.20	12.62	12.36	19.40	19.38	23.39	19.91	23.10	**25.70**	2.47

Note: the best results are highlighted.

Table 30. Standard deviation of ranks of combined conditional probabilities estimates from ranks of statistical combined conditional probabilities of fragment sequence from control text sampling.

Base	L	% of correct prediction of text fragment sequence									Advantage of NFIM (%)
		By unconditional probability	By 1st fragment	By 2nd fragment	By mean probability	Method 1	Method 2	Method 3	Method 4	**NFIM**	
"Russian"	1	35.62	24.32	22.82	18.64	18.65	15.56	16.06	17.26	**10.87**	43.15
	2	9.97	6.59	6.54	5.37	5.37	5.09	4.98	4.94	**4.13**	19.61
"German"	1	21.45	16.25	12.07	11.40	11.86	10.58	10.08	9.88	**6.97**	41.75
	2	11.78	8.25	6.97	6.61	6.62	6.03	6.00	5.66	**4.55**	24.40
"Ukrainian"	1	35.60	23.27	21.60	20.50	20.58	15.66	16.74	14.11	**8.65**	63.12
	2	1.22	0.72	0.75	0.66	0.66	0.61	0.63	0.57	**0.54**	5.56
"English"	1	29.85	18.37	17.77	14.17	14.28	12.78	11.71	12.22	**8.33**	40.58
	2	34.04	22.67	22.16	17.55	17.56	16.92	16.13	16.46	**13.40**	20.37

Note: The best results are highlighted.

12. Summary

When verifying theorems automatically, their validity can be proven by going through all constant cases. The first person to highlight this was Jacques Herbrand [32]. In the logic of predicates of the first order in far-away 1930, he proposed the procedure of verification based on the theorem which states: a range of disjuncts S is impossible when and only when there is finite impossible range of fundamental examples (*i.e.* constant cases) of disjuncts with S.

The proof of the theorem is called the answer to the question whether some formula B logically follows out of the stated range of formulae $f_1, f_2, \cdots, f_n$:

$f_1, f_2, \cdots, f_n \to B$ is the verification of general validity of the formula.

It is much easier to prove the unprovability of the formula:

$f_1, f_2, \cdots, f_n \to \neg B$ —refutation procedure.

As for the question that is being considered in this article, it means that it is important to find such method of determination of combined conditional probability by unconditional and certain conditional ones which gives a better result

than that offered in the non-force interaction theory. Moreover, it should be true for any texts.

The author did not manage to do this, despite doing his best. There is the last point, though. It seems that if from one text there are taken fragments $p^{\text{text 1}}\left(a_i/a_{i-1}\right)$ and $p^{\text{text 1}}\left(a_i/a_{i+1}\right)$, and the equivalent values of them are found in the other text

$$p^{\text{text 2}}\left(a_i/a_{i-1}\right)=p^{\text{text 1}}\left(a_i/a_{i-1}\right);$$

$$p^{\text{text 2}}\left(a_i/a_{i+1}\right)=p^{\text{text 1}}\left(a_i/a_{i+1}\right),$$

then the value $p^{\text{text 1}}\left(a_i/a_{i-1}a_{i+1}\right)$ will be closer to $p^{\text{text 2}}\left(a_i/a_{i-1}a_{i+1}\right)$ than those received by listed methods. For example, if:

$$p^{\text{text 1}}\left(a_i/a_{i-1}\right)=p^{\text{text 2}}\left(a_i/a_{i-1}\right)=0.7;$$

$$p^{\text{text 1}}\left(a_i/a_{i+1}\right)=p^{\text{text 2}}\left(a_i/a_{i+1}\right)=0.3;$$

$$p^{\text{text 1}}\left(a_i/a_{i-1}a_{i+1}\right)=0.8,$$

then

$$p^{\text{text 2}}\left(a_i/a_{i-1}a_{i+1}\right)\approx 0.8.$$

Otherwise, if the value of combined conditional probability is different for the same values of certain conditional probabilities, the question arises. Is it at all possible to find the method which orients itself at probabilities and will provide a better result than the non-force interaction method?

There was an experiment conducted. In "English" base, the input conditional probabilities are determined. In "Russian" base, the conditional probabilities equal to them are found. Certainly, finding all the probabilities in "English" base correspondent to each conditional probability of text fragments in "Russian" base failed (the deviation within 0.0005 was allowed). There was found K = 15,307 one-letter fragments where probabilities match.

By these fragments using least square method there was calculated the deviation dispersion of probability $p^{\text{English}}\left(a_i/a_{i-1}a_{i+1}\right)$ from probability $p^{\text{Russian}}\left(a_i/a_{i-1}a_{i+1}\right)$, and probability $p_9^{\text{Russian}}\left(a_i/a_{i-1}a_{i+1}\right)$ from probability $p^{\text{Russian}}\left(a_i/a_{i-1}a_{i+1}\right)$.

Here is the answer:

$$\frac{\sum_{i=1}^{K}\left(p^{\text{English}}\left(a_i/a_{i-1}a_{i+1}\right)-p^{\text{Russian}}\left(a_i/a_{i-1}a_{i+1}\right)\right)^2}{K}\approx 0.0031873;$$

$$\frac{\sum_{i=1}^{K}\left(p_9^{\text{Russian}}\left(a_i/a_{i-1}a_{i+1}\right)-p^{\text{Russian}}\left(a_i/a_{i-1}a_{i+1}\right)\right)^2}{K}\approx 0.0017463,$$

under the condition that

$$\left|p^{\text{English}}\left(a_i/a_{i-1}\right)-p^{\text{Russian}}\left(a_i/a_{i-1}\right)\right|<0.0005;$$

$$\left|p^{\text{English}}\left(a_i/a_{i+1}\right)-p^{\text{Russian}}\left(a_i/a_{i+1}\right)\right|<0.0005,$$

where K is the quantity of matches in the values of probabilities of "English" and

"Russian" bases; $p_9^{\text{Russian}}\left(a_i/a_{i-1}a_{i+1}\right)$ is estimate of combined conditional probability of sequence of text fragment a_i by the non-force interaction method in "Russian base"; $p^{\text{Russian}}\left(a_i/a_{i-1}a_{i+1}\right)$ is combined conditional probability of sequence of text fragment a_i in "Russian" base; $p^{\text{English}}\left(a_i/a_{i-1}a_{i+1}\right)$ is combined conditional probability of sequence of text fragment a_i in "English" base; $p^{\text{Russian}}\left(a_i/a_{i-1}\right)$ is conditional probability of sequence of text fragment a_i in "Russian" base if the previous fragment is a_{i-1}; $p^{\text{Russian}}\left(a_i/a_{i+1}\right)$ is conditional probability of sequence of text fragment a_i in "Russian" base if the following fragment is a_{i+1}; $p^{\text{English}}\left(a_i/a_{i-1}\right)$ is conditional probability of sequence of text fragment a_i in "English" base if the previous fragment is a_{i-1}; $p^{\text{English}}\left(a_i/a_{i+1}\right)$ is conditional probability of sequence of text fragment a_i in "English" base if the following fragment is a_{i+1}.

Deviation dispersion of the probability calculated by the non-force interaction method is almost two times smaller than the dispersion of the probability deviation taken from the other text!

Therefore, the main conclusion of the experiments is the following: the method based on the non-force interaction theory is better when analysing natural language texts as it does not just operate probabilities, but rather reflects information interactions in a human brain while producing speech (letter combinations, words, sentences). Moreover, these interactions are described with the same formulae that are used in the description of mechanical movement and direct (contact) interaction of material objects if they are given information-probabilistic interpretation. In the author's opinion, this states that there is the correspondence of the processes of information interaction in a human brain when producing natural language texts (the second signal system) to the processes of interaction at the microlevel of the Nature, which speaks about the general validity of the received non-force interaction formulae.

13. Discussion

Is the information the basis of our Nature? Is it the foundation of all interactions? The results received make it very high that the probability of the fact that information also determines the processes of interaction on a physical level (interactions of different physical nature). Moreover, information processes in a human brain correspond to the non-force interaction model received from information-probabilistic interpretation of the laws of physics.

This direction is relevant and promising, open to work with for many scientists of different countries. Of course, the author is the advocate of the development in this direction. Furthermore, like nobody else, he is interested in that other scientists tried to refute this statement through other experiments. Then, if they fail to do so, this will be the confirmation of the non-force interaction theory (***here is one more reference to Herbrand*** [32]).

14. Practical Implementation

If the consistent patterns in estimating combined conditional probability by the

non-force interaction method are preserved in different subject fields, this creates a possibility to develop artificial intelligence systems, in which by the deviation of conditional probabilities from unconditional ones the value of the non-force influence of various factors will be calculated. Then, the total of these influences will determine the most probable reaction of the system. This will allow to create simple, cheap and… effective artificial intelligence systems in which accumulated statistic information will be in the basis of reflexes (development of the most probable reactions to external influences). Allow us to illustrate.

Let us assume that in the process of previous training the adequate reactions of the system to external influences were determined. The training process was in consistent exposure on the system showing adequate reactions:

Training sampling 1: $W_{k_1} \to R_{i_1}$;

Training sampling 2: $W_{k_2} \to R_{i_2}$;

⋮

Training sampling j: $W_{k_j} \to R_{i_j}$;

⋮

Training sampling *N*: $W_{k_N} \to R_{i_N}$,

where $W_{k_j} \subseteq W$ is a subrange of the factors of influence on the system at *j*-th step of training; $R_{i_j} \in R$ is an adequate reaction of the system at *j*-th step of training; W is the range of factors of influence; R is the range of reactions.

Let us assume that at the moment of time *t*, the system is exposed to the influences that are included in the subrange $w_s \in W_0$.

$$\forall R_i \in R, \forall w_s \in W_0 : p\left(R_i / w_s\right).$$

Of course, if in the system memory, the reaction to all this subrange of influences is determined,

$$\exists R_q \in R : p\left(R_q / W_0\right) = 1,$$

the task of decision making is trivial. The decision is reaction R_q. However, usually there are so many influences that it is impossible to accumulate statistics for all these combinations. In this case, the non-force interaction method can become useful. As it is possible to calculate the value of the influence on the reaction by deviation of $p\left(R_i / w_s\right)$ from $p\left(R_i\right)$. Then, the reaction with the highest combined conditional probability estimate $p_9\left(R_q / W_0\right)$ can be chosen and implemented (see Section 6).

As it was shown in the experiments and practice, the combined conditional probability estimate of the reaction calculated by the non-force interaction method allows to choose correct reactions in different subject fields in 99% of cases [4] [6] [30] [33]. At the moment, on the basis of the non-force interaction method the systems of reactive artificial intelligence machines that produce adequate reactions to the influences of the multitude of factors have been and are being developed. The author and his students have created the systems of: predicting the results of sport games, evaluation of investment proposals in development, access the databases using natural language, automatic abstracting of

texts, automatic message forwarding, voice control of TV and phone, etc. [4]. The main advantage of these systems is not even the fact that they are more effective than others, but rather in the fact that the expenses on their development are minimum.

How do they work and why are they called the systems of a reactive type? The work started from the creation of the system of access to the database of the construction of Southern Ukrainian Nuclear Power Station using natural language [30]. It was developed the automated system which included almost all data necessary for construction: drawings (specifications), budget estimates, plans, scopes of conducted works, material and technical resources. Most of the workers had low level of computer skills (it was in 80s of the 20th century). Therefore, they were offered a system KET ("компилятор естественно-языковых текстов" in Russian, "natural language compiler" in English) which had single window where a user wrote their request in natural language. For example, "What scope of works was made last month in reactor department by subdivision PROM-1". There were standard reactions entered in the system. They were fulfilled by database management system КВАНТ-М:

- Tasks: planned scopes of works, planned cost of works, performance of works, need for resources;
- Construction objects: reactor department, generator hall, etc. (more than 100 objects);
- Performers (people responsible) (more than 100 organisations and subdivisions);
- And other.

During the training, the teacher entered a random sentence and chose corresponding reaction. In essence, the reflex to a request was developed. The reflex was created in the following way:

1) Input text was divided into fragments (from 2 to 10 symbols).

2) For each fragment, statistical probability of choice of each reaction (conditional probability) was calculated.

3) For each reaction, its statistical probability of choice (unconditional probability) was calculated.

4) The difference between conditional and unconditional probabilities for each reaction of each fragment was treated as a value of non-force (information) influence on this reaction by this fragment.

The reaction with the biggest sum of non-force influences from the user's request was chosen.

The use of the system showed its high reliability when processing users' requests. The probability of correct reaction for non-programming professionals (engineers, skilled workers, foremen) exceeded 95%. The most importantly, the system did not need to perform morphological, syntax, semantic analysis of the text as it is done by similar systems. The non-force interaction method is a really simple and reliable method of creation of the systems of natural language!

One more similar system was depicted in the publication [6]. This is the sys-

tem of TV voice control. The principles, method and algorithm are the same as described in the system of access to databases. However, at the input it is not the text that is entered, but phonemes of a language. The probability of correct reaction in this system is close to 0.99.

The system of predicting the results of sport matches [1] works in a slightly different way. As an influenced object, there is a result of a match (win, draw, loss). As influencing objects there are various factors: the result of a previous match, what were the previous results of the matches between the same competitors, on whose field the game is, how many goals are scored and missed in previous matches. Factors like weather, referee, availability of a stronger team, etc. can be added. In this case, the unconditional probability of each result (win, draw, loss) of each team is chosen from the statistics, as well as conditional probabilities of these results, but only if the selected influence factors exist. The non-force influence, which is calculated as it is described in Subsection 2.3, characterises many factors that in total give a different value of result probability. Taking into account many factors gives probable result of a match. The statistics of predicting leading championships of European countries and matches of representative teams has been accumulated for 10 years. The probability of correct predicting of results in different championships varies from 0.55 to 0.65 (English championship is the most difficult to predict), and it reaches 0.7 for the matches of representative teams.

In the system of evaluation of investment proposals in development, the values of various parameters of projects (location, destination, cost, etc.) were calculated. This system is unique because in the process of experimental research it gave 100% match of parameter values to the ones given by experts [33].

15. Conclusions

The confidence of the author in the validity of the non-force interaction theory, its versatility, is based on computer experiments and practical implementation of the systems created with the use of the results receive in the non-force interaction theory [4] [6] [30] [33]. The theory was received from computer experiments in the basis of which it was the assignment of information content to objects (see **Figure 1**) and modelling of their mechanical movement and interaction during collision (the law of conservation of momentum) [4]. The information-probabilistic interpretation of mechanical movement was the result of these experiments. The next step of work on any theory is its experimental verification and practical implementation.

This work is dedicated to experimental verification of the general validity of the formulae of dealing with the information content of interacting objects, which were received from the information-probabilistic interpretation of mechanical movement. Specifically, its correspondence to the interaction processes happen in a human brain during producing natural language texts. All computer experiments, the results of which are provided in this article definitely show that

the non-force interaction theory formulae allow to describe statistical consistent patterns in the texts of different languages more exactly than other methods selected for the experiments. In its turn, this allows to state that there is a high probability of **the interactions at both micro- and macro-levels of the Nature are informational (non-force) ones and are describes with the same formulae.**

However, it is also possible that it is not the information-probabilistic interpretation of mechanical movement that is primary, and it is not the one to be taken as the foundation of the research of natural language texts. It is possible that the existence of the systems in which the information lies in the basis of their life-sustaining activity, and their development, training and self-development, and their reflexes were the basis of the creation of "informational laws" of mechanical movement. Moreover, it is these laws that we research in **the computer simulation that we call the Universe** [1] [34] [35] [36] [37]…

Acknowledgements

Without the help of my friends, colleagues, students and family this article would never have been written. Firstly, I thank to my extremely talented and professional daughter, Yegorchenkova Natalia Yuriivna, Doctor of Science. She is both a science editor, assistant and professional advisor on the material of the article. Useful advice, constructive criticism and ideas which were aimed at improving the methodology of research are the input of my wife Olga Victorivna. All mathematical calculations were made under the supervision of a wonderful mathematician, Podchasova Tetiana Pavlivna, Doctor of Science. For advice on the same questions, I am really grateful to my colleagues, Scherba A. I., Candidate of Science, and Snytiuk Vitaliy Yevhenovych, Doctor of Science. Professional support in preparation of the article text was provided by my talented student and assistant Khlevna Yulia Leonidivna, Doctor of Science. Real support in my research and a person who in all situations backed me and my ideas up is Pavlenko Petro Mykolaiovych, Doctor of Science. For the professional translation of the article into English (and this is not the first time she performs the translation of my works) I thank to Mariana Kucherenko. I highly APPRECIATE everybody who made a contribution into this work!

Conflicts of Interest

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

References

[1] Bostrom, N. (2003) *Philosophical Quarterly*, **53**, 243-255. https://doi.org/10.1111/1467-9213.00309

[2] Wolfram, S. (2002) A New Kind of Science. Wolfram Media, Inc., Champaign. https://www.wolframscience.com/nks/

[3] Hooft, G. (2016) The Cellular Automaton Interpretation of Quantum Mechanics. Fundamental Theories of Physics, Vol. 185, Springer, Cham, 256 p. https://doi.org/10.1007/978-3-319-41285-6 https://link.springer.com/content/pdf/10.1007%2F978-3-319-41285-6.pdf

[4] Teslia, I.M. (2014) Non-Forceful Nature: Monograph. Chabanenko Yu.A., Ukraine, 176 p. https://er.chdtu.edu.ua/handle/ChSTU/2814

[5] Teslia, I. and Klapchenko, V. (2011) Probabilistic Interpretation of Mechanical motion. http://arxiv.org/find/all/1/all:+klapchenko/0/1/0/all/0/1

[6] Teslia, I., Pylypenko, V., Popovych, N. and Chornyy, O. (2014) The Non-Force Interaction Theory for Reflex System Creation with Application to TV Voice Control. 6*th International Conference on Agents and Artificial Intelligence* (ICAART 2014), Angers, 6-8 March 2014, 288-296.

[7] Landau, L.D. and Lifshitz, E.M. (1980) The Classical Theory of Fields. Course of Theoretical Physics Vol. 2, 4th Edition, Butterworth-Heinemann, Oxford.

[8] Feynman, R.P., Leighton, R.B. and Sands, M. (2005) The Feynman Lectures on Physics, Volume 1: Mainly Mechanics, Radiation, and Heat (Definitive ed.). Pearson Addison-Wesley, San Francisco.

[9] Alfredo, A. and Elizabeth, S. (2017) Wernicke's Area. *Reference Module in Neuroscience and Biobehavioral Psychology*, Elsevier, Amsterdam. https://www.researchgate.net/publication/312042389_Wernicke's_area

[10] Binder, J.R. (2015) *Neurology*, **85**, 2170-2175.

[11] Friederici, A. (2011) *Physiological Reviews*, **91**, 1357-1392. https://doi.org/10.1152/physrev.00006.2011 https://journals.physiology.org/doi/full/10.1152/physrev.00006.2011

[12] Hickok, G. (2009) *Physics of Life Reviews*, **6**, 121-143. https://doi.org/10.1016/j.plrev.2009.06.001 http://ncbi.nlm.nih.gov/pmc/articles/PMC2747108/

[13] Kuligowska, K., Kisielewicz, P. and Włodarz, A. (2018) Speech Synthesis Systems: Disadvantages and Limitations. *International Journal of Engineering and Technology* (*UAE*), **7**, 234-239.

[14] Schroeter, J. (2008) Basic Principles of Speech Synthesis. In: Benesty, J., Sondhi, M.M. and Huang, Y.A., Eds., *Springer Handbook of Speech Processing*, Springer, Berlin, Heidelberg, 413-428. https://doi.org/10.1007/978-3-540-49127-9_19 https://www.researchgate.net/publication/290619779_Basic_Principles_of_Speech_Synthesis

[15] Cheng, X.-Y. and Yan, P. (2011) Review of Modern Speech Synthesis. In: Hu, W., Ed., *Electronics and Signal Processing*, Vol. 97, Springer, Berlin, Heidelberg, 517-524. https://doi.org/10.1007/978-3-642-21697-8_65

[16] Mussabayev, R.R., Amirgaliyev, Y.N., Tairova, A.T., Mussabayev, T.R. and Koibagarov, K.C. (2016) The Technology for the Automatic Formation of the Personal Digital Voice Pattern. 2016 *IEEE* 10*th International Conference on Application of Information and Communication Technologies* (*AICT*), Baku, 12-14 October, 2016, 1-5. https://doi.org/10.1109/ICAICT.2016.7991733 https://ieeexplore.ieee.org/document/7991733

[17] Korkut, C., Haznedaroglu, A. and Arslan, L. (2020) Comparison of Deep Learning Methods for Spoken Language Identification. *SPECOM* 2020: *International Conference on Speech and Computer*, St. Petersburg, 7-9 October 2020, 223-231. https://doi.org/10.1007/978-3-030-60276-5_23 https://link.springer.com/chapter/10.1007/978-3-030-60276-5_23#citeas

[18] https://github.com/Yehorchenkov/data_for_paper_teslia

[19] http://ocw.mit.edu/ans7870/6/6.006/s08/lecturenotes/files/t8.shakespeare.txt

[20] https://modernlib.net/books/tolstoy_lev_nikolaevich/voyna_i_mir_tom_1/

[21] https://modernlib.net/books/tolstoy_lev_nikolaevich/voyna_i_mir_tom_2/

[22] https://all-the-books.ru/books/tolstoy-lev-voyna-i-mir-tom-3/

[23] https://all-the-books.ru/books/tolstoy-lev-voyna-i-mir-tom-4/

[24] http://originalbook.ru/faust-gyote-deutsch-faust-goethe/

[25] http://originalbook.ru/ecce-homo-f-nietzsche-deutsch/

[26] http://originalbook.ru/das-schlos-franz-kafka-deutsch/

[27] https://avidreaders.ru/book/kobzar.html

[28] https://royallib.com/book/kostenko_lna/zbrka_vrshv.html

[29] Teslia, I. (n.d.) Personal Site. http://teslia.kyiv.ua/?page_id=55

[30] Teslia, I. (1998) Reactive System of Processing Natural Language Texts in Automated Command and Control System of the Construction of Complex Power plants. *Radioelectronics and Information Science*, No. 4, 52-55.

[31] Lebedev, Y., Livinska, G., Rozor, I. and Sharapov, M. (2016) Mathematical Statistics: Starter Guide. PPC "Kyiv University", Kyiv, 160 p. (Ukrainian) http://applstat.univ.kiev.ua/ukr/docs/materials/matstat.pdf

[32] Louis Laurière, J. (1991) Problem-Solving and Artificial Intelligence. Mir, Moskow, 568 p. (Russian) http://library.kpi.kharkov.ua/vustavki/Lingv/38.pdf

[33] Teslia, I., Kayuk, P. and Chernova, M. (2011) Formation and Analysis of Decision Making to Evaluate Investments in Development Projects. *Management of Complex system Development*, No. 7, 60-62.

[34] Bostrom, N. (2005) *Philosophical Quarterly*, **55**, 90-97. https://doi.org/10.1111/j.0031-8094.2005.00387.x

[35] Brueckner, A. (2008) The Simulation Argument Again. *Analysis*, **68**, 224-226. https://doi.org/10.1093/analys/68.3.224

[36] Campbell, T., Owhadi, H., Sauvageau, J. and Watkinson, D. (2017) On Testing the Simulation Hypothesis. http://www.ijqf.org/wps/wp-content/uploads/2017/03/IJQF-3888.pdf

[37] Fan, R., Wang, J.-Y. and Li, H. (2008) Discussion on Simulation Application Theory. 2008 *Asia Simulation Conference*: 7*th International Conference on System Simulation and Scientific Computing*, Beijing, 10-12 October 2008, 363-367.

Journal of Modern Physics, 2022, 13, 574-619
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Two Concepts in Optics of Anisotropic Dispersive Media and Polariton Case in Coordinate-Invariant Way

Alfred Wünsche

Institut für Physik, Humboldt-Universität, Berlin, Germany
Email: alfred.wuensche@physik.hu-berlin.de

How to cite this paper: Wünsche, A. (2022) Two Concepts in Optics of Anisotropic Dispersive Media and Polariton Case in Coordinate-Invariant Way. *Journal of Modern Physics*, **13**, 574-619.
https://doi.org/10.4236/jmp.2022.134035

Received: February 21, 2022
Accepted: April 26, 2022
Published: April 29, 2022

Abstract

Two concepts of phenomenological optics of homogeneous, anisotropic and dispersive media are compared, the younger and more general concept of media with spatial dispersion and the older concept of (bi)-anisotropic media with material tensors for electric and magnetic induction which only depend on the frequency. The general algebraic form of the polarization vectors for the electric field and their one-dimensional projection operators is discussed without the degenerate cases of optic axis for which they become two-dimensional projection operators. Group velocity and diffraction coefficients in an approximate equation for the slowly varying amplitudes of beam solutions are calculated. As special case a polariton permittivity for isotropic media with frequency dispersion but without losses is discussed for the usual passive case and for the active case (occupation inversion of two energy levels that goes in direction of laser theory) and the group velocity is calculated. For this active case, regions of frequency and wave vector with group velocities greater than that of light in vacuum were found. This is not fully understood and due to large diffraction is likely only to realize in guided resonator form. The notion of "negative refraction" is shortly discussed but we did not find agreement with its assessment in the original paper.

Keywords

Spatial and Frequency Dispersion, Bi-Anisotropic Media, Uniaxial Media, Passive and Active Media, Negative Refraction, Operator Invariants, Complementary Operator, Group Velocity

To the Notations

Three-dimensional vectors: bold letters, e.g., $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},\cdots$,

$\boldsymbol{ab}$ scalar products, $[\boldsymbol{a},\boldsymbol{b}]$ vector products, $[\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}]$ volume products, $\boldsymbol{a}\cdot\boldsymbol{b}$ dyadic products,

AB operator products, $\mathsf{A}\boldsymbol{a},\tilde{\boldsymbol{a}}\mathsf{A}$ products of operators with vectors, $\tilde{\boldsymbol{a}}\mathsf{A}\boldsymbol{a}$ bilinear (and quadratic) forms.

In Euclidean spaces with a symmetric metric tensor g_{ij}, the dual tensor $b_{ik}\equiv\epsilon_{ijk}b_j$ to a vector b_j (ϵ_{ijk} Levi-Cicita symbol) can be also seen as antisymmetric operator and in coordinate-invariant form we can write this antisymmetric operator as $[\boldsymbol{b}]$ with the advantage that vector and also volume products can be written only by displacement of the squared brackets, e.g., $\boldsymbol{a}[\boldsymbol{b}]\equiv[\boldsymbol{a},\boldsymbol{b}]$, $[\boldsymbol{b}]\boldsymbol{c}\equiv[\boldsymbol{b},\boldsymbol{c}]$, $\boldsymbol{a}[\boldsymbol{b}]\boldsymbol{c}\equiv[\boldsymbol{a},\boldsymbol{b}]\boldsymbol{c}\equiv\boldsymbol{a}[\boldsymbol{b},\boldsymbol{c}]\equiv[\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}]$.

In mathematical texts I write three-dimensional operators by serif-less Capital letters. In physical texts this makes sometimes difficulties because one cannot reasonably write all operators with physical meaning by Capital letters and all vectors with physical meaning by small letters. Furthermore, in case of Greek letters, "Latex" (and also printing) does not provide serif-less letters. In these cases I write operators as a compromise by bold letters, e.g., $\boldsymbol{\varepsilon},\boldsymbol{\mu}$ such as vectors to distinguish them, in particular, from scalars. This means that in present physical text one must know which kind of quantities one has: scalars, vectors or operators.

1. Introduction

There are two concepts of phenomenological macroscopic optics of the linear constitutive equations for anisotropic dispersive media. The first and younger concept is spatial dispersion in the first time mainly developed by Russian physicist in the fiftieths, in particular, Ginzburg and Agranovich [1], Ginzburg [2], Silin and Rukhadze [3] and in shorter form considered in a new chapter in the new edition of vol. 8 of the course of Landau and Lifshits [4].

The second and much older concept is to use two material equations for the electric and magnetic induction in dependence on the electric and magnetic field using tensors of second and sometimes in addition of third rank (optic activity) depending on frequency only (dispersion) and now often called "bi-anisotropic media" (including also electrical anisotropy only). Three representative monographs of the many possible possible ones are that of Tamm [5], that from Sommerfeld [6] and that of Born and Wolf [7] and in addition the comprehensive encyclopedic article from Szivessy [8]. We cite here also the most basic works of Fyodorov, the initiator of coordinate-invariant methods, and his followers from Minsk [9] [10] [11] who in addition to this concept use last methods and where the monograph [11] contains beside theory also experimental material to different media and crystals and by impression was mainly written by Filippov. Coordinate-invariant methods do not only write the starting equations in vector or tensor form but work from beginning up to the results only with vectors, operators and tensors which have a relation to the problem but not with arbitrary coordinate representations and which are mostly of advantage

compared with often voluminous coordinate representations but with more sophisticated algebra. We also apply in this article widely coordinate-invariant methods where it is possible and used them also long ago in the past, e.g., [12] [13] [14] [15].

As a special case we discuss in detail a permittivity which we call polariton permittivity and which is related to phenomenological theory of excitons, e.g., in addition to [1] [2] [3] [4] by Knox, Agranovich, Davydov, Galanin and Pekar [16] [17] [18] [19] [20]. It admits to consider two essentially different special cases called the passive and the active case. The active case is connected with occupation inversion of at least two energy levels in the medium and is described in certain parts of frequency or connected wave vectors by amplification and leads into the neighborhood to laser theory. It contains also a very interesting phenomenon of propagation of excitations with velocities faster than light that is not fully understood.

In connection with my article [21] the notion of "negative refraction" of Pendry [22] came into the focus of my considerations[1]. I never have used it and likely never would use the notion "negative refraction" in connection with my own results in this field. In Section 11 and in Appendix D I try to represent my imaginations to the content of this notion which seems to me as incorrect ones.

Sections 2-8 are devoted to general characterization and comparison of both concepts including calculation of group velocities with and without taking into consideration the dispersion and Section 9 and Section 10 to the most simple model of a polariton permittivity. Section 11 was made to prepare short remarks to the notion of "negative refraction" in Appendix D.

2. The Concept of Spatial Dispersion

In this more general concept compared with the bi-anisotropic concept considered in next Section we write the equations of macroscopic electrodynamics in the form

$$\begin{aligned} \left[\nabla, \boldsymbol{E}(\boldsymbol{r},t)\right] + \frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{B}(\boldsymbol{r},t) = \boldsymbol{0}, \quad \nabla \boldsymbol{B}(\boldsymbol{r},t) = 0, \\ \left[\nabla, \boldsymbol{B}(\boldsymbol{r},t)\right] - \frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{D}(\boldsymbol{r},t) = \boldsymbol{0}, \quad \nabla \boldsymbol{D}(\boldsymbol{r},t) = 0, \end{aligned} \tag{2.1}$$

where $\boldsymbol{E}(\boldsymbol{r},t)$ is the macroscopic electric and $\boldsymbol{B}(\boldsymbol{r},t)$ the macroscopic magnetic field[2] The linear constitutive equation for spatially and temporally homogeneous but, in general, anisotropic dispersive media are written in the form

$$D_i(\boldsymbol{r},t) = \int \mathrm{d}^3 r' \wedge \mathrm{d}t' \hat{\varepsilon}_{ij}(\boldsymbol{r}-\boldsymbol{r}', t-t') E_j(\boldsymbol{r}', t'). \tag{2.2}$$

We now make a Fourier transformation for $\boldsymbol{E}, \boldsymbol{B}$ and $\boldsymbol{D}$ according to the

[1]Long ago I visited a lecture of U. Leonhardt about this which I did not fully understand but also did not trace the original paper to this time. However, it seemed to me that U.L. assessed this paper very positively and as correct.

[2]Sometimes called magnetic induction but in this concept $\boldsymbol{B}$ and $\boldsymbol{H}$ are identical (see next Section) and $\boldsymbol{D}(\boldsymbol{r},t)$ is the electric induction.

scheme

$$\begin{aligned}\boldsymbol{E}(\boldsymbol{r},t)&=\frac{1}{(2\pi)^4}\int \mathrm{d}^3k\wedge \mathrm{d}\omega\,\boldsymbol{E}(\boldsymbol{k},\omega)\mathrm{e}^{\mathrm{i}(\boldsymbol{kr}-\omega t)},\\ \boldsymbol{E}(\boldsymbol{k},\omega)&=\int \mathrm{d}^3r\wedge \mathrm{d}t\,\boldsymbol{E}(\boldsymbol{r},t)\mathrm{e}^{-\mathrm{i}(\boldsymbol{kr}-\omega t)},\end{aligned}\tag{2.3}$$

we find the transformed constitutive relation in the form

$$D_i(\boldsymbol{k},\omega)=\varepsilon_{ij}(\boldsymbol{k},\omega)E_j(\boldsymbol{k},\omega),\tag{2.4}$$

with the definition of the general permittivity tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)$

$$\varepsilon_{ij}(\boldsymbol{k},\omega)\equiv\int \mathrm{d}^3\rho\wedge \mathrm{d}\tau\hat{\varepsilon}_{ij}(\boldsymbol{\rho},\tau)\mathrm{e}^{-\mathrm{i}(\boldsymbol{k\rho}-\omega\tau)}.\tag{2.5}$$

In general, this tensor is non-symmetric.

After Fourier transformation of (2.1) these equations take on the form

$$\begin{aligned}\left[\boldsymbol{k},\boldsymbol{E}(\boldsymbol{k},\omega)\right]-\frac{\omega}{c}\boldsymbol{B}(\boldsymbol{k},\omega)=\boldsymbol{0},\quad \boldsymbol{kB}(\boldsymbol{k},\omega)=0,\\ \left[\boldsymbol{k},\boldsymbol{B}(\boldsymbol{k},\omega)\right]+\frac{\omega}{c}\boldsymbol{D}(\boldsymbol{k},\omega)=\boldsymbol{0},\quad \boldsymbol{kD}(\boldsymbol{k},\omega)=0,\end{aligned}\tag{2.6}$$

By elimination of $\boldsymbol{B}$ from these equations and using the constitutive Equation (2.4) we find the following operator equation for the Fourier components of the electric field (in case of $\omega\neq 0$)

$$\boldsymbol{0}=\left\{\frac{c^2}{\omega^2}\left(\boldsymbol{k}\cdot\boldsymbol{k}-\boldsymbol{k}^2\mathsf{I}\right)+\boldsymbol{\varepsilon}(\boldsymbol{k},\omega)\right\}\boldsymbol{E}(\boldsymbol{k},\omega)\equiv\mathsf{L}(\boldsymbol{k},\omega)\boldsymbol{E}(\boldsymbol{k},\omega).\tag{2.7}$$

From this equation follows equivalently to the vanishing of the divergence of $\boldsymbol{D}(\boldsymbol{r},t)$

$$0=\boldsymbol{k}\mathsf{L}(\boldsymbol{k},\omega)\boldsymbol{E}(\boldsymbol{k},\omega)=\boldsymbol{k}\boldsymbol{\varepsilon}(\boldsymbol{k},\omega)\boldsymbol{E}(\boldsymbol{k},\omega)=\boldsymbol{kD}(\boldsymbol{k},\omega).\tag{2.8}$$

Equation (2.7) is an operator equation for the Fourier transforms of the electric field to the eigenvalue "zero" which in the original form can be written

$$\boldsymbol{0}=\left\{\frac{c^2}{\dfrac{\partial^2}{\partial t^2}}\left(\nabla\cdot\nabla-\nabla^2\mathsf{I}\right)+\boldsymbol{\varepsilon}\left(-\mathrm{i}\nabla,\mathrm{i}\frac{\partial}{\partial t}\right)\right\}\boldsymbol{E}(\boldsymbol{r},t)\equiv\mathsf{L}\left(-\mathrm{i}\nabla,\mathrm{i}\frac{\partial}{\partial t}\right)\boldsymbol{E}(\boldsymbol{r},t),\tag{2.9}$$

with the differential operator (or integral operator in case of $\boldsymbol{\varepsilon}$)

$$\mathsf{L}\left(-\mathrm{i}\nabla,\mathrm{i}\frac{\partial}{\partial t}\right)\equiv\frac{c^2}{\dfrac{\partial^2}{\partial t^2}}\left(\nabla\cdot\nabla-\nabla^2\mathsf{I}\right)+\boldsymbol{\varepsilon}\left(-\mathrm{i}\nabla,\mathrm{i}\frac{\partial}{\partial t}\right).\tag{2.10}$$

For the solution of these operator equations it is favorable to consider some algebra of the operators before this.

The operator $\mathsf{L}(\boldsymbol{k},\omega)$ in the wave Equation (2.7) is defined by

$$\mathsf{L}(\boldsymbol{k},\omega)\equiv\frac{c^2}{\omega^2}\left(\boldsymbol{k}\cdot\boldsymbol{k}-\boldsymbol{k}^2\mathsf{I}\right)+\boldsymbol{\varepsilon}(\boldsymbol{k},\omega).\tag{2.11}$$

The invariants of the operator $\mathsf{L}(\boldsymbol{k},\omega)$ are ($\boldsymbol{\varepsilon}\equiv\boldsymbol{\varepsilon}(\boldsymbol{k},\omega)$)

$$\left|\mathsf{L}(\boldsymbol{k},\omega)\right| = \frac{c^4}{\omega^4}\boldsymbol{k}^2(\boldsymbol{k}\boldsymbol{\varepsilon}\boldsymbol{k}) - \frac{c^2}{\omega^2}\left(\langle\boldsymbol{\varepsilon}\rangle\boldsymbol{k}\boldsymbol{\varepsilon}\boldsymbol{k} - \boldsymbol{k}\boldsymbol{\varepsilon}^2\boldsymbol{k}\right) + |\boldsymbol{\varepsilon}|,$$
$$\left[\mathsf{L}(\boldsymbol{k},\omega)\right] = \frac{c^4}{\omega^4}\left(\boldsymbol{k}^2\right)^2 - \frac{c^2}{\omega^2}\left(\langle\boldsymbol{\varepsilon}\rangle\boldsymbol{k}^2 + \boldsymbol{k}\boldsymbol{\varepsilon}\boldsymbol{k}\right) + [\boldsymbol{\varepsilon}], \tag{2.12}$$
$$\langle\mathsf{L}(\boldsymbol{k},\omega)\rangle = -2\frac{c^2}{\omega^2}\boldsymbol{k}^2 + \langle\boldsymbol{\varepsilon}\rangle,$$

which are involved in the Cayley-Hamilton identity $\mathsf{L}^3 - \langle\mathsf{L}\rangle\mathsf{L}^2 + [\mathsf{L}]\mathsf{L} - |\mathsf{L}|\mathsf{I} = 0$ for the operator $\mathsf{L} \equiv \mathsf{L}(\boldsymbol{k},\omega)$ (see (A.1) in Appendix A).

The vanishing of the determinant of $\mathsf{L}(\boldsymbol{k},\omega)$

$$\left|\mathsf{L}(\boldsymbol{k},\omega)\right| = 0, \tag{2.13}$$

is the dispersion equation and describes a three-dimensional (hyper)-surface in the four-dimensional space of variables $(\boldsymbol{k},\omega)$. In the specialization to only frequency dispersion ($\boldsymbol{\varepsilon}(\boldsymbol{k},\omega) = \boldsymbol{\varepsilon}(\omega)$) it is identical in content but not in form with the Fresnel Equation (e.g., [6] [7]).

For the complementary operator $\bar{\mathsf{L}}(\boldsymbol{k},\omega)$ to $\mathsf{L}(\boldsymbol{k},\omega)$ we find (see Appendix A)

$$\bar{\mathsf{L}}(\boldsymbol{k},\omega) \equiv \mathsf{L}^2 - \langle\mathsf{L}\rangle\mathsf{L} + [\mathsf{L}]\mathsf{I}$$
$$= \frac{c^4}{\omega^4}\left(\boldsymbol{k}^2\right)\boldsymbol{k}\cdot\boldsymbol{k} - \frac{c^2}{\omega^2}\left(\langle\boldsymbol{\varepsilon}\rangle\boldsymbol{k}\cdot\boldsymbol{k} - \boldsymbol{\varepsilon}\boldsymbol{k}\cdot\boldsymbol{k} - \boldsymbol{k}\cdot\boldsymbol{k}\boldsymbol{\varepsilon} + (\boldsymbol{k}\boldsymbol{\varepsilon}\boldsymbol{k})\mathsf{I}\right) + \bar{\boldsymbol{\varepsilon}},$$
$$\langle\bar{\mathsf{L}}(\boldsymbol{k},\omega)\rangle = [\mathsf{L}(\boldsymbol{k},\omega)],\ \langle\bar{\boldsymbol{\varepsilon}}\rangle = [\boldsymbol{\varepsilon}],\ \boldsymbol{\varepsilon}^{-1} = \frac{\bar{\boldsymbol{\varepsilon}}}{|\boldsymbol{\varepsilon}|},\ \boldsymbol{\varepsilon} = |\boldsymbol{\varepsilon}|(\bar{\boldsymbol{\varepsilon}})^{-1},\ \bar{\bar{\boldsymbol{\varepsilon}}} = |\boldsymbol{\varepsilon}|\boldsymbol{\varepsilon}. \tag{2.14}$$

The complementary operator $\bar{\mathsf{L}}(\boldsymbol{k},\omega)$ to $\mathsf{L}(\boldsymbol{k},\omega)$ plays an important role in optics of anisotropic media. If the determinant of L is vanishing, *i.e.*, $[\mathsf{L}] = 0$ then the squared complementary operator $\bar{\mathsf{L}}^2$ is proportional to $\bar{\mathsf{L}}$, more precisely[3]

$$|\mathsf{L}| = 0: \ \Rightarrow (\bar{\mathsf{L}})^2 = \langle\bar{\mathsf{L}}\rangle\bar{\mathsf{L}},\ \ \langle\bar{\mathsf{L}}\rangle = [\mathsf{L}], \tag{2.15}$$

and Π according to the following definition

$$\Pi \equiv \frac{\bar{\mathsf{L}}}{\langle\bar{\mathsf{L}}\rangle} = \frac{\bar{\mathsf{L}}}{[\mathsf{L}]},\ \ \Rightarrow\ \ \Pi^2 = \Pi,\ \ \langle\Pi\rangle = 1, \tag{2.16}$$

is projection operator to the eigenvalue $\lambda = 0$ of L. If $\boldsymbol{a}$ and $\tilde{\boldsymbol{a}}$ are arbitrary vectors then non-vanishing vectors $\bar{\mathsf{L}}\boldsymbol{a}$ are right-hand eigenvector and non-vanishing vectors $\tilde{\boldsymbol{a}}\bar{\mathsf{L}}$ left-hand eigenvector of L to the eigenvalue $\lambda = 0$, *i.e.*

$$|\mathsf{L}| = 0,\ \ \bar{\mathsf{L}}\boldsymbol{a} \neq \boldsymbol{0},\ \ \tilde{\boldsymbol{a}}\bar{\mathsf{L}} \neq 0: \ \Rightarrow\ \ \mathsf{L}\bar{\mathsf{L}}\boldsymbol{a} = |\mathsf{L}|\boldsymbol{a} = 0,\ \ \tilde{\boldsymbol{a}}\bar{\mathsf{L}}\mathsf{L} = \tilde{\boldsymbol{a}}|\mathsf{L}| = 0. \tag{2.17}$$

This follows from the Cayley-Hamilton identity. Arbitrary right-hand eigenvalues $\bar{\mathsf{L}}\boldsymbol{a}$ are proportional to possible solutions for the Fourier transform of the electric field according to the Equation (2.7). One may introduce mutually

[3]In the general case of three-dimensional operators if $|\mathsf{L}| \neq 0$ we have according to (A.5) the relation $\bar{\mathsf{L}}^2 = [\mathsf{L}]\bar{\mathsf{L}} + |\mathsf{L}|(\mathsf{L} - \langle\mathsf{L}\rangle\mathsf{I})$ as can be straightforwardly calculated from the definitions using the Cayley-Hamilton identity.

normalized "polarization" vectors $\boldsymbol{e}$ and $\tilde{\boldsymbol{e}}$ to the electric field by the condition

$$\Pi = \boldsymbol{e}\cdot\tilde{\boldsymbol{e}}, \quad \langle\Pi\rangle = \tilde{\boldsymbol{e}}\boldsymbol{e} = 1, \quad \Pi^2 = \boldsymbol{e}\cdot\tilde{\boldsymbol{e}} = \Pi,$$

$$\boldsymbol{k}\boldsymbol{\varepsilon}\Pi = (\boldsymbol{k}\boldsymbol{\varepsilon}\boldsymbol{e})\tilde{\boldsymbol{e}}, \quad \Pi\boldsymbol{\varepsilon}\boldsymbol{k} = \boldsymbol{e}(\tilde{\boldsymbol{e}}\boldsymbol{\varepsilon}\boldsymbol{k}), \quad \Rightarrow \quad \boldsymbol{k}\boldsymbol{\varepsilon}\boldsymbol{e} = 0, \quad \tilde{\boldsymbol{e}}\boldsymbol{\varepsilon}\boldsymbol{k} = 0, \tag{2.18}$$

where in dependence on the symmetries of the operator $\mathsf{L}(\boldsymbol{k},\omega)$ the "co-vectors" $\tilde{\boldsymbol{e}}$ can be often specialized, for example, to $\tilde{\boldsymbol{e}} = \boldsymbol{e}^*$ for operators $\mathsf{L}(\boldsymbol{k},\omega) = \left(\mathsf{L}\left(\boldsymbol{k}^*,\omega^*\right)\right)^* \equiv \mathsf{L}^*(\boldsymbol{k},\omega)$.

Thus the explicit for of the projection operators for the determination of polarization vectors of the electric field are ($\boldsymbol{\varepsilon} \equiv \boldsymbol{\varepsilon}(\boldsymbol{k},\omega)$)

$$\Pi(\boldsymbol{k},\omega) = \frac{\overline{\mathsf{L}}(\boldsymbol{k},\omega)}{\left[\mathsf{L}(\boldsymbol{k},\omega)\right]}$$

$$= \frac{\frac{c^4}{\omega^4}\left(\boldsymbol{k}^2\right)\boldsymbol{k}\cdot\boldsymbol{k} - \frac{c^2}{\omega^2}\left(\langle\boldsymbol{\varepsilon}\rangle\boldsymbol{k}\cdot\boldsymbol{k} - \boldsymbol{\varepsilon}\boldsymbol{k}\cdot\boldsymbol{k} - \boldsymbol{k}\cdot\boldsymbol{k}\boldsymbol{\varepsilon} + (\boldsymbol{k}\boldsymbol{\varepsilon}\boldsymbol{k})\mathsf{I}\right) + \overline{\boldsymbol{\varepsilon}}}{\frac{c^4}{\omega^4}\left(\boldsymbol{k}^2\right)^2 - \frac{c^2}{\omega^2}\left(\langle\boldsymbol{\varepsilon}\rangle\boldsymbol{k}^2 + \boldsymbol{k}\boldsymbol{\varepsilon}\boldsymbol{k}\right) + [\boldsymbol{\varepsilon}]},$$

$$\langle\Pi(\boldsymbol{k},\omega)\rangle = 1, \quad \left(\Pi(\boldsymbol{k},\omega)\right)^2 = \Pi(\boldsymbol{k},\omega). \tag{2.19}$$

The degenerate case $\overline{\mathsf{L}}(\boldsymbol{k},\omega) = 0 \Rightarrow \langle\overline{\mathsf{L}}(\boldsymbol{k},\omega)\rangle = \left[\mathsf{L}(\boldsymbol{k},\omega)\right] = 0$ (but not true in inverse order) is the case of optic axes which we do not consider in present article in detail. However, isotropic media where all axes are "optic" axes also belong to this case.

In coordinate-invariant calculations of polarization vectors by means of the projection operator (2.19) as vectors $\boldsymbol{a}$ should be taken only vectors which possess a physical meaning of the considered system. According to (2.17) we have a great selection of possible choice of vectors $\boldsymbol{a}$ and $\tilde{\boldsymbol{a}}$ for determination of such polarization vectors but not all are advantageous. According to $\boldsymbol{k}\boldsymbol{\varepsilon}\boldsymbol{e} = 0$ the right-hand polarization vectors $\boldsymbol{e}$ are perpendicularly to the vector $\boldsymbol{k}\boldsymbol{\varepsilon}$ and one should not choose vectors which form a very small angle with this vector $\boldsymbol{k}\boldsymbol{\varepsilon}$ as, for example, the vector $\boldsymbol{k}$ since then in limiting cases $\boldsymbol{k} \to \boldsymbol{k}\boldsymbol{\varepsilon}$ it becomes undetermined. It seems to be favorable to choose for this purpose vector products of the vectors $\boldsymbol{k}\boldsymbol{\varepsilon}$ or of $\boldsymbol{k}$ with other vectors where the last choice is favorable since in this case the most terms in the numerator of the projection operator (2.19) are canceled. We choose first the vector product $[\boldsymbol{k},\boldsymbol{k}\boldsymbol{\varepsilon}]$ for which we find as (non-normalized) right-hand eigenvectors of the operator $\mathsf{L}(\boldsymbol{k},\omega)$ to the eigenvalue $\lambda = 0$

$$\Pi(\boldsymbol{k},\omega)[\boldsymbol{k},\boldsymbol{k}\boldsymbol{\varepsilon}] = \frac{\left(\overline{\boldsymbol{\varepsilon}} - \frac{c^2}{\omega^2}(\boldsymbol{k}\boldsymbol{\varepsilon}\boldsymbol{k})\mathsf{I}\right)[\boldsymbol{k},\boldsymbol{k}\boldsymbol{\varepsilon}]}{\frac{c^4}{\omega^4}\left(\boldsymbol{k}^2\right)^2 - \frac{c^2}{\omega^2}\left(\langle\boldsymbol{\varepsilon}\rangle\boldsymbol{k}^2 + \boldsymbol{k}\boldsymbol{\varepsilon}\boldsymbol{k}\right) + [\boldsymbol{\varepsilon}]}$$

$$= \frac{\left[\boldsymbol{k}\boldsymbol{\varepsilon},\boldsymbol{k}\boldsymbol{\varepsilon}^2\right] + \frac{c^2}{\omega^2}(\boldsymbol{k}\boldsymbol{\varepsilon}\boldsymbol{k})[\boldsymbol{k}\boldsymbol{\varepsilon},\boldsymbol{k}]}{\frac{c^4}{\omega^4}\left(\boldsymbol{k}^2\right)^2 - \frac{c^2}{\omega^2}\left(\langle\boldsymbol{\varepsilon}\rangle\boldsymbol{k}^2 + \boldsymbol{k}\boldsymbol{\varepsilon}\boldsymbol{k}\right) + [\boldsymbol{\varepsilon}]}, \tag{2.20}$$

where the identity (B.6) was applied. This choice becomes inappropriate in limiting or other cases when $[\boldsymbol{k},\boldsymbol{k\varepsilon}]=\boldsymbol{0}$ that means if they become parallel.

If we directly choose as vector $\boldsymbol{a}$ one of the vectors vectors $\boldsymbol{k},\boldsymbol{\varepsilon k},\boldsymbol{\varepsilon}^2\boldsymbol{k}$ then we find as (non-normalized) polarization vectors of the electric field

$$\begin{aligned}
\Pi(\boldsymbol{k},\omega)\boldsymbol{k} &= \frac{\frac{c^2}{\omega^2}\boldsymbol{k}^2\left(\left(\frac{c^2}{\omega^2}\boldsymbol{k}^2-\langle\boldsymbol{\varepsilon}\rangle\right)\boldsymbol{k}+\boldsymbol{\varepsilon k}\right)+\bar{\boldsymbol{\varepsilon}}\boldsymbol{k}}{\frac{c^4}{\omega^4}\left(\boldsymbol{k}^2\right)^2-\frac{c^2}{\omega^2}\left(\langle\boldsymbol{\varepsilon}\rangle\boldsymbol{k}^2+\boldsymbol{k\varepsilon k}\right)+[\boldsymbol{\varepsilon}]},\\
\Pi(\boldsymbol{k},\omega)\boldsymbol{\varepsilon k} &= \frac{\left(\frac{c^4}{\omega^4}\boldsymbol{k}^2\left(\boldsymbol{k\varepsilon k}\right)-\frac{c^2}{\omega^2}\left(\langle\boldsymbol{\varepsilon}\rangle\boldsymbol{k\varepsilon k}-\boldsymbol{k\varepsilon}^2\boldsymbol{k}\right)+|\boldsymbol{\varepsilon}|\right)\boldsymbol{k}}{\frac{c^4}{\omega^4}\left(\boldsymbol{k}^2\right)^2-\frac{c^2}{\omega^2}\left(\langle\boldsymbol{\varepsilon}\rangle\boldsymbol{k}^2+\boldsymbol{k\varepsilon k}\right)+[\boldsymbol{\varepsilon}]}=0,\\
\Pi(\boldsymbol{k},\omega)\boldsymbol{\varepsilon}^2\boldsymbol{k} &= \frac{\frac{c^2}{\omega^2}\left(\frac{c^2}{\omega^2}\boldsymbol{k}^2\left(\boldsymbol{k\varepsilon}^2\boldsymbol{k}\right)-[\boldsymbol{\varepsilon}]\boldsymbol{k\varepsilon k}+|\boldsymbol{\varepsilon}|\boldsymbol{k}^2\right)\boldsymbol{k}+\left(\frac{c^2}{\omega^2}\boldsymbol{k\varepsilon}^2\boldsymbol{k}+|\boldsymbol{\varepsilon}|\right)\boldsymbol{\varepsilon k}-\frac{c^2}{\omega^2}\left(\boldsymbol{k\varepsilon k}\right)\boldsymbol{\varepsilon}^2\boldsymbol{k}}{\frac{c^4}{\omega^4}\left(\boldsymbol{k}^2\right)^2-\frac{c^2}{\omega^2}\left(\langle\boldsymbol{\varepsilon}\rangle\boldsymbol{k}^2+\boldsymbol{k\varepsilon k}\right)+[\boldsymbol{\varepsilon}]}.
\end{aligned} \tag{2.21}$$

where $\boldsymbol{\varepsilon k}$ is inappropriate since it provides the zero vector and expresses that polarization vectors of the electric field are perpendicularly to $\boldsymbol{k\varepsilon}$. One may check that $\boldsymbol{k\varepsilon}\Pi\boldsymbol{a}=0$ (2.18) in all cases.

Favorable representations of polarization vectors one may often find if we use in addition the vectors to optic axes in the representation of the permittivity tensor in principal axes form, most generally

$$\boldsymbol{\varepsilon}=\varepsilon_1\boldsymbol{c}_1\cdot\tilde{\boldsymbol{c}}_1+\varepsilon_2\boldsymbol{c}_2\cdot\tilde{\boldsymbol{c}}_2+\varepsilon_3\boldsymbol{c}_3\cdot\tilde{\boldsymbol{c}}_3,\quad \tilde{\boldsymbol{c}}_i\boldsymbol{c}_j=\delta_{ij}, \tag{2.22}$$

where all involved vectors and scalars may or may not depend on wave vector and frequency depending on the symmetry of the medium. In lossless case we have the simplification $\tilde{\boldsymbol{c}}_i=\boldsymbol{c}_i^*$ and under additional symmetry of the permittivity tensor $\boldsymbol{\varepsilon}$ for homogeneous waves (real wave vector and frequency) $c_i^*=c_i$. One may choose as vectors $\boldsymbol{a}$ for the determination of polarization vectors the vectors of the optic axes $\boldsymbol{c}_i,(i=1,2,3)$ themselves or the vector products $[\boldsymbol{k},\tilde{\boldsymbol{c}}_i]$.

Another possible approach is via the vector field of the electric induction. From (2.7) using the representation (A.3) for the inverse operator $\boldsymbol{\varepsilon}$ one may derive the following wave equation for the electric induction $\boldsymbol{D}(\boldsymbol{k},\omega)$

$$\boldsymbol{0}=\left\{\frac{c^2}{\omega^2|\boldsymbol{\varepsilon}(\boldsymbol{k},\omega)|}\left(\boldsymbol{k}\cdot\boldsymbol{k}-\left(\boldsymbol{k}^2\right)\mathrm{I}\right)\bar{\boldsymbol{\varepsilon}}(\boldsymbol{k},\omega)+\mathrm{I}\right\}\boldsymbol{D}(\boldsymbol{k},\omega)\equiv\mathrm{L}^{\boldsymbol{D}}(\boldsymbol{k},\omega)\boldsymbol{D}(\boldsymbol{k},\omega). \tag{2.23}$$

where we define the operator $\mathrm{L}^{\boldsymbol{D}}(\boldsymbol{k},\omega)$ by

$$\mathrm{L}^{\boldsymbol{D}}(\boldsymbol{k},\omega)\equiv\frac{c^2}{\omega^2|\boldsymbol{\varepsilon}(\boldsymbol{k},\omega)|}\left(\boldsymbol{k}\cdot\boldsymbol{k}-\left(\boldsymbol{k}^2\right)\mathrm{I}\right)\bar{\boldsymbol{\varepsilon}}(\boldsymbol{k},\omega)+\mathrm{I}-\frac{\boldsymbol{\varepsilon}(\boldsymbol{k},\omega)\cdot\boldsymbol{k}}{\boldsymbol{k\varepsilon}(\boldsymbol{k},\omega)\boldsymbol{k}},\quad \boldsymbol{k}\mathrm{L}^{\boldsymbol{D}}(\boldsymbol{k},\omega)=0. \tag{2.24}$$

We have substituted taking into account $\boldsymbol{kD}=0$ the three-dimensional unit operator I by a two-dimensional unit operator (or projection operator) I'

$$\mathsf{I}' \equiv \mathsf{I} - \frac{\boldsymbol{\varepsilon k}\cdot\boldsymbol{k}}{\boldsymbol{k\varepsilon k}}, \quad \mathsf{I}'^2 = \mathsf{I}', \quad \langle \mathsf{I}' \rangle = 2, \tag{2.25}$$

in such a way the operator L^D possesses the properties

$$\boldsymbol{k}\mathsf{L}^D(\boldsymbol{k},\omega) = \boldsymbol{0}, \quad \mathsf{L}^D(\boldsymbol{k},\omega)\boldsymbol{\varepsilon}(\boldsymbol{k},\omega)\boldsymbol{k} = \boldsymbol{0}. \tag{2.26}$$

It is now a two-dimensional operator by multiplication from the left in the plane perpendicular to $\boldsymbol{k}$ and from right in the plane perpendicular to $\boldsymbol{\varepsilon k}$. Landau and Lifshits [4] prefer for the treatment of some problems more directly the electric induction $\boldsymbol{D}$ but, clearly, without formalizing this with introduction of an operator L^D. The use of $\boldsymbol{D}$ instead of $\boldsymbol{E}$ possesses advantages (orthogonality to $\boldsymbol{k}$) but also disadvantages and we do not consider this.

3. The Concept of Bi-Anisotropic Constitutive Equations

The concept of bi-anisotropic media with the special case of bi-isotropic media is more specially than the concept of spatial dispersion discussed in last Section. The basic equations of macroscopic optics are written in this concept in the following way for the Fourier transforms

$$\begin{aligned} &\left[\boldsymbol{k}, \boldsymbol{E}(\boldsymbol{k},\omega)\right] - \frac{\omega}{c}\boldsymbol{B}(\boldsymbol{k},\omega) = \boldsymbol{0}, \quad \boldsymbol{k}\boldsymbol{B}(\boldsymbol{k},\omega) = 0, \\ &\left[\boldsymbol{k}, \boldsymbol{H}(\boldsymbol{k},\omega)\right] + \frac{\omega}{c}\boldsymbol{D}'(\boldsymbol{k},\omega) = \boldsymbol{0}, \quad \boldsymbol{k}\boldsymbol{D}'(\boldsymbol{k},\omega) = 0, \end{aligned} \tag{3.1}$$

where by definition

$$\boldsymbol{D}'(\boldsymbol{k},\omega) \equiv \boldsymbol{E}(\boldsymbol{k},\omega) + 4\pi\boldsymbol{P}'(\boldsymbol{k},\omega), \quad \boldsymbol{H}(\boldsymbol{k},\omega) \equiv \boldsymbol{B}(\boldsymbol{k},\omega) - 4\pi\boldsymbol{M}(\boldsymbol{k},\omega), \tag{3.2}$$

and where $\boldsymbol{P}'$ is the polarization in a narrow sense and $\boldsymbol{M}$ the magnetization and with constitutive equations of the following form for the Fourier transforms

$$\boldsymbol{D}'(\boldsymbol{k},\omega) = \boldsymbol{\varepsilon}(\omega)\boldsymbol{E}(\boldsymbol{k},\omega), \quad \boldsymbol{B}(\boldsymbol{k},\omega) = \boldsymbol{\mu}(\omega)\boldsymbol{H}(\boldsymbol{k},\omega), \tag{3.3}$$

where we do not assume that $\boldsymbol{\varepsilon}(\omega)$ and $\boldsymbol{\mu}(\omega)$ are symmetric tensors (e.g., magneto-optic effects). Usually, $\boldsymbol{H}$ is called the magnetic field and $\boldsymbol{B}$ the magnetic induction also $\boldsymbol{B}$ is the averaged microscopic magnetic field [4]. These notions are made for the symmetries between $\boldsymbol{E} \rightleftharpoons \boldsymbol{H}$ and $\boldsymbol{D} \rightleftharpoons \boldsymbol{B}$ in the field equations but this may be confusing.

In considered case it is favorable to introduce the notion of refraction vectors $\boldsymbol{n}$ by the definition ($\omega \neq 0$)

$$\boldsymbol{n} \equiv \frac{c}{\omega}\boldsymbol{k}. \tag{3.4}$$

The basic equations for the Fourier transforms of fields (3.1) simplify the slightly and are

$$\begin{aligned} &\left[\boldsymbol{n}, \boldsymbol{E}\right] - \boldsymbol{B} = \boldsymbol{0}, \quad \boldsymbol{n}\boldsymbol{B} = 0, \\ &\left[\boldsymbol{n}, \boldsymbol{H}\right] + \boldsymbol{D}' = \boldsymbol{0}, \quad \boldsymbol{n}\boldsymbol{D}' = 0, \end{aligned} \tag{3.5}$$

together with the constitutive equations

$$\boldsymbol{D}' = \boldsymbol{\varepsilon}\boldsymbol{E}, \quad \boldsymbol{B} = \boldsymbol{\mu}\boldsymbol{H}, \tag{3.6}$$

where we omitted to write the arguments in the fields and in the material tensors, e.g., $\boldsymbol{E}(\boldsymbol{k},\omega) \equiv \boldsymbol{E}$, $\boldsymbol{\varepsilon}(\omega) \equiv \boldsymbol{\varepsilon}$.

From (3.5) follow for a given refraction vectors $\boldsymbol{n}$ the orthogonalities

$$\boldsymbol{B}\boldsymbol{E} = [\boldsymbol{n},\boldsymbol{E}]\boldsymbol{E} = 0, \quad \boldsymbol{D}'\boldsymbol{H} = -[\boldsymbol{n},\boldsymbol{H}]\boldsymbol{H} = 0. \tag{3.7}$$

It is also important to mention here that the Equations (3.5) together with (3.6) remain unchanged under the simultaneous permutations

$$\boldsymbol{E} \leftrightarrow \boldsymbol{H}, \quad \boldsymbol{D}' \leftrightarrow \boldsymbol{B}, \quad \boldsymbol{\varepsilon} \leftrightarrow \boldsymbol{\mu}, \quad \boldsymbol{n} \leftrightarrow -\boldsymbol{n}, \tag{3.8}$$

but, clearly, all this is well known.

First, we derive a wave equation for the Fourier components $\boldsymbol{E}$ of the electric field. For this purpose we use the formula (A.3) for the inverse operator and the mathematical identity (B.6) and, furthermore, the transposition of the application of an operator to a vector $\mathsf{A}\boldsymbol{x} = \boldsymbol{x}\mathsf{A}^{\mathrm{T}}$ where A^{T} is the transposed operator to A and find

$$\begin{aligned}\boldsymbol{0} &= [\boldsymbol{n},\boldsymbol{H}] + \boldsymbol{D}' = \left[\boldsymbol{n},\boldsymbol{\mu}^{-1}\boldsymbol{B}\right] + \boldsymbol{D}' = \frac{\left[\boldsymbol{n},\bar{\boldsymbol{\mu}}[\boldsymbol{n},\boldsymbol{E}]\right]}{|\boldsymbol{\mu}|} + \boldsymbol{D}' = \frac{\left[\boldsymbol{n},[\boldsymbol{n}\boldsymbol{\mu},\boldsymbol{E}\boldsymbol{\mu}]\right]}{|\boldsymbol{\mu}|} + \boldsymbol{D}' \\ &= \frac{\left[\boldsymbol{n},\left[\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{n},\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{E}\right]\right]}{|\boldsymbol{\mu}|} + \boldsymbol{D}' = \left\{\frac{\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{n}\cdot\boldsymbol{n}\boldsymbol{\mu}^{\mathrm{T}} - \left(\boldsymbol{n}\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{n}\right)\boldsymbol{\mu}^{\mathrm{T}}}{|\boldsymbol{\mu}|} + \boldsymbol{\varepsilon}\right\}\boldsymbol{E}.\end{aligned} \tag{3.9}$$

We write this Equation ($\boldsymbol{E} \equiv \boldsymbol{E}\left(\frac{\omega}{c}\boldsymbol{n},\omega\right)$)

$$\mathsf{L}^{\boldsymbol{E}}(\boldsymbol{n})\boldsymbol{E} = \boldsymbol{0}, \quad \Rightarrow \quad \boldsymbol{n}\mathsf{L}^{\boldsymbol{E}}(\boldsymbol{n})\boldsymbol{E} = \boldsymbol{n}\boldsymbol{\varepsilon}\boldsymbol{E} = 0, \tag{3.10}$$

with an operator $\mathsf{L}^{\boldsymbol{E}}(\boldsymbol{n})$ defined by

$$\mathsf{L}^{\boldsymbol{E}}(\boldsymbol{n}) \equiv \frac{\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{n}\cdot\boldsymbol{n}\boldsymbol{\mu}^{\mathrm{T}} - \left(\boldsymbol{n}\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{n}\right)\boldsymbol{\mu}^{\mathrm{T}}}{|\boldsymbol{\mu}|} + \boldsymbol{\varepsilon}. \tag{3.11}$$

Due to symmetry (3.8) in the starting Equations (3.5) and (3.6) one may immediately write down an analogous equation for $\boldsymbol{H}$

$$\mathsf{L}^{\boldsymbol{H}}(\boldsymbol{n})\boldsymbol{H} = \boldsymbol{0}, \quad \Rightarrow \quad \boldsymbol{n}\mathsf{L}^{\boldsymbol{H}}(\boldsymbol{n})\boldsymbol{H} = \boldsymbol{n}\boldsymbol{\mu}\boldsymbol{H} = 0, \tag{3.12}$$

with an operator $\mathsf{L}^{\boldsymbol{H}}(\boldsymbol{n})$ defined by

$$\mathsf{L}^{\boldsymbol{H}}(\boldsymbol{n}) \equiv \frac{\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{n}\cdot\boldsymbol{n}\boldsymbol{\varepsilon}^{\mathrm{T}} - \left(\boldsymbol{n}\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{n}\right)\boldsymbol{\varepsilon}^{\mathrm{T}}}{|\boldsymbol{\varepsilon}|} + \boldsymbol{\mu}. \tag{3.13}$$

Both operators $\mathsf{L}^{\boldsymbol{E}}(\boldsymbol{n})$ and $\mathsf{L}^{\boldsymbol{H}}(\boldsymbol{n})$ possess the form of the operator L discussed in Appendix C with the following substitutions in case of $\mathsf{L}^{\boldsymbol{E}}(\boldsymbol{n})$ which we consider now

$$\mathsf{A} \to \boldsymbol{\mu}^{\mathrm{T}}, \quad \mathsf{B} \to \boldsymbol{\varepsilon}, \quad \boldsymbol{x} = \tilde{\boldsymbol{x}} \to \boldsymbol{n}. \tag{3.14}$$

According to (C.7) and (C.10) its invariants are $\boldsymbol{\mu} \equiv \boldsymbol{\mu}(\omega)$, $\boldsymbol{\varepsilon} \equiv \boldsymbol{\varepsilon}(\omega)$

$$\left\langle \mathsf{L}^{E}(\boldsymbol{n})\right\rangle = \frac{\boldsymbol{n}\boldsymbol{\mu}^{2}\boldsymbol{n} - \langle\boldsymbol{\mu}\rangle \boldsymbol{n}\boldsymbol{\mu}\boldsymbol{n} + |\boldsymbol{\mu}|\langle\boldsymbol{\varepsilon}\rangle}{|\boldsymbol{\mu}|},$$

$$\left[\mathsf{L}^{E}(\boldsymbol{n})\right] = \frac{(\boldsymbol{n}\boldsymbol{\mu}\boldsymbol{n})\boldsymbol{n}^{2} - \left(\left\langle\boldsymbol{\mu}\right\rangle\langle\boldsymbol{\varepsilon}\rangle - \left\langle\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\varepsilon}\right\rangle\right)\boldsymbol{n}\boldsymbol{\mu}\boldsymbol{n} - \langle\boldsymbol{\varepsilon}\rangle\boldsymbol{n}\boldsymbol{\mu}^{2}\boldsymbol{n} + \boldsymbol{n}\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\varepsilon}\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{n}\right) + |\boldsymbol{\mu}|[\boldsymbol{\varepsilon}]}{|\boldsymbol{\mu}|},$$

$$\left|\mathsf{L}^{E}(\boldsymbol{n})\right| = \frac{(\boldsymbol{n}\boldsymbol{\mu}\boldsymbol{n})(\boldsymbol{n}\boldsymbol{\varepsilon}\boldsymbol{n}) - \left(\left\langle\boldsymbol{\mu}^{\mathrm{T}}\overline{\boldsymbol{\varepsilon}}\right\rangle\boldsymbol{n}\boldsymbol{\mu}\boldsymbol{n} + \boldsymbol{n}\boldsymbol{\mu}^{\mathrm{T}}\overline{\boldsymbol{\varepsilon}}\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{n}\right) + |\boldsymbol{\mu}||\boldsymbol{\varepsilon}|}{|\boldsymbol{\mu}|}.$$

(3.15)

We wrote here the general case $\boldsymbol{\mu} \neq \boldsymbol{\mu}^{\mathrm{T}}$ (and $\boldsymbol{\varepsilon} \neq \boldsymbol{\varepsilon}^{\mathrm{T}}$) but it was not necessary to write the sign "T" at $\boldsymbol{\mu}$ for transposition in all cases because, e.g., $\boldsymbol{n}\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{n} = \boldsymbol{n}\boldsymbol{\mu}\boldsymbol{n}$ and for all invariants holds, e.g., $\left\langle\boldsymbol{\mu}^{\mathrm{T}}\right\rangle = \langle\boldsymbol{\mu}\rangle$ (however, e.g., $\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\varepsilon} = \left(\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\mu}\right)^{\mathrm{T}}$)[4]. The complementary operator $\overline{\mathsf{L}^{E}}(\boldsymbol{n})$ obtainable from (C.11) is fairly complicated and we do not write it down.

The dispersion equation for a bi-anisotropic medium is

$$\left|\mathsf{L}^{E}(\boldsymbol{n})\right| = 0, \tag{3.16}$$

or equivalently the analogous equation for the operator $\left|\mathsf{L}^{H}(\boldsymbol{n})\right|$. Polarization vectors $\boldsymbol{e}$ together with left-hand eigenvectors $\tilde{\boldsymbol{e}}$ to the operator $\mathsf{L}^{E}(\boldsymbol{n})$ for the electric field can be obtained by the projection operators

$$\Pi^{E}(\boldsymbol{n}) = \frac{\overline{\mathsf{L}^{E}}(\boldsymbol{n})}{\left\langle\overline{\mathsf{L}^{E}}(\boldsymbol{n})\right\rangle} = \frac{\overline{\mathsf{L}^{E}}(\boldsymbol{n})}{\left[\mathsf{L}^{E}(\boldsymbol{n})\right]} = \boldsymbol{e}\cdot\tilde{\boldsymbol{e}}, \quad \left\langle\Pi^{E}(\boldsymbol{n})\right\rangle = \tilde{\boldsymbol{e}}\boldsymbol{e} = 1, \tag{3.17}$$

but this is complicated and we will it only do for the special case of bi-anisotropic uniaxial media in next Section.

4. Bi-Anisotropic Media as Special Case of Spatial Dispersion

The approach to the linear optics of media by spatial dispersion is much more general than the approach by bi-anisotropic media which is mainly interesting for its symmetries between electric and magnetic quantities and I am not a fan of the last for reason which will become clear in the following. Spatial dispersion is often discussed with expansion of the tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)$ into powers of the wave vector as follows [1] [2] [3] [4]

$$\varepsilon_{ij}(\boldsymbol{k},\omega) = \varepsilon_{ij}(\omega) + \mathrm{i}\frac{c}{\omega}\gamma_{ijk}(\omega)k_{k} + \frac{c^{2}}{\omega^{2}}\alpha_{ijkl}(\omega)k_{k}k_{l} + \cdots,$$

$$\begin{aligned}\varepsilon_{ij}^{-1}(\boldsymbol{k},\omega) &= \varepsilon_{ij}^{-1}(\omega) + \mathrm{i}\frac{c}{\omega}\delta_{ijk}(\omega)k_{k} + \frac{c^{2}}{\omega^{2}}\beta_{ijkl}(\omega)k_{k}k_{l} + \cdots \\ &= \varepsilon_{ij}^{-1}(\omega) - \mathrm{i}\frac{c}{\omega}\varepsilon_{im}^{-1}(\omega)\gamma_{mnk}(\omega)\varepsilon_{nj}^{-1}(\omega)k_{k} \\ &\quad - \frac{c^{2}}{\omega^{2}}\varepsilon_{im}^{-1}(\omega)\left(\alpha_{mnkl}(\omega) + \gamma_{mpk}(\omega)\varepsilon_{pq}^{-1}(\omega)\gamma_{qnl}(\omega)\right)\varepsilon_{nj}^{-1}(\omega)k_{k}k_{l} - \cdots.\end{aligned}$$

(4.1)

[4]All these problems could be removed if one defines $\boldsymbol{B} = \boldsymbol{H}\boldsymbol{\mu}$ instead of (3.6) but this possesses the danger to cause some confusion.

The concept of bi-anisotropic media with the constitutive Equations (3.3) can be expressed as a special case of the concept of spatial dispersion with the following dependence of tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)$ on the wave vector $\boldsymbol{k}$

$$\begin{aligned}\varepsilon_{ij}(\boldsymbol{k},\omega) &= \varepsilon_{ij}(\omega)+\frac{c^2}{\omega^2}\alpha_{ijkl}(\omega)k_k k_l \\ &= \varepsilon_{ij}(\omega)+\frac{c^2}{\omega^2}\epsilon_{ikm}\epsilon_{jln}\left(\delta_{mn}-\mu^{-1}_{mn}(\omega)\right)k_k k_l.\end{aligned} \tag{4.2}$$

This can be expressed in a representation without indices in the form of Equation (2.7) with the general permittivity $\boldsymbol{\varepsilon}(\boldsymbol{k},\omega)$ by (A^{T} means transposition of A; definition $\boldsymbol{k}\equiv\frac{\omega}{c}\boldsymbol{n}$ would shorten the following representation)

$$\boldsymbol{\varepsilon}(\boldsymbol{k},\omega)=\boldsymbol{\varepsilon}(\omega)+\frac{c^2}{\omega^2}\left(\frac{\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{k}\cdot\boldsymbol{k}\boldsymbol{\mu}^{\mathrm{T}}-(\boldsymbol{k}\boldsymbol{\mu}\boldsymbol{k})\boldsymbol{\mu}^{\mathrm{T}}}{|\boldsymbol{\mu}|}-\left(\boldsymbol{k}\cdot\boldsymbol{k}-\boldsymbol{k}^2\mathrm{I}\right)\right). \tag{4.3}$$

In the general concept of spatial dispersion a bi-anisotropic medium appears as second-order effect in the expansion of the general tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)$ in powers of the wave vector $\boldsymbol{k}$. It does, however, not possess the general form of a tensor of forth rank $\alpha_{ijkl}(\omega)$ with only symmetry in the last both indices (k,l) and, furthermore, sum terms which are linear in the wave vector $\boldsymbol{k}$ are completely absent (e.g., optic gyrotropy). The reason that the tensor $\alpha_{ijkl}(\omega)$ does not possess in the concept of bi-anisotropy the general form of such a tensor comes from the neglect of electric quadrupole terms and also of higher electric and magnetic multipole terms in the expansion of the general polarization $\boldsymbol{P}(\boldsymbol{k},\omega)$ in powers of the wave vector $\boldsymbol{k}$. Apart from the first term $\boldsymbol{P}(\omega)$ which mostly provides the greatest contribution and is uniquely defined the higher contributions are difficult to separate from each other since in multipole expansions only the first multipole moment which is non-vanishing is uniquely defined whereas the others depend on the chosen origin of the multipole expansion. The magnitude of the different terms from the multipole effects is difficult to assess but one has to assume that within terms of the same order they should be comparable.

The concept of bi-isotropy with $\varepsilon(\omega)$ and $\mu(\omega)$ as scalars in the constitutive equations is old and goes back under other names to the development of macroscopic electrodynamics by the Maxwell equations and its generalization to bi-anisotropy by transition to second-rank tensors $\varepsilon_{ij}(\omega)$ and $\mu_{ij}(\omega)$ is natural. However, the last leads usually to very complicated formulae if one calculates propagation and reflection and refraction problems (amplitudes included), moreover, if this is made by coordinate methods. The most comprehensive and unrivaled representation was given by Szivessy [8] and long ago I thought that it remains the last which works mainly with coordinate methods. However, more than in most other sources in this respect is made in the book of Fyodorov [9] with coordinate-invariant methods which he initiated and developed. In the book [10] of the same author the concept of bi-anisotropy (Fyodorov calls it "crystals with electric and magnetic anisotropy") is extended to inclusion of op-

tic gyrotropy that even by coordinate-invariant treatment leads as a rule to very complicated formulae. The last chapter in this book contains linear algebra in three-dimensional Euclidean space in a form which is very useful for the application of coordinate-invariant calculations in three-dimensional spaces (chap. IV, pp. 362-450)[5].

An extended concept of bi-anisotropy in the basic equations is maintained, in particular, in the very versatile monograph of de Groot [23] and in nonlinear optics by Bloembergen [24] (called the "Netherland school" in [14] with inclusion of some other authors). Furthermore, there are articles to the calculation of the dyadic Green functions to the Huygens principle for bi-anisotropic media ([15] [25] and, e.g., Weiglhofer [26] with many citations published in: [27]).

It is necessary to report here also about an unprecedented scientific plagiarism in form of a book from 1983 by Hollis C. Chen, a professor of the Ohio University in U.S.A, about which I was informed by Fyodor Ivanovich Fyodorov in the middle of the eighties. This book and papers of Chen cannot be cited in normal way under references and I make some remarks to this case in the following footnote[6].

[5]For inclusion of higher than second-rank tensors the concept of coordinate-invariant treatment without using tensor indices fails but in three-dimensional case also third-rank tensors which are anti-symmetric in two indices can be included since they may be mapped onto second-rank pseudo-tensors.

[6]In the beginning of the eighties I sent my papers with application of coordinate-invariant methods (about 10) to Fyodorov who as it proved did not know them. Since this time we were in loose correspondence up to the end of the eighties and I used a big Optics Conference in Minsk in the eighties especially to meet him there personally and this took place in the main building of Academy of Sciences of Belarus on the main boulevard in Minsk. Once in the eighties I received a letter from Fyodorov about an unprecedented plagiarism in a book of H. C. Chen "Theory of Electromagnetic Waves, A Coordinate-Free Approach" from McGraw-Hill (1983). Fyodorov is not cited there and all is made to camouflage the real authorship of these methods. When I tried to see this book I could not find it and, as usual in such cases, a search programme in the libraries of GDR was started. It was not found in GDR and then in such cases it could be searched in West-Germany. In West-Berlin (about 15 km airline from me but unreachable that time) it was found in the Library of the Technical University and I got it for one month. Altogether, it lasted almost three quarters of a year to get it. All what Fyodorov said turned out to be true. The main part of Chen's book is almost a free translation of main chapters of the book of Fyodorov [9]. However, what Fyodorov obviously did not know was that two chapters of Chen's book are a plagiarism of my paper [15] to Huygens principle. Obviously also that Chen did not know my paper [25]. This journal was hardly known in the world and ceased to exist after the turn in GDR and he also did not use my papers to amplitude relations for reflection and refraction at anisotropic media in "Ann. d. Physik", likely because my notations were too different from that of Fyodorov and were more adapted to Landau and Lifshits [4]. When I tried to inform a branch office of McGraw-Hill in Hamburg (FRG) by a letter about this and to get an exemplar without payment (I could not pay to this time West-German currency) my chief at this time Witlof Brunner demanded to write not about the "plagiarism" that was not acceptable for me and I renounced to send this letter. Later, after the turn in GDR, I found this same exemplar of the book of Chen which I earlier have had for a month in the reading room of the Library of the Technical University in West-Berlin and made a copy. Obviously, to this time it was already taken from market (I tried earlier to get it through a colleague, H. Haake, from Essen in West Germany whom I met at a Workshop in Poland but he wrote me then that it was impossible to order). Fyodorov and his scientific colleagues reached that in scientific media in the West and in Russian newspapers was written about the plagiarism. Weiglhofer [26] (see [27]) and also some others did not know all this and cite the fraud of Chen instead of the genuine authors (but in [26] a book of Chen from 1992 is cited which obviously later could appear in U.S.A.). My chief to this time, W. Brunner, was indifferent and uninterested in all this and was not ready to support me.

5. Optic Uniaxial Bi-Anisotropic Media

We consider now the special case of optic uniaxial bi-anisotropic media which is determined by the following tensors $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$

$$\begin{aligned}\boldsymbol{\varepsilon}(\omega) &= \varepsilon^e(\omega)\boldsymbol{c}\cdot\boldsymbol{c} + \varepsilon^o(\omega)(\mathsf{I}-\boldsymbol{c}\cdot\boldsymbol{c}),\\ \boldsymbol{\mu}(\omega) &= \mu^e(\omega)\boldsymbol{c}\cdot\boldsymbol{c} + \mu^o(\omega)(\mathsf{I}-\boldsymbol{c}\cdot\boldsymbol{c}), \quad (\boldsymbol{c}^2=1).\end{aligned} \tag{5.1}$$

where the tensors $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ considered as operators are symmetric and commute (as consequence of axial symmetry with the same axes for the electric and magnetic properties)

$$\boldsymbol{\mu}(\omega) = \boldsymbol{\mu}^{\mathrm{T}}(\omega), \quad \boldsymbol{\varepsilon}(\omega) = \boldsymbol{\varepsilon}^{\mathrm{T}}(\omega), \quad \boldsymbol{\mu}(\omega)\boldsymbol{\varepsilon}(\omega) = \boldsymbol{\varepsilon}(\omega)\boldsymbol{\mu}(\omega). \tag{5.2}$$

With $\boldsymbol{c}$ we have denoted a unit vector in direction of the common optic axis of the permittivity tensor $\boldsymbol{\varepsilon}$ and the permeability tensor $\boldsymbol{\mu}$ of the uniaxial bi-anisotropic medium (notations of Fyodorov [9]) and $\varepsilon^e, \varepsilon^o$ and μ^e, μ^o (upper indices "*e*" and "*o*" stand for "extraordinary" and "ordinary") are frequency depend material scalars. The complementary operators and the invariants, for example, for $\boldsymbol{\varepsilon} \equiv \boldsymbol{\varepsilon}(\omega)$ are (similarly, $\boldsymbol{\mu}(\omega)$)

$$\begin{aligned}\overline{\boldsymbol{\varepsilon}} &= \varepsilon^o\left(\varepsilon^o\boldsymbol{c}\cdot\boldsymbol{c} + \varepsilon^e(\mathsf{I}-\boldsymbol{c}\cdot\boldsymbol{c})\right),\\ \langle\boldsymbol{\varepsilon}\rangle &= \varepsilon^e + 2\varepsilon^o, \quad [\boldsymbol{\varepsilon}] = \varepsilon^o\left(2\varepsilon^e + \varepsilon^o\right) = \langle\overline{\boldsymbol{\varepsilon}}\rangle, \quad |\boldsymbol{\varepsilon}| = \varepsilon^e\left(\varepsilon^o\right)^2.\end{aligned} \tag{5.3}$$

Using this together with $\left|\mathsf{L}^E(\boldsymbol{n})\right|$ in (3.15) and $\mathsf{L}^H(\boldsymbol{n})$ in (3.13) one may specialize this determinant to

$$\mu^e\left(\mu^o\right)^2\left|\mathsf{L}^E(\boldsymbol{n})\right| = \left(\boldsymbol{n}\boldsymbol{\mu}\boldsymbol{n} - \varepsilon^o\mu^e\mu^o\right)\left(\boldsymbol{n}\boldsymbol{\varepsilon}\boldsymbol{n} - \mu^o\varepsilon^e\varepsilon^o\right) = \varepsilon^e\left(\varepsilon^o\right)^2\left|\mathsf{L}^H(\boldsymbol{n})\right|. \tag{5.4}$$

The dispersion equation that means the vanishing of the determinants $\left|\mathsf{L}^E(\boldsymbol{n})\right|$ or $\left|\mathsf{L}^H(\boldsymbol{n})\right|$ decomposes into a product of two separate equations as follows

$$\begin{aligned}0 &= \boldsymbol{n}\boldsymbol{\mu}\boldsymbol{n} - \varepsilon^o\mu^e\mu^o \quad \leftrightarrow \quad \frac{(\boldsymbol{n}\boldsymbol{c})^2}{\varepsilon^o\mu^o} + \frac{[\boldsymbol{n},\boldsymbol{c}]^2}{\varepsilon^o\mu^e} = 1,\\ 0 &= \boldsymbol{n}\boldsymbol{\varepsilon}\boldsymbol{n} - \mu^o\varepsilon^e\varepsilon^o \quad \leftrightarrow \quad \frac{(\boldsymbol{n}\boldsymbol{c})^2}{\mu^o\varepsilon^o} + \frac{[\boldsymbol{n},\boldsymbol{c}]^2}{\mu^o\varepsilon^e} = 1,\end{aligned} \tag{5.5}$$

which for real positive parameters $\varepsilon^o, \varepsilon^e, \mu^o, \mu^e$ represent two rotation ellipsoids with axes lengths which are the square roots of the denominators in (5.5) and with equal axis length in direction of the optic axis. This means that the two rotation ellipsoids touches in axis direction. The determination of polarization vectors via projection operators as described in Section 2 seems to be too tedious and we choose a more special approach. Due to $\boldsymbol{n}\boldsymbol{\varepsilon}\boldsymbol{E} = 0$ polarization vectors of the electric field have to be perpendicular to the vector $\boldsymbol{n}\boldsymbol{\varepsilon}$ and are therefore representable by the vector product of $\boldsymbol{n}\boldsymbol{\varepsilon}$ with a vector which possesses a component in the plane perpendicular to $\boldsymbol{n}\boldsymbol{\varepsilon}$.

We consider this for the first dispersion equation $\boldsymbol{n}\boldsymbol{\mu}\boldsymbol{n} = \varepsilon^o\mu^e\mu^o$ in (5.5). Using this equation a (non-normalized) polarization vector $\boldsymbol{e}'$ for the electric

field with the proposition $\boldsymbol{e}' = [\boldsymbol{n\mu}, \boldsymbol{c}]$ according to (3.5) and (3.6) has to satisfy the equation

$$\begin{aligned}\mathbf{0} &= \left\{\frac{\boldsymbol{\mu n}\cdot\boldsymbol{n\mu} - (\boldsymbol{n\mu n})\boldsymbol{\mu}}{|\boldsymbol{\mu}|} + \boldsymbol{\varepsilon}\right\}\boldsymbol{e}' = \frac{(\boldsymbol{\mu n}\cdot\boldsymbol{n\mu} - (\boldsymbol{n\mu n})\boldsymbol{\mu})[\boldsymbol{n\mu}, \boldsymbol{c}]}{|\boldsymbol{\mu}|} + \boldsymbol{\varepsilon e}' \\ &= -\frac{\varepsilon^o \mu^e \mu^o}{\mu^e (\mu^o)^2}\boldsymbol{\mu}[\boldsymbol{n\mu}, \boldsymbol{c}] + \boldsymbol{\varepsilon e}' = -\frac{\varepsilon^o}{\mu^o}\left(\mu^e \boldsymbol{c}\cdot\boldsymbol{c} + \mu^o(\mathrm{I} - \boldsymbol{c}\cdot\boldsymbol{c})\right)\mu^o[\boldsymbol{n}, \boldsymbol{c}] + \boldsymbol{\varepsilon e} \\ &= -\varepsilon^o \mu^o [\boldsymbol{n}, \boldsymbol{c}] + \boldsymbol{\varepsilon e}'\end{aligned} \tag{5.6}$$

from which follows

$$\begin{aligned}\boldsymbol{e}' &= \varepsilon^o \mu^o \boldsymbol{\varepsilon}^{-1}[\boldsymbol{n}, \boldsymbol{c}] = \varepsilon^o \mu^o \left(\frac{1}{\varepsilon^e}\boldsymbol{c}\cdot\boldsymbol{c} + \frac{1}{\varepsilon^o}(\mathrm{I} - \boldsymbol{c}\cdot\boldsymbol{c})\right)[\boldsymbol{n}, \boldsymbol{c}] \\ &= \mu^o [\boldsymbol{n}, \boldsymbol{c}] = \frac{\mu^o}{\varepsilon^o}[\boldsymbol{n\varepsilon}, \boldsymbol{c}].\end{aligned} \tag{5.7}$$

For a (non-normalized) polarization vector $\boldsymbol{h}'$ of the magnetic field follows then from (3.5) using (5.7) and in addition the first of the Equations (5.5)

$$\begin{aligned}\boldsymbol{h}' &= \boldsymbol{\mu}^{-1}\boldsymbol{b}' = \boldsymbol{\mu}^{-1}[\boldsymbol{n}, \boldsymbol{e}'] = \mu^o \left(\frac{1}{\mu^e}\boldsymbol{c}\cdot\boldsymbol{c} + \frac{1}{\mu^o}(\mathrm{I} - \boldsymbol{c}\cdot\boldsymbol{c})\right)[\boldsymbol{n}, [\boldsymbol{n}, \boldsymbol{c}]] \\ &= -\frac{\mu^o}{\mu^e}[\boldsymbol{n}, \boldsymbol{c}]^2 \boldsymbol{c} + (\boldsymbol{nc})[\boldsymbol{c}, [\boldsymbol{n}, \boldsymbol{c}]] = (\boldsymbol{nc})\boldsymbol{n} - \varepsilon^o \mu^o \boldsymbol{c} = \frac{1}{\mu^e}[\boldsymbol{n\mu}, [\boldsymbol{n}, \boldsymbol{c}]].\end{aligned} \tag{5.8}$$

To get the analogous relations for the second dispersion equation $\boldsymbol{n\varepsilon n} = \mu^o \varepsilon^e \varepsilon^o$ in (5.5) one has only to apply the symmetry relations (3.8). Thus we find in this case a (non-normalized) polarization vector $\boldsymbol{h}$

$$\boldsymbol{h}' = \varepsilon^o [\boldsymbol{n}, \boldsymbol{c}] = \frac{\varepsilon^o}{\mu^o}[\boldsymbol{n\mu}, \boldsymbol{c}], \tag{5.9}$$

and a (non-normalized) polarization vector $\boldsymbol{e}$

$$\boldsymbol{e}' = (\boldsymbol{nc})\boldsymbol{n} - \mu^o \varepsilon^o \boldsymbol{c} = \frac{1}{\mu^e}[\boldsymbol{n\varepsilon}, [\boldsymbol{n}, \boldsymbol{c}]]. \tag{5.10}$$

For non-normalized polarization vectors scalar factors are unimportant and can be omitted.

We now give preference to (non-normalized) polarization vectors of the electric field and omit there the unfavorable factors. Then we have for the first dispersion equations

$$\begin{aligned}&\boldsymbol{n\mu n} = \varepsilon^o \mu^e \mu^o : \\ &\boldsymbol{e} = [\boldsymbol{n\varepsilon}, \boldsymbol{c}] = \varepsilon^o [\boldsymbol{n}, \boldsymbol{c}], \quad \boldsymbol{h} = \frac{\varepsilon^o}{\mu^e \mu^o}[\boldsymbol{n\mu}, [\boldsymbol{n}, \boldsymbol{c}]] = \frac{\varepsilon^o}{\mu^o}\left((\boldsymbol{nc})\boldsymbol{n} - \varepsilon^o \mu^o \boldsymbol{c}\right),\end{aligned} \tag{5.11}$$

and for the second dispersion equation

$$\begin{aligned}&\boldsymbol{n\varepsilon n} = \mu^o \varepsilon^e \varepsilon^o : \\ &\boldsymbol{e} = [\boldsymbol{n\varepsilon}, [\boldsymbol{n}, \boldsymbol{c}]] = \varepsilon^e \left((\boldsymbol{nc})\boldsymbol{n} - \mu^o \varepsilon^o \boldsymbol{c}\right), \quad \boldsymbol{h} = \frac{\varepsilon^e \varepsilon^o}{\mu^o}[\boldsymbol{n\mu}, \boldsymbol{c}] = \varepsilon^e \varepsilon^o [\boldsymbol{n}, \boldsymbol{c}].\end{aligned} \tag{5.12}$$

It is not difficult to make normalization of the polarization vectors. We left

the factors in such a way that to each dispersion Equations (5.11) and (5.12) separately the vector $\boldsymbol{h}$ follows from $\boldsymbol{e}$ without changing a factor. It is also easy to make a transition to normalized polarization vectors that we do not write down.

The transition to the special case of "only electrically" uniaxial media can be made by definition in (5.13) by the substitution

$$\boldsymbol{\varepsilon}(\omega)=\varepsilon^{e}(\omega)\boldsymbol{c}\cdot\boldsymbol{c}+\varepsilon^{o}(\omega)(\mathrm{I}-\boldsymbol{c}\cdot\boldsymbol{c}),\quad \boldsymbol{\mu}(\omega)=\mathrm{I},\quad\rightarrow\quad \mu^{e}(\omega)=\mu^{o}(\omega)=1. \tag{5.13}$$

The two dispersion equations become asymmetric to each other and are the first for ordinary and the second for extraordinary waves

$$\begin{aligned} 0&=\boldsymbol{n}^{2}-\varepsilon^{o}\quad\leftrightarrow\quad\frac{(\boldsymbol{nc})^{2}+[\boldsymbol{n},\boldsymbol{c}]^{2}}{\varepsilon^{o}}=\frac{\boldsymbol{n}^{2}}{\varepsilon^{o}}=1,\\ 0&=\boldsymbol{n\varepsilon n}-\varepsilon^{e}\varepsilon^{o}\quad\leftrightarrow\quad\frac{(\boldsymbol{nc})^{2}}{\varepsilon^{o}}+\frac{[\boldsymbol{n},\boldsymbol{c}]^{2}}{\varepsilon^{e}}=1. \end{aligned} \tag{5.14}$$

The (non-normalized) polarization vectors for the electric and magnetic field (5.11) and (5.12) become for ordinary waves

$$\begin{aligned} &\boldsymbol{n}^{2}=\varepsilon^{o}:\\ &\boldsymbol{e}=[\boldsymbol{n\varepsilon},\boldsymbol{c}]=\varepsilon^{o}[\boldsymbol{n},\boldsymbol{c}],\quad \boldsymbol{h}=\varepsilon^{o}[\boldsymbol{n},[\boldsymbol{n},\boldsymbol{c}]]=\varepsilon^{o}\big((\boldsymbol{nc})\boldsymbol{n}-\varepsilon^{o}\boldsymbol{c}\big), \end{aligned} \tag{5.15}$$

and for extraordinary waves

$$\begin{aligned} &\boldsymbol{n\varepsilon n}=\varepsilon^{e}\varepsilon^{o}:\\ &\boldsymbol{e}=[\boldsymbol{n\varepsilon},[\boldsymbol{n},\boldsymbol{c}]]=\varepsilon^{e}\big((\boldsymbol{nc})\boldsymbol{n}-\varepsilon^{o}\boldsymbol{c}\big),\quad \boldsymbol{h}=\varepsilon^{e}\varepsilon^{o}[\boldsymbol{n},\boldsymbol{c}]. \end{aligned} \tag{5.16}$$

Thus (electrically) ordinary waves are polarized perpendicular to the axis plane spanned by the axis vector $\boldsymbol{c}$ and the refraction vector $\boldsymbol{n}$ and extraordinary waves within this plane. Amplitude relations for reflection and refraction at the boundary between an isotropic and a uniaxial medium can be found, e.g., in [9], in [11] (a little too complicated) and in [13].

6. Group Velocity and Diffraction Coefficients

In a preliminary summary about the two discussed concepts one can say that spatial dispersion is the more general concept but the concept of bi-anisotropy leads to interesting symmetries between the electric and magnetic properties of media and is up to now often the only concept represented in excellent monographs, e.g., [7]. However, in the last concept it is difficult to include some phenomena such as, for example, natural optical activity although this is tried to make in the book [10] of Fyodorov. The concept of bi-anisotropy is used in the whole work of the Minsk Group [9] [10] [11]. Practically, in all older works about classical optics the concept of bi-anisotropy is used but not under this name and this development is comprehensively represented in the encyclopedic article of Szivessy [8]. One cannot be sure that all terms of a same level of spatial dispersion in a certain order in the wave vectors are included in this more symmetric bi-anisotropic concept or are included in doubled way. In every case one

has to calculate anew such quantities as the group velocity in comparison to the classical optics with only one frequency-dependent permittivity tensor $\boldsymbol{\varepsilon}(\omega)$ with and without taking into account frequency dispersion and this is mostly not easy.

We now consider the concept of spatial dispersion with the permittivity tensor $\boldsymbol{\varepsilon}(\boldsymbol{k},\omega)$ in the wave Equation (2.7) with the operator $\mathsf{L}(\boldsymbol{k},\omega)$ given in (2.11) together with their invariants in (2.12). The dispersion equation that is the vanishing of the determinant of $\mathsf{L}(\boldsymbol{k},\omega)$ can be resolved in a function $\omega=\omega(\boldsymbol{k})$ with different possible branches and if we insert this function into the dispersion equation one obtains identities of the form

$$0=\left|\mathsf{L}(\boldsymbol{k},\omega)\right|, \quad \Rightarrow \quad \omega=\omega(\boldsymbol{k}), \tag{6.1}$$

where $\left|\mathsf{L}(\boldsymbol{k},\omega(\boldsymbol{k}))\right|$ depends only on the wave vector $\boldsymbol{k}$. We introduce two important notions and prepare its calculation for specialized cases. If we differentiate the identity (6.1) one and two times with respect to $\boldsymbol{k}$ according to (we abbreviate $\mathsf{L}\equiv\mathsf{L}(\boldsymbol{k},\omega(\boldsymbol{k}))$)

$$\begin{aligned}0&=\frac{\partial|\mathsf{L}|}{\partial k_i}+\frac{\partial|\mathsf{L}|}{\partial\omega}\frac{\partial\omega}{\partial k_i},\\ 0&=\frac{\partial^2|\mathsf{L}|}{\partial k_i\partial k_j}+\frac{\partial^2|\mathsf{L}|}{\partial k_i\partial\omega}\frac{\partial\omega}{\partial k_j}+\frac{\partial^2|\mathsf{L}|}{\partial k_j\partial\omega}\frac{\partial\omega}{\partial k_i}+\frac{\partial^2|\mathsf{L}|}{\partial\omega^2}\frac{\partial\omega}{\partial k_i}\frac{\partial^2\omega}{\partial k_i\partial k_j}.\end{aligned} \tag{6.2}$$

We define the group velocity

$$v_i\equiv\frac{\partial\omega}{\partial k_i}=-\frac{\dfrac{\partial|\mathsf{L}|}{\partial k_i}}{\dfrac{\partial|\mathsf{L}|}{\partial\omega}}, \tag{6.3}$$

and in addition the symmetric diffraction coefficients

$$\begin{aligned}W_{ij}\equiv\frac{\partial^2\omega}{\partial k_i\partial k_j}&=-\frac{\dfrac{\partial^2|\mathsf{L}|}{\partial k_i\partial k_j}+\dfrac{\partial^2|\mathsf{L}|}{\partial k_i\partial\omega}\dfrac{\partial\omega}{\partial k_j}+\dfrac{\partial^2|\mathsf{L}|}{\partial k_j\partial\omega}\dfrac{\partial\omega}{\partial k_i}}{\dfrac{\partial^2|\mathsf{L}|}{\partial\omega^2}}\\ &=-\frac{1}{\dfrac{\partial^2|\mathsf{L}|}{\partial\omega^2}}\left\{\frac{\partial^2|\mathsf{L}|}{\partial k_i\partial k_j}-\frac{\dfrac{\partial^2|\mathsf{L}|}{\partial k_i\partial\omega}\dfrac{\partial|\mathsf{L}|}{\partial k_j}+\dfrac{\partial^2|\mathsf{L}|}{\partial k_j\partial\omega}\dfrac{\partial|\mathsf{L}|}{\partial k_i}}{\dfrac{\partial|\mathsf{L}|}{\partial\omega}}\right\}=W_{ji},\end{aligned} \tag{6.4}$$

where $W_{ij}=W_{ji}$ is a symmetric bilinear form. Both become important for the beam propagation in second-order approximation.

The formula for the group velocity in (10.1) using (2.18) can be written

$$v_i=-\frac{\left\langle\overline{\mathsf{L}}\dfrac{\partial\mathsf{L}}{\partial k_i}\right\rangle}{\left\langle\overline{\mathsf{L}}\dfrac{\partial\mathsf{L}}{\partial\omega}\right\rangle}=-\frac{\boldsymbol{e}^*\dfrac{\partial\mathsf{L}}{\partial k_i}\boldsymbol{e}}{\boldsymbol{e}^*\dfrac{\partial\mathsf{L}}{\partial\omega}\boldsymbol{e}}, \quad \overline{\mathsf{L}}\propto\boldsymbol{e}\cdot\boldsymbol{e}^*, \tag{6.5}$$

where is taken into account that $\bar{\mathsf{L}}$ is proportional to the dyadic product of (in general, non-normalized) polarization vectors $\boldsymbol{e}\cdot\boldsymbol{e}^*$ of the electric field (see (2.18); $\tilde{\boldsymbol{e}}=\boldsymbol{e}^*$ since we consider the lossless case). We find

$$\begin{aligned}\frac{\partial L_{kl}}{\partial k_i}&=\frac{c^2}{\omega^2}\left(k_k\delta_{il}+\delta_{ik}k_l-2k_i\delta_{kl}\right)+\frac{\partial\varepsilon_{kl}}{\partial k_i},\\ \frac{\partial L_{kl}}{\partial\omega}&=-2\frac{c^2}{\omega^3}\left(k_kk_l-\boldsymbol{k}^2\delta_{kl}\right)+\frac{\partial\varepsilon_{kl}}{\partial\omega},\end{aligned}\tag{6.6}$$

from which follows

$$\begin{aligned}\left\langle\bar{\mathsf{L}}\frac{\partial\mathsf{L}}{\partial k_i}\right\rangle&=\frac{c^2}{\omega^2}\left(\left(\bar{\mathsf{L}}\boldsymbol{k}\right)_i+\left(\boldsymbol{k}\bar{\mathsf{L}}\right)_i-2\left\langle\bar{\mathsf{L}}\right\rangle k_i\right)+\left\langle\bar{\mathsf{L}}\frac{\partial\boldsymbol{\varepsilon}}{\partial k_i}\right\rangle,\\ \left\langle\bar{\mathsf{L}}\frac{\partial\mathsf{L}}{\partial\omega}\right\rangle&=-2\frac{c^2}{\omega^3}\left(\boldsymbol{k}\bar{\mathsf{L}}\boldsymbol{k}-\left\langle\bar{\mathsf{L}}\right\rangle\boldsymbol{k}^2\right)+\left\langle\bar{\mathsf{L}}\frac{\partial\boldsymbol{\varepsilon}}{\partial\omega}\right\rangle,\quad\left\langle\bar{\mathsf{L}}\right\rangle=\left[\mathsf{L}\right]\end{aligned}\tag{6.7}$$

Neglect of spatial dispersion means that we do not take into account the term $\left\langle\bar{\mathsf{L}}\frac{\partial\boldsymbol{\varepsilon}}{\partial k_i}\right\rangle$ in the numerator and neglect of frequency dispersion the term $\left\langle\bar{\mathsf{L}}\frac{\partial\boldsymbol{\varepsilon}}{\partial\omega}\right\rangle$ in the denominator of the formula for the group velocity (6.5).

Under neglect of dispersion one finds from $\left|\mathsf{L}\left(\boldsymbol{k},\omega\right)\right|$ explicitly given in (2.12)

$$v_i=\omega\frac{\frac{c^2}{\omega^2}\left(\boldsymbol{k}^2\left(\left(\boldsymbol{\varepsilon}\boldsymbol{k}\right)_i+\left(\boldsymbol{k}\boldsymbol{\varepsilon}\right)_i\right)+2\left(\boldsymbol{k}\boldsymbol{\varepsilon}\boldsymbol{k}\right)k_i\right)-\left\langle\boldsymbol{\varepsilon}\right\rangle\left(\left(\boldsymbol{\varepsilon}\boldsymbol{k}\right)_i+\left(\boldsymbol{k}\boldsymbol{\varepsilon}\right)_i\right)+\left(\left(\boldsymbol{\varepsilon}^2\boldsymbol{k}\right)_i+\left(\boldsymbol{k}\boldsymbol{\varepsilon}^2\right)_i\right)}{2\left(2\frac{c^2}{\omega^2}\left(\boldsymbol{k}^2\left(\boldsymbol{k}\boldsymbol{\varepsilon}\boldsymbol{k}\right)-\left\langle\boldsymbol{\varepsilon}\right\rangle\boldsymbol{k}\boldsymbol{\varepsilon}\boldsymbol{k}+\boldsymbol{k}\boldsymbol{\varepsilon}^2\boldsymbol{k}\right)\right)}.\tag{6.8}$$

From this follows after scalar multiplication with k_i follows, e.g., [4] [9]

$$\boldsymbol{k}\boldsymbol{v}=\omega,\quad\Rightarrow\quad\frac{\boldsymbol{k}}{\omega}\frac{\partial\omega}{\partial\boldsymbol{k}}\equiv\boldsymbol{n}\boldsymbol{s}=1,\tag{6.9}$$

with definition of the refraction vector $\boldsymbol{n}$ and of the ray vector $\boldsymbol{s}$ by (notations as in [4])

$$\boldsymbol{n}\equiv\frac{c\boldsymbol{k}}{\omega},\quad\boldsymbol{s}\equiv\frac{\boldsymbol{v}}{c}.\tag{6.10}$$

One should not forget that the relation (6.9) is derived under neglect of the dispersion of the permittivity tensor $\boldsymbol{\varepsilon}\left(\boldsymbol{k},\omega\right)$ and the differences between ray vector and group velocity in regions near to resonance frequencies or zeros of the permittivity tensor can become very important and even the direction of the group velocity can be changed by this additional terms.

7. Electrically and Magnetically Isotropic Media and Group Velocity

The constitutive equations for bi-isotropic or electrically and magnetically isotropic media are

$$\boldsymbol{D}(\boldsymbol{k},\omega)=\varepsilon(\omega)\boldsymbol{E}(\boldsymbol{k},\omega),\quad \boldsymbol{B}(\boldsymbol{k},\omega)=\mu(\omega)\boldsymbol{H}(\boldsymbol{k},\omega), \tag{7.1}$$

with scalar functions $\varepsilon(\omega)$ and $\mu(\omega)$. The equation for the electric field (3.10) in the concept of bi-anisotropy with the specialized operator (3.11) after multiplication with $\mu(\omega)$ becomes

$$\boldsymbol{0}=\mu(\omega)\mathsf{L}^{E}(\boldsymbol{n})\boldsymbol{E}\left(\frac{\omega}{c}\boldsymbol{n},\omega\right)=\left\{\boldsymbol{n}\cdot\boldsymbol{n}-\boldsymbol{n}^{2}\mathsf{I}+\varepsilon(\omega)\mu(\omega)\mathsf{I}\right\}\boldsymbol{E}\left(\frac{\omega}{c}\boldsymbol{n},\omega\right) \tag{7.2}$$

or equivalently by transition to the more general concept of spatial dispersion with $\boldsymbol{\varepsilon}(\boldsymbol{k},\omega)=\varepsilon(\omega)\mu(\omega)\mathsf{I}$

$$\boldsymbol{0}=\mathsf{L}(\boldsymbol{k},\omega)\boldsymbol{E}(\boldsymbol{k},\omega)=\left\{\frac{c^{2}}{\omega^{2}}\left(\boldsymbol{k}\cdot\boldsymbol{k}-\boldsymbol{k}^{2}\mathsf{I}\right)+\varepsilon(\omega)\mu(\omega)\mathsf{I}\right\}\boldsymbol{E}(\boldsymbol{k},\omega). \tag{7.3}$$

The dispersion equation for transversal waves polarized for both $\boldsymbol{E}$ and $\boldsymbol{B}$ in the plane perpendicular to wave vector $\boldsymbol{k}$

$$\frac{c^{2}}{\omega^{2}}\boldsymbol{k}^{2}\equiv\boldsymbol{n}^{2}=\varepsilon(\omega)\mu(\omega),\quad \boldsymbol{k}\boldsymbol{E}(\boldsymbol{k},\omega)=0, \tag{7.4}$$

and for longitudinal waves

$$\varepsilon(\omega)\mu(\omega)=0,\quad \boldsymbol{E}(\omega)\neq\boldsymbol{0},\quad \boldsymbol{B}(\omega)=0. \tag{7.5}$$

The longitudinal waves correspond in present approximation to pure temporal oscillations of the electric field with arbitrary possible direction of polarization (since $\boldsymbol{k}=\boldsymbol{0}$). We are interested here merely in the transversal waves.

The dispersion Equation (7.4) for transversal waves can be resolved in the form $\omega=\omega(\boldsymbol{k})$ (6.1) with different branches for $\omega(\boldsymbol{k})$. In Sections 9 and 10 we will consider in detail an example where the dispersion Equation (7.4) can be explicitly resolved in the form (6.1) with different branches, the permittivity for polaritons. By differentiation of the dispersion equation in the form (6.1) with respect to the wave vector $\boldsymbol{k}$ one may derive a general formula for the group velocity $\boldsymbol{v}\equiv\frac{\partial\omega}{\partial\boldsymbol{k}}$ for bi-isotropic media and also for higher coefficients $\frac{\partial^{2}\omega}{\partial k_{i}\partial k_{j}}$ and so on which play a role in higher approximations of propagation of beam-like waves in such media (diffraction). For the group velocity $\boldsymbol{v}$ one finds the general formula ($\boldsymbol{k}^{2}=|\boldsymbol{k}|^{2}$)

$$\boldsymbol{v}\equiv\frac{\partial\omega}{\partial\boldsymbol{k}}=\frac{2c^{2}\boldsymbol{k}}{\frac{\partial}{\partial\omega}\left(\omega^{2}\varepsilon(\omega)\mu(\omega)\right)}=c\frac{c|\boldsymbol{k}|}{\omega\left(\varepsilon(\omega)\mu(\omega)+\frac{\omega}{2}\frac{\partial}{\partial\omega}\left(\varepsilon(\omega)\mu(\omega)\right)\right)}\frac{\boldsymbol{k}}{|\boldsymbol{k}|}. \tag{7.6}$$

Without taking into account the frequency dispersion of the permittivities we find

$$\boldsymbol{v}'\equiv\frac{\partial\omega}{\partial\boldsymbol{k}}=\frac{2c^{2}\boldsymbol{k}}{\varepsilon(\omega)\mu(\omega)\frac{\partial}{\partial\omega}\left(\omega^{2}\right)}=c\frac{c|\boldsymbol{k}|}{\omega\varepsilon(\omega)\mu(\omega)}\frac{\boldsymbol{k}}{|\boldsymbol{k}|}=\frac{c}{\sqrt{\varepsilon(\omega)\mu(\omega)}}\frac{\boldsymbol{k}}{|\boldsymbol{k}|}. \tag{7.7}$$

One may introduce a dispersion factor $\alpha_{\mathrm{disp}}(\omega)$ by

$$\alpha_{\text{disp}}(\omega) \equiv \frac{\varepsilon(\omega)\mu(\omega)\frac{\partial}{\partial\omega}(\omega^2)}{\frac{\partial}{\partial\omega}(\omega^2\varepsilon(\omega)\mu(\omega))} = \frac{\varepsilon(\omega)\mu(\omega)}{\varepsilon(\omega)\mu(\omega)+\frac{\omega}{2}\frac{\partial}{\partial\omega}(\varepsilon(\omega)\mu(\omega))}. \tag{7.8}$$

From (7.7) follows using the dispersion Equation (7.4)

$$\frac{\boldsymbol{k}}{\omega}\frac{\partial\omega}{\partial\boldsymbol{k}} = \frac{1}{1+\frac{\omega}{2\varepsilon(\omega)\mu(\omega)}\frac{\partial}{\partial\omega}(\varepsilon(\omega)\mu(\omega))} = \frac{1}{\frac{\omega}{2}\frac{\partial}{\partial\omega}(\log(\omega^2\varepsilon(\omega)\mu(\omega)))} \equiv \alpha_{\text{disp}}(\omega), \tag{7.9}$$

where we have given also a representation of the dispersion coefficient $\alpha_{\text{disp}}(\omega)$ by a logarithmic derivative (useful or not?). Under neglect of dispersion that means if we do not take into account the first derivative of $\varepsilon(\omega)\mu(\omega)$ with respect to frequency ω we have ($\boldsymbol{s}$ is called ray vector, e.g., [4], §97, or [9] $(\boldsymbol{n},\boldsymbol{s})\to(\boldsymbol{m},\boldsymbol{p})$)

$$\frac{\boldsymbol{k}}{\omega}\frac{\partial\omega}{\partial\boldsymbol{k}} \equiv \boldsymbol{n}\underbrace{\frac{\boldsymbol{v}}{c}}_{=\boldsymbol{s}} = 1, \quad \Rightarrow \quad \alpha_{\text{disp}}(\omega) = 1. \tag{7.10}$$

Therefore, the dispersion coefficient $\alpha_{\text{disp}}(\omega)$ says by which factor one has to modify the group velocity in comparison to neglect of dispersion if one take it into account. It goes also into some other formulae as, for example, the energy of the wave solution. On this very general level of treatment we cannot say whether or not $\alpha_{\text{disp}}(\omega)$ is in every case positive for possible real-valued functions $\varepsilon(\omega)\mu(\omega)$. In last case of negative $\alpha_{\text{disp}}(\omega)$ the genuine group velocity and the ray vector would have even opposite directions. The introduction of ray vectors $\boldsymbol{s}$ in addition to the refraction vectors $\boldsymbol{n}$ is appropriate to formulate duality (or symmetry) relations between electric and magnetic quantities which leave invariant the basic equations of macroscopic optics [4], (§ 97) such as (for $\boldsymbol{\mu}(\omega)=1$)

$$\boldsymbol{E}\leftrightarrow\boldsymbol{D}, \quad \varepsilon_{ij}\leftrightarrow\varepsilon_{ij}^{-1}, \quad \boldsymbol{n}\leftrightarrow\boldsymbol{s}, \tag{7.11}$$

but one should not forget that these are only approximate relations and are only true under neglect of dispersion of the permittivities.

A second coefficient which we will consider is the relation of the modulus $|\boldsymbol{v}|$ of the group velocity $\boldsymbol{v}$ to the light velocity c for which one derives from (7.7) the relation

$$\frac{\boldsymbol{v}}{c} = \frac{\sqrt{\varepsilon(\omega)\mu(\omega)}}{\varepsilon(\omega)\mu(\omega)+\frac{\omega}{2}\frac{\partial}{\partial\omega}(\varepsilon(\omega)\mu(\omega))}\frac{\boldsymbol{k}}{|\boldsymbol{k}|} = \beta(\omega)\frac{\boldsymbol{k}}{|\boldsymbol{k}|}, \quad (\omega>0), \tag{7.12}$$

with definition

$$\beta(\omega)\equiv\frac{\boldsymbol{kv}}{|\boldsymbol{k}|c}=\frac{\sqrt{\varepsilon(\omega)\mu(\omega)}}{\varepsilon(\omega)\mu(\omega)+\frac{\omega}{2}\frac{\partial}{\partial\omega}\big(\varepsilon(\omega)\mu(\omega)\big)}=\frac{\alpha_{\text{disp}}(\omega)}{\sqrt{\varepsilon(\omega)\mu(\omega)}}. \tag{7.13}$$

For negative or complex $\varepsilon(\omega)\mu(\omega)$ it becomes imaginary or complex and is then not to interpret in easy way.

Let us write down at this opportunity the general form of the second-order coefficients $\frac{\partial^2\omega}{\partial k_i\partial k_j}$ for bi-isotropic media which are

$$\begin{aligned}\frac{\partial^2\omega}{\partial k_i\partial k_j}&=\frac{2c^2\delta_{ij}-\frac{\partial^2}{\partial\omega^2}\big(\omega^2\varepsilon(\omega)\mu(\omega)\big)v_iv_j}{\frac{\partial}{\partial\omega}\big(\omega^2\varepsilon(\omega)\mu(\omega)\big)}\\&=\frac{2c^2}{\frac{\partial}{\partial\omega}\big(\omega^2\varepsilon(\omega)\mu(\omega)\big)}\\&\quad\cdot\left\{\delta_{ij}-\frac{k_ik_j}{|\boldsymbol{k}|^2}+\left(1-\frac{2\omega^2\varepsilon(\omega)\mu(\omega)\frac{\partial^2}{\partial\omega^2}\big(\omega^2\varepsilon(\omega)\mu(\omega)\big)}{\left(\frac{\partial}{\partial\omega}\big(\omega^2\varepsilon(\omega)\mu(\omega)\big)\right)^2}\right)\frac{k_ik_j}{|\boldsymbol{k}|^2}\right\}.\end{aligned} \tag{7.14}$$

As was to expect they are a linear combination of $\delta_{ij}-\frac{k_ik_j}{|\boldsymbol{k}|^2}$ and $\frac{k_ik_j}{|\boldsymbol{k}|^2}$ the only second-rank symmetric tensors which can be built from vectors $\boldsymbol{k}$ alone and which are covariant under transformations of the rotation group $SO(3)$. The group velocity and the diffraction coefficients are involved in the expansion of the equation for the slowly varying amplitudes of beams with respect to spatial and temporal derivatives. We will shortly consider the corresponding equations in next Section.

8. Approximate Beam Equations for Homogeneous Isotropic Media

We consider an isotropic medium with permittivity $\varepsilon(\omega)$ and for simplicity with $\mu(\omega)=1$. The wave equation for the electric field with the involved operator $\mathsf{L}(\boldsymbol{k},\omega)$ in such a medium is

$$\boldsymbol{0}=\mathsf{L}\left(-\mathrm{i}\nabla,\mathrm{i}\frac{\partial}{\partial t}\right)\boldsymbol{E}(\boldsymbol{r},t),\quad \mathsf{L}(\boldsymbol{k},\omega)\equiv\frac{c^2}{\omega^2}\left(\boldsymbol{k}\cdot\boldsymbol{k}-\boldsymbol{k}^2\mathsf{I}\right)+\varepsilon(\omega)\mathsf{I}. \tag{8.1}$$

The necessary equation for solutions is the vanishing of the determinant

$$\boldsymbol{0}=\left|\mathsf{L}\left(-\mathrm{i}\nabla,\mathrm{i}\frac{\partial}{\partial t}\right)\right|\boldsymbol{E}(\boldsymbol{r},t). \tag{8.2}$$

We make now the proposition of slowly varying amplitudes $\boldsymbol{E}_0(\boldsymbol{r},t)$

$$\boldsymbol{E}(\boldsymbol{r},t)=\boldsymbol{E}_0(\boldsymbol{r},t)\mathrm{e}^{\mathrm{i}(\boldsymbol{k}_0\boldsymbol{r}-\omega_0t)}+\boldsymbol{E}_0^*(\boldsymbol{r},t)\mathrm{e}^{-\mathrm{i}(\boldsymbol{k}_0\boldsymbol{r}-\omega_0t)}. \tag{8.3}$$

Inserting this into (8.2) we find first

$$\begin{aligned}\mathbf{0} &= \mathrm{e}^{\mathrm{i}(\boldsymbol{k}_0 r-\omega_0 t)}\mathsf{L}\left(\boldsymbol{k}_0-\mathrm{i}\boldsymbol{\nabla},\omega_0+\mathrm{i}\frac{\partial}{\partial t}\right)\boldsymbol{E}_0(\boldsymbol{r},t)\\ &+\mathrm{e}^{-\mathrm{i}(\boldsymbol{k}_0 r-\omega_0 t)}\mathsf{L}\left(-\boldsymbol{k}_0-\mathrm{i}\boldsymbol{\nabla},\omega_0-\mathrm{i}\frac{\partial}{\partial t}\right)\boldsymbol{E}_0^*(\boldsymbol{r},t),\end{aligned} \tag{8.4}$$

We suppose that both sum parts are separated in a way that we can set them equal to zero independently. For example, we may think that we include into one part all frequencies $0<\omega<+\infty$ and into the other part all frequencies $-\infty<\omega<0$. With this assumption we have the equation

$$\begin{aligned}\mathbf{0} &= \mathsf{L}\left(\boldsymbol{k}_0-\mathrm{i}\boldsymbol{\nabla},\omega_0+\mathrm{i}\frac{\partial}{\partial t}\right)\boldsymbol{E}_0(\boldsymbol{r},t)\\ &=\left\{\mathsf{L}_0-\mathrm{i}\left(\left(\frac{\partial\mathsf{L}}{\partial\boldsymbol{k}}\right)_0\boldsymbol{\nabla}-\left(\frac{\partial\mathsf{L}}{\partial\omega}\right)_0\frac{\partial}{\partial t}\right)-\frac{1}{2}\left(\left(\frac{\partial^2\mathsf{L}}{\partial k_i\partial k_j}\right)_0\nabla_i\nabla_j\right.\right.\\ &\left.\left.-2\left(\frac{\partial^2\mathsf{L}}{\partial k_i\partial\omega}\right)_0\nabla_i\frac{\partial}{\partial t}+\left(\frac{\partial^2\mathsf{L}}{\partial\omega^2}\right)_0\frac{\partial^2}{\partial t^2}\right)+\cdots\right\}\boldsymbol{E}_0(\boldsymbol{r},t),\end{aligned} \tag{8.5}$$

where we wrote the first terms in an expansion in powers of the differential operators. From this equation follows as necessary condition for all components of the solutions $\boldsymbol{E}_0(\boldsymbol{r},t)$ the vanishing of the determinant and with the analogous expansion as in (8.5)

$$\begin{aligned}\mathbf{0} &= \left|\mathsf{L}\left(\boldsymbol{k}_0-\mathrm{i}\boldsymbol{\nabla},\omega_0+\mathrm{i}\frac{\partial}{\partial t}\right)\right|\boldsymbol{E}_0(\boldsymbol{r},t)\\ &=\left\{|\mathsf{L}_0|-\mathrm{i}\left(\left(\frac{\partial|\mathsf{L}|}{\partial k_i}\right)_0\nabla_i-\left(\frac{\partial|\mathsf{L}|}{\partial\omega}\right)_0\frac{\partial}{\partial t}\right)-\frac{1}{2}\left(\left(\frac{\partial^2|\mathsf{L}|}{\partial k_i\partial k_j}\right)_0\nabla_i\nabla_j\right.\right.\\ &\left.\left.-2\left(\frac{\partial^2|\mathsf{L}|}{\partial k_i\partial\omega}\right)_0\nabla_i\frac{\partial}{\partial t}+\left(\frac{\partial^2|\mathsf{L}|}{\partial\omega^2}\right)_0\frac{\partial^2}{\partial t^2}\right)+\cdots\right\}\boldsymbol{E}_0(\boldsymbol{r},t),\end{aligned} \tag{8.6}$$

plus the corresponding complex conjugate equation. With the general formula for the differentiation of the determinant $\frac{\partial}{\partial\lambda}|\mathsf{A}|=\left\langle\bar{\mathsf{A}}\frac{\partial}{\partial\lambda}\mathsf{A}\right\rangle$ of an operator A with respect to a parameter λ this equation may be also written (we do not insert a more complicated formula for the second derivative of a determinant with respect to two parameters which also exists)

$$\begin{aligned}\mathbf{0} &=\left\{-\mathrm{i}\left(\left\langle\bar{\mathsf{L}}_0\left(\frac{\partial\mathsf{L}}{\partial k_i}\right)_0\right\rangle\nabla_i-\left\langle\bar{\mathsf{L}}_0\left(\frac{\partial\mathsf{L}}{\partial\omega}\right)_0\right\rangle\frac{\partial}{\partial t}\right)-\frac{1}{2}\left(\left(\frac{\partial^2|\mathsf{L}|}{\partial k_i\partial k_j}\right)_0\nabla_i\nabla_j\right.\right.\\ &\left.\left.-2\left(\frac{\partial^2|\mathsf{L}|}{\partial k_i\partial\omega}\right)_0\nabla_i\frac{\partial}{\partial t}+\left(\frac{\partial^2|\mathsf{L}|}{\partial\omega^2}\right)_0\frac{\partial^2}{\partial t^2}\right)+\cdots\right\}\boldsymbol{E}_0(\boldsymbol{r},t).\end{aligned} \tag{8.7}$$

with

$$|\mathsf{L}_0|\equiv|\mathsf{L}(\boldsymbol{k}_0,\omega_0)|=0, \tag{8.8}$$

and where index 0 means that the derivatives are to take at $(\boldsymbol{k},\omega)=(\boldsymbol{k}_0,\omega_0)$. By

division of this equation with $\left\langle \left(\bar{\mathsf{L}}_0 \frac{\partial \mathsf{L}}{\partial \omega} \right)_0 \right\rangle$ one finds

$$\begin{aligned} 0 = \Bigg\{ & \mathrm{i}\left(\frac{\partial}{\partial t} - \frac{\left\langle \bar{\mathsf{L}}_0 \left(\frac{\partial \mathsf{L}}{\partial k_i} \right)_0 \right\rangle}{\left\langle \bar{\mathsf{L}}_0 \left(\frac{\partial \mathsf{L}}{\partial \omega} \right)_0 \right\rangle} \nabla_i \right) \\ & - \frac{1}{2} \frac{\left(\frac{\partial^2 |\mathsf{L}|}{\partial k_i \partial k_j} \right)_0 \nabla_i \nabla_j - 2 \left(\frac{\partial^2 |\mathsf{L}|}{\partial k_i \partial \omega} \right)_0 \nabla_i \frac{\partial}{\partial t} + \left(\frac{\partial^2 |\mathsf{L}|}{\partial \omega^2} \right)_0 \frac{\partial^2}{\partial t^2}}{\left\langle \left(\bar{\mathsf{L}}_0 \frac{\partial \mathsf{L}}{\partial \omega} \right)_0 \right\rangle} + \cdots \Bigg\} \boldsymbol{E}_0(\boldsymbol{r}, t). \end{aligned} \tag{8.9}$$

The dispersion equation $\left| \mathsf{L}(\boldsymbol{k}, \omega) \right| = 0$ can be resolved in the form $\omega = \omega(\boldsymbol{k})$ for the different branches of the solution. In application to the slowly varying amplitudes with average wave vector $\boldsymbol{k}_0$ and frequency $\omega = \omega_0$ this means the resolution

$$\begin{aligned} \mathbf{0} &= \left\{ \omega_0 + \mathrm{i} \frac{\partial}{\partial t} - \omega(\boldsymbol{k}_0 - \mathrm{i}\nabla) \right\} \boldsymbol{E}_0(\boldsymbol{r}, t) \\ &= \left\{ \mathrm{i}\left(\frac{\partial}{\partial t} + \boldsymbol{v}_0 \nabla \right) + \frac{1}{2} \nabla \mathsf{W}_0 \nabla + \cdots \right\} \boldsymbol{E}_0(\boldsymbol{r}, t), \quad \left(\omega_0 = \omega_0(\boldsymbol{k}_0) \right), \end{aligned} \tag{8.10}$$

with the group velocity $\boldsymbol{v}_0$ and the quadratic form W_0 defined by

$$\boldsymbol{v}_0 \equiv \left(\frac{\partial \omega}{\partial \boldsymbol{k}} \right)_0, \quad \mathsf{W}_0 \equiv \left(\frac{\partial^2 \omega}{\partial \boldsymbol{k} \cdot \partial \boldsymbol{k}} \right)_0, \quad \left(\text{or} \quad W_{0,ij} \equiv \left(\frac{\partial^2 \omega}{\partial k_i \partial k_j} \right)_0 \right). \tag{8.11}$$

Obviously (8.9) and (8.10) are identical and one may find the correspondences.

We define the polarization vectors $\boldsymbol{e}_0$ and $\boldsymbol{e}_0^*$ which are right-hand and left-hand eigen-vectors to the operator L_0 to eigenvalue 0 according to

$$0 = \mathsf{L}_0 \boldsymbol{e}_0, \quad 0 = \boldsymbol{e}_0^* \mathsf{L}_0, \quad \boldsymbol{k}_0 \boldsymbol{e}_0 = 0, \quad \boldsymbol{e}_0^* \boldsymbol{k}_0 = 0. \tag{8.12}$$

By comparison of (8.9) with (8.10) we find for the group velocity v_0

$$v_{0,i} \equiv \left(\frac{\partial \omega}{\partial k_i} \right)_0 = -\frac{\left(\frac{\partial |\mathsf{L}|}{\partial k_i} \right)_0}{\left(\frac{\partial |\mathsf{L}|}{\partial \omega} \right)_0} = -\frac{\left\langle \bar{\mathsf{L}}_0 \left(\frac{\partial \mathsf{L}}{\partial k_i} \right)_0 \right\rangle}{\left\langle \bar{\mathsf{L}}_0 \left(\frac{\partial \mathsf{L}}{\partial \omega} \right)_0 \right\rangle} = -\frac{\boldsymbol{e}_0^* \left(\frac{\partial \mathsf{L}}{\partial k_i} \right)_0 \boldsymbol{e}_0}{\boldsymbol{e}_0^* \left(\frac{\partial \mathsf{L}}{\partial \omega} \right)_0 \boldsymbol{e}_0}, \tag{8.13}$$

where we applied $\mathsf{L}_0 \propto \boldsymbol{e}_0 \cdot \boldsymbol{e}_0^*$ meaning that L_0 is proportional to the dyadic product of the polarization vectors $\boldsymbol{e}_0$ and $\boldsymbol{e}_0^*$ (see also formulae (2.18) and (2.19)).

The beam solutions in their Fourier decomposition contain components to wave vectors and frequencies around the average wave vectors and frequency $(\boldsymbol{k}_0, \omega_0)$ and therefore the solution cannot possess solution which are exactly proportional to polarization vector $\boldsymbol{e}_0$. Therefore we make now the following proposition for solutions of the beam Equation (8.5)

$$\boldsymbol{E}_0(\boldsymbol{r},t)=\boldsymbol{e}_0A_0(\boldsymbol{r},t)+\left[\boldsymbol{e}_0,\boldsymbol{A}'(\boldsymbol{r},t)\right], \tag{8.14}$$

where $\boldsymbol{e}_0A_0(\boldsymbol{r},t)$ is the main part with polarization $\boldsymbol{e}_0$ and $\left[\boldsymbol{e}_0,\boldsymbol{A}'(\boldsymbol{r},t)\right]$ a small additional part with polarization perpendicular to $\boldsymbol{e}_0$. Both parts have to satisfy Equation (8.5) that means for the main part the following approximate scalar equation up to second-order derivatives of the slowly varying amplitude $A_0(\boldsymbol{r},t)$

$$0=\left\{\mathrm{i}\left(\frac{\partial}{\partial t}+\boldsymbol{v}_0\nabla\right)+\frac{1}{2}\nabla\mathsf{W}_0\nabla\right\}A_0(\boldsymbol{r},t),\quad\left(\nabla\mathsf{W}_0\nabla\equiv\nabla_iW_{0,ij}\nabla_j\right). \tag{8.15}$$

The additional part $\left[\boldsymbol{e}_0,\boldsymbol{A}'(\boldsymbol{r},t)\right]$ of the beam solution is not independent of the main part $\boldsymbol{e}_0A_0(\boldsymbol{r},t)$. Inserting both parts into Equation (8.5) we get approximately using $\mathsf{L}_0\boldsymbol{e}_0=\boldsymbol{0}$

$$\begin{aligned}\boldsymbol{0}&=\left\{\mathsf{L}_0-\mathrm{i}\left(\left(\frac{\partial\mathsf{L}}{\partial\boldsymbol{k}}\right)_0\nabla-\left(\frac{\partial\mathsf{L}}{\partial\omega}\right)_0\frac{\partial}{\partial t}\right)+\cdots\right\}\left(\boldsymbol{e}_0A_0(\boldsymbol{r},t)+\left[\boldsymbol{e}_0,\boldsymbol{A}_0'(\boldsymbol{r},t)\right]\right)\\&=\left\{\mathsf{L}_0\left[\boldsymbol{e}_0,\boldsymbol{A}_0'(\boldsymbol{r},t)\right]-\mathrm{i}\left(\left(\frac{\partial\mathsf{L}}{\partial\boldsymbol{k}}\right)_0\nabla-\left(\frac{\partial\mathsf{L}}{\partial\omega}\right)_0\frac{\partial}{\partial t}\right)\boldsymbol{e}_0A_0(\boldsymbol{r},t)+\cdots\right\},\end{aligned} \tag{8.16}$$

that has to be resolved to $\left[\boldsymbol{e}_0,\boldsymbol{A}_0'(\boldsymbol{r},t)\right]$. As approximation we use only the two explicitly written sum terms in the second line. First we find using the dispersion equation $\frac{c^2}{\omega_0^2}\boldsymbol{k}_0^2=\varepsilon_0$

$$\begin{aligned}\mathsf{L}_0\left[\boldsymbol{e}_0,\boldsymbol{A}_0'(\boldsymbol{r},t)\right]&=\left(\frac{c^2}{\omega_0^2}\left(\boldsymbol{k}_0\cdot\boldsymbol{k}_0-\boldsymbol{k}_0^2\mathsf{I}\right)+\varepsilon_0\mathsf{I}\right)\left[\boldsymbol{e}_0,\boldsymbol{A}_0'(\boldsymbol{r},t)\right]\\&=\frac{c^2}{\omega_0^2}\left[\boldsymbol{k}_0,\boldsymbol{e}_0,\boldsymbol{A}_0'(\boldsymbol{r},t)\right]\cdot\boldsymbol{k}_0=\varepsilon_0\frac{\left[\boldsymbol{k}_0,\boldsymbol{e}_0,\boldsymbol{A}_0'(\boldsymbol{r},t)\right]\cdot\boldsymbol{k}_0}{\boldsymbol{k}_0^2},\end{aligned} \tag{8.17}$$

that is proportional to the average wave vector $\boldsymbol{k}_0$. Furthermore follows for the operator part of the second sum term in (8.16) which acts onto $A_0(\boldsymbol{r},t)$ and here written with indices

$$\begin{aligned}&\left\{\left(\frac{\partial L_{ij}}{\partial k_k}\right)_0\nabla_k-\left(\frac{\partial L_{ij}}{\partial\omega}\right)_0\frac{\partial}{\partial t}\right\}e_{0,j}\\&=\left\{\frac{c^2}{\omega_0^2}\left(\delta_{ik}k_{0,j}+k_{0,i}\delta_{jk}-2k_{0,k}\delta_{ij}\right)\nabla_k+2\frac{c^3}{\omega_0^3}\left(k_{0,i}k_{0,j}-\boldsymbol{k}_0^2\delta_{ij}\right)\frac{\partial}{\partial t}\right\}e_{0,j}\\&=\frac{c^2}{\omega_0^2}\left\{k_{0,i}\boldsymbol{e}_0\nabla-2e_{0,i}\left(\boldsymbol{k}_0\nabla+\frac{c}{\omega_0}\frac{\partial}{\partial t}\right)\right\}.\end{aligned} \tag{8.18}$$

This possesses two sum terms proportional to the vector $\boldsymbol{k}_0$ and to the main polarization $\boldsymbol{e}_0$ and shows a typical difficulty consisting in the correct neglect of terms deriving equation for additional components with no contradictions. The second sum term is proportional to the polarization $\boldsymbol{e}_0$ of the main component and has to be neglected. If we do so we find from (8.17) and (8.18) the following formula for the additional component in direction of $\boldsymbol{k}_0$

$$\left[\boldsymbol{k}_0,\boldsymbol{e}_0,\boldsymbol{A}_0'(\boldsymbol{r},t)\right]=\mathrm{i}\left(\boldsymbol{e}_0\nabla\right)A_0(\boldsymbol{r},t). \tag{8.19}$$

In special case of vacuum we have $\boldsymbol{v}_0 = c\frac{\boldsymbol{k}_0}{|\boldsymbol{k}_0|}$ and $\mathsf{W}_0 = \frac{c}{|\boldsymbol{k}_0|}\left(\mathsf{I} - \frac{\boldsymbol{k}_0 \cdot \boldsymbol{k}_0}{|\boldsymbol{k}_0|^2}\right)$ and Equation (8.15) for the main component becomes

$$0 = \left\{ \mathrm{i}\left(\frac{\partial}{\partial t} + c\frac{\boldsymbol{k}_0}{|\boldsymbol{k}_0|}\nabla \right) + \frac{c}{2|\boldsymbol{k}_0|}\nabla\left(\mathsf{I} - \frac{\boldsymbol{k}_0 \cdot \boldsymbol{k}_0}{|\boldsymbol{k}_0|^2} \right)\nabla \right\} \boldsymbol{e}_0 A_0(\boldsymbol{r}, t). \tag{8.20}$$

In Section 9 we derive a more complicated case for media with the polariton permittivity.

Thus the approximate equations for beam solutions taking into account diffraction in first order consists of the Equation (8.15) for the main part of the slowly varying amplitude plus the Equation (8.19) for a "small" additional part in direction of $\boldsymbol{k}_0$ which can be determined alone from the main part by differentiations. We wanted to show how the group velocity and the diffraction coefficients are involved in approximate beam equations but a detailed consideration of these equations and of the solution of (8.15) requires much place and is here not intended.

9. Permittivity to Polariton Dispersion in Isotropic Media

We consider in this Section the following special permittivity $\varepsilon(\omega)$ and permeability $\mu(\omega)$ of an isotropic medium with two real parameters ω_l and ω_t (or λ and ω_t) called polariton permittivity, e.g., [6] (§17, 18)

$$\begin{aligned} \varepsilon(\omega) &= 1 - \frac{\lambda}{\omega^2 - \omega_t^2} = 1 - \frac{\lambda}{2\omega_t}\left(\frac{1}{\omega - \omega_t} - \frac{1}{\omega + \omega_t} \right) = 1 - \frac{\omega_l^2 - \omega_t^2}{\omega^2 - \omega_t^2} \\ &= \frac{\omega^2 - \omega_l^2}{\omega^2 - \omega_t^2}, \quad (\mu(\omega) = 1), \quad \lambda \equiv \omega_l^2 - \omega_t^2, \end{aligned} \tag{9.1}$$

where for $\omega \approx (\geq)\omega_t$ the second sum term in round brackets can be neglected. In **Figure 1** this permittivity is illustrated for the two principal cases with different properties which we call the passive case $\omega_l \geq \omega_t$ and the active case $\omega_t > \omega_l$ (occupation inversion) and which are also characterized by (for $\omega > 0$)

$$\begin{aligned} &\lambda > 0, \quad \omega_l > \omega_t: \quad \frac{\partial \varepsilon}{\partial \omega}(\omega) > 0, \quad (\text{passive case}), \\ &\lambda < 0, \quad \omega_l < \omega_t: \quad \frac{\partial \varepsilon}{\partial \omega}(\omega) < 0, \quad (\text{active case}). \end{aligned} \tag{9.2}$$

The indices "l" and "t" in (9.1) mean "longitudinal" and "transversal". Polaritons (or real excitons) are a mixing of excitons in a medium and of photons in the vacuum (e.g., [1] [2] [17]) and correspond to the possible real light excitations in a medium. In this simple model ω_t is the frequency to a transition between two energy levels in the medium or a lattice oscillation and $\lambda = \omega_l^2 - \omega_t^2$ is beside other parameters proportional to the difference of the occupation of the two involved levels. For models of the medium in thermal equilibrium with temperature $T \neq 0$ the permittivity $\varepsilon(\omega)$ has to be generalized (see general form of the permittivity of an isotropic medium, e.g., in [4] chap XII, [3]). With

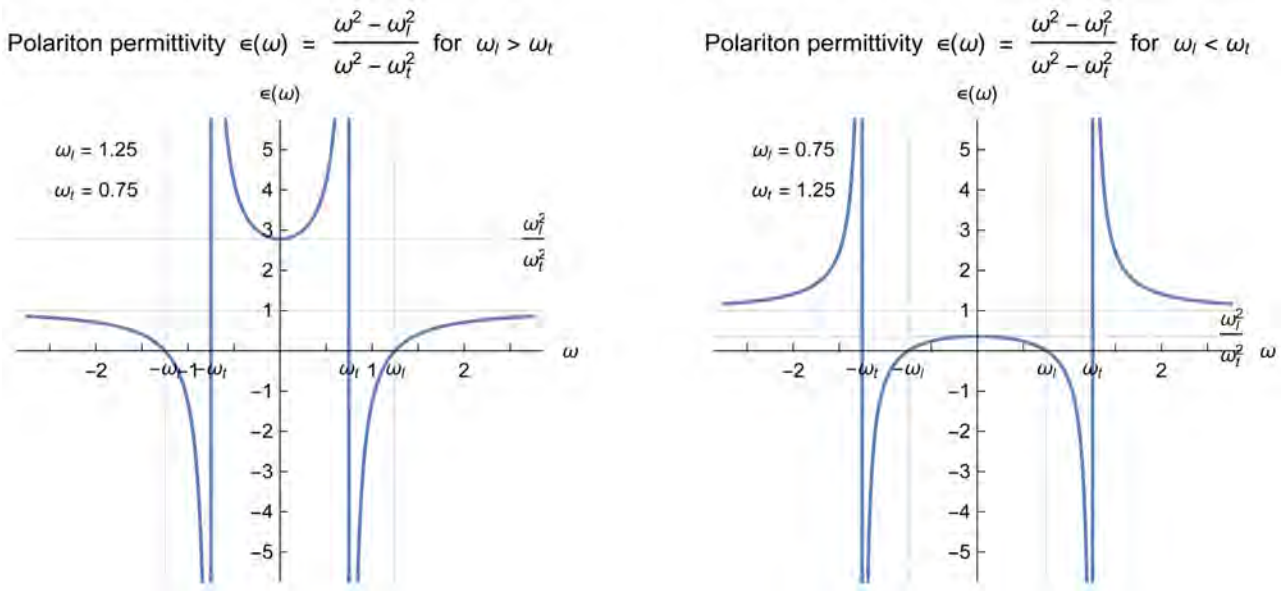


Figure 1. Polariton permittivity $\varepsilon(\omega)$ for $\omega_l > \omega_t$ (passive case) and for $\omega_l < \omega_t$ (active case). Apart from the jump from plus infinity to minus infinity (or vice versa) all derivatives of $\varepsilon(\omega)$ with respect to frequency are positive $\frac{\partial\varepsilon}{\partial\omega}(\omega) \geq 0$ in the passive case and negative in the active case $\frac{\partial\varepsilon}{\partial\omega}(\omega) \leq 0$ (for $\omega > 0$ right half-plane.

respect to the propagation of light beams in such a medium it is equivalent to a medium which possesses the right-hand form for the product $\varepsilon(\omega)\mu(\omega)$ instead for $\varepsilon(\omega)$ alone and can be included into the last case.

One may think that the distinction in passive and active case in (9.2) corresponds in certain way to the usual distinction in normal and anomalous dispersion but beside analogies there are also essential differences. Normal dispersion is usually discussed for the passive case alone and appears for real frequencies if one adds in the denominators for $\varepsilon(\omega)$ in (9.1) an imaginary part to take into account losses in the medium and if we consider then the real part of arising permittivity and is present in (small) parts between the (main) regions of normal dispersion. In the model (9.1) these regions are reduced to the points $\omega^2 = \omega_t^2$ and normal dispersion is present in the whole region with exclusion of these points. In contrast, in the active case we have in the whole region "anomalous" dispersion also with exclusion of the points $\omega^2 = \omega_t^2$ only and imaginary parts in the denominator do not play a role. This picture can change in some way for thermal equilibrium but the distinction in (9.2) is meant without losses. The occupation inversion in active case corresponds in some sense to a negative absolute temperature (notion occasionally used in second half of last century) but a thermal equilibrium in this case is only possible for a finite number of energy levels.

For $\omega_l \equiv \omega_p$ and $\omega_t = 0$ we have the permittivity of a cold isotropic plasma

$$\omega_l \equiv \omega_p, \quad \omega_t = 0: \quad \Rightarrow \quad \varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2}, \quad \omega_p \equiv \sqrt{\frac{4\pi n_e e^2}{m_e}}, \tag{9.3}$$

with ω_p the plasma frequency given here for an electron plasma (indices "e"; e, n_e, m_e mean electron charge, electron density, electron mass and electron charge). This is again idealized for temperature $T = 0$.

Longitudinal waves are in the idealized form (9.1) of the permittivity only possible for the frequency ω_l

$$\omega = \omega_l, \tag{9.4}$$

and are pure oscillations with no dispersion. Our main interest concern transversal waves which we now consider. The dispersion relation for transversal waves (7.4) specialized for the polariton case (9.1) becomes

$$\boldsymbol{k}^2 = \frac{\omega^2}{c^2}\varepsilon(\omega)\mu(\omega) = \frac{\omega^2}{c^2}\frac{\omega^2 - \omega_l^2}{\omega^2 - \omega_t^2}, \tag{9.5}$$

and depends only on the product $\varepsilon(\omega)\mu(\omega)$. For such waves the relation between the energy flow density $\boldsymbol{S}$ and the energy density w in the lossless case for quasi-plane and quasi-monochromatic waves and in the transition to the limiting case of plane monochromatic waves with real wave vector $\boldsymbol{k}$ and real frequency ω (homogeneous waves)

$$\boldsymbol{S} = \boldsymbol{v}w, \tag{9.6}$$

remains in every case the same and depends only from the dispersion relation ($\boldsymbol{v}$ is group velocity; see next Section). However, the splitting of the energy density w in (9.6) into a part from the electric field and into a part from the magnetic field depends on $\varepsilon(\omega)$ and $\mu(\omega)$ separately and therefore also the calculation of the corresponding energy flow density $\boldsymbol{S}$ which can be made from the energy density w by (9.6). As illustrated in Figure 1 it is interesting to extend the permittivity (9.1) to $\lambda = \omega_l^2 - \omega_t^2 < 0$ which corresponds to a model medium with inverse occupation density of the two involved levels. The condition $\omega_l > \omega_t$ in the form of the permittivity (9.1) which belongs to the passive case is satisfied for taking into account only one transition with frequency ω_t between two energy levels and for sufficiently low temperatures. It may be converted into $\omega_l < \omega_t$ for pumping to a higher energy level of a laser medium to get inversion of occupation but to keep their difference $\omega_t^2 - \omega_l^2$ constant can be only a very rough approximation for the laser action near the threshold. Such a permittivity falls under the active case (not to be confused with notion (natural) optical activity!). By far, not all consequences for the active case are clear and are well understood.

Let us begin with a general consideration to dispersion equations in a homogeneous isotropic and infinitely extended medium. The general case is that both wave vector $\boldsymbol{k}$ and frequency ω are complex quantities. For the vacuum with the dispersion equation $c^2\boldsymbol{k}^2 = \omega^2$ with the splitting of wave vector and frequency in real and imaginary parts

$$\boldsymbol{k} = \boldsymbol{k}' + \mathrm{i}\boldsymbol{k}'', \quad \omega = \omega' + \mathrm{i}\omega'', \tag{9.7}$$

this leads to a complex equation with the following separation into a real and

imaginary part

$$\begin{aligned} &c^2\left(\boldsymbol{k}'+\mathrm{i}\boldsymbol{k}''\right)^2=\left(\omega'+\mathrm{i}\omega''\right)^2, \quad \Rightarrow \\ &c^2\left(\boldsymbol{k}'^2-\boldsymbol{k}''^2\right)=\omega'^2-\omega''^2, \quad c^2\left(\boldsymbol{k}'\boldsymbol{k}''\right)=\omega'\omega''. \end{aligned} \tag{9.8}$$

These are 2 scalar equations for 5 real variables (for example, $\left|\boldsymbol{k}'\right|,\left|\boldsymbol{k}''\right|,\cos(\varphi)\equiv\dfrac{\boldsymbol{k}'\boldsymbol{k}''}{\left|\boldsymbol{k}'\right|\left|\boldsymbol{k}''\right|},\omega',\omega''$) that restricts the number of free variables to 3 real variables. It is impossible to represent this in a single graphical representation and one has to make compromises. For example, for real frequency ($\omega'=\omega,\omega''=0$) we find from (9.8) the orthogonality $\boldsymbol{k}'\boldsymbol{k}''=0$ of real to imaginary part of the wave vector $\boldsymbol{k}$. It is known that such waves are generated in the vacuum under total reflection within an isotropic medium with $\boldsymbol{k}'$ parallel to the boundary plane and with $\boldsymbol{k}''$ in direction of the normal vector to the boundary plane corresponding to exponential decrease. Such waves are called inhomogeneous waves (not to confuse with inhomogeneous media!). All this is well known and understood. Which of the components $\left(\boldsymbol{k}',\boldsymbol{k}'',\omega',\omega''\right)$ are involved into a process can be only determined if one knows the boundary together with the boundary conditions. If we apply this to the polariton permittivity (9.1) with the complex dispersion Equation (9.5) we find the following equation

$$c^2\left(\boldsymbol{k}'^2-\boldsymbol{k}''^2+\mathrm{i}2\boldsymbol{k}'\boldsymbol{k}''\right)=\left(\omega'^2-\omega''^2+\mathrm{i}2\omega'\omega''\right)\frac{\omega'^2-\omega''^2-\omega_l^2+\mathrm{i}2\omega'\omega''}{\omega'^2-\omega''^2-\omega_t^2+\mathrm{i}2\omega'\omega''}. \tag{9.9}$$

Separated into real and imaginary part this leads to the two equations

$$\begin{aligned} 0=&\left(\omega'^2-\omega''^2-\omega_t^2\right)\left(\omega'^2-\omega''^2-c^2\left(\boldsymbol{k}'^2-\boldsymbol{k}''^2\right)\right)\\ &-4\omega'\omega''\left(\omega'\omega''-c^2\left(\boldsymbol{k}'\boldsymbol{k}''\right)\right)-\left(\omega_l^2-\omega_t^2\right)\left(\omega'^2-\omega''^2\right),\\ 0=&\left(\omega'^2-\omega''^2-\omega_t^2\right)\left(\omega'\omega''-c^2\left(\boldsymbol{k}'\boldsymbol{k}''\right)\right)\\ &+\omega'\omega''\left(\omega'^2-\omega''^2-c^2\left(\boldsymbol{k}'^2-\boldsymbol{k}''^2\right)\right)-\left(\omega_l^2-\omega_t^2\right)\omega'\omega'', \end{aligned} \tag{9.10}$$

which are of forth degree with respect to the real components of wave vector and frequency. Both equations have to be satisfied at the same time. This means that we have different possibilities of two-dimensional graphical representations for inhomogeneous waves and this is relatively complicated in such generality.

If we choose real wave vectors as free variable then the dispersion Equation (9.5) leads to a bi-quadratic equation for the frequency $\omega=\omega\left(\boldsymbol{k}\right)$ in dependence on the modulus $\left|\boldsymbol{k}\right|$ of the wave vector as follows ($\boldsymbol{k}^2\equiv\left|\boldsymbol{k}\right|^2$)

$$0=\omega^4-\left(\omega_l^2+c^2\boldsymbol{k}^2\right)\omega^2+\omega_t^2c^2\boldsymbol{k}^2, \tag{9.11}$$

which resolved provides two branches of squared solutions

$$\begin{aligned} \omega_\pm^2\left(\boldsymbol{k}\right)&=\frac{1}{2}\left\{\omega_l^2+c^2\boldsymbol{k}^2\pm\sqrt{\left(\omega_l^2+c^2\boldsymbol{k}^2\right)^2-4\omega_t^2c^2\boldsymbol{k}^2}\right\}\\ &=\frac{1}{2}\left\{\omega_l^2+c^2\boldsymbol{k}^2\pm\sqrt{\omega_l^4+2\left(\omega_l^2-2\omega_t^2\right)c^2\boldsymbol{k}^2+\left(c^2\boldsymbol{k}^2\right)^2}\right\}, \end{aligned} \tag{9.12}$$

or of frequency solutions[7]

$$\begin{aligned}\omega_{\pm}^{(\pm)}(\boldsymbol{k}) &= (\pm)\frac{1}{2}\left\{\sqrt{\omega_l^2 + 2\omega_t c|\boldsymbol{k}| + c^2|\boldsymbol{k}|^2} \pm \sqrt{\omega_l^2 - 2\omega_t c|\boldsymbol{k}| + c^2|\boldsymbol{k}|^2}\right\} \\ &= (\pm)\frac{1}{2}\left\{\sqrt{(\omega_t + c|\boldsymbol{k}|)^2 + \omega_l^2 - \omega_t^2} \pm \sqrt{(\omega_t - c|\boldsymbol{k}|)^2 + \omega_l^2 - \omega_t^2}\right\}.\end{aligned} \tag{9.13}$$

This means that we have to given $|\boldsymbol{k}|$ two different solutions for $\omega_{(\pm)}^2$ signified by the upper indices "(±)" where one has to pay attention mainly to the two different lower signs "±" to the two sum terms with square roots. In the hatched region the solution (9.13) for the frequency in dependence on the modulus of the wave vector becomes complex and can be better represented in the form

$$\omega_{\pm}^{(\pm)}(\boldsymbol{k}) = (\pm)\frac{1}{2}\left\{\sqrt{(\omega_t + c|\boldsymbol{k}|)^2 - (\omega_t^2 - \omega_l^2)} \pm i\sqrt{\omega_t^2 - \omega_l^2 - (\omega_t - c|\boldsymbol{k}|)^2}\right\}. \tag{9.14}$$

Figure 2 represents the different branches of solutions $\omega = \omega(\boldsymbol{k})$ in dependence on the modulus $|\boldsymbol{k}|$ of the wave vector for the both principal cases $\omega_l > \omega_t$ and $\omega_l < \omega_t$. The left-hand picture is well known mostly in form of the right upper quadrant (see, e.g., [17], chap. III, Figure 8 or [2], chap. 11, Figure 11.4).

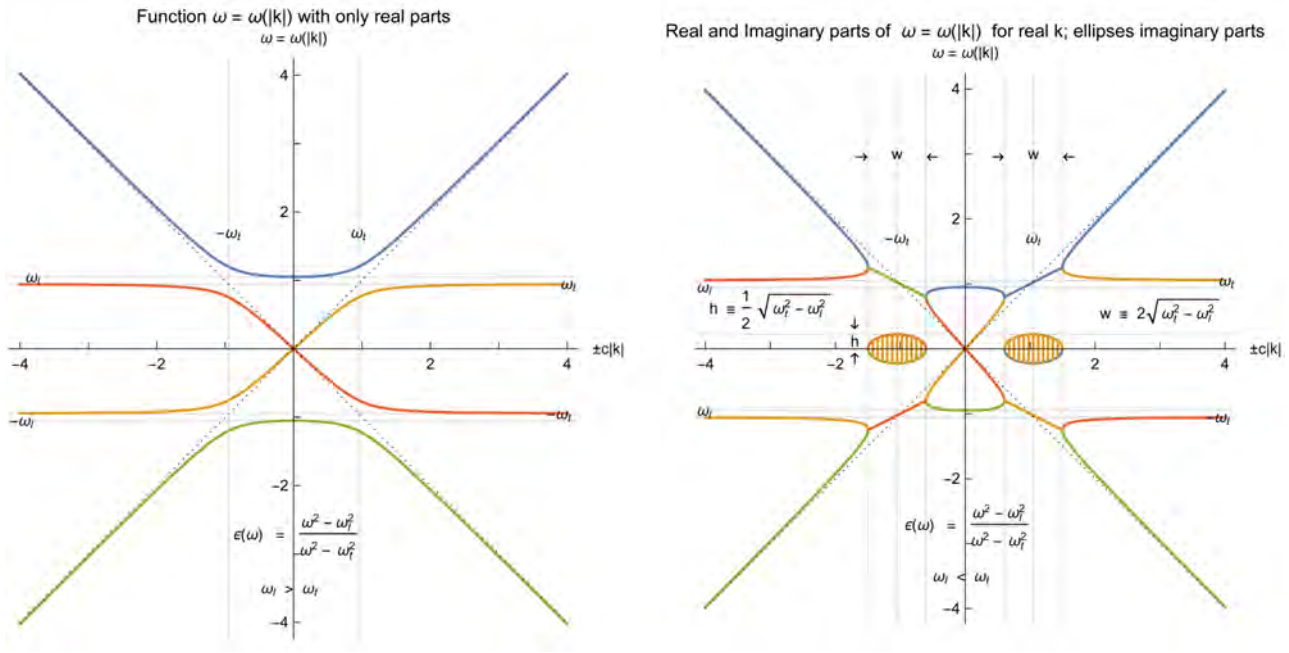


Figure 2. Frequencies in dependence on real wave vector for $\varepsilon(\omega) = \dfrac{\omega^2 - \omega_l^2}{\omega^2 - \omega_t^2}$, $\mu(\omega) = 1$ for $\omega_l > \omega_t$ and $\omega_l < \omega_t$. The contours of the hatched parts in the right-hand picture are imaginary and describe amplification. For the figures we have chosen the values $\omega_l = 1.05, \omega_t = 0.95$ in passive case (to the left) and $\omega_l = 0.95, \omega_t = 1.05$ in active case (to the right).

[7]A bi-quadratic equation of the form $z^4 - 2pz^2 + q^2 = 0$, $(p = p^*, q = q^*, \Rightarrow q^2 \geq 0)$, with real p and q and therefore non-negative q^2 possesses 4 solutions which may be written in the following form $z_{\pm}^{(\pm)} = (\pm)\dfrac{1}{\sqrt{2}}\left(\sqrt{p+q} \pm \sqrt{p-q}\right) = (\pm)\dfrac{\sqrt{2}\,q}{\sqrt{p+q} \mp \sqrt{p-q}}$, with only real or imaginary sum terms $\sqrt{p \pm q}$. Clearly, these solutions are also true in general case but usually then do not provide separation into real and imaginary parts.

Asymptotically, for large $|\boldsymbol{k}|$ we have in both cases a branch $\omega = \pm c|\boldsymbol{k}|$ the same as for light beams in vacuum and a branch with constant $\omega = \pm\omega_t$ where ω_t is the resonance frequency to the energy difference of the considered two levels. The case $\omega_l - \omega_t > 0$ corresponds to lower occupation of higher level in comparison the lower level and the case $\omega_t - \omega_l > 0$ to inverse occupation. The last case can be achieved by pumping this level and as was to expect it possesses properties of amplification in a certain region of wave vectors. This is exactly separated by the two sum terms with square roots in the solutions in the form (9.13) or (9.14). We have hatched the imaginary parts inside their contours. Their real parts are in the figure to find over, respectively, under the hatched contours and look like straight lines but are such only in the limiting case $\omega_t - \omega_l \to 0$ as follows from the first sum term in (9.14).

In Figure 3 the case $\omega_t > \omega_l$ is presented enlarged for the right upper quadrant of Figure 2 in two numerical cases which show the dependence on the parameters ω_t and ω_l and, in particular, on the difference $\sqrt{\omega_t^2 - \omega_l^2}$. Microscopic models for the permittivity (9.1) show that the difference $\omega_t^2 - \omega_l^2$ is among other parameters as factors of the active transition levels proportional to the density occupation inversion $\sigma \propto \omega_t^2 - \omega_l^2 \equiv -\lambda$ of these levels. This means that in our idealized model the product of height h with width w of the amplifier contour (see Figure 2) is proportional to the square root of the density of inverse occupation of the considered active levels whereas h and w themselves are proportional to its square root. In laser theory to our knowledge the height h is

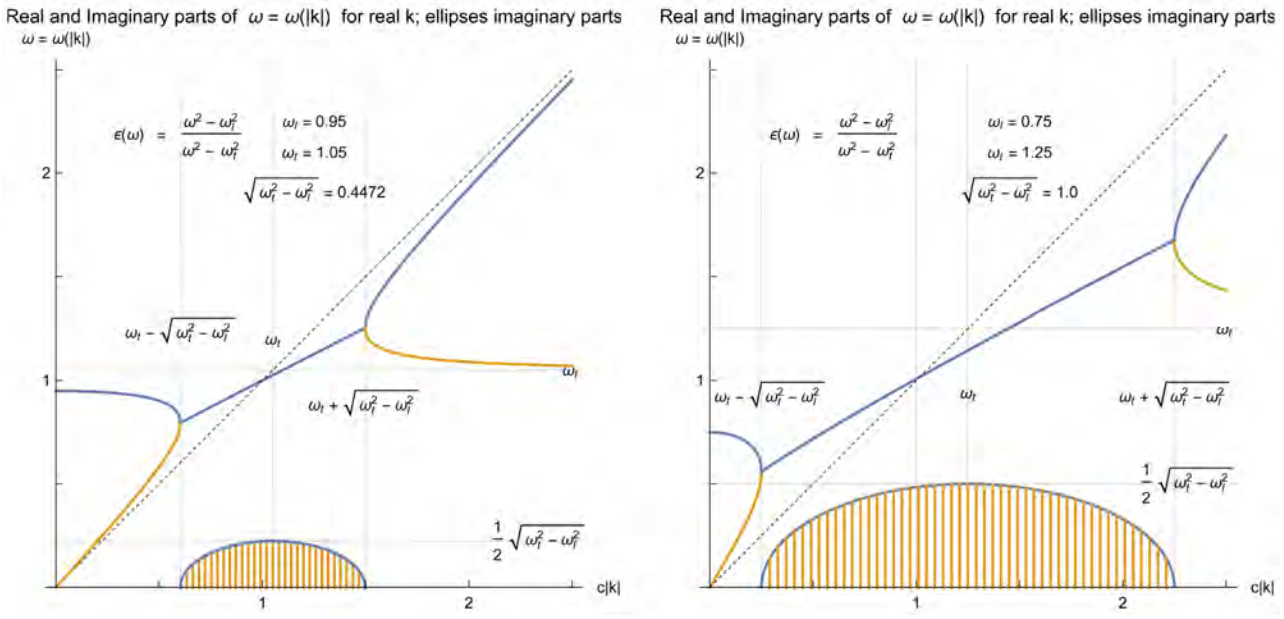


Figure 3. Function $\omega = \omega(\boldsymbol{k})$ for $\varepsilon(\omega) = \dfrac{\omega^2 - \omega_l^2}{\omega^2 - \omega_t^2}$, $\mu(\omega) = 1$ with $\omega_l < \omega_t$ (amplification) in two numerical cases. The left-hand figure is practically the right upper quadrant of Figure 2 but amplified for better visibility and for making a comparison with a similar figure with other parameters. For the figures we have chosen the values $\omega_t = 1.05, \omega_l = 0.95$ and $\omega_t = 1.25, \omega_l = 0.75$.

usually assumed or calculated as direct proportional to the density of inverse occupation whereas for its width are made complicated considerations about natural line width and its enlargement.

For application to laser theory it is necessary to add feedback by a resonator. The unspecific losses of the involved resonator modes in the concerning region cut off an upper part of the amplification contour and thus lower it whereas their frequencies are determined mainly from the real part (first sum term in (9.14))

$$\begin{aligned}&\omega_{\mathrm{Re}}\left(|k|\right)=\frac{1}{2}\sqrt{\left(\omega_t+c|\boldsymbol{k}|\right)^2-\left(\omega_t^2-\omega_l^2\right)}\\&=\frac{\sqrt{3\omega_t^2+\omega_l^2}}{2}+\frac{\omega_t}{\sqrt{3\omega_t^2+\omega_l^2}}\left(c|\boldsymbol{k}|-\omega_t\right)\left(1-\frac{\omega_t^2-\omega_l^2}{4\left(3\omega_t^2+\omega_l^2\right)}\left(c|\boldsymbol{k}|-\omega_t\right)+\cdots\right),\\&\left(\omega_t-\sqrt{\omega_t^2-\omega_l^2}\le c|\boldsymbol{k}|\le\omega_t+\sqrt{\omega_t^2-\omega_l^2}\right).\end{aligned}\tag{9.15}$$

For a long resonator in comparison to the transverse dimensions only the longitudinal modes without reflection at the side wands play a role. For such a resonator with ideal mirrors at the end the field at the mirrors has to be vanishing and the possible resonator modes have to possess a multiple m of the half the wavelengths λ, ($|\boldsymbol{k}|\equiv\frac{2\pi}{\lambda}$), which fit into the resonator length L. Thus for the possible wave vectors and frequencies we find in this idealized case ($m_0\equiv m_{\min}$)

$$\begin{aligned}&|\boldsymbol{k}|=m\frac{\pi}{L},\quad \omega_{\mathrm{Re}}\left(|k|\right)=\frac{1}{2}\sqrt{\left(\omega_t+m\pi\frac{c}{L}\right)^2-\left(\omega_t^2-\omega_l^2\right)},\\&\omega_t-\sqrt{\omega_t^2-\omega_l^2}\le m\pi\frac{c}{L}\le\omega_t+\sqrt{\omega_t^2-\omega_l^2},\quad \left(m=m_0,m_0+1,\cdots,m_{\max}\right).\end{aligned}\tag{9.16}$$

The resonator losses as said lower the amplification contour and therefore narrow the possible values for m and change also a little the relation (9.15) for the possible wave vectors and frequencies. With losses both the wave vectors and frequencies may become slightly complex. The imaginary part of the frequency determines also a line width by only classical considerations. For $\frac{\omega_t^2-\omega_l^2}{\omega_t^2}\le 1$ one finds in approximation from (9.15)

$$\frac{\omega_t^2-\omega_l^2}{\omega_t^2}\le 1:\quad \omega_{\mathrm{Re}}\left(\boldsymbol{k}\right)\approx\omega_t+\frac{1}{2}\left(c|\boldsymbol{k}|-\omega_t\right),\tag{9.17}$$

and the density of possible frequencies in the corresponding frequency interval is doubled in comparison to the density within a resonator with vacuum. If we assume that there is a process (pumping) which keeps constant the density of the inverse occupation then it may be considered as a very simple and idealized classical model of laser action, at least, near the threshold. However, this model cannot provide information at which level of occupation inversion the equilibrium between pumping and radiation is reached. In this sense it is similar to thermal equilibrium where without additional information it cannot be said how it was reached. Clearly, quantum-mechanical generalization makes further mod-

ifications also to the line widths.[8]

Generalizations of the permittivity (9.1) in different directions are possible and interesting, for example, by an additional constant sum term on the right-hand side taking into account summarily the contribution of all other resonances of two levels or taking into account losses by imaginary terms in the denominator. We consider now shortly the case of taking into account two resonance frequencies $\omega_{t,1}$ and $\omega_{t,2}$ leading to a permittivity of the form [6] (§18, Equation (6))

$$\varepsilon(\omega)=1-\frac{\lambda_1}{\omega^2-\omega_{t,1}^2}-\frac{\lambda_2}{\omega^2-\omega_{t,2}^2}, \tag{9.18}$$

with two further parameters λ_2 and $\omega_{t,2}$ which determine the strength of a second resonance. This can be also written in the following form

$$\varepsilon(\omega)=\frac{\left(\omega^2-\omega_{l,-}^2\right)\left(\omega^2-\omega_{l,+}^2\right)}{\left(\omega^2-\omega_{t,1}^2\right)\left(\omega^2-\omega_{t,2}^2\right)}, \tag{9.19}$$

with the definitions

$$\omega_{l,\pm}^2\equiv\frac{1}{2}\left\{\omega_{t,1}^2+\lambda_1+\omega_{t,2}^2+\lambda_2\pm\sqrt{\left(\omega_{t,1}^2+\lambda_1-\omega_{t,2}^2-\lambda_2\right)^2+4\lambda_1\lambda_2}\right\}, \tag{9.20}$$

where $\omega_{l,\mp}^2$ is real-valued or may become even complex-to complex-conjugate-valued and where $\omega_{l,-}$ and $\omega_{l,+}$ cannot be properly assigned to $\omega_{t,1}$ and $\omega_{t,2}$. The dispersion Equation (7.4) resolved to an equation for ω^2 in dependence on the squared wave vector $|\boldsymbol{k}|^2$ becomes a bi-cubic equation in $|\boldsymbol{k}|$ which is already difficult to solve for ω and to discuss. In dependence on the 4 parameters $\omega_{t,1},\omega_{t,2}$ and λ_1 and λ_2 one would have to distinguish many principal cases.

10. Group Velocity to the Polariton Permittivity in Passive Case and Group Velocities Faster Than Light Velocity in Active Case

We now consider the group velocity of transversal waves for the polariton permittivity (9.1). By differentiation of the dispersion Equation (9.5) with respect to $\boldsymbol{k}$ we calculate for the group velocity in direction $\frac{\boldsymbol{k}}{|\boldsymbol{k}|}$ of the wave vector and in dependence of its modulus on the frequency ω only

[8]Sometime in the eighties I asked H. Paul from our Institute to discuss with me the polariton model and he agreed. My first aim was to see did he know whether or not this model or something similar was already discussed in literature since I did not find neither theoretical considerations nor experimental hints for some results. He also did not know something and likely nothing existed. Thank! My second aim was to show him that the height as well as the width of the amplification contours in form of ellipses are proportional to the square root of the inverse occupation density and I made a drawing of this contour but nothing of the kind of the more general and complicated figures made by computer and presented here. I could not expect that we may discuss the formulae in detail (more than the permittivity he did not want to see) and soon H. Paul finished the discussion saying approximately: I believe you only if you have also the nonlinear terms in the equations. This, however, was not my intention, my possibility and my official task at that time.

$$\boldsymbol{v} \equiv \frac{\partial \omega}{\partial \boldsymbol{k}} = \frac{2c^2\boldsymbol{k}}{\frac{\partial}{\partial \omega}\left(\omega^2 \varepsilon(\omega)\right)} = \frac{c^2}{\omega}\frac{\left(\omega^2-\omega_t^2\right)^2}{\omega^4 - 2\omega_t^2\omega^2 + \omega_l^2\omega_t^2}\boldsymbol{k}$$
$$= c\underbrace{\sqrt{\frac{\omega^2-\omega_l^2}{\omega^2-\omega_t^2}}}_{\frac{c}{\omega}|\boldsymbol{k}|=\sqrt{\varepsilon(\omega)}}\frac{\left(\omega^2-\omega_t^2\right)^2}{\left(\omega^2-\omega_t^2\right)^2+\left(\omega_l^2-\omega_t^2\right)\omega_t^2}\frac{\boldsymbol{k}}{|\boldsymbol{k}|} \equiv \varphi(\omega)\frac{\boldsymbol{k}}{|\boldsymbol{k}|}. \tag{10.1}$$

It possesses in this model in every case the direction of the wave vector also if it is complex-valued with different directions of real and imaginary part but the frequency-dependent coefficients $\varphi(\omega)$ can also become complex-valued (for real values of ω) due to presence of the square root. Without taking into account the dispersion of $\varepsilon(\omega)$ at the considered frequency ω (setting $\frac{\partial \varepsilon(\omega)}{\partial \omega} \to 0$) it is

$$\boldsymbol{v}' = \frac{2c^2\boldsymbol{k}}{\varepsilon(\omega)\frac{\partial}{\partial \omega}\left(\omega^2\right)} = c\frac{c\boldsymbol{k}}{\omega\varepsilon(\omega)} = c\frac{c|\boldsymbol{k}|}{\omega\varepsilon(\omega)}\frac{\boldsymbol{k}}{|\boldsymbol{k}|} = \frac{c}{\sqrt{\varepsilon(\omega)}}\frac{\boldsymbol{k}}{|\boldsymbol{k}|} = c\sqrt{\frac{\omega^2-\omega_t^2}{\omega^2-\omega_l^2}}\frac{\boldsymbol{k}}{|\boldsymbol{k}|}, \tag{10.2}$$

but the dispersion cannot be switched off. For such points where $\varepsilon(\omega)$ possesses a minimum (or maximum) the derivative vanishes (*i.e.*, $\frac{\partial \varepsilon(\omega)}{\partial \omega}(\omega) = 0$) and the group velocity with and without taking into account the dispersion are equal. The dispersion factor $\alpha_{\text{disp}}(\omega) \equiv \alpha(\omega)$ is

$$\alpha(\omega) = \frac{\varepsilon(\omega)\frac{\partial}{\partial \omega}\left(\omega^2\right)}{\frac{\partial}{\partial \omega}\left(\omega^2\varepsilon(\omega)\right)} = \frac{\left(\omega^2-\omega_l^2\right)\left(\omega^2-\omega_t^2\right)}{\left(\omega^2-\omega_t^2\right)^2+\left(\omega_l^2-\omega_t^2\right)\omega_t^2} \equiv \frac{\left(\omega^2-\omega_l^2\right)\left(\omega^2-\omega_t^2\right)}{\left(\omega^2-\omega_-^2\right)\left(\omega^2-\omega_+^2\right)},$$
$$\omega_\mp^2 \equiv \omega_t\left(\omega_t \mp \sqrt{\omega_t^2-\omega_l^2}\right), \quad \omega_\mp \equiv \frac{\sqrt{\omega_t\left(\omega_t+\omega_l\right)} \mp \sqrt{\omega_t\left(\omega_t-\omega_l\right)}}{\sqrt{2}}. \tag{10.3}$$

Due to an extremum of $\varepsilon(\omega)$ for $\omega = 0$ its derivative with respect to frequency vanishes there and the dispersion factor becomes $\alpha(\omega = 0) = 1$.

One may distinguish 2 different cases and a limiting case between them represented by

$$|\boldsymbol{v}| = \varphi(\omega) = c\frac{\sqrt{1-\frac{\omega_l^2-\omega_t^2}{\omega^2-\omega_t^2}}}{1+\frac{\omega_l^2-\omega_t^2}{\omega^2-\omega_t^2}\frac{\omega_t^2}{\omega^2-\omega_t^2}} \to \begin{cases} < c, & \left(\omega_l > \omega_t\right) \\ = c, & \left(\omega_l = \omega_t\right), \\ > c, & \left(\omega_l < \omega_t\right) \end{cases} \tag{10.4}$$

which in relation to the light velocity obviously are determined by different properties. In last case of $\omega_l < \omega_t$ the group velocity can even be opposite to the direction of the wave vector in certain regions of the frequency.

If one wants to have the full dependence $\boldsymbol{v} \equiv \boldsymbol{v}(\boldsymbol{k})$ of the group velocity on the wave vector one has to distinguish the 4 branches in (9.13) or in (9.14) and finds by differentiation with respect to $\boldsymbol{k}$

$$\begin{aligned}\boldsymbol{v}_{\pm}^{(\pm)}(\boldsymbol{k}) &= (\pm)\frac{c}{2}\left\{\frac{\omega_t + c|\boldsymbol{k}|}{\sqrt{\left(\omega_t + c|\boldsymbol{k}|\right)^2 + \omega_l^2 - \omega_t^2}} \mp \frac{\omega_t - c|\boldsymbol{k}|}{\sqrt{\left(\omega_t - c|\boldsymbol{k}|\right)^2 + \omega_l^2 - \omega_t^2}}\right\}\frac{\boldsymbol{k}}{|\boldsymbol{k}|} \\ &= (\pm)\frac{c}{2}\left\{\frac{\omega_t + c|\boldsymbol{k}|}{\sqrt{\left(\omega_t + c|\boldsymbol{k}|\right)^2 - \left(\omega_t^2 - \omega_l^2\right)}} \pm i\frac{\omega_t - c|\boldsymbol{k}|}{\sqrt{\left(\omega_t^2 - \omega_l^2\right) - \left(\omega_t - c|\boldsymbol{k}|\right)^2}}\right\}\frac{\boldsymbol{k}}{|\boldsymbol{k}|} \\ &\equiv c\beta(\omega)\frac{\boldsymbol{k}}{|\boldsymbol{k}|}.\end{aligned} \tag{10.5}$$

which we have written in two favorable representations for the passive case $\omega_l > \omega_t$ and the active case $\omega_t > \omega_l$.

The calculation of the second-order coefficients W_{ij} in the Equation (8.15) for beam propagation in isotropic media can be calculated, for example, from the principal structure of the group velocity

$$v_i \equiv \frac{\partial\omega}{\partial k_i} = \varphi(\omega)\frac{k_i}{|\boldsymbol{k}|}, \quad \varphi(\omega) \equiv c\frac{\left(\omega^2 - \omega_l^2\right)^{\frac{1}{2}}\left(\omega^2 - \omega_t^2\right)^{\frac{3}{2}}}{\left(\omega^2 - \omega_t^2\right)^2 + \left(\omega_l^2 - \omega_t^2\right)\omega_t^2}, \tag{10.6}$$

by further differentiation with respect to variables k_j. Using

$$\frac{\partial|\boldsymbol{k}|}{\partial k_i} = \frac{k_i}{|\boldsymbol{k}|}, \quad \Rightarrow \quad \frac{\partial^2|\boldsymbol{k}|}{\partial k_i \partial k_j} = \frac{\partial}{\partial k_j}\left(\frac{k_i}{|\boldsymbol{k}|}\right) = \frac{1}{|\boldsymbol{k}|}\left(\delta_{ij} - \frac{k_i k_j}{|\boldsymbol{k}|^2}\right), \tag{10.7}$$

and due to $\omega \equiv \omega(\boldsymbol{k})$ and (10.7)

$$\frac{\partial\varphi}{\partial k_j}(\omega) = \frac{\partial\varphi}{\partial\omega}(\omega)\frac{\partial\omega}{\partial k_j} = \frac{\partial\varphi}{\partial\omega}(\omega)\varphi(\omega)\frac{k_j}{|\boldsymbol{k}|} = \frac{1}{2}\frac{\partial}{\partial\omega}\left(\left(\varphi(\omega)\right)^2\right)\frac{k_j}{|\boldsymbol{k}|}, \tag{10.8}$$

we find the general structure

$$W_{ij} \equiv \frac{\partial^2\omega}{\partial k_i \partial k_j} = \varphi(\omega)\frac{1}{|k|}\left(\delta_{ij} - \frac{k_i k_j}{|\boldsymbol{k}|^2}\right) + \frac{1}{2}\frac{\partial}{\partial\omega}\left(\left(\varphi(\omega)\right)^2\right)\frac{k_i k_j}{|\boldsymbol{k}|^2} \tag{10.9}$$

and, finally, with the special function $\varphi(\omega)$ in (10.6)

$$\begin{aligned}W_{ij} \equiv \frac{\partial^2\omega}{\partial k_i \partial k_j} &= \frac{c^2}{\omega}\frac{\left(\omega^2 - \omega_t^2\right)^2}{\left(\omega^2 - \omega_t^2\right)^2 + \left(\omega_l^2 - \omega_t^2\right)\omega_t^2} \\ &\cdot\left\{\delta_{ij} - \frac{k_i k_j}{|\boldsymbol{k}|^2} + \left(\omega_l^2 - \omega_t^2\right)\frac{\omega^2\left(\omega^4 + 2\omega_t^2\omega^2 - 3\omega_l^2\omega_t^2\right)}{\left(\left(\omega^2 - \omega_t^2\right)^2 + \left(\omega_l^2 - \omega_t^2\right)\omega_t^2\right)^2}\frac{k_i k_j}{|\boldsymbol{k}|^2}\right\}.\end{aligned} \tag{10.10}$$

This result was checked by a slightly modified calculation which I do not present here. It possesses two sum terms proportional to the tensors $\delta_{ij} - \frac{k_i k_j}{|\boldsymbol{k}|^2}$ and $\frac{k_i k_j}{|\boldsymbol{k}|^2}$ describing transversal and longitudinal diffraction (or diffusion) of the light beam during propagation.

In the special case of an isotropic cold plasma with the permittivity $\varepsilon(\omega)=1-\frac{\omega_p^2}{\omega^2}$ setting $\omega_l=\omega_p, \omega_t=0$ the formulae (10.1) for the group velocity and (10.10) for the diffraction coefficients simplify to

$$v_i = c\frac{c|\boldsymbol{k}|}{\omega}\frac{k_i}{|\boldsymbol{k}|} = c\sqrt{1-\frac{\omega_p^2}{\omega^2}}\frac{k_i}{|\boldsymbol{k}|}, \quad \Rightarrow \quad \boldsymbol{kv} = \omega\left(1-\frac{\omega_p^2}{\omega^2}\right), \tag{10.11}$$

that due to taking into account the frequency dispersion does not agree with (6.9). For the diffraction coefficients one finds

$$W_{ij} = \frac{c^2}{\omega}\left\{\delta_{ij} - \frac{k_ik_j}{|\boldsymbol{k}|^2} + \frac{\omega_p^2}{\omega^2}\frac{k_ik_j}{|\boldsymbol{k}|^2}\right\}, \tag{10.12}$$

and it contains also a term proportional to $\frac{k_ik_j}{|\boldsymbol{k}|^2}$ leading to a diffusion in longitudinal direction of the beam.

Without taking into account the dispersion we would obtain from (10.2)

$$W'_{ij} = \frac{c^2}{\sqrt{\varepsilon(\omega)}}\left(\delta_{ij} - \frac{k_ik_j}{|\boldsymbol{k}|^2}\right), \tag{10.13}$$

that is without a "longitudinal" contribution proportional to $\frac{k_ik_j}{|\boldsymbol{k}|^2}$.

In Figure 4 we represent the permittivity $\varepsilon(\omega)$ (blue curves) together with the group velocity in relation to the light velocity without taking into account dispersion of the permittivity $\alpha(\omega)\left(\equiv\alpha_{\text{disp}}(\omega)\right)$ (yellow curves) and with taking it into account as $\beta(\omega)\equiv\frac{|\boldsymbol{v}|}{c}$ (red curves) for the passive ($\omega_l>\omega_t$, to the left) and the active case ($\omega_t>\omega_l$ to the right). In the passive case we see that the group velocity remains smaller than the light velocity in every case and that it is real-valued in regions where $\varepsilon(\omega)$ is positive and that taking into account the dispersion the deviations in comparison to neglect of dispersion are important, in particular, in the neighborhood of the resonance frequency ω_t and in the neighborhood of the "(longitudinal)" frequency ω_l. In addition, the group velocity possesses in every case the same direction as the wave vector. In the region $\omega_t\le\omega\le\omega_l$ no relation of group velocity to light velocity (red) is drawn because the group velocity is there imaginary. All is so as expected. However, the same picture for the active case (to the right) contains an unexpected surprise.

In the active case $\omega_l\le\omega_t$ (right-hand picture in Figure 4) the group velocity is here in certain regions of frequency larger than the light velocity or it is in opposite direction to the wave vector. The regions of ω where it is larger than the light velocity but in direction of the wave vector $\boldsymbol{k}$ are

$$0<\omega<\omega_-, \quad \omega_+<\omega<\infty. \tag{10.14}$$

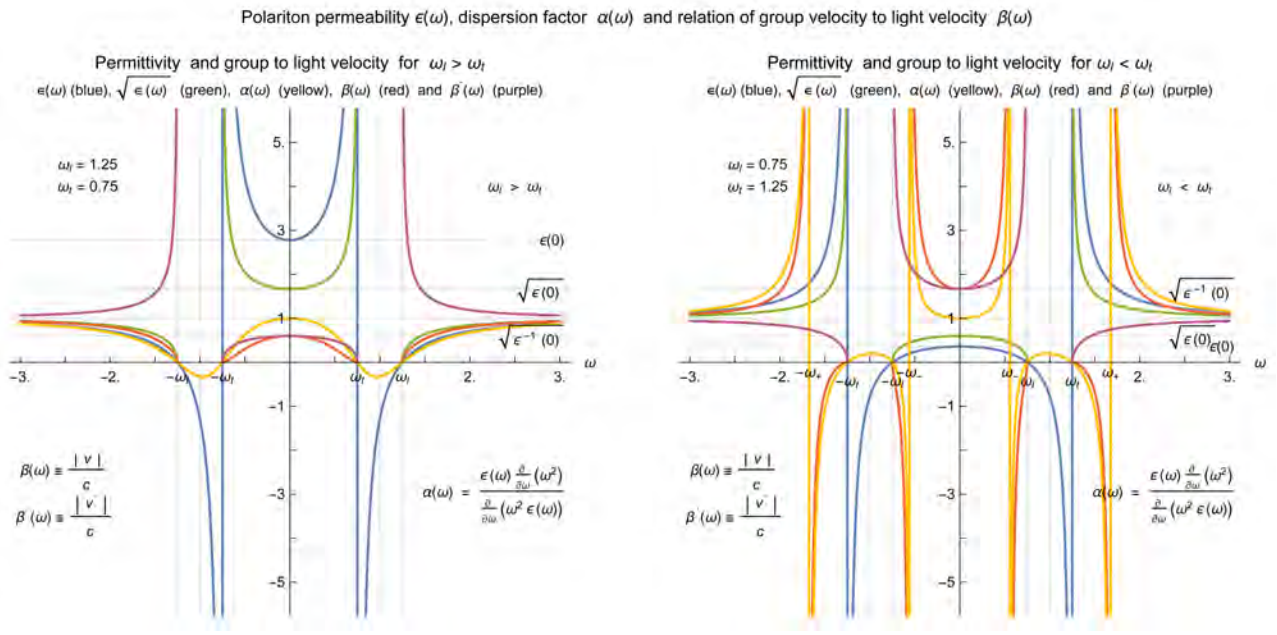


Figure 4. Polariton permittivity $\varepsilon(\omega)=\frac{\omega^2-\omega_l^2}{\omega^2-\omega_t^2},\left(\mu(\omega)=1\right)$, $\alpha(\omega)$ and $\beta(\omega)=\frac{|\boldsymbol{v}|}{c}$ for $\omega_l>\omega_t$ and for $\omega_l<\omega_t$. In regions $|\omega_t|<|\omega|<|\omega_l|$ for $\omega_l>\omega_t$ and $|\omega_l|<|\omega|<|\omega_t|$ for $\omega_l<\omega_t$ the relation of the group velocity to the light velocity (red curves with and purple without dispersion) becomes imaginary that is not drawn. In passive case $\omega_l>\omega_t$ wave vector and group velocity possess in every case the same direction and the last is smaller than the light velocity. In active case $\omega_l<\omega_t$ the group velocity is greater than the light velocity in the regions $0<\omega<\omega_-$ and $\omega_+<\omega$ but is in the same direction as the wave vector. In the region $\omega_t<\omega<\omega_+$ it is even in the opposite direction to the corresponding light velocity but only with taking into account the dispersion. The difference between $\beta(\omega)$ (red) and $\beta'(\omega)$ (purple is that last is calculated under neglect of the frequency dispersion of the permittivity. We have chosen for the pictures $\omega_l=1.25,\omega_t=0.75$ in passive case and $\omega_l=0.75,\omega_t=1.25$ in active case, the same as in Figure 1.

We will not come here with a quick physical explanation of this phenomenon although we considered the polariton permittivity with frequency $\omega_l<\omega<\omega_t$ already much earlier mainly with respect to its description of amplification in a certain region of wave vectors and frequencies. This phenomenon of group velocity greater than light velocity may likely come from a correlation within all parts of the medium by preparation of the occupation inversion of a level which is made already before it gets this property. The diffraction coefficients W_{ij} according to (10.10) possess the same zeros ($\omega_\pm$ in (10.3) and in Figure 4)) in the denominators as the group velocity v_i in (10.1) and can become very large. It is possible that the expansion (8.15) in the equation for the slowly varying amplitude (8.15) does not converge and that this equation is not applicable but a more general treatment should not change this basically. A practical use is likely very difficult to make and far from now. The most chances has its use for guided waves in long resonators and, in principle, such use due to the amplification properties of these similar media is already realized by lasers but not due to group velocity larger than light velocity as subsidiary effect. On the other side we cannot fully exclude that this phenomenon is already known and somewhere

discussed in literature. Hypothetical particles in free space which move with a velocity greater than that of light are called tachyons and were occasionally considered from the sixtieths on, in detail, e.g., by Terletski [28] (Chapters V. and VI.) and by others but in our case they may be only quasi-particles within an active medium with occupation inversion if we associate them with the group velocity greater than that of light. It is known that the existence of tachyons with imaginary mass is not excluded by the Relativity theory but experimentally such are not found. However, in our case they cannot exist as such which usually form beams due to very large diffraction and in this sense they are not the same as were discussed as tachyons.

There is a second fully unexpected phenomenon. In the regions $\omega_- < \omega < \omega_l$ and $\omega_t < \omega_+$ the group velocity $\boldsymbol{v}$ is in opposite direction to the direction of the wave vector $\boldsymbol{k}$ but only if we take into account the dispersion of the permittivity that let us believe at first in an error of signs but all was calculated with the same reliable formulae as in the passive case and the formula (10.3) for the dispersion factor does not involve square roots where the sign is unclear. It seems that this opposite group velocity to refraction vector violets the causality but we cannot exclude an explanation by the established correlation between different parts of the medium. So the described phenomenon requires further attention.

In search for references to the unexpected phenomena I came via the interesting book of Vaas [29] (part II, Section 7, e.g., p. 145, 146) to the reference of authors Nimtz and Haibel [30]. In [30] it is claimed that G. Nimtz (together with H. Aichmann, p. 111) transmitted in 1994 a symphony using the tunneling effect through a sub-dimensioned wave guide (full length about 12 cm) with approximately the five-fold velocity of the light velocity in vacuum[9]. In the transmission of information (e.g. music) the frequency modulation of a signal is used and has to be detected from it. Despite the large spatial distortion of a beam via the propagation the frequency is more stable and not very distorted over longer length of a beam that may explain some results. All this is a challenging theme and one has to wait for further clarification.

11. Remark to Reflection and Refraction of Beams at Isotropic and Bi-Isotropic Media

The following considerations concern only passive cases which are not very problematic with respect to basic discussions. In reflection and refraction problems of beams at a boundary between isotropic and (or) bi-isotropic media maximum 4 different beams with the same average frequency ω can be related to each other since they must possess the same tangential component of the

[9]Long ago I heard or saw in popular sources something about experiments of G. Nimtz but did not take the super-luminal velocities seriously and did not try to find the original papers. Now, when I find as it seems to me by reliable mathematics such possibilities of group velocities greater than light velocity in vacuum I changed my opinion and look for physical explanations. However, I have to emphasize that I cannot support by own calculation the effects concerning tunneling and the conditions for them since I did not make such.

average wave vectors $\boldsymbol{k}_{\nu}^{i,r}$ or refraction vector $\boldsymbol{n}_{\nu}^{i,r} \equiv \frac{c}{\omega}\boldsymbol{k}_{\nu}^{i,r}$ (lower indices "1" and "2" stand for the both media and upper indices "*i*" for incident and "*r*" for reflected or refracted wave. The group velocities of the corresponding beams are calculated in (7.6). This is represented in Figure 5 on the left-hand picture and the corresponding beam propagation on the right-hand picture. If $\boldsymbol{N}$ is a normal unit vector to the boundary at the considered point of beam reflection and refraction then the tangential component of all refraction vectors is

$$\overline{\boldsymbol{n}} \equiv \left[\boldsymbol{N},\left[\boldsymbol{n},\boldsymbol{N}\right]\right] = \boldsymbol{n} - \boldsymbol{n}\boldsymbol{N}\cdot\boldsymbol{N}, \quad \left(\boldsymbol{N}^2 = 1\right), \tag{11.1}$$

where for $\boldsymbol{n}$ an arbitrary of the involved refraction vectors can be inserted. Nothing changes in this picture if the refraction vectors in one or both media satisfy a dispersion equation $\boldsymbol{n}^2 = \varepsilon(\omega)\mu(\omega)$ in comparison to $\mu(\omega) = 1$. Also in the active case of isotropic media the group velocities $\boldsymbol{v}$ are in every case parallel to the corresponding refraction vectors $\boldsymbol{n}$ although we found that in this case the possibility exists that they are in opposite direction to the refraction vectors and may possess super-luminal velocities that is not yet fully understood and affirmed. Mostly one has only an incident beam from one side of the two media and only three waves (incident, reflected and refracted) are coupled at the boundary. All this is necessary to take into account if one discusses the work of Pendry [22] (see Appendix D).

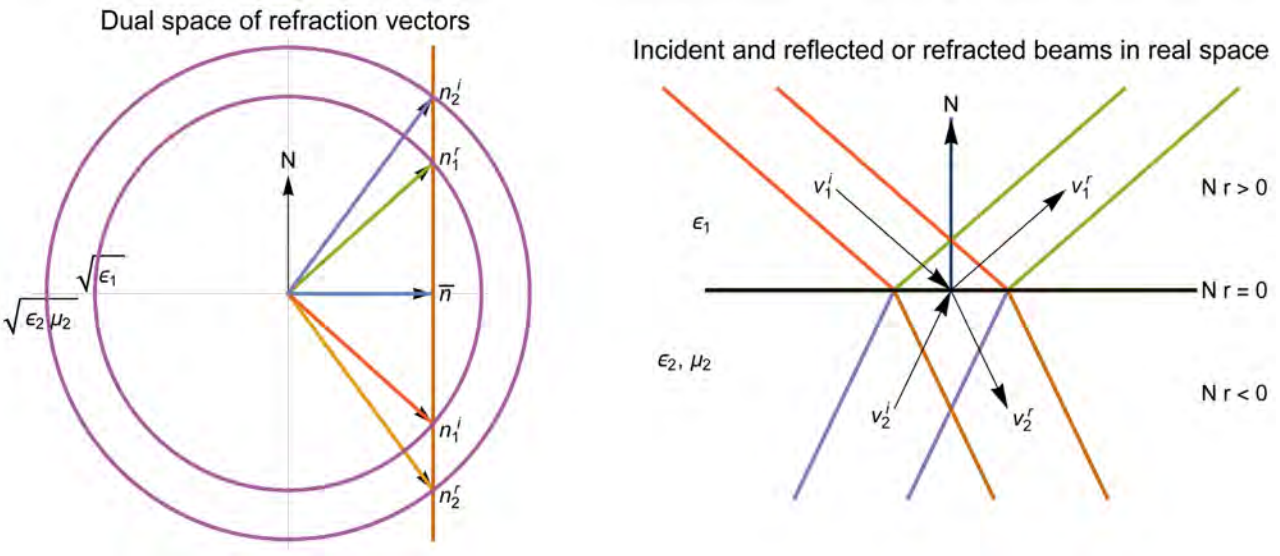


Figure 5. Refraction vectors $\boldsymbol{n}$ with equal tangential components and group velocities $\boldsymbol{v}$ at a boundary. The two media possess the lower indices 1 and 2 where medium 2 may possess an electric permittivity and a magnetic permeability. The upper indices "*i*" and "*r*" mean "incident" and "reflected or refracted" waves seen from the different sides of the boundary. It does not play a role whether a positive product $\varepsilon_2\mu_2$ is obtained from both positive or both negative values of ε_2 and μ_2. All refraction vectors $\boldsymbol{n}$ possess the same tangential component $\overline{\boldsymbol{n}}$ to the boundary $\boldsymbol{N}\boldsymbol{r} = 0$ and only their normal components are different in general case. For isotropic media the directions of the refraction vectors $\boldsymbol{n}$ and the corresponding group velocities $\boldsymbol{v}$ are the same. With respect to polarization of the 4 waves one may distinguish the two cases of polarization perpendicular and within the incidence plane spanned by the normal unit vector $\boldsymbol{N}$ to the boundary plane and an arbitrary of the refraction vectors $\boldsymbol{n}$.

The normal components $(\boldsymbol{nN})\cdot\boldsymbol{N}$ of all involved refraction vectors $\boldsymbol{n}$ can be obtained from the dispersion equations $\boldsymbol{n}_1^2=\varepsilon_1(\omega)$ and $\boldsymbol{n}_2^2=\varepsilon_2(\omega)\mu_2(\omega)$, respectively

$$\boldsymbol{n}_1^{i,r}\boldsymbol{N}=\pm\sqrt{\varepsilon_1(\omega)-\overline{\boldsymbol{n}}^2},\quad \boldsymbol{n}_2^{i,r}\boldsymbol{N}=\pm\sqrt{\varepsilon_2(\omega)\mu_2(\omega)-\overline{\boldsymbol{n}}^2}. \tag{11.2}$$

Furthermore, in Figure 5 two case of polarizations are possible, with the electric field polarized in the incidence plane spanned by vectors $\boldsymbol{N}$ and $\overline{\boldsymbol{n}}$ and perpendicular to it. The discussion of the amplitude relations for the involved wave is not necessary here for the intended purpose.

12. Conclusion

We compared two concepts of representing the constitutive equations in classical macroscopic optics of homogeneous anisotropic media, first the more general concept of spatial dispersion and then the more special concept of bi-anisotropic media with two constitutive equations for the electric and the magnetic induction and made this in coordinate-invariant way. Then this was specialized to uniaxial and, finally, to isotropic media where a possible equation for quasi-plane and quasi-monochromatic beam propagation in approximation with the first two terms of an expansion of the slowly varying beam amplitudes with respect to derivatives in space and time was derived taking into account diffraction. This was then applied for the isotropic case to the polariton permittivity (9.1) with a detailed discussion of the passive and the active case. The active case is obtained from the passive case by changing a sign in the susceptibility and describes amplification in certain regions of the frequency that somehow goes connected with feedback in resonators in direction of (a prestep of) laser action in the stationary regime. Losses in form of imaginary parts in the permittivity, we did not include for more simplicity and calculability of the formulae.

When we calculated the group velocity for beams in the active case we found as a surprise for some regions of frequency the possibility of velocities faster than the light velocity in vacuum and as a yet greater surprise the possibility of an opposite direction to the direction of the wave vectors (but different signs with and without taking into account dispersion). A certain physical explanation is possibly by the preparation of occupation inversion in establishing the active case and therefore a correlation between all parts of the medium which is not contained in the equations. We found after this that it is not fully unknown from literature. We have to think furthermore about these phenomena and have to find access to more literature about this.

In the process of working with the topics into our viewpoint came also the notion of "negative refraction" but we could not find some positive aspects and the possibility of its realization in isotropic media and make a few remarks to this in the text and in Appendix D.

Primarily, I intended to include also such peculiar cases as inhomogeneous waves (for example, total reflection) and, in particular, the cases of optic axes with calculation of the two-dimensional projection operators for the electric

field and the conical approximation of the dispersion surface in the neighborhood of optic axes in coordinate-invariant treatment. However, for extent and necessary time hoping to realize it we move this to a possible later time.

Conflicts of Interest

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

References

[1] Agranovich, V.M. and Ginzburg, V.L. (1979) Spatial Dispersion in Crystal Optics and the Theory of Excitons. Second Edition, Nauka, Moskva.

[2] Ginzburg, V.L. (1981) Theoretical Physics and Astrophysics. Nauka, Moskva.

[3] Silin, V.P. and Rukhadze, A.A. (1961) Elektromagnitnye svoistva plasma i plasmapodobnykh sred, (Electromagnetic Properties of Plasma and Plasma-Like Media). Gosatomisdat, Moskva.

[4] Landau, L.D. and Lifshitz, E.M. (1982) Electrodynamics of Continuous Media. 2nd Edition, Revised and Complemented by E.M. Lifshits and L.P. Pitayevski, Nauka, Moskva.

[5] I.E. Tamm, (1976) Osnovy teorii elektrichestva (Foundations of the Theory of Electricity). 9th Edition, Nauka, Moskva.

[6] Sommerfeld, A. (1959) Optik. 2nd Edition, Akademische Verlagsgesellschaft, Geest and Portig, Leipzig.

[7] Born, M. and Wolf, E. (1999) Principles of Optics. Seventh Edition, Cambridge University Press, Cambridge.

[8] Szivessy, G. (1928) Kristalloptik. In: Geiger, H. und Scheel, K., Eds., *Handbuch der Physik*, *Band XX*, Springer, Berlin, 635-904. https://doi.org/10.1007/978-3-642-90780-7_11

[9] Fyodorov, F.I. (1958) Optika anisotropnykh sred (Optics of Anisotropic Media). Izdatelstvo Akademya nauk BSSR, Minsk. (In Russian)

[10] Fyodorov, F.I. (1976) Teoriya girotropii (Theroy of Gyrotropy). Nauka i tekhnika, Minsk.

[11] Fyodorov, F.I. and Filippov, V.V. (1976) Otrazhenye i prelomlyenye sveta prosrachnymi fristallami (Reflection and Refraction of Light by Transparent Crystals). Nauka i Tekhnika, Minsk.

[12] Wünsche, A. (1970) *Annals of Physics* (*Leipzig*), **25**, 201-214. https://doi.org/10.1002/andp.19704800207

[13] Wünsche, A. (1978) *Annals of Physics* (*Leipzig*), **35**, 303-320. https://doi.org/10.1002/andp.19784900407

[14] Wünsche, A. (2021) *Journal of Modern Physics*, **12**, 1866-1921. https://doi.org/10.4236/jmp.2021.1213108

[15] Wünsche, A. (1970) *Annals of Physics* (*Leipzig*), **25**, 179-200. https://doi.org/10.1002/andp.19704800206

[16] Knox, R.S. (1963) Theory of Excitons. Academic Press, New York.

[17] Agranovich, V.M. (1968) Teoriya Eksitonov (Theory of Excitons). Nauka, Moskva.

[18] Davydov, A.S. (1968) Teoria molekularnykh eksitonov (Theory of Molecular Excitons). Nauka, Moskva.

[19] Agranovich, V.M. and Galanin, M.D. (1978) Perenos energii elektronnovo vozbuzhdyeniya v kondensirovannykh sredakh, (Transmission of Energy of Electronic Excitations in Condensed Media). Nauka, Moskva.

[20] Pekar, S.I. (1982) Kristallooptika i dobavochnyje volny (Crystal Optics and Additional Waves). Naukova Dumka, Kiev.

[21] Wünsche, A. (2021) Relativistic-Covariant Energy-Momentum Tensor for Homogeneous Anisotropic Dispersive Media. https://doi.org/10.4236/jmp.2021.1213108

[22] Pendry, J.B. (2000) *Physical Review Letters*, **89**, 3966-3969. https://doi.org/10.1103/PhysRevLett.85.3966

[23] de Groot, S.R. and Suttorp, L.G. (1972) Foundations of Electrodynamics. North Holland Publ. Comp., Amsterdam.

[24] Bloembergen, N. (1965) Nonlinear Optics. W.A. Benjamin, Inc., New York.

[25] Wünsche, A. (1971) *der Wissenschaften zu Berlin*, **13**, 754-769. https://doi.org/10.1515/9783112531341-007

[26] Weiglhofer, W.S. (1995) Frequency-Dependent Dyadic Green Functions for Bianisotropic Media. In: Barrett, T.W. and Grimes, D.M., Eds., *Advanced Electromagnetism*: *Foundations*, *Theory and Applications*, World Scientific Publishing, Singapore, 376-389. https://doi.org/10.1142/9789812831323_0013

[27] Barrett, T.W. and Grimes, D.M. (1995) Advanced Electromagnetism, Foundations, Theory and Applications. World Scientific, Singapore. https://doi.org/10.1142/2599

[28] Terletsky, Ya.P. (1966) Paradoksy teoriya otnositel"nosti, (Paradoxes of Relativity Theory). Nauka, Moskva.

[29] Vaas, R. (2013) Tunnel durch Raum und Zeit. 5th Edition, Kosmos, Stuttgart.

[30] Nimtz, G. and Haibel, H. (2004) Tunneleffekt—Räume ohne Zeit. Wiley VCH, Weinheim.

Appendix A. Some Important Relations for Three-Dimensional Operators

The most important relation for general three-dimensional operators A is the Cayley-Hamilton identity

$$\mathsf{A}^3 - \langle \mathsf{A} \rangle \mathsf{A}^2 + [\mathsf{A}]\mathsf{A} - |\mathsf{A}|\mathsf{I} = 0, \tag{A.1}$$

with the invariants with respect to similarity transformations which are the trace $\langle \mathsf{A} \rangle \equiv A^j_j$, the second invariant $[\mathsf{A}]$ and the determinant $|\mathsf{A}|$ according to (for spaces with metric tensor $g_{ij} = g_{ji}$ we may set $A_{ik} \equiv g_{ij} A^j_k$)

$$\langle \mathsf{A} \rangle \equiv A_{ii}, \quad [\mathsf{A}] \equiv \frac{1}{2}\left(\langle \mathsf{A} \rangle^2 - \langle \mathsf{A}^2 \rangle \right), \quad |\mathsf{A}| \equiv \frac{1}{6}\left(\langle \mathsf{A} \rangle^3 - 3\langle \mathsf{A} \rangle \langle \mathsf{A}^2 \rangle + 2\langle \mathsf{A}^3 \rangle \right). \tag{A.2}$$

As consequence of the Cayley-Hamilton identity the inverse operator to an arbitrary operator A can be represented by

$$\mathsf{A}^{-1} \equiv \frac{\overline{\mathsf{A}}}{|\mathsf{A}|}, \quad \overline{\mathsf{A}} \equiv \mathsf{A}^2 - \langle \mathsf{A} \rangle \mathsf{A} + [\mathsf{A}]\mathsf{I}, \quad \Rightarrow \quad \overline{\mathsf{A}}\mathsf{A} = \mathsf{A}\overline{\mathsf{A}} = |\mathsf{A}|\mathsf{I}, \tag{A.3}$$

where $\overline{\mathsf{A}}$ is the so-called complementary (or associated) operator to A. The invariants of this operator are

$$\langle \overline{\mathsf{A}} \rangle = [\mathsf{A}], \quad [\overline{\mathsf{A}}] = |\mathsf{A}| \langle \mathsf{A} \rangle, \quad |\overline{\mathsf{A}}| = |\mathsf{A}|^2, \tag{A.4}$$

and it possesses the properties

$$\overline{\mathsf{A}}^2 = [\mathsf{A}]\overline{\mathsf{A}} + |\mathsf{A}|\left(\mathsf{A} - \langle \mathsf{A} \rangle \mathsf{I} \right), \quad \overline{\overline{\mathsf{A}}} \equiv \overline{\left(\overline{\mathsf{A}} \right)} = \frac{\mathsf{A}}{|\mathsf{A}|}. \tag{A.5}$$

All relations in (A.3), (A.4) and (A.5) are general relations for arbitrary three-dimensional operators A.

For the invariants of the sum of two operators A and B one derives the following identities (the complementary operators $\overline{\mathsf{A}}$ to operator A we define later)

$$\begin{aligned} &\langle \mathsf{A} + \mathsf{B} \rangle = \langle \mathsf{A} \rangle + \langle \mathsf{B} \rangle, \\ &[\mathsf{A} + \mathsf{B}] = [\mathsf{A}] + \langle \mathsf{A} \rangle \langle \mathsf{B} \rangle - \langle \mathsf{AB} \rangle + [\mathsf{B}], \\ &|\mathsf{A} + \mathsf{B}| = |\mathsf{A}| + \langle \overline{\mathsf{A}}\mathsf{B} \rangle + \langle \mathsf{A}\overline{\mathsf{B}} \rangle + |\mathsf{B}|. \end{aligned} \tag{A.6}$$

Special cases are $\mathsf{B} = \beta \mathsf{I}$

$$\begin{aligned} &\langle \mathsf{A} + \beta \mathsf{I} \rangle = \langle \mathsf{A} \rangle + 3\beta, \\ &[\mathsf{A} + \beta \mathsf{I}] = [\mathsf{A}] + 2\beta \langle \mathsf{A} \rangle + 3\beta^2, \\ &|\mathsf{A} + \beta \mathsf{I}| = |\mathsf{A}| + \beta [\mathsf{A}] + \beta^2 \langle \mathsf{A} \rangle + \beta^3. \end{aligned} \tag{A.7}$$

The complementary operator of the sum of two operators is

$$\overline{\mathsf{A} + \mathsf{B}} \equiv \overline{\mathsf{A}} + \mathsf{AB} + \mathsf{BA} + \overline{\mathsf{B}} - \left(\langle \mathsf{B} \rangle \mathsf{A} + \langle \mathsf{A} \rangle \mathsf{B} \right) + \left(\langle \mathsf{A} \rangle \langle \mathsf{B} \rangle - \langle \mathsf{AB} \rangle \right) \mathsf{I}. \tag{A.8}$$

The projection operator Π for determination of eigenvectors of an operator A to non-degenerate eigenvalue α is

$$\Pi = \frac{\overline{\mathsf{A} - \alpha \mathsf{I}}}{\langle \overline{\mathsf{A} - \alpha \mathsf{I}} \rangle} = \frac{\mathsf{A}^2 - \left(\langle \mathsf{A} \rangle - \alpha \right) \mathsf{A} + \left([\mathsf{A}] - \alpha \langle \mathsf{A} \rangle + \alpha^2 \right) \mathsf{I}}{[\mathsf{A}] - 2\alpha \langle \mathsf{A} \rangle + 3\alpha^2}, \quad \Pi^2 = \Pi. \tag{A.9}$$

For the differentiation of a determinant $|\mathsf{A}|$ of an operator A with respect to a parameter λ we find from (A.2) and (A.3) the identity

$$\frac{\partial|\mathsf{A}|}{\partial\lambda}=\left\langle\overline{\mathsf{A}}\frac{\partial\mathsf{A}}{\partial\lambda}\right\rangle=\frac{1}{2}\left\langle\frac{\partial\overline{\mathsf{A}}}{\partial\lambda}\mathsf{A}\right\rangle, \tag{A.10}$$

where the complementary operator $\overline{\mathsf{A}}$ to operator A is defined in (A.3).

Appendix B. Identities for Vector and Volume Products in Connection with Operators

We derive here mathematical identities for volume and vector products in connection with operators which are almost unknown or less known in case that they are already somewhere published. A part of them is used in the main text of our considerations.

We consider the volume product $[\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}]$ of three vectors $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$ and apply now to each vector the same operator A that means we consider the volume product $[\mathsf{A}\boldsymbol{x},\mathsf{A}\boldsymbol{y},\mathsf{A}\boldsymbol{z}]$. Due to complete antisymmetry of the volume product with respect to permutations of neighbored vectors the volume product $[\mathsf{A}\boldsymbol{x},\mathsf{A}\boldsymbol{y},\mathsf{A}\boldsymbol{z}]$ is proportional to the volume product $[\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}]$ with a proportionality factor which we denote by $|\mathsf{A}|$ and call the determinant of A that means

$$[\mathsf{A}\boldsymbol{x},\mathsf{A}\boldsymbol{y},\mathsf{A}\boldsymbol{z}]=|\mathsf{A}|[\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}]. \tag{B.1}$$

One may convince oneself that this is really a good possibility to define the determinant of a three-dimensional operator. Now we write the chain of identities for a general volume product

$$[\mathsf{A}\boldsymbol{x},\mathsf{A}\boldsymbol{y}]\mathsf{A}\boldsymbol{z}=[\mathsf{A}\boldsymbol{x},\mathsf{A}\boldsymbol{y},\mathsf{A}\boldsymbol{z}]=[\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}]|\mathsf{A}|=[\boldsymbol{x},\boldsymbol{y}]|\mathsf{A}|\boldsymbol{z}=[\boldsymbol{x},\boldsymbol{y}]\overline{\mathsf{A}}\mathsf{A}\boldsymbol{z}, \tag{B.2}$$

where we substituted the determinant $|\mathsf{A}|$ according to $|\mathsf{A}|\to|\mathsf{A}|\mathsf{I}=\overline{\mathsf{A}}\mathsf{A}$ with $\overline{\mathsf{A}}$ the complementary operator to A (see also (A.3)). Since $\boldsymbol{z}$ is an arbitrary vector we may omit $\mathsf{A}\boldsymbol{z}$ in the identity (B.2) and obtain the identity for vector products

$$[\mathsf{A}\boldsymbol{x},\mathsf{A}\boldsymbol{y}]=[\boldsymbol{x},\boldsymbol{y}]\overline{\mathsf{A}},\quad\Leftrightarrow\quad[\mathsf{A}\boldsymbol{x},\mathsf{A}\boldsymbol{y}]\mathsf{A}=|\mathsf{A}|[\boldsymbol{x},\boldsymbol{y}]. \tag{B.3}$$

From this also follows almost immediately

$$\left[\overline{\mathsf{A}}\boldsymbol{x},\overline{\mathsf{A}}\boldsymbol{y}\right]=[\boldsymbol{x},\boldsymbol{y}]\left(\overline{\overline{\mathsf{A}}}\right)=|\mathsf{A}|[\boldsymbol{x},\boldsymbol{y}]\mathsf{A},\quad\left(\left(\overline{\overline{\mathsf{A}}}\right)\equiv\overline{\overline{\mathsf{A}}}=|\mathsf{A}|\mathsf{A}\right). \tag{B.4}$$

If we let act the operator A onto vectors $\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}},\tilde{\boldsymbol{z}}$ to the left then from

$$[\tilde{\boldsymbol{x}}\mathsf{A},\tilde{\boldsymbol{y}}\mathsf{A},\tilde{\boldsymbol{z}}\mathsf{A}]=|\mathsf{A}|[\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}},\tilde{\boldsymbol{z}}], \tag{B.5}$$

follows in analogous way

$$[\tilde{\boldsymbol{x}}\mathsf{A},\tilde{\boldsymbol{y}}\mathsf{A}]=\overline{\mathsf{A}}[\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}],\quad\Leftrightarrow\quad\mathsf{A}[\tilde{\boldsymbol{x}}\mathsf{A},\tilde{\boldsymbol{y}}\mathsf{A}]=|\mathsf{A}|[\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}]. \tag{B.6}$$

If we substitute in (B.1) according to $\mathsf{A}\to\mathsf{A}-\alpha\mathsf{I}$ with arbitrary scalar α and use for determinants

$$|\mathsf{A}-\alpha\mathsf{I}|=|\mathsf{A}|-[\mathsf{A}]\alpha+\langle\mathsf{A}\rangle\alpha^2-\alpha^3, \tag{B.7}$$

then by collecting on both sides of the obtained identity the terms to equal pow-

ers of α we find in addition to (B.1) two further identities of the form

$$\begin{aligned}&[\mathsf{A}\boldsymbol{x},\mathsf{A}\boldsymbol{y},\boldsymbol{z}]+[\mathsf{A}\boldsymbol{x},\boldsymbol{y},\mathsf{A}\boldsymbol{z}]+[\boldsymbol{x},\mathsf{A}\boldsymbol{y},\mathsf{A}\boldsymbol{z}]=[\mathsf{A}][\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}],\\&[\mathsf{A}\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}]+[\boldsymbol{x},\mathsf{A}\boldsymbol{y},\boldsymbol{z}]+[\boldsymbol{x},\mathsf{A}\boldsymbol{y},\mathsf{A}\boldsymbol{z}]=\langle\mathsf{A}\rangle[\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}].\end{aligned}\tag{B.8}$$

If we remove from (B.8) the vector $\boldsymbol{z}$ or make the substitution $\mathsf{A}\to\mathsf{A}-\alpha\mathsf{I}$ in the identity (B.3) using

$$\overline{\mathsf{A}-\alpha\mathsf{I}}=\overline{\mathsf{A}}-\left(\langle\mathsf{A}\rangle\mathsf{I}-\mathsf{A}\right)\alpha+\mathsf{I}\alpha^2,\tag{B.9}$$

and collect all terms on both sides to equal powers of α we find in addition to (B.3) the identities

$$\begin{aligned}&[\mathsf{A}\boldsymbol{x},\mathsf{A}\boldsymbol{y}]+\left([\mathsf{A}\boldsymbol{x},\boldsymbol{y}]+[\boldsymbol{x},\mathsf{A}\boldsymbol{y}]\right)\mathsf{A}=[\mathsf{A}][\boldsymbol{x},\boldsymbol{y}],\\&[\mathsf{A}\boldsymbol{x},\boldsymbol{y}]+[\boldsymbol{x},\mathsf{A}\boldsymbol{y}]+[\boldsymbol{x},\boldsymbol{y}]\mathsf{A}=\langle\mathsf{A}\rangle[\boldsymbol{x},\boldsymbol{y}],\end{aligned}\tag{B.10}$$

or equivalently to the last

$$[\mathsf{A}\boldsymbol{x},\boldsymbol{y}]+[\boldsymbol{x},\mathsf{A}\boldsymbol{y}]=[\boldsymbol{x},\boldsymbol{y}]\left(\langle\mathsf{A}\rangle\mathsf{I}-\mathsf{A}\right).\tag{B.11}$$

All these mathematical identities may be also derived by means of the Levi-Civita pseudo-tensors. It is also clear that analogous identities can be derived for the action of operators onto the vectors to the left similar to (B.4) and (B.5) which we do not write down here.

We make a short remark to vector and volume products. It is very convenient to define $[\boldsymbol{y}]$ as an antisymmetric covariant second-rank pseudo-tensor to the contravariant vector $\boldsymbol{y}$. Then one can use identities for vector and volume products by displacement of the squared brackets such as[10]

$$\begin{aligned}&[\boldsymbol{y}]\boldsymbol{z}=[\boldsymbol{y},\boldsymbol{z}],\quad \boldsymbol{x}[\boldsymbol{y}]=[\boldsymbol{x},\boldsymbol{y}],\\&\boldsymbol{x}[\boldsymbol{y}]\boldsymbol{z}=\boldsymbol{x}[\boldsymbol{y},\boldsymbol{z}]=[\boldsymbol{x},\boldsymbol{y}]\boldsymbol{z}=[\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}].\end{aligned}\tag{B.12}$$

The second line written with contra- and covariant indices and with the Levi-Civity pseudo-tensor ε_{jkl} written is for example

$$[\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}]=\varepsilon_{jkl}x^jy^kz^l=x^j\left(\varepsilon_{jkl}y^k\right)z^l=x^j[\boldsymbol{y}]_{jl}z^l=x^j[\boldsymbol{y},\boldsymbol{z}]_j=[\boldsymbol{x},\boldsymbol{y}]_lz^l,\tag{B.13}$$

and $[\boldsymbol{y}]_{jl}\equiv\varepsilon_{jkl}y^k$ is the antisymmetric pseudo-tensor to vector y^k. An "antisymmetric" operator $[\boldsymbol{y}]^i_l$ which transforms vectors (or pseudo-vectors, depending on the kind of $\boldsymbol{y}$) can be made from $[\boldsymbol{y}]_{jk}$ only for Euclidean or pseudo-Euclidean spaces which possess a symmetrical metric tensor $g_{jk}=g_{kj},\left(g^{ij}g_{jk}=\delta^i_k\right)$ by $[\boldsymbol{y}]^i_l\equiv g^{ij}[\boldsymbol{y}]_{jl}$. With all this which may be presented in more precise form I made good experience for a long period of scientific work.

Appendix C. Algebra to a Special Operator for Bi-Anisotropic Media

We consider the algebra to the following three-dimensional operator L which

[10]This is likely also the intention when Fyodorov writes instead of my $[\boldsymbol{y}]$ the form $\boldsymbol{y}^\times$ with the consequence for the vector product "$[\boldsymbol{yz}]=\boldsymbol{y}^\times\boldsymbol{z}$" (similar to "$\boldsymbol{y}\times\boldsymbol{z}$" (Gibbs); besides, F. writes vector products also with squared brackets but without a comma between the vectors) however, with the disadvantage that it works only to the right onto vectors.

we need to obtain, in particular, to the calculation of the determinant of the operator for a general bi-anisotropic medium

$$\mathsf{L} \equiv \frac{\mathsf{A}\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}\mathsf{A}-(\tilde{\boldsymbol{x}}\mathsf{A}\boldsymbol{x})\mathsf{A}}{|\mathsf{A}|}+\mathsf{B}\equiv\mathsf{M}+\mathsf{B}, \tag{C.1}$$

where A and B are general three-dimensional operators. It is a little more general than we need it since, in principal, we use in the main text only the special case of equality $\tilde{\boldsymbol{x}}=\boldsymbol{x}$ of the vectors $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$. With M we abbreviate the operator

$$\mathsf{M} \equiv \frac{\mathsf{A}\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}\mathsf{A}-(\tilde{\boldsymbol{x}}\mathsf{A}\boldsymbol{x})\mathsf{A}}{|\mathsf{A}|}. \tag{C.2}$$

The determinant $|\mathsf{L}|$ of the operator L can be calculated by

$$|\mathsf{L}|=|\mathsf{M}|+\langle\bar{\mathsf{M}}\mathsf{B}\rangle+\langle\mathsf{M}\bar{\mathsf{B}}\rangle+|\mathsf{B}|, \tag{C.3}$$

where overlining an operator means the transition to the complementary operator (see Section 2). This formula is known [9] [10] and surely some others and is easily to obtain by coordinate-invariant calculation. Thus we have first to consider some algebra of the more complicate part M in (C.1).

First we have to calculate the powers M^2 and M^3 and their traces that is straightforward to make and that we do not write down. In particular, we find for the invariants of M

$$\begin{aligned}\langle\mathsf{M}\rangle &= \frac{\tilde{\boldsymbol{x}}\mathsf{A}^2\boldsymbol{x}-\langle\mathsf{A}\rangle\tilde{\boldsymbol{x}}\mathsf{A}\boldsymbol{x}}{|\mathsf{A}|},\\ [\mathsf{M}] &= \frac{(\tilde{\boldsymbol{x}}\mathsf{A}\boldsymbol{x})\tilde{\boldsymbol{x}}\boldsymbol{x}}{|\mathsf{A}|},\\ |\mathsf{M}| &= 0,\end{aligned} \tag{C.4}$$

that means the determinant of M is vanishing which simplifies the application of the formula (C.3). That the determinant $|\mathsf{M}|$ is vanishing follows also because the operator M possesses the eigenvalue $\mu=0$ to right-hand eigenvector $\boldsymbol{x}$ and to left-hand eigenvector $\tilde{\boldsymbol{x}}$. For the complementary operator $\bar{\mathsf{M}}$ to M using (A.3) we find then

$$\bar{\mathsf{M}}=\frac{(\tilde{\boldsymbol{x}}\mathsf{A}\boldsymbol{x})\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}}{|\mathsf{A}|},\quad \langle\bar{\mathsf{M}}\rangle=[\mathsf{M}]=\frac{(\tilde{\boldsymbol{x}}\mathsf{A}\boldsymbol{x})\tilde{\boldsymbol{x}}\boldsymbol{x}}{|\mathsf{A}|}, \tag{C.5}$$

and, furthermore, it is easy to check

$$\bar{\mathsf{M}}\mathsf{M}=\mathsf{M}\bar{\mathsf{M}}=|\mathsf{M}|\mathsf{I}=0. \tag{C.6}$$

Now, according to the formula (C.3) with the explicit form (C.5) of M and $|\mathsf{M}|=0$ we find

$$\begin{aligned}|\mathsf{L}| &= \frac{(\tilde{\boldsymbol{x}}\mathsf{A}\boldsymbol{x})(\tilde{\boldsymbol{x}}\mathsf{B}\boldsymbol{x})-\langle\mathsf{A}\bar{\mathsf{B}}\rangle(\tilde{\boldsymbol{x}}\mathsf{A}\boldsymbol{x})+\tilde{\boldsymbol{x}}\mathsf{A}\bar{\mathsf{B}}\mathsf{A}\boldsymbol{x}+|\mathsf{A}||\mathsf{B}|}{|\mathsf{A}|}\\ &= \frac{(\tilde{\boldsymbol{x}}\mathsf{A}\boldsymbol{x})(\tilde{\boldsymbol{x}}\mathsf{B}\boldsymbol{x})-\tilde{\boldsymbol{x}}\left(\langle\mathsf{A}\bar{\mathsf{B}}\rangle\mathsf{A}-\mathsf{A}\bar{\mathsf{B}}\mathsf{A}\right)\boldsymbol{x}+|\mathsf{A}\mathsf{B}|}{|\mathsf{A}|}.\end{aligned} \tag{C.7}$$

This agrees with the unessentially more special results of Fyodorov [9] (Equations (36.9), (36.10)) which, however, were calculated for the operator $\mathsf{B}^{-1}\mathsf{L}$ (in our notation) from multiple vector products which brings an additional factor $|\mathsf{B}|$ into the denominator of (C.7) but since he applies this immediately to the wave equation for which the determinant has to vanish he can omit this factor $|\mathsf{AB}|$ in the denominator in Equation (36.10). The numerator in (C.7) is symmetric with respect to permutation of the operators $\mathsf{A}\rightleftharpoons\mathsf{B}$. This can be achieved using a general operator identity from which results the following equivalent representation

$$|\mathsf{L}|=\frac{(\tilde{\boldsymbol{x}}\mathsf{A}\boldsymbol{x})(\tilde{\boldsymbol{x}}\mathsf{B}\boldsymbol{x})-\tilde{\boldsymbol{x}}\left(\overline{\mathsf{AB}}+\overline{\mathsf{BA}}-[\mathsf{AB}]\mathsf{I}-\left([\mathsf{A}]\overline{\mathsf{B}}+[\mathsf{B}]\overline{\mathsf{A}}-[\mathsf{A}][\mathsf{B}]\mathsf{I}\right)\right)\boldsymbol{x}+|\mathsf{AB}|}{|\mathsf{A}|}, \quad \text{(C.8)}$$

which though a little longer in the numerator but shows this property. The mentioned operator identity is

$$\begin{aligned}[\mathsf{A}]\overline{\mathsf{B}}=&\,\mathsf{ABAB}+\mathsf{AB}^2\mathsf{A}+\mathsf{BABA}-\langle\mathsf{B}\rangle\mathsf{ABA}-\langle\mathsf{AB}\rangle(\mathsf{AB}+\mathsf{BA})\\&+\left(\langle\mathsf{AB}\rangle\langle\mathsf{B}\rangle-\langle\mathsf{AB}^2\rangle\right)\mathsf{A}+[\mathsf{AB}]\mathsf{I},\end{aligned} \quad \text{(C.9)}$$

and can be derived from a more general operator identity with 3 operators A, 1 operator B and 1 operator C in each sum term and in symmetric way which generalizes the Cayley-Hamilton identity by substitutions and specialization $\mathsf{C}=\mathsf{B}$. A derivation by means of the Levi-Civita pseudo-tensors is also possible. Since it is long we do not derive it here but hope to find opportunity to do this in future. Finally, we give here the other two invariants of L which are

$$\begin{aligned}\langle\mathsf{L}\rangle&=\frac{\left(\tilde{\boldsymbol{x}}\mathsf{A}^2\boldsymbol{x}-\langle\mathsf{A}\rangle\tilde{\boldsymbol{x}}\mathsf{A}\boldsymbol{x}\right)+|\mathsf{A}|\langle\mathsf{B}\rangle}{|\mathsf{A}|},\\ [\mathsf{L}]&=\frac{(\tilde{\boldsymbol{x}}\mathsf{A}\boldsymbol{x})(\tilde{\boldsymbol{x}}\boldsymbol{x})-\left(\tilde{\boldsymbol{x}}\mathsf{ABA}\boldsymbol{x}-\langle\mathsf{B}\rangle\tilde{\boldsymbol{x}}\mathsf{A}^2\boldsymbol{x}+\left(\langle\mathsf{A}\rangle\langle\mathsf{B}\rangle-\langle\mathsf{AB}\rangle\right)\tilde{\boldsymbol{x}}\mathsf{A}\boldsymbol{x}\right)+|\mathsf{A}|[\mathsf{B}]}{|\mathsf{A}|},\end{aligned} \quad \text{(C.10)}$$

which are not symmetric in A and B and the complementary operator to L is

$$\begin{aligned}\overline{\mathsf{L}}=\frac{1}{|\mathsf{A}|}\{&(\tilde{\boldsymbol{x}}\mathsf{A}\boldsymbol{x})\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}+\mathsf{A}\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}\mathsf{AB}+\mathsf{BA}\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}\mathsf{A}-\tilde{\boldsymbol{x}}\mathsf{A}\boldsymbol{x}(\mathsf{AB}+\mathsf{BA})\\&-\langle\mathsf{B}\rangle\mathsf{A}\boldsymbol{x}\cdot\tilde{\boldsymbol{x}}\mathsf{A}+\tilde{\boldsymbol{x}}\mathsf{A}\boldsymbol{x}\left(\langle\mathsf{B}\rangle\mathsf{A}+\langle\mathsf{A}\rangle\mathsf{B}\right)-\left(\tilde{\boldsymbol{x}}\mathsf{A}^2\boldsymbol{x}\right)\mathsf{B}\\&-\left(\tilde{\boldsymbol{x}}\mathsf{ABA}\boldsymbol{x}-\langle\mathsf{B}\rangle\tilde{\boldsymbol{x}}\mathsf{A}^2\boldsymbol{x}+\left(\langle\mathsf{A}\rangle\langle\mathsf{B}\rangle-\langle\mathsf{AB}\rangle\right)\tilde{\boldsymbol{x}}\mathsf{A}\boldsymbol{x}\right)\mathsf{I}\}+\overline{\mathsf{B}},\quad\langle\overline{\mathsf{L}}\rangle=[\mathsf{L}].\end{aligned} \quad \text{(C.11)}$$

Appendix D. Is the Notion "Negative Refraction" Useful in Geometric and Wave Optics?

In connection with the paper of Pendry [22] and the mass of reactions to it I will express here also my thoughts though I did not follow this development from the beginning and cannot exclude that similar thoughts are already discussed in literature.

In wave optics in the treatment of beams we have a mean value $\boldsymbol{k}_0$ of the

wave vector and a mean value ω_0 of the frequency and both together with the complex conjugate part but well separated are involved in the main factors $e^{\pm i(\boldsymbol{k}_0\boldsymbol{r}-\omega_0 t)}$. From the wave vector one may form the refraction vector $\boldsymbol{n}$ by the definition

$$\boldsymbol{k} \equiv \frac{\omega}{c}\boldsymbol{n}, \quad |\boldsymbol{n}|^2 = \boldsymbol{n}^2, \quad \Rightarrow \quad \boldsymbol{k} = \frac{\omega}{c}|\boldsymbol{n}|\frac{\boldsymbol{n}}{|\boldsymbol{n}|} = |\boldsymbol{k}|\frac{\boldsymbol{n}}{|\boldsymbol{n}|}, \tag{D.1}$$

where both signs of $|\boldsymbol{n}|$ (refraction index; in general it is complex) can be chosen without something changing at the wave. Only the direction described by the unit vector $\frac{\boldsymbol{n}}{|\boldsymbol{n}|}$ in connection with the product $\boldsymbol{k} = |\boldsymbol{k}|\frac{\boldsymbol{n}}{|\boldsymbol{n}|}$ is invariant but not one of the two factors alone. If one change the sign of only one factor then one describes a wave propagating in the opposite direction and for isotropic media in no other than of these directions.

The second important fact is that under reflection and refraction of a beam at a surface $\boldsymbol{N}\boldsymbol{r} = a$ with the normal unit vector $\boldsymbol{N}$ the tangential components $\bar{\boldsymbol{k}} \equiv \big[\boldsymbol{N}[\boldsymbol{k}, \boldsymbol{N}]\big]$ of all coupled wave vectors $\boldsymbol{k}$

$$\boldsymbol{k} = \big[\boldsymbol{N}, [\boldsymbol{k}, \boldsymbol{N}]\big] + \boldsymbol{k}\boldsymbol{N}\cdot\boldsymbol{N} \equiv \bar{\boldsymbol{k}} + \boldsymbol{k}\boldsymbol{N}\cdot\boldsymbol{N}, \quad \big(\boldsymbol{N}^2 = 1\big), \tag{D.2}$$

in both media have to be the same and only the normal components $(\boldsymbol{k}\boldsymbol{N})\boldsymbol{N}$ can be different with both possible signs and the same is true for the refraction vectors $\boldsymbol{n}$. Even both signs of the normal component of the wave vector in the second medium are possible one as the refracted beam from the incident beam in the first medium and the second as an incident beam in the second medium which generates the refracted beam in the first medium which is identical in its direction with the reflected beam from the incident beam in the first medium. Both these beams possess a group velocity for isotropic media in the same direction as the wave vector independent on positive or negative signs of the electric permittivity and the magnetic permeability and whether or not the wave vectors are real or complex quantities. This excludes the possibility of refraction as drawn in Figure 1 in [22]. Only in case that the group velocity is in opposite direction to the wave or refraction vector in second medium we have in our **Figure 5** a right-hand picture which is somehow similar to mentioned of Pendry. This, however, is not connected with negative $\varepsilon(\omega)$ and $\mu(\omega)$ but merely with an active medium with all its problems as discussed (e.g., super-luminal velocities, super-diffraction of beams) and is hardly realizable. The situation changes but not basically if the medium is anisotropic and, obviously, this is not meant.

Wave vectors and equally refraction vectors possess only a modulus and a certain direction in space but not a certain sign of each separately and the notion in the heading seems to me not being really useful.

Journal of Modern Physics, 2022, 13, 620-670
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Minimum Resolution of the Minkowski, Schwarzschild and Kerr Differential Modules

J.-F. Pommaret

CERMICS, Ecole des Ponts ParisTech, France
Email: jean-francois.pommaret@wanadoo.fr

How to cite this paper: Pommaret, J.-F. (2022) Minimum Resolution of the Minkowski, Schwarzschild and Kerr Differential Modules. *Journal of Modern Physics*, **13**, 620-670.
https://doi.org/10.4236/jmp.2022.134036

Received: March 18, 2022
Accepted: April 26, 2022
Published: April 29, 2022

Abstract

Our recent arXiv preprints and published papers on the solution of the Riemann-Lanczos and Weyl-Lanczos problems have brought our attention on the importance of revisiting the algebraic structure of the Bianchi identities in Riemannian geometry. We also discovered in the meantime that, in our first GB book of 1978, we had already used a new way for studying the compatibility conditions (CC) of an operator that may not be necessarily formally integrable (FI) in order to construct canonical formally exact differential sequences on the jet level. The purpose of this paper is to prove that the combination of these two facts clearly shows the specific importance of the Spencer operator and the Spencer δ-cohomology, totally absent from mathematical physics today. The results obtained are unavoidable because they only depend on elementary combinatorics and diagram chasing. They also provide for the first time the purely intrinsic interpretation of the respective numbers of successive first, second, third and higher order generating CC. However, if they of course agree with the linearized Killing operator over the Minkowski metric, they largely disagree with recent publications on the respective numbers of generating CC for the linearized Killing operator over the Schwarzschild and Kerr metrics. Many similar examples are illustrating these new techniques, providing in particular a few resolutions in which the orders of the successive operators may go "up and down" surprisingly, like in the conformal situation for various dimensions.

Keywords

Formal Integrability, Involutivity, Compatibility Conditions, Spencer Operator, Janet Sequence, Spencer Sequence, Differential Module, Homological Algebra, Extension Module

1. Introduction

The present study is mainly local and we only use standard notations of diffe-

rential geometry. For simplicity, we shall also adopt the same notation for a vector bundle $(E,F,\cdots)$ and its set of sections $(\xi,\eta,\zeta,\cdots)$. Now, if X is the ground manifold with dimension n and local coordinates $(x^1,\cdots,x^n)$ and E is a vector bundle over X with local coordinates (x,y), we shall denote by $J_q(E)$ the q-jet bundle of E with local coordinates (x,y_q) and sections ξ_q transforming like the q-derivatives $j_q(\xi)$ of a section $\xi=\xi_0$ of E. If F with section η is another vector bundle over X and $\Phi:J_q(E)\to F$ is an epimorphism with kernel the linear system $R_q\subset J_q(E)$, we shall associate the differential operator $\mathcal{D}=\Phi\circ j_q:E\to F:\xi\to\eta$ and set $\Theta=ker(\mathcal{D})$. All the operators considered will be locally defined over a differential field K with n derivations $(\partial_1,\cdots,\partial_n)$ and we shall indicate the order of an operator under its arrow. It is well known and we shall provide many explicit examples, that, if we want to solve, at least locally the linear inhomogeneous system $\mathcal{D}\xi=\eta$, one usually needs *compatibility conditions* (CC) of the form $\mathcal{D}_1\eta=0$ defined by another differential operator $\mathcal{D}_1:F=F_0\to F_1:\eta\to\zeta$ that may be of high order in general but still locally defined over K. However, two types of "*phenomena*" can arise for exhibiting such CC but, though they can be quite critical in actual practice, we do not know any other reference on the possibility to solve them effectively because most people rely on the work of E. Cartan.

1) As shown in ([1], Introduction) or ([2]) with the Janet system $(\xi_{33}-x^2\xi_{11}=0,\xi_{22}=0)$ over the differential field $K=\mathbb{Q}(x)$ and in ([3]), it may be possible to find no CC of order one, no CC of order two, one CC of order three, then nothing new but one additional CC of order six and so on with no way to know when to stop. For the fun, when we started computer algebra around 1990, we had to ask a special permit to the head of our research department for running the computer a full night and were not even able after a day to go any further on. Hence, a first basic problem is to establish a preliminary list of *generating* CC and know their maximum order.

2) Once the previous problem is solved, we do know a generating $\mathcal{D}_1$ of order q_1 and may start anew with it in order to obtain a generating $\mathcal{D}_2$ of order q_2 and so on as a way to work out a differential sequence. Contrary to what can be found in the Poincaré sequence for the exterior derivative where all the successive operators are of order one, things may not be so simple in actual practice and "*jumps*" may appear, that is the orders may go up and down in a apparently surprising manner that only the use of "*acyclicity*" through the Spencer cohomology can explain. As we shall see with more details in the case of the conformal Killing operator of order 1, the successive orders are $(1,3,1)$ when $n=3$, $(1,2,2,1)$ when $n=4$, $(1,2,1,2,1)$ when $n=5$ ([4]).

A we have shown in our seven books, the only possibility to escape from these two types of problems is to start with an involutive operator $\mathcal{D}$ and construct in an intrinsic way two *canonical* differential sequences, namely the *linear Janet sequence* ([5], p. 185, 391 for a global definition):

$$0\to\Theta\to E\underset{q}{\overset{\mathcal{D}}{\to}}F_0\underset{1}{\overset{\mathcal{D}_1}{\to}}F_1\underset{1}{\overset{\mathcal{D}_2}{\to}}\cdots\underset{1}{\overset{\mathcal{D}_{n-1}}{\to}}F_{n-1}\underset{1}{\overset{\mathcal{D}_n}{\to}}F_n\to 0 \tag{1}$$

and the *linear Spencer sequence* ([5], p. 185 for a global definition):

$$0 \to \Theta \xrightarrow{j_q} C_0 \xrightarrow[1]{D_1} C_1 \xrightarrow[1]{D_2} C_2 \xrightarrow[1]{D_3} \cdots \xrightarrow[1]{D_{n-1}} C_{n-1} \xrightarrow[1]{D_n} C_n \to 0 \tag{2}$$

As in *both cases*, the central operator is the Spencer operator but *not the exterior derivative*, contrary to what is done in ([6] [7] [8]) and the corresponding references, in particular ([6], Ref. [8]), we do not agree on the effectivity of their definition of "*involutivity*" ([6], pp. 1608-1609). In fact, the most important property of theses two sequences is that they are *formally exact* on the jet level as follows. Introducing the (composite) *r- prolongation* by means of the *formal derivatives* d_i:

$$\rho_r(\Phi): J_{q+r}(E) \to J_r\left(J_q(E)\right) \to J_r(F_0): \left(x, y_{q+r}\right) \to \left(x, z_\nu = d_\nu \Phi, 0 \le |\nu| \le r\right)$$

with kernel $R_{q+r} = \rho_r\left(R_q\right) = J_r\left(R_q\right) \cap J_{q+r}(E) \subset J_r\left(J_q(E)\right)$, we have the long exact sequences:

$$0 \to R_{q+r} \to J_{q+r}(E) \to J_r(F_0) \tag{3}$$

$$0 \to R_{q+q_1+r} \to J_{q+q_1+r}(E) \to J_{q_1+r}(F_0) \to J_r(F_1) \tag{4}$$

$$0 \to R_{q+q_1+q_2+r} \to J_{q+q_1+q_2+r}(E) \to J_{q_1+q_2+r}(F_0) \to J_{q_2+r}(F_1) \to J_r(F_2) \tag{5}$$

and so on till the similar ones stopping at $J_r(F_n), \forall r \ge 0$. As shown by the counterexample exhibited in ([9], p. 119-126), *all these sequences may be absolutely useful till the last one*. We shall also define the symbol $g_q = R_q \cap S_q T^* \otimes E$ and its *r*-prolongations $g_{q+r} = \rho_r\left(g_q\right)$ only depends on g_q in a purely algebraic way, that is no differentiation is involved. On the contrary, we shall say that R_q or $\mathcal{D}$ is *formally integrable* (FI) if R_{q+r} is a vector bundle $\forall r \ge 0$ *and* all the epimorphisms $\pi_{q+r}^{q+r+1}: J_{q+r+1}(E) \to J_{q+r}(E): \left(x, y_{q+r+1}\right) \to \left(x, y_{q+r}\right)$ are inducing epimorphisms $R_{q+r+1} \to R_{q+r}$ of constant rank $\forall r \ge 0$, which is a true purely differential property.

Of course, for people familiar with functional analysis, the definition of Θ could seem strange and uncomplete as it is not clear where to look for solutions. In our opinion (See [10] and review Zbl 1079.93001) it is mainly for this reason that differential modules or simply *D*-modules have been introduced but we shall explain why such a procedure leads in fact to a (rather) *vicious circle* as follows. Working locally for simplicity with $dim(E) = m$, $dim(F) = p$, we may turn the definition backwards by introducing the non-commutative ring $D = K\left[d_1, \cdots, d_n\right] = K[d]$ of differential polynomials $(P, Q, \cdots)$ with coefficients in *K*. Then, instead of acting on the "*left*" of column vectors of sections by differentiations as in the previous differential setting, *we shall use the same operator matrix* still denoted by $\mathcal{D}$ but now acting on the "*right*" of row vectors by composition. Introducing the canonical projection onto the residual module *M*, we obtain the exact sequence $D^p \xrightarrow[q]{\mathcal{D}} D^m \to M \to 0$ of differential modules also called "*free resolution*" of *M* because D^m and D^p are clearly free differential modules. However, as *D* is filtred by the order of operators, then $I = im(\mathcal{D}) \subset D^m$ is filtred too and, as we shall clearly see on the motivating

examples, the induced filtration of $M = D^m/I$ can only been obtained in *any* applications if and only if R_q or $\mathcal{D}$ is FI. Accordingly, all the difficulty will be to use the following key theorem (For Spencer cohomology and acyclicity or involutivity, see [1] [2] [5] [9]-[14]):

THEOREM 1.1: There is a finite *Prolongation/Projection* (PP) algorithm providing two integers $r, s \geq 0$ by successive increase of each of them such that the new system $R_{q+r}^{(s)} = \pi_{q+r}^{q+r+s}\left(R_{q+r+s}\right)$ has the same solutions as R_q but is FI with a 2-acyclic or involutive symbol and first order CC. The order of a generating $\mathcal{D}_1$ is thus bounded by $r+s+1$ as we used $r+s$ prolongations.

EXAMPLE 1.2: In the Janet example we have $R_2 \to R_3^{(1)} \to R_4^{(2)} \to R_5^{(2)}$ with $8 < 11 < 12 = 12$ and $dim_K(M) = 12 \Rightarrow rk_D(M) = 0$. The final system is trivially involutive because it is FI with a zero symbol, a fact highly not evident *a prori* because it needs 5 prolongations and the maximum order of the CC is thus equal to $3+2+1=6$. We obtain therefore a minimum resolution of the form $0 \to D \xrightarrow[4]{\mathcal{D}_2} D^2 \xrightarrow[6]{\mathcal{D}_1} D^2 \xrightarrow[2]{\mathcal{D}} D \to M \to 0$ (See the introduction of [1] or [2] for details).

When a system is FI, we have a *projective limit*

$R = R_\infty \to \cdots \to R_q \to \cdots \to R_1 \to R_0$.

As we are dealing with a differential field *K*, there is a bijective correspondence:

$$M_q = hom_K(R_q, K) \quad \Leftrightarrow \quad R_q = hom_K(M_q, K) \tag{6}$$

and we obtain the *injective limit* $0 \subseteq M_0 \subseteq M_1 \subseteq \ldots \subseteq M_q \subseteq \cdots \subseteq M_\infty = M$ providing the *filtration* of *M*. We have in particular $d_i M_q \subseteq M_{q+1}$ and $M = DM_q$ for $q \gg 0$.

THEOREM 1.3: $R = hom_K(M, K)$ is a differential module for the Spencer operator.

Proof: As the ring *D* is generated by *K* and $T = \left\{a^i d_i \mid a^i \in K\right\}$, we just need to define:

$$(af)(m) = a(f(m)) = f(am),$$
$$(d_i f)(m) = \partial_i(f(m)) - f(d_i m), \forall a \in K, \forall m \in M, \forall d_i \in T, \forall f \in R$$

and obtain $d_i a = a d_i + \partial_i a$ in the operator sense. Choosing $m \in M$ to be the residue of $d_\mu y^k = y_\mu^k$ and setting $f(y_q) = \xi_q : f(y_\mu^k) = \xi_\mu^k \in K$, we obtain in actual practice *exactly* the Spencer operator: $d : R \to T^* \otimes R : f \to dx^i \otimes d_i f$ with $(d_i f)_\mu^k = \partial_i \xi_\mu^k - \xi_{\mu+1_i}^k$ or $d\xi_{q+1} = j_1(\xi_q) - \xi_{q+1}$ or simply $d = \partial - \delta$ with a slight abuse of language. We notice that a "section" $\xi_q \in R_q$ has in general, particularly for the non-commutative case (See [4] for examples), *nothing to do* with a "solution", a concept missing in ([6] [7] [8]).

□

As we shall see in the motivating examples, once a differential module *M* or the dual system $R = hom_K(M, K)$ is given, there may be quite different differential sequences or quite different resolutions and the problem will be to choose the one that could be the best in the application considered. During the last

world war, many mathematicians discovered that a few concepts, called *extension modules*, were not depending on the sequence used in order to compute them but *only* on M. A (very) delicate theorem of (differential) homological algebra even proves that no others can exist ([15]). Let us explain in a way as simple as possible these new concepts.

As a preliminary crucial definition, if $P = a^{\mu} d_{\mu} \in D$, we shall define its (formal) *adjoint* by the formula $ad(P) = (-1)^{|\mu|} d_{\mu} a^{\mu}$ where we have set $|\mu| = \mu_1 + \cdots + \mu_n$ whenever $\mu = (\mu_1, \cdots, \mu_n)$ is a multi-index. Such a definition can be extended by linearity in order to define the formal adjoint $ad(\mathcal{D})$ to be the transposed operator matrix obtained after taking the adjoint of each element. The main property is that

$ad(PQ) = ad(Q)ad(P), \forall P, Q \in D \Rightarrow ad(\mathcal{D}_1 \circ \mathcal{D}) = ad(\mathcal{D}) \circ ad(\mathcal{D}_1)$.

EXAMPLE 1.4: With $\partial_{22}\xi = \eta^2, \partial_{12}\xi = \eta^1$ for $\mathcal{D}$, we get $\partial_1\eta^2 - \partial_2\eta^1 = \zeta$ for $\mathcal{D}_1$. Then $ad(\mathcal{D}_1)$ is defined by $\mu^2 = -\partial_1\lambda, \mu^1 = \partial_2\lambda$ while $ad(\mathcal{D})$ is defined by $\nu = \partial_{12}\mu^1 + \partial_{22}\mu^2$ but the CC of $ad(\mathcal{D}_1)$ are generated by $\nu' = \partial_1\mu^1 + \partial_2\mu^2$. In the operator framework, we have the differential sequences:

$$\begin{array}{ccccc} \xi & \overset{\mathcal{D}}{\rightarrow} & \eta & \overset{\mathcal{D}_1}{\rightarrow} & \zeta \\ & & & & \\ \nu & \overset{ad(\mathcal{D})}{\leftarrow} & \mu & \overset{ad(\mathcal{D}_1)}{\leftarrow} & \lambda \\ & \swarrow & & & \\ \nu' & & & & \end{array} \tag{7}$$

where the upper sequence *is* formally exact at η but the lower sequence *is not* formally exact at μ.

Passing to the module framework, we obtain the sequences:

$$\begin{array}{ccccccc} D & \overset{\mathcal{D}_1}{\rightarrow} & D^2 & \overset{\mathcal{D}}{\rightarrow} & D & \rightarrow M \rightarrow 0 \\ & & & & & & \\ D & \overset{ad(\mathcal{D}_1)}{\leftarrow} & D^2 & \overset{ad(\mathcal{D})}{\leftarrow} & D & & \end{array} \tag{8}$$

where the lower sequence is not exact at D^2. The "*extension modules*" have been introduced in order to study this kind of "*gaps*".

Therefore, it may be important or useful to prove that certain extension modules vanish, that is $ad(\mathcal{D})$ generates the CC of $ad(\mathcal{D}_1)$ whenever $\mathcal{D}_1$ generates the CC of $\mathcal{D}$. Such a problem is even essential for checking controllability in control theory ([2]) but we also remind the reader that it is not so easy to exhibit the CC of the Maxwell or Morera parametrizations when $n = 3$ and that a direct checking for $n = 4$ should be strictly impossible ([16]). It has been proved by L. P. Eisenhart in 1926 (Compare to [5]) that the solution space Θ of the Killing system has $n(n+1)/2$ *infinitesimal generators* $\{\theta_\tau\}$ linearly independent over the constants *if and only if* ω had constant Riemannian curvature, namely zero in our case. As we have a Lie group of transformations preserving the metric, the three theorems of Sophus Lie assert than $[\theta_\rho, \theta_\sigma] = c^\tau_{\rho\sigma}\theta_\tau$ where the *structure constants* c define a Lie algebra $\mathcal{G}$. We have therefore $\xi \in \Theta \Leftrightarrow \xi = \lambda^\tau\theta_\tau$ with $\lambda^\tau = cst$. Hence, we may replace the

Killing system by the system $\partial_i \lambda^\tau = 0$, getting therefore the differential sequence:

$$0 \to \Theta \to \wedge^0 T^* \otimes \mathcal{G} \overset{d}{\to} \wedge^1 T^* \otimes \mathcal{G} \overset{d}{\to} \cdots \overset{d}{\to} \wedge^n T^* \otimes \mathcal{G} \to 0 \tag{9}$$

which is the tensor product of the Poincaré sequence for the exterior derivative by the Lie algebra $\mathcal{G}$. Finally, as the extension modules do not depend on the resolution used and that most of them do vanish because the Poincaré sequence is self adjoint (up to sign), that is $ad(d)$ generates the CC of $ad(d)$ at any position, exactly like d generates the CC of d at any position. We invite the reader to compare with the situation of the Maxwell equations in electromagnetism ([13]). However, we have proved in ([17] [18] [19] [20]) why neither the Janet sequence nor the Poincaré (de Rham in USA!) sequence can be used in physics and must be replaced by another resolution of Θ called *Spencer sequence* (See [14] for details and compare to [21]).

We are now in a position to tell about the unpleasant story that has motivated such a paper. In October 23-27, 2017, I was invited to lecture at the Albert Einstein Institute (AEI, Potsdam/Berlin). Though the series of lectures was already planned and written (arXiv: 1802.09610 published in [18]), the day before the first lecture the group leader decided that I should lecture on compatibility conditions (CC). I suddenly understood that General Relativity (GR) at AEI was no longer a Science but became a Religion that does not admit *any* criticism by lecturing visitors, with a similar comment for Gauge Theory (GT) ([1] [13] [14]). The following elementary example will explain the title of this paper and the problems raised by its content in such a framework.

With $n=2$, $D=\mathbb{Q}(a)[d_1,d_2]$, y an indeterminate and a an arbitrary constant parameter, let us consider the second order system $R_2 \subset J_2(E)$ defined by the PD equations $y_{22}=0$, $y_{12}+ay_1=0$. When $a=0$, we have the involutive system $y_{22}=0$, $y_{12}=0$ defined over $D=\mathbb{Q}[d_1,d_2]$ and the minimum resolution $0 \to D \underset{1}{\to} D^2 \underset{2}{\to} D \to M \to 0$ already considered with *one first order* CC.

However, if $a \neq 0$, after a few crossed derivatives, we may obtain the successive strict inclusions $R_2^{(2)} \subset R_2^{(1)} \subset R_2$ providing first the intermediate subsystem $y_{22}=0$, $y_{12}=0$, $y_1=0$ and then the final involutive subsystem $y_{22}=0$, $y_{12}=0$, $y_{11}=0$, $y_1=0$. *Not only the final subsystem is surprisingly no longer depending on the parameter* but the corresponding minimum resolution, namely $0 \to D \underset{2}{\to} D^2 \underset{2}{\to} D \to M \to 0$, is quite different as it now involves *one second order* CC because $(d_{12}+ad_1)d_{22}-d_{22}(d_{12}+ad_1)=0$. One can also consider the linear inhomogeneous second order system $y_{22}=u$, $y_{12}=v-aw$, $y_{11}=w_1$, $y_1=w$ with $a^2w=u_1-v_2+av$ and discover that the 4 canonical CC determined by the corresponding Janet tabular are in fact generated by a single second order CC (See [1] [3] and [5] for other more sophisticated explicit examples).

As a kind of training exercise, I solved the PP problem for the linearized Killing operator over the Minkowski and Schwarzschild (S) metrics with parameter

(*a*) in ([22]) and over the Kerr (K) metric with parameters (a,m) in ([23]). As I hope to have convinced the reader that the previous example is *exactly* similar, my claim in this paper is that the search for CC has nothing to do with GR and is a purely mathematical problem of "Formal Integrability".

After this long introduction, the content of the paper will become clear:

In Section 2 we provide the mathematical tools from homological algebra and differential geometry needed for finding the generating CC of various orders.

Then, Section 3 will provide motivating examples in order to illustrate these new concepts.

They are finally applied to the Killing systems for the S and K metrics in Section 4 in such a way that the results obtained, though surprising they are, cannot be avoided because they will only depend on diagram chasing and elementary combinatorics. They largely disagree with ([6] [7] [8]) because the techniques used in these papers are not intrinsic. As the final involutive systems do not depend any longer on the S or K parameters like in the above example, the worst conclusion concerns the physical usefulness of solving such a problem but… this is surely another story!

2. Mathematical Tools

A) HOMOLOGICAL ALGEBRA

We now need a few definitions and results from homological algebra ([2] [10] [15]). In all that follows, $A,B,C,\cdots$ are modules over a ring or vector spaces over a field and the linear maps are making the diagrams commutative. We introduce the notations $rk=rank$, $nb=number$, $dim=dimension$, $ker=kernel$, $im=image$, $coker=cokernel$. When $\Phi:A\to B$ is a linear map (homomorphism), we may consider the so-called ker/coker exact sequence where $coker(\Phi)=B/im(\Phi)$:

$$0\to ker(\Phi)\to A\overset{\Phi}{\to}B\to coker(\Phi)\to 0$$

In the case of vector spaces over a field k, we successively have:

$$rk(\Phi)=dim(im(\Phi)),dim(ker(\Phi)))=dim(A)-rk(\Phi),$$
$$dim(coker(\Phi))=dim(B)-rk(\Phi)$$

with $dim(coker(\Phi))=nb(compatibility\ conditions)$. We obtain thus by substraction:

$$dim(ker(\Phi))-dim(A)+dim(B)-dim(coker(\Phi))=0$$

In the case of modules, using localization, we may replace the dimension by the rank and obtain the same relations because of the additive property of the rank. The following result is essential:

SNAKE THEOREM 2.A.1: When one has the following commutative diagram resulting from the two central vertical short exact sequences by exhibiting the three corresponding horizontal ker/coker exact sequences:

$$\begin{array}{ccccccccccc} & & 0 & & 0 & & 0 & & & & \\ & & \downarrow & & \downarrow & & \downarrow & & & & \\ 0 & \to & K & \to & A & \to & A' & \to & Q & \to & 0 \\ & & \downarrow & & \downarrow \Phi & & \downarrow \Phi' & & \downarrow & & \\ 0 & \to & L & \to & B & \to & B' & \to & R & \to & 0 \\ & & \downarrow & & \downarrow \Psi & & \downarrow \Psi' & & \downarrow & & \\ 0 & \to & M & \to & C & \to & C' & \to & S & \to & 0 \\ & & & & \downarrow & & \downarrow & & \downarrow & & \\ & & & & 0 & & 0 & & 0 & & \end{array} \tag{10}$$

then there exists a connecting map $M \to Q$ both with a long exact sequence:

$$0 \to K \to L \to M \to Q \to R \to S \to 0.$$

Proof: We start constructing the connecting map by using the following succession of elements:

$$\begin{array}{ccccccc} & & a & \cdots & a' & \to & q \\ & & \vdots & & \downarrow & & \vdots \\ & & b & \to & b' & \cdots & 0 \\ & & \downarrow & & \vdots & & \\ m & \to & c & \cdots & 0 & & \end{array}$$

Indeed, starting with $m \in M$, we may identify it with $c \in C$ in the kernel of the next horizontal map. As Ψ is an epimorphism, we may find $b \in B$ such that $c = \Psi(b)$ and apply the next horizontal map to get $b' \in B'$ in the kernel of Ψ' by the commutativity of the lower square. Accordingly, there is a unique $a' \in A'$ such that $b' = \Phi'(a')$ and we may finally project a' to $q \in Q$. The map is well defined because, if we take another lift for c in B, it will differ from b by the image under Φ of a certain $a \in A$ having zero image in Q by composition. The remaining of the proof is similar and left to the reader as an exercise. The above explicit procedure will not be repeated.

□

We may now introduce *cohomology theory* through the following definition:

DEFINITION 2.A.2: If one has any sequence $A \xrightarrow{\Phi} B \xrightarrow{\Psi} C$, then one may introduce $coboundary = im(\Phi) \subseteq ker(\Psi) = cocycle \subseteq B$ and the cohomology at B is the quotient *cocycle*/*coboundary*.

THEOREM 2.A.3: The following commutative diagram where the two central vertical sequences are long exact sequences and the horizontal lines are ker/coker exact sequences:

$$\begin{array}{ccccccccccccc} & & & & 0 & & 0 & & 0 & & & & \\ & & & & \downarrow & & \downarrow & & \downarrow & & & & \\ 0 & \to & K & \to & A & \to & A' & \to & Q & \to & 0 & & \\ & & & & \downarrow & & \downarrow \Phi & & \downarrow \Phi' & & \downarrow & & \\ 0 & \to & L & \to & B & \to & B' & \to & R & \to & 0 & & \\ \cdots & \cdots & \downarrow & \cdots & \downarrow \Psi & \cdots & \downarrow \Psi' & \cdots & \downarrow & \cdots & \cdots & \cdots & cut \\ 0 & \to & M & \to & C & \to & C' & \to & S & \to & 0 & & \\ & & & & \downarrow & & \downarrow \Omega & & \downarrow \Omega' & & \downarrow & & \\ 0 & \to & N & \to & D & \to & D' & \to & T & \to & 0 & & \\ & & & & \downarrow & & \downarrow & & \downarrow & & & & \\ & & & & 0 & & 0 & & 0 & & & & \end{array} \tag{11}$$

induces an isomorphism between the cohomology at M in the left vertical column and the kernel of the morphism $Q \to R$ in the right vertical column.

Proof: Let us "cut" the preceding diagram into the following two commutative and exact diagrams by taking into account the relations $im(\Psi) = ker(\Omega)$, $im(\Psi') = ker(\Omega')$:

$$\begin{array}{ccccccccccc}
 & & 0 & & 0 & & 0 & & & & \\
 & & \downarrow & & \downarrow & & \downarrow & & & & \\
0 & \to & K & \to & A & \to & A' & \to & Q & \to & 0 \\
 & & \downarrow & & \downarrow \Phi & & \downarrow \Phi' & & \downarrow & & \\
0 & \to & L & \to & B & \to & B' & \to & R & \to & 0 \\
 & & \downarrow & & \downarrow \Psi & & \downarrow \Psi' & & & & \\
0 & \to & cocycle & \to & im\Psi & \to & im\Psi' & & & & \\
 & & & & \downarrow & & \downarrow & & & & \\
 & & & & 0 & & 0 & & & &
\end{array}$$

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & cocycle & \to & ker\Omega & \to & ker\Omega' & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & M & \to & C & \to & C' & & \\
 & & \downarrow & & \downarrow \Omega & & \downarrow \Omega' & & \\
0 & \to & N & \to & D & \to & D' & & \\
 & & & & \downarrow & & \downarrow & & \\
 & & & & 0 & & 0 & &
\end{array}$$

Using the snake theorem, we successively obtain:

$$\begin{array}{lll}
\Rightarrow & \exists & 0 \to K \to L \overset{\Psi}{\to} cocycle \to Q \to R \qquad exact \\
\Rightarrow & \exists & 0 \to coboundary \to cocycle \to ker(Q \to R) \to 0 \quad exact \\
\Rightarrow & & cohomology\ at\ M \simeq ker(Q \to R)
\end{array}$$

□

B) DIFFERENTIAL GEOMETRY

Comparing the sequences obtained in the previous examples, we may state:

DEFINITION 2.B.1: A differential sequence is said to be *formally exact* if it is exact on the jet level composition of all the prolongations involved. A formally exact sequence is said to be *strictly exact* if all the operators/systems involved are FI (See [1] [5] [11] [14] [24] [25] for more details). A strictly exact sequence is called *canonical* if all the operators/systems are involutive. Forty years ago, we did provide the link existing between the only known canonical sequences, namely the Janet and Spencer sequences ([5], See in particular the pages 185 and 391).

With canonical projection $\Phi_0 = \Phi : J_q(E) \Rightarrow J_q(E)/R_q = F_0$, the various prolongations are described by the following commutative and exact "*introductory diagram*" often used in the sequel:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & g_{q+r+1} & \to & S_{q+r+1}T^*\otimes E & \xrightarrow{\sigma_{r+1}(\Phi)} & S_{r+1}T^*\otimes F_0 & \to & h_{r+1} & \to 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0 & \to & R_{q+r+1} & \to & J_{q+r+1}(E) & \xrightarrow{\rho_{r+1}(\Phi)} & J_{r+1}(F_0) & \to & Q_{r+1} & \to 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0 & \to & R_{q+r} & \to & J_{q+r}(E) & \xrightarrow{\rho_r(\Phi)} & J_r(F_0) & \to & Q_r & \to 0 \\
 & & & & \downarrow & & \downarrow & & \downarrow & \\
 & & & & 0 & & 0 & & 0 &
\end{array} \tag{12}$$

Chasing along the diagonal of this diagram while applying the standard "*snake*" lemma, we obtain the useful "*long exact connecting sequence*" also often used in the sequel:

$$\boxed{0 \to g_{q+r+1} \to R_{q+r+1} \to R_{q+r} \to h_{r+1} \to Q_{r+1} \to Q_r \to 0} \tag{13}$$

which is thus connecting in a tricky way FI (*lower left*) with CC (*upper right*).

We finally recall the "*fundamental diagram I*" that we have presented in many books and papers, relating the (upper) *canonical Spencer sequence* to the (lower) *canonical Janet sequence*, that only depends on the left commutative square $\mathcal{D}=\Phi\circ j_q$ with $\Phi=\Phi_0$ when one has an involutive system $R_q\subseteq J_q(E)$ over E with $dim(X)=n$ and $j_q:E\to J_q(E)$ is the derivative operator up to order q while the epimorphisms $\Phi_1,\cdots,\Phi_n$ are successively induced by Φ:

$$\begin{array}{ccccccccccccc}
 & & & & 0 & & 0 & & & & 0 & & \\
 & & & & \downarrow & & \downarrow & & & & \downarrow & & \\
 & & & \xrightarrow{j_q} & & \xrightarrow{D_1} & & \xrightarrow{D_2} & & \xrightarrow{D_n} & & & \\
0 & \to & \Theta & \to & C_0 & \to & C_1 & \to & \cdots & \to & C_n & & \to 0 \\
 & & & & \downarrow & & \downarrow & & & & \downarrow & & \\
 & & & \xrightarrow{j_q} & & \xrightarrow{D_1} & & \xrightarrow{D_2} & & \xrightarrow{D_n} & & & \\
0 & \to & E & \to & C_0(E) & \to & C_1(E) & \to & \cdots & \to & C_n(E) & & \to 0 \\
 & & \| & & \downarrow\Phi_0 & & \downarrow\Phi_1 & & & & \downarrow\Phi_n & & \\
 & & & \xrightarrow{\mathcal{D}} & & \xrightarrow{\mathcal{D}_1} & & \xrightarrow{\mathcal{D}_2} & & \xrightarrow{\mathcal{D}_n} & & & \\
0 & \to & \Theta & \to & E & \to & F_0 & \to & F_1 & \to \cdots \to & F_n & & \to 0 \\
 & & & & & & \downarrow & & \downarrow & & \downarrow & & \\
 & & & & & & 0 & & 0 & & 0 & &
\end{array} \tag{14}$$

This result will be used in order to compare the M, S and K metrics when $n=4$ but it is important to notice that this *whole* diagram does not depend any longer on the parameter (m) of S or on the parameters (a,m) of K ([22] [23]).

PROPOSITION 2.B.2: If $R_q\subset J_q(E)$ and $R_{q+1}\subset J_{q+1}(E)$ are two systems of respective orders q and $q+1$, then $R_{q+1}\subset\rho_1(R_q)$ if and only if $\pi_q^{q+1}(R_{q+1})\subset R_q$ *and* $dR_{q+1}\subset T^*\otimes R_q$.

Proof: First we notice that *necessarily* we must have $\pi_q^{q+1}(R_{q+1})\subset R_q$ because, as $\rho_1(R_q)$ may not project *onto* R_q, it is nevertheless defined by (maybe)

more equations of strict order q than R_q. Now, if $\xi_{q+1} \in R_{q+1} \subset J_{q+1}(E)$ is such that $d\xi_{q+1} \in T^* \otimes R_q$, then $\xi_q \in R_q \Rightarrow j_1(\xi_q) \in J_1(R_q)$. As $J_1(R_q)$ is an affine bundle over R_q modelled on $T^* \otimes R_q$ (or simply $T^* \otimes R_q \subset J_1(R_q)$) and $J_{q+1}(E) \subset J_1(J_q(E))$, we have thus
$\xi_{q+1} = j_1(\xi_q) - d\xi_{q+1} \in J_1(R_q) \cap J_{q+1}(E) = \rho_1(R_q)$.

The converse way is similar.

□

The next key idea has been discovered in ([5]) as a way to define the so-called Janet bundles and thus for a totally different reason.

DEFINITION 2.B.3: Let us "cut" the preceding introductory diagram by means of a central vertical line and define $R'_r = im(\rho_r(\Phi)) \subseteq J_r(F_0)$ with $R'_0 = F_0$. Chasing in this diagram, we notice that $\pi_r^{r+1} : J_{r+1}(E) \to J_r(E)$ induces an epimorphism $\pi_r^{r+1} : R'_{r+1} \to R'_r, \forall r \geq 0$. However, a chase in this diagram proves that *the kernel of this epimorphism is not* $im(\sigma_{r+1}(\Phi))$ unless R_q is FI (*care*). For this reason, we shall define it to be *exactly* g'_{r+1}.

THEOREM 2.B.4: $R'_{r+1} \subseteq \rho_1(R'_r)$ and $dim(\rho_1(R'_r)) - dim(R'_{r+1})$ is the number of new generating CC of order $r+1$.

Proof: First of all, we have the following commutative and exact diagram obtained by applying the Spencer operator to the top long exact sequence:

$$\begin{array}{ccccccccc}
0 \to & R_{q+r+1} & \to & J_{q+r+1}(E) & \to & J_{r+1}(F_0) & \to & Q_{r+1} & \to 0 \\
 & \downarrow d & & \downarrow d & & \downarrow d & & \downarrow d & \\
0 \to & T^* \otimes R_{q+r} & \to & T^* \otimes J_{q+r}(E) & \to & T^* \otimes J_r(F_0) & \to & T^* \otimes Q_r & \to 0
\end{array}$$

"Cutting" the diagram in the middle as before while using the last definition, we obtain the induced map $R'_{r+1} \overset{d}{\to} T^* \otimes R'_r$ and the first inclusion follows from the last proposition. Such a procedure cannot be applied to the top row of the introductory diagram through the use of δ instead of d because of the comment done on the symbol in the last definition.

Now, using only the definition of the prolongation for the system and its symbol, we have the following commutative and exact diagram:

$$\begin{array}{ccccccccc}
 & 0 & & 0 & & 0 & & 0 & \\
 & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0 \to & \rho_1(g'_r) & \to & S_{r+1}T^* \otimes F_0 & \overset{\sigma_1(\Psi)}{\to} & T^* \otimes Q_r & \to & Q'_1 & \to 0 \\
 & \downarrow & & \downarrow & & \downarrow & & \| & \\
0 \to & \rho_1(R'_r) & \to & J_{r+1}(F_0) & \overset{\rho_1(\Psi)}{\to} & J_1(Q_r) & \to & Q'_1 & \to 0 \\
 & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0 \to & R'_r & \to & J_r(F_0) & \overset{\Psi}{\to} & Q_r & \to & 0 & \\
 & \downarrow & & \downarrow & & \downarrow & & & \\
 & 0 & & 0 & & 0 & & &
\end{array}$$

and obtain the following commutative and exact diagram:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & g'_{r+1} & \to & \rho_1(g'_r) & \to & A & \to & 0 \\
 & & \downarrow & & \downarrow & & \| & & \\
0 & \to & R'_{r+1} & \to & \rho_1(R'_r) & \to & A & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & R'_r & = & R'_r & \to & 0 & & \\
 & & \downarrow & & \downarrow & & & & \\
 & & 0 & & 0 & & & &
\end{array}$$

The computation of $y = dim(A) = dim(\rho_1(R'_r)) - dim(R'_{r+1})$ only depends on $x = dim(Q'_1)$ and is rather tricky as follows (See the motivating examples):

$$dim(Q_r) = dim(R_{q+r}) + dim(J_r(F_0)) - dim(J_{q+r}(E))$$

$$dim(\rho_1(R'_r)) = dim(J_{r+1}(F_0)) + x - dim(J_1(Q_r))$$

$$dim(R'_{r+1}) = dim(J_{q+r+1}(E)) - dim(R_{q+r+1})$$

As we shall see with the motivating examples and with the S or K metrics, the computation is easier when the system is FI but can be much more difficult when the system is not FI.

However, the number of linearly independent CC of order $r+1$ coming from the CC of order r is $dim(J_1(Q_r)) - x$ while the total number of CC of order $r+1$ is:

$$\begin{aligned} dim(Q_{r+1}) &= dim(R_{q+r+1}) + dim(J_{r+1}(F_0)) - dim(J_{q+r+1}(E)) \\ &= dim(J_{r+1}(F_0)) - dim(R'_{r+1}) \end{aligned}$$

The number of new CC of *strict* order $r+1$ is equal to y because $dim(J_{r+1}(F_0))$ disappears by difference. For a later use in GR, we point out the fact that, if the given system $R_q \subset J_q(E)$ depends on parameters that *must* be contained in the ground differential field K (only (m) for the S metric but (a,m) for the K metric), all the dimensions considered may *highly* depend on them even if the underlying procedure is of course the same.

As an alternative proof, we may say that the number of CC of strict order $r+1$ obtained from the CC of order r is equal to $dim(S_{r+1}T^* \otimes F_0) - dim(\rho_1(g'_r))$ while the total number of CC of order $r+1$ is equal to $dim(S_{r+1}T^* \otimes F_0) - dim(g'_{r+1})$. The number of new CC of strict order $r+1$ is thus also equal to $y = dim(\rho_1(g'_r)) - dim(g'_{r+1})$ because $dim(S_{r+1}T^* \otimes F_0)$ also disappears by difference. However, unless R_q is FI, we have in general $g'_r \neq im(\sigma_r(\Phi))$ and it thus better to use the systems rather than their symbols.

□

COROLLARY 2.B.5: The system $R'_r \subset J_r(F_0)$ becomes FI with a 2-acyclic or involutive symbol and $R'_{r+1} = \rho_1(R'_r) \subset J_{r+1}(F_0)$ when r is large enough.

Proof: According to the last diagram, we have $g'_{r+1} \subseteq \rho_1(g'_r)$ and g'_{r+1} is

thus defined by more linear equations than $\rho_1(g'_r)$. We are facing a purely algebraic problem over commutative polynomial rings and well known noetherian arguments are showing that $g'_{r+1} = \rho_1(g'_r)$ or, equivalently, $y = 0$ when r is large enough. Chasing in the last diagram, we obtain therefore $R'_{r+1} = \rho_1(R'_r)$ for r large enough and R'_r is a vector bundle because because R_{q+r} is a vector bundle. If we denote by M' the differential module obtained from the system $R'_r \subset J_r(F_0)$ exactly like we have denoted by M the differential module obtained from the system $R_q \subset J_q(E)$, we have the short exact sequence $0 \to M' \to D^m \to M \to 0$. Accordingly, $M' \simeq I \subset D^m$ is a torsion-free differential module and there cannot exist any specialization as an epimorphism $M' \to M'' \to 0$ with $rk_D(M') = rk_D(M'')$ because the kernel should be a torsion differential module and thus should vanish. This comment is strengthening the fact that the knowledge of M and thus of I can only be done through Theorem 1.1. Therefore, if (r,s) are the ones produced by this theorem, then the order of the CC system must be $r+s+1$. We obtain $3+2+1=6$ for the Janet system with systems R'_r of successive dimensions 2, 8, 20, 39, 66, 102, 147 and ask the reader to find $dim(R'_7) = 202$ (Hint: [1]).

□

We are now ready for working out the generating CC $D_1 : F_0 \to F_1$ and start afresh in a simpler way because this new operator is FI (Compare to [5], Proposition 2.9, p 173). However, contrary to what the reader could imagine, it is precisely at this point that troubles may start and the best example is the conformal Killing operator. Indeed, it is known that the order of the generating CC for a system of order q which is FI is equal to $s+1$ if the symbol g_{q+s} becomes 2-acyclic *before* becoming involutive. This fact will be illustrated in a forthcoming motivating example but we recall that the conformal Killing symbol $\hat{g}_1 \subset T^* \otimes T$ is such that $\hat{g}_2$ is 2-acyclic when $n \geq 4$ while $\hat{g}_3 = 0$, a fact explaining why the Weyl operator is of order 2 *but* the Bianchi-type operator is also of order 2, a result still neither known nor even acknowledged today ([4] [9]).

3. Motivating Examples

We now provide three motivating examples in order to illustrate both the usefulness and the *limit* of the previous procedure.

EXAMPLE 3.1: With $m=1, n=3, K=\mathbb{Q}$, we revisit the nice example of Macaulay ([26]) presented in ([3]), namely the homogeneous second order linear system $R_2 \subset J_2(E)$ defined by $\xi_{33} = 0$, $\xi_{13} - \xi_2 = 0$ which is far from being formally integrable. We let the reader prove the strict inclusions $R_2^{(2)} \subset R_2^{(1)} \subset R_2 \subset J_2(E)$ with successive dimensions $6 < 7 < 8 < 10$. The respective symbols are involutive but only the final system $R_2^{(2)}$ is involutive. It follows that the generating CC of the operator defined by R_2 are at most of order 3 but there is indeed only *one* single generating second order CC ([3]). Elementary combinatorics allows to prove the formulas $dim(g_{r+2}) = r+4$, $dim(R_{r+2}) = 4r+8$, $\forall r \geq 0$. We have the short exact sequences:

$$0 \to R_2 \to J_2(E) \to F_0 \to 0, \quad 0 \to 8 \to 10 \to 2 \to 0$$

$$0 \to R_3 \to J_3(E) \to J_1(F_0) \to 0, \quad 0 \to 12 \to 20 \to 8 \to 0$$

and the following commutative diagram:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
0 \to & g_5 & \to & S_5T^* \otimes E & \to & \boxed{S_3T^* \otimes F_0} & \to & T^* \otimes F_1 & \to 0 \\
 & \downarrow & & \downarrow & & \downarrow & & \| & \\
0 \to & R_5 & \to & J_5(E) & \to & J_3(F_0) & \to & J_1(F_1) & \to 0 \\
 & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0 \to & R_4 & \to & J_4(E) & \to & J_2(F_0) & \to & F_1 & \to 0 \\
 & & & \downarrow & & \downarrow & & \downarrow & \\
 & & & 0 & & 0 & & 0 &
\end{array}$$

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
0 \to & 7 & \to & 21 & \to & 20 & \to & 3 & \to 0 \\
 & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0 \to & 20 & \to & 56 & \to & 40 & \to & 4 & \to 0 \\
 & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0 \to & 16 & \to & 35 & \to & 20 & \to & 1 & \to 0 \\
 & & & \downarrow & & \downarrow & & \downarrow & \\
 & & & 0 & & 0 & & 0 &
\end{array}$$

First of all, we have $R'_0 = F_0$, $R'_1 = J_1(F_0)$, $dim(R'_2) = 35 - 16 = 19$, $dim(R'_3) = 56 - 20 = 36$.

It follows that we have successively:

$dim(\rho_1(R'_0)) - dim(R'_1) = 8 - 8 = 0 \Rightarrow$ 0 CC of order 1.

$dim(\rho_1(R'_1)) - dim(R'_2) = 20 - 19 = 1 \Rightarrow$ 1 new CC of order 2.

$dim(\rho_1(R'_2)) - dim(R'_3) = 36 - 36 = 0 \Rightarrow$ 0 new CC of order 3 and so on with:

$$dim(R'_{r+3}) = dim(J_{r+5}(E)) - dim(R_{r+5}) = (r+6)(r+7)(r+8)/6 - (4r+20)$$

$$\begin{aligned} dim(\rho_1(R'_{r+2})) &= dim(J_{r+3}(F_0)) - dim(J_{r+1}(F_1)) \\ &= 2(r+4)(r+5)(r+6)/6 - (r+2)(r+3)(r+4)/6 \end{aligned}$$

and check that $dim(R'_{r+3}) = dim(\rho_1(R'_{r+2})) = (r+4)(r^2+17r+54)/6$, $\forall r \geq 0$.

Then, counting the dimensions, it is easy to check that the two prolongation sequences are exact on the jet level but that the upper symbol sequence is not exact at $S_3T^* \otimes F_0$ with coboundary space of imension $21 - 7 = 14$, cocycle space of dimension $20 - 3 = 17$ and thus cohomology space of dimension $17 - 14 = 3$ that is $dim(R_4 / R_4^{(1)})$ as we check that $7 - 20 + 16 - 3 = 0$. The reader may use the snake theorem to find this result directly through a chase not evident at first sight.

We have then $dim(R_{r+2}^{(1)}) = dim(R_{r+3}) - dim(g_{r+3}) = 3r + 7$ and similarly

$dim\left(g_{r+2}^{(1)}\right)=r+3$ leading to $dim\left(R_{r+2}^{(2)}\right)=dim\left(R_{r+3}^{(1)}\right)-dim\left(g_{r+3}^{(1)}\right)=2r+6$ with $dim\left(g_{r+2}^{(2)}\right)=2$, $\forall r\geq 0$. This result is of course coherent with the fact that the involutive system with the same solutions as R_2 is $R_2^{(2)}$ which is defined by $\xi_{33}=0$, $\xi_{23}=0$, $\xi_{22}=0$, $\xi_{13}-\xi_2=0$.

EXAMPLE 3.2: With $m=1,n=3,q=2,K=\mathbb{Q}$ and the commutative ring $D=K\left[d_1,d_2,d_3\right]$ of PD operators with coefficients in K, we revisit another example of Macaulay ([26]), namely the homogeneous second order formally integrable linear system $R_2\subset J_2\left(E\right)$ defined in operator form by $P\xi\equiv\xi_{33}=0$, $Q\xi\equiv\xi_{23}-\xi_{11}=0$, $R\xi\equiv\xi_{22}=0$ and an epimorphism $R_2\to J_1\left(E\right)\to 0$. As for the systems, we have $dim\left(R_2\right)=7$, $dim\left(R_{r+3}\right)=8$, $\forall r\geq 0$. As for the symbols, we have $dim\left(g_2\right)=3$, $dim\left(g_3\right)=1$, $g_{r+4}=0$, $\forall r\geq 0$. This finite type system has the very particular feature that g_3 is 2-acyclic but not 3-acyclic (thus involutive) with the short exact δ-sequence:

$$3\times 1=3=1\times 3\quad\Rightarrow\quad 0\to\wedge^2T^*\otimes g_3\xrightarrow{\delta}\wedge^3T^*\otimes g_2\to 0$$

and we have the three linearly independent equations:

$$\begin{cases}\xi_{11,123}=\xi_{111,23}+\xi_{112,31}+\xi_{113,12}=\xi_{111,23}\\ \xi_{12,123}=\xi_{112,23}+\xi_{122,31}+\xi_{123,12}=\xi_{111,12}\\ \xi_{13,123}=\xi_{113,23}+\xi_{123,31}+\xi_{133,12}=\xi_{111,31}\end{cases}$$

Collecting these results, we get the two following commutative and exact diagrams:

$$\begin{array}{ccccccccc}
 & & & & 0 & 0 & 0 & & \\
 & & & & \downarrow & \downarrow & \downarrow & & \\
 & & 0 & \to & S_4T^*\otimes E & \to & S_2T^*\otimes F_0 & \to & F_1 & \to 0\\
 & & \downarrow & & \downarrow & & \downarrow & & \| & \\
0 & \to & R_4 & \to & J_4\left(E\right) & \to & J_2\left(F_0\right) & \to & F_1 & \to 0\\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0 & \to & R_3 & \to & J_3\left(E\right) & \to & J_1\left(F_0\right) & \to & 0 & \\
 & & \downarrow & & \downarrow & & \downarrow & & & \\
 & & 0 & & 0 & & 0 & & &
\end{array}$$

$$\begin{array}{ccccccccc}
 & & & & 0 & 0 & 0 & & \\
 & & & & \downarrow & \downarrow & \downarrow & & \\
 & & 0 & \to & S_5T^*\otimes E & \to & S_3T^*\otimes F_0 & \to & T^*\otimes F_1 & \to 0\\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0 & \to & R_5 & \to & J_5\left(E\right) & \to & J_3\left(F_0\right) & \to & J_1\left(F_1\right) & \to 0\\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0 & \to & R_4 & \to & J_4\left(E\right) & \to & J_2\left(F_0\right) & \to & F_1 & \to 0\\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
 & & 0 & & 0 & & 0 & & 0 &
\end{array}$$

We obtain from these diagrams $R_1'=J_1\left(F_0\right)\Rightarrow g_1'=T^*\otimes F_0$, $\rho_1\left(R_1'\right)=J_2\left(F_0\right)\Rightarrow R_2'\subset\rho_1\left(R_1'\right)$ with a strict inclusion because $27<30$ and

we have at least $30-27=3$ generating second order CC. However, from the second diagram, we obtain $dim\left(\rho_1\left(R_2'\right)\right)=60-12=48=56-8=dim\left(R_3'\right)$ and thus $R_3'=\rho_1\left(R_2'\right)$, a result showing that there are no new generating CC of order 3.

As $dim\left(E\right)=1$, we have $S_qT^*\otimes E\simeq S_qT^*$ and the commutative diagram of δ-sequences:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & S_6T^* & \to & T^*\otimes S_5T^* & \to & \wedge^2T^*\otimes S_4T^* & \to & \wedge^3T^*\otimes S_3T^* & \to 0 \\
 & & \| & & \| & & \| & & \downarrow & \\
0 & \to & g_4' & \to & T^*\otimes g_3' & \to & \wedge^2T^*\otimes g_2' & \to & \wedge^3T^*\otimes g_1' & \to 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
 & & 0 & & 0 & & 0 & & 0 &
\end{array}$$

Using the fact that the upper sequence is known to be exact and $dim\left(g_1'\right)=9<10=dim\left(S_3T^*\right)$, an easy chase proves that the lower sequence *cannot* be exact and thus g_2' *cannot* be 2-acyclic.

The generating CC of $\mathcal{D}_1$ is thus a second order operator $\mathcal{D}_2:F_1\to F_2$ where F_2 is defined by the long exact prolongation sequence:

$$0\to R_6\to J_6\left(E\right)\to J_4\left(F_0\right)\to J_2\left(F_1\right)\to F_2\to 0$$

or by the long exact symbol sequence (*by chance* if one refers to the previous example!):

$$0\to S_6T^*\otimes E\to S_4T^*\otimes F_0\to S_2T^*\otimes F_1\to F_2\to 0$$

showing that $dim\left(F_2\right)=dim\left(S_6T^*\right)-3dim\left(S_4T^*\right)+3dim\left(S_2T^*\right)=28-45+18=1$ in a coherent way with ([9] [14]).

We have thus obtained the following formally exact differential sequence which is nevertheless *not* a Janet sequence because R_2 is FI but *not* involutive as g_2 is finite type with $g_4=0$:

$$0\to\Theta\to E\underset{2}{\overset{\mathcal{D}}{\to}}F_0\underset{2}{\overset{\mathcal{D}_1}{\to}}F_1\underset{2}{\overset{\mathcal{D}_2}{\to}}F_2\to 0$$

$$0\to\Theta\to 1\to 3\underset{2}{\to}3\to 1\to 0$$

$$0\to D\to D^3\to D^3\to D\overset{p}{\to}M\to 0$$

Surprisingly, the situation is *even quite worst* if we start with $R_3\subset J_3\left(E\right)$ which has nevertheless a 2-acyclic symbol g_3 which is *not* 3-acyclic (thus involutive because $n=3$). Indeed, we know from the second section or by repeating the previous procedure for this new third order operator $\mathcal{D}$ that the generating CC are described by a *first order* operator $\mathcal{D}_1$. However, the symbol of this operator is only 1-acyclic but *not* 2-acyclic (exercise). Hence, one can prove that the corresponding CC are described by a new second order operator $\mathcal{D}_2$ which is involutive… by chance, giving rise to a Janet sequence with first

order operators as follows $\mathcal{D}_3, \mathcal{D}_4, \mathcal{D}_5$ ([9], p 119-125):

$$0 \to \Theta \to 1 \underset{3}{\overset{\mathcal{D}}{\to}} 12 \underset{1}{\overset{\mathcal{D}_1}{\to}} 21 \underset{2}{\overset{\mathcal{D}_2}{\to}} 46 \underset{1}{\overset{\mathcal{D}_3}{\to}} 72 \underset{1}{\overset{\mathcal{D}_4}{\to}} 48 \underset{1}{\overset{\mathcal{D}_5}{\to}} 12 \to 0$$

One could also finally use the involutive system $R_4 \subset J_4(E)$ in order to construct the canonical Janet sequence and consider the first order involutive system $R_5 \subset J_1(R_4)$ in order to obtain the canonical Spencer sequence with $C_r = \wedge^r T^* \otimes R_4$ and dimensions $(8,24,24,8)$:

$$0 \to \Theta \overset{j_4}{\to} C_0 \underset{1}{\overset{D_1}{\to}} C_1 \underset{1}{\overset{D_2}{\to}} C_2 \underset{1}{\overset{D_3}{\to}} C_3 \to 0$$

To recapitulate, this example clearly proves that the differential sequences obtained largely depend on whether we use R_2, R_3 or R_4 but also whether we look for a sequence of Janet or Spencer type.

We invite the reader to treat similarly the example $\xi_{33} - \xi_{11} = 0$, $\xi_{23} = 0$, $\xi_{22} - \xi_{11} = 0$.

EXAMPLE 3.3: In our opinion, the best striking use of acyclicity is the construction of differential sequences for the Killing and conformal Killing operators which are both defined over the ground differential field $K = \mathbb{Q}$ for the Minkowski metric in dimension 4 or the Euclidean metric in dimension 5. We have indeed ([9] [20]):

$$0 \to \Theta \to 4 \underset{1}{\overset{\mathcal{D}}{\to}} 10 \underset{2}{\overset{\mathcal{D}_1}{\to}} 20 \underset{1}{\overset{\mathcal{D}_2}{\to}} 20 \underset{1}{\overset{\mathcal{D}_3}{\to}} 6 \to 0$$

with $E = T, F_0 = S_2 T^*$ and, successively, the Killing, Riemann and Bianchi operators acting on the left of column vectors. The differential module counterpart over $D = K[d]$ is the resolution of the differential Killing module M:

$$0 \to D^6 \underset{1}{\overset{\mathcal{D}_3}{\to}} D^{20} \underset{1}{\overset{\mathcal{D}_2}{\to}} D^{20} \underset{2}{\overset{\mathcal{D}_1}{\to}} D^{10} \underset{1}{\overset{\mathcal{D}}{\to}} D^4 \overset{p}{\to} M \to 0$$

with the same operators as before but acting now on the right of row vectors by composition.

The conformal situation for $n = 4$ is quite unexpected with a second order Bianchi-type operator:

$$0 \to \Theta \to 4 \underset{1}{\overset{\mathcal{D}}{\to}} 9 \underset{2}{\overset{\mathcal{D}_1}{\to}} 10 \underset{2}{\overset{\mathcal{D}_2}{\to}} 9 \underset{1}{\overset{\mathcal{D}_3}{\to}} 4 \to 0$$

$$0 \to D^4 \underset{1}{\overset{\mathcal{D}_3}{\to}} D^9 \underset{2}{\overset{\mathcal{D}_2}{\to}} D^{10} \underset{2}{\overset{\mathcal{D}_1}{\to}} D^9 \underset{1}{\overset{\mathcal{D}}{\to}} D^4 \overset{p}{\to} M \to 0$$

The conformal situation for $n = 5$ is even quite different with the conformal differential sequence:

$$0 \to \Theta \to 5 \underset{1}{\overset{\mathcal{D}}{\to}} 14 \underset{2}{\overset{\mathcal{D}_1}{\to}} 35 \underset{1}{\overset{\mathcal{D}_2}{\to}} 35 \underset{2}{\overset{\mathcal{D}_3}{\to}} 14 \underset{1}{\overset{\mathcal{D}_4}{\to}} 5 \to 0$$

Though these results and "jumps" highly depend on acyclicity, in particular the fact that the conformal symbol $\hat{g}_2$ is 2-acyclic for $n = 4$ but 3-acyclic for $n \geq 5$, and have been confirmed by computer algebra, they are still neither known nor acknowledged ([4] [9]).

4. Applications

Considering the *classical Killing* operator $\mathcal{D}:\xi\to\mathcal{L}(\xi)\omega=\Omega\in S_2T^*=F_0$ where $\mathcal{L}(\xi)$ is the Lie derivative with respect to ξ and $\omega\in S_2T^*$ is a non-degenerate metric with $det(\omega)\neq 0$. Accordingly, it is a lie operator with $\mathcal{D}\xi=0,\mathcal{D}\eta=0\Rightarrow\mathcal{D}[\xi,\eta]=0$ and we denote simply by $\Theta\subset T$ the set of solutions with $[\Theta,\Theta]\subset\Theta$. Now, as we have explained many times, the main problem is to describe the CC of $\mathcal{D}\xi=\Omega\in F_0$ in the form $\mathcal{D}_1\Omega=0$ by introducing the so-called *Riemann* operator $\mathcal{D}_1:F_0\to F_1$. We advise the reader to follow closely the next lines and to imagine why it will not be possible to repeat them for studying the *conformal Killing* operator. Introducing the well known Levi-Civita isomorphism $j_1(\omega)=(\omega,\partial_x\omega)\simeq(\omega,\gamma)$ by defining the Christoffel symbols $\gamma_{ij}^k=\frac{1}{2}\omega^{kr}(\partial_i\omega_{rj}+\partial_j\omega_{ir}-\partial_r\omega_{ij})$ where (ω^{rs}) is the inverse matrix of (ω_{ij}) and the *formal Lie derivative*, we get the second order system $R_2\subset J_2(T)$:

$$\begin{cases}\Omega_{ij}\equiv\left(L(\xi_1)\omega\right)_{ij}=\omega_{rj}(x)\xi_i^r+\omega_{ir}(x)\xi_j^r+\xi^r\partial_r\omega_{ij}(x)=0\\ \Gamma_{ij}^k\equiv\left(L(\xi_2)\gamma\right)_{ij}^k=\xi_{ij}^k+\gamma_{rj}^k(x)\xi_i^r+\gamma_{ir}^k(x)\xi_j^r+\gamma_{ir}^k(x)\xi_j^r-\gamma_{ij}^r(x)\xi_r^k+\xi^r\partial_r\gamma_{ij}^k(x)=0\end{cases}$$

with sections $\xi_2:x\to(\xi^k(x),\xi_i^k(x),\xi_{ij}^k(x))$ transforming like $j_2(\xi):x\to(\xi^k(x),\partial_i\xi^k(x),\partial_{ij}\xi^k(x))$. The system $R_1\subset J_1(T)$ has a symbol $g_1\simeq\wedge^2T^*\subset T^*\otimes T$ depending only on ω with $dim(g_1)=n(n-1)/2$ and is finite type because its first prolongation is $g_2=0$. It cannot be thus involutive and we need to use one additional prolongation. Indeed, using one of the main results to be found in ([1] [5] [9] [10] [14]), we know that, when R_1 is FI, then the CC of $\mathcal{D}$ are of order $s+1$ where s is the number of prolongations needed in order to get a 2-acyclic symbol, that is $s=1$ in the present situation, a result that should lead to CC of order 2 if R_1 were FI. However, it is known that R_2 is FI, thus involutive, *if and only if* ω has constant Riemannian curvature, a result first found by L.P. Eisenhart in 1926 which is only a particular example of the *Vessiot structure equations* discovered b E. Vessiot in 1903 ([27]), though in a quite different setting (See [1] [5] [9] [14] for an explicit modern proof and compare to the references ([22] [23]) of ([6]).

We may introduce the (formal) linearization $\Gamma\in S_2T^*\otimes T$ of the Christoffel symbols by linearizing the relations $\omega_{kr}\gamma_{ij}^k=\frac{1}{2}(\partial_i\omega_{rj}+\partial_j\omega_{ir}-\partial_r\omega_{ij})$ in such a way that $\Omega=0\Rightarrow\Gamma=0$ with:

$$\omega_{kr}\Gamma_{ij}^k=\frac{1}{2}(d_i\Omega_{rj}+d_j\Omega_{ir}-d_r\Omega_{ij})-\gamma_{ij}^k\Omega_{kr}$$

We may also introduce the Riemann tensor $\rho_{l,ij}^k$ and its (formal) linearization:

$$R_{l,ij}^k\equiv\left(L(\xi_1)\rho\right)_{l,ij}^k=-\rho_{l,ij}^s\xi_s^k+\rho_{s,ij}^k\xi_l^s+\rho_{l,sj}^k\xi_i^s+\rho_{l,is}^k\xi_j^s+\xi^r\partial_r\rho_{l,ij}^k$$

in order to obtain the Ricci tensor $\rho_{ij}=\rho_{i,rj}^r=\rho_{ji}$ and its linearization:

$$R_{ij}=R_{i,rj}^r=\rho_{rj}\xi_i^r+\rho_{ir}\xi_j^r+\xi^r\partial_r\rho_{ij}=R_{ji}$$

allowing to introduce the Einstein tensor $\epsilon_{ij} = \rho_{ij} - \frac{1}{2}\omega_{ij}\omega^{rs}\rho_{rs} = \rho_{ij} - \frac{1}{2}\omega_{ij}\rho$ with linearization:

$$E_{ij} \equiv \left(L\left(\xi_1\right)\epsilon\right)_{ij} = R_{ij} - \frac{1}{2}\omega_{ij}\omega^{rs}R_{rs} - \frac{1}{2}\rho\Omega_{ij} + \frac{1}{2}\omega_{ij}\omega^{ru}\omega^{sv}\rho_{rs}\Omega_{uv}$$

and we must notice (*care*) that the linearization of $\rho = \omega^{rs}\rho_{rs}$ is $R = \omega^{rs}R_{rs} - \rho_{rs}\omega^{ru}\omega^{sv}\Omega_{uv}$.

These formulas become particularly simple when ω is a solution of Einstein equations in vacuum, that is when $\varepsilon_{ij} = 0 \Leftrightarrow \rho_{ij} = 0 \Rightarrow \rho = 0$.

LEMMA 4.1: When $n = 4$ and the fixed euclidean metric for simplicity, we have the useful formula:

$$E_{00} = -\left(R_{12,12} + R_{13,13} + R_{23,23}\right)$$

Proof: We have $2E_{00} = 2R_{00} - \left(R_{00} + R_{11} + R_{22} + R_{33}\right) = R_{00} - \left(R_{11} + R_{22} + R_{33}\right)$ and $R_{00} = \left(R_{10,10} + R_{20,20} + R_{30,30}\right)$. However, we have also:

$$R_{11} = R_{01,01} + R_{21,21} + R_{31,31}$$

$$R_{22} = R_{02,02} + R_{12,12} + R_{32,32}$$

$$R_{33} = R_{03,03} + R_{13,13} + R_{23,23}$$

Summing, we obtain $\left(R_{11} + R_{22} + R_{33}\right) = \left(R_{01,01} + R_{02,02} + R_{03,03}\right) + 2\left(R_{12,12} + R_{13,13} + R_{23,23}\right)$. It follows that $E_{00} = -\left(R_{12,12} + R_{13,13} + R_{23,23}\right)$ and the three other E_{ii} are obtained by circular permutations of $(0,1,2,3)$. We let the reader treat the general situation as an exercise.

□

A) MINKOWSKI METRIC:

We have considered this situation in many books or papers and refer the reader to our arXiv page or to the recent references ([22] [28]). All the operators are first order between the vector bundles $E = T$, $F_0 = T^* \otimes T / g_1 \simeq S_2T^*$, $F_1 = H^2\left(g_1\right)$, $F_2 = H^3\left(g_1\right)$, $F_3 = H^4\left(g_1\right)$ that are only depending on g_1 with dimensions 4, 10, 20, 20, 6 when $n = 4$ and Euler-Poincaré characteristic $rk_D\left(M\right) = 4 - 10 + 20 - 20 + 6 = 0$. The case of an arbitrary *n*, provided in ([20]), depends on various chases in commutative diagrams that will be exhibited later on for comparing the respective dimensions. This is *not* a Janet sequence because R_1 is FI but g_1 is *not* involutive.

B) SCHWARZSCHILD METRIC:

With the standard Boyer-Lindquist local coordinates $\left(x^0 = t, x^1 = r, x^2 = \theta, x^3 = \phi\right)$ and a constant parameter *m*, we may introduce the field of constants $k = \mathbb{Q}\left(m\right)$ and all the systems or differential modules considered in the sequel will be defined over the ground differential field $K = k\left(t, r, \sin\left(\theta\right), \cos\left(\theta\right)\right)$ with differential structure obtained by setting $\partial_2 \sin\left(\theta\right) = \cos\left(\theta\right)$, $\partial_2 \cos\left(\theta\right) = -\sin\left(\theta\right)$ together with $\sin^2\left(\theta\right) + \cos^2\left(\theta\right) = 1$ instead of using the so-called "rational coordinates" ([23]). With speed of light $c = 1$ and $A = 1 - \frac{m}{r}$, we shall introduce the diagonal Schwarzschild metric $\omega = \left(A\left(r\right), -1/A\left(r\right), -r^2, -r^2\sin^2\left(\theta\right)\right)$ with $det\left(\omega\right) = -r^4\sin^2\left(\theta\right)$. Following

closely the motivating examples already presented, our challenge is to prove that the purely mathematical formal study of the corresponding Killing system $R_1 \subset J_1(T)$ can be achieved as a simple exercise of formal integrability, with no extra physical technical tool, contrary to ([6] [7] [8]). As the computations will be explicitly done, the numbers of CC obtained will bring serious doubts about the validity of the results obtained in the above references, later confirmed with the K metric. First of all we obtain easily the following 10 first order Killing equations $mod(\Omega)$:

$$R_1 \subset J_1(T) \begin{cases} \Omega_{33} \rightarrow \boxed{\xi^3_3} + \frac{1}{r}\xi^1 + \cot(\theta)\xi^2 = 0 \\ \Omega_{23} \rightarrow \boxed{\xi^2_3} + \sin^2(\theta)\xi^3_2 = 0 \\ \Omega_{13} \rightarrow \boxed{\xi^1_3} + Ar^2\sin^2(\theta)\xi^3_1 = 0 \\ \Omega_{03} \rightarrow \boxed{\xi^0_3} - \frac{r^2}{A}\sin^2(\theta)\xi^3_0 = 0 \\ \Omega_{22} \rightarrow \boxed{\xi^2_2} + \frac{1}{r}\xi^1 = 0 \\ \Omega_{12} \rightarrow \boxed{\xi^1_2} + Ar^2\xi^2_1 = 0 \\ \Omega_{02} \rightarrow \boxed{\xi^0_2} - \frac{r^2}{A}\xi^2_0 = 0 \\ \Omega_{11} \rightarrow \boxed{\xi^1_1} - \frac{m}{2Ar^2}\xi^1 = 0 \\ \Omega_{01} \rightarrow \boxed{\xi^1_0} - A^2\xi^0_1 = 0 \\ \Omega_{00} \rightarrow \boxed{\xi^0_0} + \frac{m}{2Ar^2}\xi^1 = 0 \end{cases}$$

where we have framed the leading jets.

This is a finite type system because we get $\Gamma^k_{ij} \equiv \xi^k_{ij} + \cdots = 0$ with *only one prolongation*!

The only 9 non-zero Christoffel symbols on 40 are:

$$\begin{array}{lll} \gamma^1_{00} = \frac{mA}{2r^2}, & \gamma^0_{01} = \frac{m}{2Ar^2}, & \gamma^1_{11} = -\frac{m}{2Ar^2} \\ \gamma^2_{12} = \frac{1}{r}, & \gamma^3_{13} = \frac{1}{r}, & \gamma^1_{22} = -Ar, \\ \gamma^3_{23} = \cot(\theta), & \gamma^1_{33} = -Ar\sin^2(\theta), & \gamma^2_{33} = -\sin(\theta)\cos(\theta) \end{array}$$

We obtain for example:

$$\Gamma^1_{22} \equiv \xi^1_{22} + \left(1 - \frac{3m}{2r}\right)\xi^1 = 0,$$

$$\Gamma^1_{33} \equiv \xi^1_{33} + \sin(\theta)\cos(\theta)\xi^1_2 + \left(1 - \frac{3m}{2r}\right)\sin^2(\theta)\xi^1 = 0$$

$$\Rightarrow \quad \Gamma^1_{33} - \sin^2(\theta)\Gamma^1_{22} \equiv \xi^1_{33} + \sin(\theta)\cos(\theta)\xi^1_2 - \sin^2(\theta)\xi^1_{22} = 0$$

after only one prolongation (care).

Then, using r as a summation index, we shall see that we have *in general* for

the linearization of the Riemann and Ricci tensors:

$$R_{kl,ij} \equiv \rho_{rl,ij}\xi^r_k + \rho_{kr,ij}\xi^r_l + \rho_{kl,rj}\xi^r_i + \rho_{kl,ir}\xi^r_j + \xi^r\partial_r\rho_{kl,ij} \neq 0$$

$$R_{ij} \equiv \rho_{rj}\xi^r_i + \rho_{ir}\xi^r_j + \xi^r\partial_r\rho_{ij} \neq 0$$

The only 6 non-zero components of the Riemann tensor are:

$$\boxed{\begin{array}{lll} \rho_{01,01} = +\dfrac{m}{r^3}, & \rho_{02,02} = -\dfrac{mA}{2r}, & \rho_{03,03} = -\dfrac{mA\sin^2(\theta)}{2r} \\ \rho_{12,12} = +\dfrac{m}{2Ar}, & \rho_{13,13} = +\dfrac{m\sin^2(\theta)}{2Ar}, & \rho_{23,23} = -mr\sin^2(\theta) \end{array}}$$

but we must not forget hat we have indeed $\rho_{ij} = 0$ for the 10 components of the Ricci tensor, in particular $\rho_{ii} = 0$ for the diagonal components with $i = 0,1,2,3$. We have in particular:

$$\begin{aligned} \rho_{22} &= \rho^0_{2,02} + \rho^1_{2,12} + \rho^3_{2,32} \\ &= \frac{1}{A}\rho_{02,02} - A\rho_{12,12} - \frac{1}{r^2\sin^2(\theta)}\rho_{23,23} \\ &= -\frac{m}{2r} - \frac{m}{2r} + \frac{m}{r} = 0 \end{aligned}$$

We also obtain $mod(\Omega)$:

$$R_{01,01} \equiv 2\rho_{01,01}\left(\xi^0_0 + \xi^1_1\right) + \xi^r\partial_r\left(\rho_{01,01}\right) = \xi^1\partial_1\rho_{01,01}$$

$$\Rightarrow \boxed{R_{01,01} \equiv -\frac{3m}{r^4}\xi^1 = 0 \Rightarrow \xi^1 = 0}$$

and similarly:

$$R_{01,02} \equiv \rho_{01,01}\xi^1_2 + \rho_{02,02}\xi^2_1 + \xi^r\partial_r\rho_{01,02} = \frac{3m}{2r^3}\xi^1_2 = 0 \Rightarrow \xi^1_2 = 0$$

$$R_{02,12} \equiv \rho_{12,12}\xi^1_0 + \rho_{02,02}\xi^0_1 + \xi^r\partial_r\rho_{02,12} = \frac{m}{2rA}\left(\xi^1_0 - A^2\xi^0_1\right) = 0$$

$$\boxed{\begin{aligned} R_{23,12} &\equiv \rho_{21,12}\xi^1_3 + \rho_{23,32}\xi^3_1 + \xi^r\partial_r\rho_{23,12} \\ &= -\frac{m}{2rA}\xi^1_3 + mr\sin^2(\theta)\xi^3_1 \\ &= -\frac{3m}{2rA}\xi^1_3 = 0 \end{aligned}}$$

$$\Rightarrow R_{01,03} = \frac{3m}{2r^3}\xi^1_3 = -\frac{A}{r^2}R_{23,12}$$

$$\rho_{01,23} = 0 \Rightarrow R_{01,23} = \rho_{01,23}\left(\xi^0_0 + \xi^1_1 + \xi^2_2 + \xi^3_3\right) + \xi^r\partial_r\rho_{01,23} = 0$$

and so on, in order to avoid using computer algebra. However, the main consequence of this remark is to explain the existence of the 15 second order CC. Indeed, denoting by "~" a linear proportional dependence $mod(\Omega)$, we have the successive three cases:

$(R_{00}, R_{11}, R_{22}, R_{33})$	$R_{01,01} \sim R_{02,02} \sim R_{03,03} \sim R_{12,12} \sim R_{13,13} \sim R_{23,23} \to \xi^1 = 0$
(R_{12})	$R_{01,02} \sim R_{13,23} \to \xi^1_2 = 0$
(R_{13})	$R_{01,03} \sim R_{12,23} \to \xi^1_3 = 0$
(R_{02})	$R_{01,12} \sim R_{03,23} \to \xi^0_2 = 0$
(R_{03})	$R_{01,13} \sim R_{02,23} \to \xi^0_3 = 0$
(R_{01}, R_{23})	$R_{01,23} \to 0, R_{02,03} \to 0, R_{02,12} \to 0, R_{02,13} \to 0, R_{03,13} \to 0, R_{12,13} \to 0$

as a way to obtain the 5 equalities to zero on the right and thus a total of $20-5=15$ second order CC obtained by elimination. However, the present partition $15=5+4+6$ is quite different from the partition $15=10+5$ used by the authors quoted in the Introduction which is obtained by taking into account the vanishing assumption of the 10 components of the Ricci tensor. As such a result questions once more the mathematical foundations of general relativity, in particular the existence of gravitational waves, we provide a few additional technical comments.

The main point is a tricky formula which is not evident at all. Indeed, using the well known properties of the Lie derivative, we have the following geometric objects (*not necessarily tensors*) and their linearizations (*generally tensors*):

$$\omega_{ij} \to \Omega_{ij} \in S_2T^*,\ \gamma^k_{ij} \to \Gamma^k_{ij} \in S_2T^* \otimes T,$$

$$\rho_{kl,ij} \to R_{kl,ij} \in \wedge^2 T^* \otimes T^* \otimes T,\ \beta_{kl,ijr} \to B_{kl,ijr} \in \wedge^3 T^* \otimes T^* \otimes T$$

Then, using r as a summation index, we shall see that we have *in general*:

$$R_{kl,ij} \equiv \rho_{rl,ij}\xi^r_k + \rho_{kr,ij}\xi^r_l + \rho_{kl,rj}\xi^r_i + \rho_{kl,ir}\xi^r_j + \xi^r\partial_r\rho_{kl,ij} \neq 0$$

$$\rho_{kl,ij} = \omega_{kr}\rho^r_{l,ij} \Rightarrow R_{kl,ij} = \omega_{kr}R^r_{l,ij} + \rho^r_{l,ij}\Omega_{kr} \Rightarrow \omega^{rs}R_{ri,sj} = R_{ij} + \omega^{rs}\rho^t_{i,rj}\Omega_{st}$$

$$\rho_{ij} = \rho^r_{i,rj} \Rightarrow R_{ij} = R^r_{i,rj} \neq \omega^{rs}R_{ri,sj}$$

We prove these results using local coordinates and the formal Lie derivative obtained while replacing $j_1(\xi)$ by ξ_1 (See [1] [5] [9] [14] for details). First of all, from the tensorial property of the Riemann tensor and the Killing equations $\Omega_{us} = \omega_{ku}\xi^k_s + \omega_{ks}\xi^k_u + \xi^r\partial_r\omega_{us}$, we have:

$$R^k_{l,ij} \equiv \left(L(\xi_1)\rho\right)^k_{l,ij} = -\rho^s_{l,ij}\xi^k_s + \rho^k_{s,ij}\xi^s_l + \rho^k_{l,sj}\xi^s_i + \rho^k_{l,is}\xi^s_j + \xi^r\partial_r\rho^k_{l,ij}$$

$$\omega_{ku}\left(-\rho^s_{v,ij}\xi^k_s + \xi^r\partial_r\rho^s_{v,ij}\right) = \rho^s_{v,ij}\omega_{ks}\xi^k_u + \left(\xi^r\partial_r\omega_{us}\right)\rho^s_{v,ij} + \omega_{ku}\xi^r\partial_r\rho^k_{v,ij} - \rho^s_{v,ij}\Omega_{su}$$

$$= \rho_{sv,ij}\xi^s_u + \xi^r\partial_r\rho_{uv,ij} - \rho^s_{v,ij}\Omega_{su}$$

and thus $\omega_{ku}R^k_{v,ij} = R_{uv,ij} - \rho^s_{v,ij}\Omega_{su}$.

$$\rho^1_{1,11} = 0 \Rightarrow \rho_{11} = \rho^0_{1,01} + \rho^2_{1,21} + \rho^3_{1,31} = \frac{1}{A}\rho_{01,01} - \frac{1}{r^2}\rho_{12,12} - \frac{1}{r^2\sin^2(\theta)}\rho_{13,13} = 0$$

We have for example, *in this particular case*:

$$\rho^0_{1,01}=\frac{1}{A}\rho_{01,01}=\frac{m}{Ar^3}$$

$$\Rightarrow R^0_{1,01}=\left(\mathcal{L}(\xi)\rho\right)^0_{1,01}=2\rho^0_{1,01}\xi^1_1+\xi^1\partial_1\rho^0_{1,01}=\frac{2m}{Ar^3}\xi^1_1+\xi^1\partial_1\left(\frac{m}{Ar^3}\right)$$

The only use of $R_{01,01}$ is allowing to get $\xi_1=0$ in the previous list, but we have also *exactly*:

$$\omega^{rs}R_{r1,s2}=\frac{1}{A}R_{01,02}-\frac{1}{r^2\sin^2(\theta)}R_{31,32}=-\frac{m}{2r^3}\Omega_{12}=R_{12}+\omega^{11}\rho^2_{1,12}\Omega_{12}\Rightarrow R_{12}=0$$

The use of $R_{01,02}$ or $R_{13,23}$ is allowing to get $\xi^1_2=0$ in the previous list with:

$$R^0_{1,02}=-\frac{3m}{2r^3}\xi_{1,2}+\frac{m}{2r^3}\Omega_{12},\ R^3_{1,32}=+\frac{3m}{2r^3}\xi_{1,2}-\frac{m}{r^3}\Omega_{12}$$

and thus also *exactly*:

$$\omega^{rs}\rho^t_{1,r2}\Omega_{st}=-\omega^{11}\omega^{22}\rho_{12,12}\Omega_{12}=-\frac{m}{2r^3}\Omega_{12}\Rightarrow R_{12}=R^0_{1,02}+R^3_{1,32}+\frac{m}{2r^3}\Omega_{12}=0$$

It follows that the 4 central second order CC of the list successively amounts to $R_{12}=0,R_{13}=0,R_{02}=0,R_{03}=0$, a result breaking the intrinsic/coordinate-free interpretation of the 10 Einstein equations and the situation is even worst for the other components of the Ricci tensor. Indeed, R_{01} and R_{23} only depend on the vanishing of $R_{02,12},R_{03,13}$ and $R_{02,03},R_{12,13}$ among the bottom CC of the list, while the diagonal terms $R_{00},R_{11},R_{22},R_{33}$ only depend, *as we just saw*, on the 6 non zero components of the Riemann tensor. We have thus obtained the totally unusual partition $10=4+4+2$ along the successive blocks of the former list with:

$$\{R_{ij}\}=\{R_{00},R_{11},R_{22},R_{33}\}+\{R_{12},R_{13},R_{02},R_{03}\}+\{R_{01},R_{23}\}$$

Finally, we notice that $R_{01,23}=0,R_{02,31}=0\Rightarrow R_{03,12}=0$ from the identity in $\wedge^3T^*\otimes T^*$:

$$R_{01,23}+R_{02,31}+R_{03,12}=0$$

and there is no way to have two identical indices in the first jets appearing through the (formal) Lie derivative just described. As for the third order CC, setting $\xi^1_1=\frac{A'}{2A}\xi^1\in j_2(\Omega)$, we have at least the first prolongations of the previous second order CC to which we have to add the three new generating ones:

$$\boxed{d_1\xi^1-\xi^1_1=0,d_2\xi^1-\xi^1_2=0,d_3\xi^1-\xi^1_3=0}$$

provided by the Spencer operator, leading to the crossed terms $d_i\xi^1_j-d_j\xi^1_i=0$ for $i,j=1,2,3$ because the Spencer operator is not FI.

Setting now $\xi^1=U,\xi^1_2=V_2,\xi^1_3=V_3,\xi^0_2=W_2,\xi^0_3=W_3$ with $(U,V,W)\in j_2(\Omega)$, we have to look for the CC of the system $R^{(1)}_1=R_1$ already presented, then the system $R^{(2)}_1$ with $dim\left(R^{(2)}_1\right)=5$ and finally $R^{(3)}_1$ with $dim\left(R^{(3)}_1\right)=4$ which

is formally integrable but not involutive because it is of finite type. Beside the only zero order equation $\xi^1 = U$, we have the following 15 first order ones:

$$\begin{aligned}
&\xi^0_0 = -\frac{A'}{2A}U,\ \xi^0_1 = \frac{1}{A^2}d_0U,\ \xi^0_2 = W_2,\ \xi^0_3 = W_3,\\
&\xi^1_0 = d_0U,\ \xi^1_1 = \frac{A'}{2A}U,\ \xi^1_2 = V_2,\ \xi^1_3 = V_3,\\
&\xi^2_0 = \frac{A}{r^2}W_2,\ \xi^2_1 = -\frac{1}{Ar^2}V_2,\ \xi^3_0 = \frac{A}{r^2\sin^2(\theta)}W_3,\ \xi^3_1 = -\frac{1}{Ar^2\sin^2(\theta)}V_3,\\
&\xi^2_2 = -\frac{1}{r}U,\ \xi^2_3 + \sin^2(\theta)\xi^3_2 = 0,\ \xi^3_3 + \cot(\theta)\xi^2 = -\frac{1}{r}U
\end{aligned}$$

Among the CC we *must* have $d_2V_3 - d_3V_2 = 0$ which is among the differential consequences of the Spencer operator as we saw but we *must* also have $d_2W_3 - d_3W_2 = 0$ and both seem to be new third order CC, together with the CC obtained by eliminating ξ^2 and ξ^3 from the three last equations *after two prolongations* as in ([23]):

$$d_3V_3 + \sin(\theta)\cos(\theta)V_2 + \sin^2(\theta)d_2V_2 + 2\sin^2(\theta)U = 0$$

However, *things are not so simple*, even if we have in mind that $(V,W) \in j_2(\Omega)$, because the central sign in the previous formula is opposite to the sign found *after one prolongation* in the formula:

$$\xi^1_{33} + \sin(\theta)\cos(\theta)\xi^1_2 - \sin^2(\theta)\xi^1_{22} = 0$$

and it is at this moment that we need introduce new differential geometric methods!

First of all, we have:

$$\rho^k_{l,ij} = \partial_i\gamma^k_{lj} - \partial_j\gamma^k_{li} + \gamma^r_{lj}\gamma^k_{ri} - \gamma^r_{li}\gamma^k_{rj}$$

and thus, because $\Gamma \in S_2T^* \otimes T$ is a tensor::

$$\begin{aligned}
R^k_{l,ij} &= d_i\Gamma^k_{lj} - d_j\Gamma^k_{li} + \gamma^r_{lj}\Gamma^k_{ri} - \gamma^r_{li}\Gamma^k_{rj} + \gamma^k_{ri}\Gamma^r_{lj} - \gamma^k_{rj}\Gamma^r_{li}\\
&= \left(d_i\Gamma^k_{lj} - \gamma^r_{li}\Gamma^k_{rj} + \gamma^k_{ri}\Gamma^r_{lj}\right) - \left(d_j\Gamma^k_{li} - \gamma^r_{lj}\Gamma^k_{ri} + \gamma^k_{rj}\Gamma^r_{li}\right)\\
&= \left(d_i\Gamma^k_{lj} - \gamma^r_{li}\Gamma^k_{rj} - \gamma^r_{ji}\Gamma^k_{lr} + \gamma^k_{ri}\Gamma^r_{lj}\right) - \left(d_j\Gamma^k_{li} - \gamma^r_{lj}\Gamma^k_{ri} - \gamma^r_{ij}\Gamma^k_{lr} + \gamma^k_{rj}\Gamma^r_{li}\right)\\
&= \nabla_i\Gamma^k_{lj} - \nabla_j\Gamma^k_{li}
\end{aligned}$$

by introducing the covariant derivative ∇. We recall that $\nabla_r\omega_{ij} = 0, \forall r,i,j$ or, equivalently, that $(id,-\gamma): \xi \in T \to \left(\xi^k, \xi^k_i = -\gamma^k_{ir}\xi^r\right) \in R_1$ is a R_1-connection with $\omega_{sj}\gamma^s_{ir} + \omega_{is}\gamma^s_{jr} = \partial_r\omega_{ij}$, a result allowing to move down the index k in the previous formulas (See [9] for more details).

We may thus take into account the Bianchi identities implied by the cyclic sums on (ijr)

$$\beta_{kl,ijr} \equiv \nabla_r\rho_{kl,ij} + \nabla_i\rho_{kl,jr} + \nabla_j\rho_{kl,ri} = 0 \quad \Leftrightarrow \quad \beta \equiv \sum_{cycl}(\partial\rho - \gamma\rho) = 0$$

and their respective linearizations $B_{kl,ijr} = 0$ as described below. We shall see later on that β and B are sections of the vector bundle F_2 defined by the short exact sequence:

$$0 \to F_2 \to \wedge^3 T^* \otimes g_1 \stackrel{\delta}{\to} \wedge^4 T^* \otimes T \to 0$$

with

$$dim(F_2) = \big(n(n-1)(n-2)/6\big)\big(n(n-1)/2\big) - \big(n(n-1)(n-2)(n-3)/24\big)n$$
$$= n^2\left(n^2-1\right)(n-2)/24$$

because $dim(g_1) = n(n-1)/2$ for *any* nondegenerate metric, that is $24-4=20$ when $n=4$.

Such results cannot be even imagined by somebody not aware of the δ-acyclicity ([1] [9] [13]).

We have the linearized cyclic sums of covariant derivatives both with their respective symbolic descriptions, not to be confused with the non-linear corresponding ones:

$$B_{kl,rij} \equiv \nabla_r R_{kl,ij} + \nabla_i R_{kl,jr} + \nabla_j R_{kl,ri} = 0 \quad mod(\Gamma)$$
$$\Leftrightarrow \sum_{cycl}\big(dR - \gamma R - \rho\Gamma\big) = 0$$
$$\Leftrightarrow B \equiv \sum_{cycl}\big(\nabla R\big) = \sum_{cycl}\big(\rho\Gamma\big)$$

In order to recapitulate these new concepts obtained after one, two or three prolongations, we have successively $\omega \to \gamma \to \rho \to \beta$ and the respective linearizations $\Omega \to \Gamma \to R \to B$.

The 24 Bianchi identities are related by the 4 linear relations like $B_{01,023} - B_{02,013} + B_{03,012} = 0$ when $n=4$ because $B_{00,123} = 0$. These relations are existinging between the 24 components of the Lanczos tensor because $B \in F_2 \subset ker(\delta)$ in the previous short exact sequence ([20]).

With more details, we number the 20 linearly independent Bianchi identities as follows:

$$\boxed{1}B_{01,012}, \boxed{2}B_{01,013},$$
$$\boxed{3}B_{02,123}, \boxed{4}B_{02,012}, \boxed{5}B_{02,013}, \boxed{6}B_{02,023},$$
$$\boxed{7}B_{03,123}, \boxed{8}B_{03,012}, \boxed{9}B_{03,013}, \boxed{10}B_{03,023},$$
$$\boxed{11}B_{12,123}, \boxed{12}B_{12,012}, \boxed{13}B_{12,013}, \boxed{14}B_{12,023},$$
$$\boxed{15}B_{13,123}, \boxed{16}B_{13,012}, \boxed{14}B_{13,013}, \boxed{18}B_{13,023},$$
$$\boxed{19}B_{23,123}, \boxed{20}B_{23,023}$$

to which we add the 4 linearly dependent:

$$\boxed{21}B_{01,123}, \boxed{22}B_{01,023}, \boxed{23}B_{23,012}, \boxed{24}B_{23,013}$$

We successively study a few situations without any, with one or with two vanishing linearized Riemann components, taking into account that the four Einstein equations are described by:

$\boxed{12}$, $\boxed{17}$, $\boxed{20}$ for the index 0, $\boxed{4}$, $\boxed{9}$, $\boxed{19}$ for the index 1, $\boxed{1}$, $\boxed{10}$, $\boxed{15}$ for the index 2, $\boxed{2}$, $\boxed{6}$, $\boxed{11}$ for the index 3.

BIANCHI $\boxed{1}$: $\boxed{B_{01,012} \equiv \nabla_0 R_{01,12} + \nabla_1 R_{01,20} + \nabla_2 R_{01,01} = -\frac{3m}{2r^3}\left(\Gamma^0_{02} + \Gamma^1_{12}\right)}$.

First of all, we have $R_{01,12}=-\frac{3m}{2r^3}\xi_2^0$, $R_{01,02}=\frac{3m}{2r^3}\xi_2^1$, $R_{01,01}=-\frac{3m}{r^4}\xi^1$ and obtain:

$$\begin{aligned}\nabla_0 R_{01,12}&=-\frac{3m}{2r^3}\xi_{02}^0-\gamma_{01}^0R_{01,02}\\&=-\frac{3m}{2r^3}\left(\xi_{02}^0+\frac{m}{2Ar^2}\xi_2^1\right)\\&=-\frac{3m}{2r^3}\Gamma_{02}^0\end{aligned}$$

$$\begin{aligned}\nabla_1 R_{01,20}&=\left(-\frac{3m}{2r^3}\xi_{12}^1+\frac{9m}{2r^4}\xi_2^1\right)-\left(2\gamma_{01}^0+\gamma_{11}^1+\gamma_{12}^2\right)R_{01,20}\\&=-\frac{3m}{2r^3}\left(\xi_{12}^1-\frac{m}{2Ar^2}\xi_2^1\right)+\frac{6m}{r^4}\xi_2^1\\&=-\frac{3m}{2r^3}\Gamma_{12}^1+\frac{6m}{r^4}\xi_2^1\end{aligned}$$

$$\begin{aligned}\nabla_2 R_{01,01}&=-\frac{3m}{r^4}\xi_2^1-2\gamma_{12}^2R_{02,01}\\&=-\frac{3m}{r^4}\xi_2^1-\frac{2}{r}\frac{3m}{2r^3}\xi_2^1\\&=-\frac{6m}{r^4}\xi_2^1\end{aligned}$$

and we notice that $\Gamma_{02}^0+\Gamma_{12}^1=\xi_{02}^0+\xi_{12}^1=d_2\left(\xi_0^0+\xi_1^1\right)=0\Rightarrow B_{01,012}=0$. As a by-product, we have $d_0W_2+\frac{m}{2Ar^2}V_2=0$, $d_1V_2-\frac{m}{2Ar^2}V_2=0\Rightarrow d_1V_2+d_0W_2=0$.

BIANCHI $\boxed{3}$: $\boxed{B_{02,123}\equiv\nabla_1R_{02,23}+\nabla_2R_{02,31}+\nabla_3R_{02,12}=\frac{3mA}{2r}\Gamma_{13}^0}$.

First of all, we have $R_{02,23}=\frac{3mA}{2r}\xi_3^0$, $R_{02,31}=0$, $R_{02,12}=0$, $R_{01,31}=\frac{3m}{2r^3}\xi_3^0$ and obtain:

$$\begin{aligned}\nabla_1 R_{02,23}&=d_1R_{02,23}-\gamma_{01}^rR_{r2,23}-\gamma_{12}^rR_{0r,23}-\gamma_{12}^rR_{02,r3}-\gamma_{13}^rR_{02,2r}\\&=\frac{3mA}{2r}\xi_{13}^0+\left(\frac{3m^2}{2r^3}-\frac{3mA}{2r^2}\right)\xi_3^0-\left(\gamma_{01}^0-\frac{3}{r}\right)R_{02,23}\\&=\frac{3mA}{2r}\xi_{13}^0-\left(\frac{12m}{2r^2}-\frac{27m^2}{4r^3}\right)\xi_3^0\\&=\frac{3mA}{2r}\Gamma_{13}^0-\frac{9mA}{2r^2}\xi_3^0\end{aligned}$$

$$\begin{aligned}\nabla_2 R_{02,31}&=0-\gamma_{02}^rR_{r2,31}-\gamma_{22}^rR_{0r,31}-\gamma_{23}^rR_{02,r1}-\gamma_{12}^rR_{02,3r}\\&=ArR_{01,31}-\cot(\theta)R_{02,31}-\frac{1}{r}R_{02,32}\\&=\frac{3mA}{2r^2}\xi_3^0+\frac{3mA}{2r^2}\xi_3^0\\&=\frac{3mA}{r^2}\xi_3^0\end{aligned}$$

$$\begin{aligned}\nabla_3 R_{02,12} &= 0-\gamma^r_{03}R_{r2,12}-\gamma^r_{23}R_{0r,12}-\gamma^r_{13}R_{02,r2}-\gamma^r_{23}R_{02,1r}\\ &= -\gamma^3_{23}R_{03,12}-\gamma^3_{13}R_{02,32}-\gamma^3_{23}R_{02,13}\\ &= \frac{3mA}{2r^2}\xi^0_3\end{aligned}$$

and we may use the fact that

$$\Gamma^0_{13}\equiv\xi^0_{13}-\frac{1}{Ar}\left(1-\frac{3m}{2r}\right)\xi^0_3=0\Rightarrow d_1W_3-\frac{1}{Ar}\left(1-\frac{3m}{2r}\right)W_3=0\,.$$

BIANCHI [4]: $\boxed{B_{02,012}\equiv\nabla_0R_{02,12}+\nabla_1R_{02,20}+\nabla_2R_{02,01}=\frac{3m}{2r^3}\Gamma^1_{22}}$.

First of all we have $R_{02,12}=0$, $R_{02,02}=\frac{3mA}{2r^2}\xi^1$, $R_{02,01}=\frac{3m}{2r^3}\xi^1_2$, $R_{12,12}=-\frac{3m}{2Ar^2}\xi^1$ and obtain:

$$\begin{aligned}\nabla_0R_{02,12} &= d_0R_{02,12}-\gamma^r_{00}R_{r2,12}-\gamma^r_{02}R_{0r,12}-\gamma^r_{01}R_{02,r2}-\gamma^r_{02}R_{02,1r}\\ &= 0-\gamma^1_{00}R_{12,12}-\gamma^0_{01}R_{02,02}\\ &= \left(-\frac{mA}{2r^2}\right)\left(-\frac{3m}{2Ar^2}\right)\xi^1-\left(\frac{m}{2Ar^2}\right)\left(\frac{3mA}{2r^2}\right)\xi^1\\ &= 0\end{aligned}$$

$$\begin{aligned}\nabla_1R_{02,20} &= d_1\left(-\frac{3mA}{2r^2}\xi^1\right)-2\gamma^r_{01}R_{r2,20}-2\gamma^r_{12}R_{0r,20}\\ &= -\frac{3mA}{2r^2}\xi^1_1-\partial_1\left(\frac{3mA}{2r^2}\right)\xi^1+2\gamma^0_{01}R_{02,02}+2\gamma^2_{12}R_{02,02}\\ &= \left(-\frac{3mA}{2r^2}\right)\left(\frac{m}{2Ar^2}\right)\xi^1-\partial_1\left(\frac{3mA}{2r^2}\right)\xi^1+\frac{3m^2}{2r^4}\xi^1+\frac{3mA}{r^3}\xi^1\\ &= -\frac{3m^2}{4r^4}\xi^1+\left(\frac{3m}{r^3}-\frac{3m^2}{r^4}+\frac{3m}{r^3}-\frac{3m^2}{r^4}\right)\xi^1\\ &= \frac{6m}{r^3}\xi^1-\frac{27m^2}{4r^4}\xi^1\end{aligned}$$

$$\begin{aligned}\nabla_2R_{02,01} &= \frac{3m}{2r^3}\xi^1_{22}-\gamma^r_{02}R_{r2,01}-\gamma^r_{22}R_{0r,01}-\gamma^r_{02}R_{02,r1}-\gamma^r_{12}R_{02,0r}\\ &= \frac{3m}{2r^3}\xi^1_{22}-\gamma^1_{22}R_{01,01}-\gamma^2_{12}R_{02,02}\\ &= \frac{3m}{2r^3}\xi^1_{22}-\frac{9mA}{2r^3}\xi^1\\ &= \frac{3m}{2r^3}\Gamma^1_{22}-\frac{6m}{r^3}\xi^1+\frac{27m^2}{4r^4}\xi^1\end{aligned}$$

This delicate checking proves that $B_{02,012}\equiv\frac{3m}{2r^3}\Gamma^1_{22}=0\Rightarrow d_2V_2+\left(1-\frac{3m}{2r}\right)U=0$ is a differential consequence of $R_{02,12}=0$. We let the reader prove as an exercise that $B_{03,013}\equiv\frac{3m}{2r^3}\Gamma^1_{33}=0\Rightarrow d_3V_3+\sin(\theta)\cos(\theta)V_2+\left(1-\frac{3m}{2r}\right)\sin^2(\theta)U=0$ in order to recover [19] by eliminating U.

BIANCHI $\boxed{6}$: $\boxed{B_{02,023} \equiv \nabla_0 R_{02,23} + \nabla_2 R_{02,30} + \nabla_3 R_{02,02} = \frac{3mA}{2r}\Gamma^0_{03}}$.

First of all, we have $R_{02,23} = \frac{3mA}{2r}\xi^0_3$, $R_{02,30} = 0$, $R_{02,02} = \frac{3mA}{2r^2}\xi^1$ and obtain:

$$\begin{aligned}\nabla_0 R_{02,23} &= d_0 R_{02,23} - \gamma^r_{00} R_{r2,23} - \gamma^r_{02} R_{0r,23} - \gamma^r_{02} R_{02,r3} - \gamma^r_{03} R_{02,2r}\\ &= d_0 R_{02,23} - \gamma^1_{00} R_{12,23}\\ &= \frac{3mA}{2r}\xi^0_{03} + \frac{3m^2}{4r^3}\xi^1_3\\ &= \frac{3mA}{2r}\Gamma^0_{03}\end{aligned}$$

$$\begin{aligned}\nabla_2 R_{02,30} &= 0 - \gamma^1_{22} R_{01,30}\\ &= ArR_{01,30}\\ &= -\frac{3mA}{2r^2}\xi^1_3\end{aligned}$$

$$\nabla_3 R_{02,02} = \frac{3mA}{2r^2}\xi^1_3$$

where $\Gamma^0_{03} \equiv \xi^0_{03} + \frac{m}{2Ar^2}\xi^1_3 = d_3\left(\xi^0_0 + \frac{m}{2Ar^2}\xi^1\right) = 0$ and $R_{12,23} = -\frac{3m}{2Ar}\xi^1_3$, $R_{01,03} = \frac{3m}{2r^3}\xi^1_3$.

Again, only this final result proves that $B_{02,023} \equiv \frac{3mA}{2r}\Gamma^0_{03} \Rightarrow d_0 W_3 + \frac{m}{2Ar^2}V_3 = 0$ is a differential consequence of $R_{02,03} = 0$.

BIANCHI $\boxed{12}$: $\boxed{B_{12,012} \equiv \nabla_0 R_{12,12} + \nabla_1 R_{12,20} + \nabla_2 R_{12,01} = -\frac{3m}{2r^3}\Gamma^0_{22}}$.

First of all, $\Gamma^0_{22} \equiv \xi^0_{22} + \frac{r}{A}\xi^1_0$, $R_{12,12} = -\frac{3m}{2Ar^2}\xi^1$, $R_{12,20} = 0$, $R_{12,01} = -\frac{3m}{2r^3}\xi^0_2$ and we obtain:

$$\begin{aligned}\nabla_0 R_{12,12} &= d_0 R_{12,12} - 2\gamma^r_{01} R_{r2,12} - 2\gamma^r_{02} R_{1r,12}\\ &= d_0 R_{12,12} - 2\gamma^0_{01} R_{02,12}\\ &= -\frac{3m}{2Ar^2}\xi^1_0\end{aligned}$$

$$\begin{aligned}\nabla_1 R_{12,20} &= 0 - \left(\gamma^1_{11} + 2\gamma^2_{12} + \gamma^0_{01}\right) R_{12,20}\\ &= -2\gamma^2_{12} R_{12,20}\\ &= 0\end{aligned}$$

$$\begin{aligned}\nabla_2 R_{12,01} &= -\frac{3m}{2r^3}\xi^0_{22} - \gamma^2_{12} R_{12,02}\\ &= -\frac{3m}{2r^3}\xi^0_{22}\\ &= -\frac{3m}{2r^3}\Gamma^0_{22} + \frac{3m}{2Ar^2}\xi^1_0\end{aligned}$$

a result leading to $d_2W_2+\frac{r}{A}d_0U=0$. Again, *only the final sum has an intrinsic mathematical meaning* with $B_{12,012}=-\frac{3m}{2r^3}\Gamma^0_{22}$.

BIANCHI $\boxed{22}$: $B_{01,023}\equiv\nabla_0R_{01,23}+\nabla_2R_{01,30}+\nabla_3R_{01,02}=0$

First of all, we have $R_{01,23}=0$, $R_{01,03}=\frac{3m}{2r^3}\xi^1_3$, $R_{01,02}=\frac{3m}{2r^3}\xi^1_2$, $R_{02,03}=0$ and, as $\Omega_{23}=0$:

$$\begin{aligned}\Gamma^1_{23}&\equiv\xi^1_{23}+\gamma^1_{33}\xi^3_2+\gamma^1_{22}\xi^2_3-\gamma^3_{23}\xi^1_3\\&=\xi^1_{23}-Ar\left(\xi^2_3+\sin^2(\theta)\xi^3_2\right)-\cot(\theta)\xi^1_3\\&=\xi^1_{23}-\cot(\theta)\xi^1_3\end{aligned}$$

$$\begin{aligned}\Gamma^2_{13}&\equiv\xi^2_{13}+\gamma^2_{r3}\xi^r_1+\gamma^2_{1r}\xi^r_3-\gamma^r_{13}\\&=\xi^2_{13}+\gamma^2_{33}\xi^3_1+\left(\gamma^2_{12}-\gamma^3_{13}\right)\xi^2_3\\&=\xi^2_{13}-\sin(\theta)\cos(\theta)\xi^3_1\end{aligned}$$

$$\begin{aligned}\Gamma^3_{12}&\equiv\xi^3_{12}+\gamma^3_{r2}\xi^r_1+\gamma^3_{1r}\xi^r_2-\gamma^r_{12}\xi^3_r\\&=\xi^3_{12}+\gamma^3_{23}\xi^3_1+\left(\gamma^3_{13}-\gamma^2_{12}\right)\xi^3_2\\&=\xi^3_{12}+\cot(\theta)\xi^3_1\end{aligned}$$

$$\begin{aligned}\nabla_0R_{01,23}&=d_0R_{01,23}-\gamma^r_{00}R_{r1,23}-\gamma^r_{01}R_{0r,23}-\gamma^r_{02}R_{01,r3}-\gamma^r_{03}R_{01,2r}\\&=d_0R_{01,23}-\gamma^1_{00}R_{01,23}-\gamma^0_{01}R_{00,13}\\&=0\end{aligned}$$

$$\begin{aligned}\nabla_2R_{01,30}&=d_2R_{01,30}-\gamma^2_{12}R_{02,30}-\gamma^3_{23}R_{01,30}\\&=-\frac{3m}{2r^3}\xi^1_{23}+\frac{3m}{2r^3}\cot(\theta)\xi^1_3\\&=-\frac{3m}{2r^3}\left(\xi^1_{23}-\cot(\theta)\xi^1_3\right)\\&=-\frac{3m}{2r^3}\Gamma^1_{23}\end{aligned}$$

$$\begin{aligned}\nabla_3R_{01,02}&=d_3R_{01,02}-\gamma^3_{13}R_{03,02}-\gamma^3_{23}R_{01,03}\\&=\frac{3m}{2r^3}\xi^1_{23}-\frac{3m}{2r^3}\cot(\theta)\xi^1_3\\&=\frac{3m}{2r^3}\left(\xi^1_{23}-\cot(\theta)\xi^1_3\right)\\&=\frac{3m}{2r^3}\Gamma^1_{23}\end{aligned}$$

We could also say that $d_2R_{01,03}=\frac{3m}{2r^3}\left(d_2V_3-\xi^1_{23}\right)+\frac{3m}{2r^3}\xi^1_{23}$ and obtain therefore finally the formula $B_{01,023}=0\Rightarrow d_2V_3-d_3V_2=0$ is a differential consequence of $R_{01,23}=0$.

We also check in particular:

$$\begin{aligned}\nabla_0R_{01,23}&\Rightarrow\rho_{r1,23}\Gamma^r_{00}+\rho_{0r,23}\Gamma^r_{01}+\rho_{01,r3}\Gamma^r_{02}+\rho_{01,2r}\Gamma^r_{03}\\&\Rightarrow 0\end{aligned}$$

$$\begin{aligned}
\nabla_2 R_{01,30} &\Rightarrow \rho_{r1,30}\Gamma^r_{02} + \rho_{0r,30}\Gamma^r_{12} + \rho_{01,r0}\Gamma^r_{23} + \rho_{01,3r}\Gamma^r_{02} \\
&\Rightarrow \rho_{03,30}\Gamma^3_{12} + \rho_{01,10}\Gamma^1_{23} \\
&\Rightarrow \frac{mA}{2r}\sin^2(\theta)\Gamma^3_{12} - \frac{m}{r^3}\Gamma^1_{23} \\
&\Rightarrow \frac{mA}{2r}\sin^2(\theta)\xi^3_{12} - \frac{m}{2r^3}\cot(\theta)\xi^1_3 - \frac{m}{r^3}\Gamma^1_{23} \\
&\Rightarrow -d_2\left(\frac{m}{2r^3}\xi^1_3\right) - \frac{mA}{r}\sin(\theta)\cos(\theta)\xi^3_1 + \frac{m}{r^3}\cot(\theta)\xi^1_3 \\
&\quad -\frac{m}{2r^3}\cot(\theta)\xi^1_3 - \frac{m}{r^3}\Gamma^1_{23} \\
&\Rightarrow -\frac{m}{2r^3}\left(\xi^1_{23} - \cot(\theta)\xi^1_3\right) - \frac{m}{2r^3}\Gamma^1_{23} \\
&\Rightarrow -\frac{3m}{2r^3}\Gamma^1_{23}
\end{aligned}$$

$$\begin{aligned}
\nabla_3 R_{01,02} &\Rightarrow \rho_{r1,02}\Gamma^r_{03} + \rho_{0r,02}\Gamma^r_{13} + \rho_{01,r2}\Gamma^r_{03} + \rho_{01,0r}\Gamma^r_{23} \\
&\Rightarrow \rho_{02,02}\Gamma^2_{13} + \rho_{01,01}\Gamma^1_{23} \\
&\Rightarrow -\frac{mA}{2r}\Gamma^2_{13} + \frac{m}{r^3}\Gamma^1_{23} \\
&\Rightarrow -\frac{mA}{2r}\left(\xi^2_{13} - \sin^2(\theta)\cot(\theta)\xi^3_1\right) + \frac{m}{r^3}\Gamma^1_{23} \\
&\Rightarrow \frac{m}{2r^3}\left(\xi^1_{23} - \cot(\theta)\xi^1_3\right) + \frac{m}{r^3}\Gamma^1_{23} \\
&\Rightarrow \frac{3m}{2r^3}\Gamma^1_{23}
\end{aligned}$$

As a tricky exercise too, we advise the reader to treat similarly the case of $\boxed{13}+\boxed{16}$ or $\boxed{21}$ in order to obtain $B_{01,123}=0 \Rightarrow d_2W_3 - d_3W_2 = 0$ because $R_{01,23}=0$ *and* $R_{03,12}=0$ (care).

REMARK 4.B.1: Though a few conditions like $\boxed{d_2V_3 - d_3V_2 = 0}$ $\boxed{22}$ or $\boxed{d_2W_3 - d_3W_2 = 0}$ $\boxed{21}$ *look like* to be third order CC for Ω, we have thus proved that they come indeed from the first prolongations of the second order CC. The same comment is also valid for a few other striking CC. Using previous results, we have successively 6 other relations:

$$\xi^1_{12} + \xi^0_{02} = d_2\left(\xi^1_1 + \xi^0_0\right) = 0 \Rightarrow \boxed{d_1V_2 + d_0W_2 = 0}\ \boxed{1}$$

$$\xi^1_{13} + \xi^0_{03} = d_3\left(\xi^1_1 + \xi^0_0\right) = 0 \Rightarrow \boxed{d_1V_3 + d_0W_3 = 0}\ \boxed{2}$$

$$\begin{aligned}
&\xi^1_{33} + \sin(\theta)\cos(\theta)\xi^1_2 - \sin^2(\theta)\xi^1_{22} = 0 \\
&\Rightarrow \boxed{d_3V_3 + \sin(\theta)\cos(\theta)V_2 - \sin^2(\theta)d_2V_2 = 0}\ \boxed{19}
\end{aligned}$$

$$\begin{aligned}
&\xi^0_{33} + \sin(\theta)\cos(\theta)\xi^0_2 - \sin^2(\theta)\xi^0_{22} = 0 \\
&\Rightarrow \boxed{d_3W_3 + \sin(\theta)\cos(\theta)W_2 - \sin^2(\theta)d_2W_2 = 0}\ \boxed{20}
\end{aligned}$$

because $\boxed{d_2W_2 + \frac{r}{A}d_0U = 0}$ $\boxed{12}$ and

$$\boxed{d_3W_3+\sin(\theta)\cos(\theta)W_2+\frac{r}{A}\sin^2(\theta)d_0U=0} \quad \boxed{17}$$

$$\xi^1_{02}-A^2\xi^0_{12}=d_2\left(\xi^1_0-A^2\xi^0_1\right)=0\Rightarrow\boxed{d_0V_2-A^2d_1W_2=0} \quad \boxed{24}$$

$$\xi^1_{03}-A^2\xi^0_{13}=d_3\left(\xi^1_0-A^2\xi^0_1\right)=0\Rightarrow\boxed{d_0V_3-A^2d_1W_3=0} \quad \boxed{23}$$

From the 24 *B*, we have thus used 8 of them and are left with $24-8=16$ expressions involving the $4\times 4=16$ different first derivatives of the 4 functions (V,W), namely *B* $\boxed{3}$ to *B* $\boxed{18}$. Now, we notice that, among these 24 *B*, *only* 4 of them do contain three components $R_{kl,ij}$ that are not vanishing for the S-metric, namely $\boxed{1}$, $\boxed{2}$, $\boxed{19}$ and $\boxed{20}$. They are providing the terms d_rE_{rr} for $r=0,1,2,3$ in the divergence type condition for the linearized Einstein equations implied by the linearized Bianchi identities over the Schwarzschild metric. Accordingly, it does not seem possible to obtain any other third order CC apart from these 4 divergence conditions.

It remains to apply these results to the successive prolongations of the Killing equations, as we know from the intrinsic study achieved in ([22] [23]) that we have the successive Lie algebroids:

$$R_1^{(4)}=R_1^{(3)}\subset R_1^{(2)}\subset R_1^{(1)}=R_1\subset J_1(T)$$

with respective dimensions $4=4<5<10=10<20$ and $R_1^{(3)}$ does not depend any longer on the *S*-parameter *m*.

The challenge will be to prove that… *the only knowledge of these numbers is sufficient*!

In an equivalent way as $g_2=0\Rightarrow g_{r+2}=0,\forall r\geq 0$, we obtain successively:

$$dim(R_1)=dim(J_1(T))-dim(S_2T^*)=20-10=10,$$

$$dim(R_2)=dim(J_2(T))-dim(J_1(S_2T^*))=60-50=10,$$

$$dim(R_3)=dim(R_2)-5=5\Rightarrow dim(R_{r+4})=dim(R_4)=dim(R_3)-1=4$$

and shall use these results from now on.

First of all, using the *introductory diagram* when $q=1,r=1$, we may apply the Spencer δ-map to the symbol top row in order to obtain the diagram:

$$\begin{array}{ccccccccc}
 & & & & 0 & & 0 & & \\
 & & & & \downarrow & & \downarrow & & \\
0 & \to & \boxed{\wedge^2T^*\otimes g_1} & \to & \wedge^2T^*\otimes T^*\otimes T & \to & \wedge^2T^*\otimes F_0 & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & \wedge^3T^*\otimes T & = & \wedge^3T^*\otimes T & \to & 0 & & \\
 & & \downarrow & & \downarrow & & & & \\
 & & 0 & & 0 & & & &
\end{array}$$

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & & & \\
 & & \downarrow & & \downarrow & & & & \\
0 & \to & S_3T^*\otimes T & \to & S_2T^*\otimes F_0 & \to & \boxed{h_2} & \to & 0 \\
 & & \downarrow & & \downarrow & & & & \\
0 & \to & T^*\otimes S_2T^*\otimes T & \to & T^*\otimes T^*\otimes F_0 & \to & 0 & & \\
 & & \downarrow & & \downarrow & & & &
\end{array}$$

Using the Spencer δ-cohomology $H^r(g_1)=Z^r(g_1)/B^r(g_1)$ at $\cdots\to\boxed{\wedge^r T^*\otimes g_1}\to\cdots$, we obtain:

PROPOSITION 4.B.2: $h_2\simeq H^2(g_1)\Rightarrow n(n+1)/2\le dim(Q_2)\le n^2(n^2-1)/12$ whenever $n\ge 3$.

Proof: As there cannot be any CC of order one and thus $Q_1=0$, we have the long exact connecting sequence $0\to R_3\to R_2\to h_2\to Q_2\to 0$ and counting the dimensions with $F_0=S_2T^*$, we have:

$$dim(Q_2)\le dim(h_2)=dim(S_2T^*\otimes S_2T^*)-dim(S_3T^*\otimes T)=n^2(n^2-1)/12$$

This result is confirmed by a circular chase proving that the left bottom δ-map is an epimorphism and a snake chase in the last diagram providing the short exact sequence:

$$0\to h_2\to\wedge^2T^*\otimes g_1\xrightarrow{\delta}\wedge^3T^*\otimes T\to 0,\quad 0\to 20\to 36\to 16\to 0$$

Indeed, as $det(\omega)\neq 0$ we may use the metric for providing an isomorphism $T\simeq T^*:(\xi^r)\to(\xi_i=\omega_{ri}\xi^r)$ in such a way that $g_1\simeq\wedge^2T^*$ is defined by $\xi_{i,j}+\xi_{j,i}=0$ for both the M, S and K metrics.

However, introducing the conformal Killing system of infinitesimal Lie equations with symbol $\hat{g}_1$ defined by the $(n(n+1)/2)-1$ linear equations $\omega_{rj}\xi^r_i+\omega_{ir}\xi^r_j-\frac{2}{n}\omega_{ij}\xi^r_r=0$ that do not depend on any conformal factor, we have the *fundamental diagram II* ([1] [12]):

$$\begin{array}{ccccccccc}
 & & & & & & 0 & & \\
 & & & & & & \downarrow & & \\
 & & & & & 0 & S_2T^* & & \\
 & & & & & \downarrow & \downarrow & & \\
 & & & & 0 & \to Z^2(g_1) & \to H^2(g_1) & \to 0 & \\
 & & & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & \to & T^*\otimes\hat{g}_2 & \xrightarrow{\delta} & Z^2(\hat{g}_1) & \to & H^2(\hat{g}_1) & \to 0 \\
 & & & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & S_2T^* & \xrightarrow{\delta} & T^*\otimes T^* & \xrightarrow{\delta} & \wedge^2T^* & \to & 0 & & \\
 & & & & \downarrow & & \downarrow & & & & \\
 & & & & 0 & & 0 & & & &
\end{array}$$

showing that we have the splitting sequence $0\to S_2T^*\to H^2(g_1)\to H^2(\hat{g}_1)\to 0$ providing a totally unusual interpretation of the successive Ricci, Riemann and Weyl tensors and the corresponding splitting. However, it must be noticed that the *Weyl*-type operator is of order 3 when $n=3$ because $n^2(n^2-1)/12-n(n+1)/2=n(n+1)(n+2)(n-3)/12$ but of order 2 for $n\ge 4$ ([4] [9]). Similar results could be obtained for the *Bianchi*-type operator as we shall see.

□

Using now the same procedure for the *introductory diagram* with $r=2$, we get the diagram:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & S_4T^*\otimes T & \to & S_3T^*\otimes F_0 & \to & h_3 & \to & 0 \\
 & & \downarrow & & \downarrow & \searrow & \downarrow & & \\
0 & \to & T^*\otimes S_3T^*\otimes T & \to & T^*\otimes S_2T^*\otimes F_0 & \to & \boxed{T^*\otimes h_2} & \to & 0 \\
 & & \downarrow & & \downarrow & & & & \\
0 & \to & \wedge^2T^*\otimes S_2T^*\otimes T & \to & \wedge^2T^*\otimes T^*\otimes F_0 & \to & 0 & & \\
 & & \downarrow & & \downarrow & & & & \\
0\to & \boxed{\wedge^3T^*\otimes g_1} & \to & \wedge^3T^*\otimes T^*\otimes T & \to & \wedge^3T^*\otimes F_0 & \to & 0 & \\
 & \downarrow & & \downarrow & & \downarrow & & & \\
0\to & \wedge^4T^*\otimes T & = & \wedge^4T^*\otimes T & \to & 0 & & & \\
 & \downarrow & & \downarrow & & & & & \\
 & 0 & & 0 & & & & &
\end{array}$$

Using a snake chase *and* Theorem 3.2.3, we obtain the short exact sequence:

$$0\to h_3\to T^*\otimes h_2\to H^3(g_1)\to 0,\quad 0\to 60\to 80\to 20\to 0$$

A chase around the upper south-east arrow on the right is leading to the following corollary where $g_2'\subset S_2T^*\otimes F_0$ is the symbol of the system $R_2'\subset J_2(F_0)$ which is the image of $J_3(T)$ and Q_1' is the cokernel of the central bottom map:

COROLLARY 4.B.3: There is a long exact connecting diagram:

$$\begin{array}{ccccccccccc}
 & & & & 0 & & 0 & & & & \\
 & & & & \downarrow & & \downarrow & & & & \\
0 & \to & S_4T^*\otimes T & \to & S_3T^*\otimes F_0 & \to & T^*\otimes h_2 & \to & H^3(g_1) & \to & 0 \\
 & & \downarrow & & \| & & \downarrow & & \downarrow & & \\
0 & \to & \rho_1(g_2') & \to & S_3T^*\otimes F_0 & \to & T^*\otimes Q_2 & \to & Q_1' & \to & 0 \\
 & & & & \downarrow & & \downarrow & & \downarrow & & \\
 & & & & 0 & & 0 & & 0 & &
\end{array}$$

allowing to use the Bianchi identities as $B\in F_2\simeq H^3(g_1)$ and we have $dim(Q_1')\le dim\left(H^3(g_1)\right)$.

Proof: Using the notations of the *introductory diagram* and the fact that $Q_1=0$, we have the following two commutative and exact diagrams obtained by choosing $F_1=Q_2$ for the first, then $F_1=Q_3$ for the second and so on, in a systematic manner as in the motivating examples:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & 0 & \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0\to & \rho_1(g_2') & \to & S_3T^*\otimes F_0 & \to & T^*\otimes Q_2 & \to & Q_1' & \to 0 \\
 & \downarrow & & \downarrow & & \downarrow & \searrow & \| & & \\
0\to & \rho_1(R_2') & \to & J_3(F_0) & \to & J_1(Q_2) & \to & Q_1' & \to 0 \\
 & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
0\to & R_2' & \to & J_2(F_0) & \to & Q_2 & \to & 0 & & \\
 & \downarrow & & \downarrow & & \downarrow & & & & \\
 & 0 & & 0 & & 0 & & & &
\end{array}$$

First, we have the short exact sequence $0 \to R_2 \to J_2(T) \to J_1(F_0) \to 0$ with $10-60+50=0$ and get $R'_0 = F_0$, $R'_1 = J_1(F_0)$ and $dim(\rho_1(R'_0)) - dim(R'_1) = 0$, that is no CC of order 1.

Now, using the long exact sequence:

$$0 \to R_3 \to J_3(T) \to J_2(F_0) \to Q_2 \to 0$$

$$R'_2 \subset \rho_1(R'_1) = \rho_1(J_1(F_0)) = J_2(F_0)$$

$$\Rightarrow dim(\rho_1(R'_1)) - dim(R'_2) = 150 - 135 = dim(Q_2) = 15$$

because $dim(R'_2) = dim(J_3(T)) - dim(R_3) = 140 - 5 = 135$ and there are 15 second order CC.

□

Then, with $dim(Q'_1) = x$, we obtain by counting the dimensions:

$$dim(\rho_1(R'_2)) = dim(J_3(F_0)) + x - dim(J_1(Q_2)) = 350 + x - 75 = x + 275,$$

$$dim(R'_3) = dim(J_4(T)) - dim(R_4) = dim(J_3(F_0)) - dim(Q_3) = 276$$

$$\Rightarrow y = dim(\rho_1(R'_2)) - dim(R'_3) = x + 275 - 276 = x - 1$$

that is $y \geq 3$ because $x \geq 4$ and thus $x = 4 \Rightarrow y = 3$ if we only take into account the 4 divergence condition of the Einstein equations. The situation will be worst for the Kerr metric with $y = 6$.

After one prolongation, we get:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
0 & \to & \rho_1(g'_3) & \to & S_4T^* \otimes F_0 & \to & T^* \otimes Q_3 & \to & Q'_1 & \to 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & || \\
0 & \to & \rho_1(R'_3) & \to & J_4(F_0) & \to & J_1(Q_3) & \to & Q'_1 & \to 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
0 & \to & R'_3 & \to & J_3(F_0) & \to & Q_3 & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

From this second diagram we obtain the commutative and exact diagram:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & & & \\
 & & \downarrow & & \downarrow & & & & \\
0 & \to & R'_4 & \to & J_4(F_0) & \to & Q_4 & \to & 0 \\
 & & \downarrow & & || & & \downarrow & & \\
0 & \to & \rho_1(R'_3) & \to & J_4(F_0) & \to & J_1(Q_3) & \to & Q'_1 & \to 0 \\
 & & & & \downarrow & & & & & \\
 & & & & 0 & & & & &
\end{array}$$

Indeed, setting again $dim(Q'_1) = x$, we obtain now similarly:

$$dim(\rho_1(R'_3)) = dim(J_4(F_0)) + x - dim(J_1(Q_3)) = 700 + x - 370 = x + 330,$$

$$dim(R'_4) = dim(J_5(T)) - dim(R_5) = dim(J_4(F_0)) - dim(Q_4) = 500$$

$$\Rightarrow y = dim\left(\rho_1\left(R_3'\right)\right) - dim\left(R_4'\right) = x + 330 - 500 = x - 170$$

that is $y = 0 \Leftrightarrow x = 170$. We find *exactly* $dim\left(F_2\right) = 170$ like in ([22], p. 1996) and the condition $y = 0$ just means that the CC of order 4 are generated by the CC of order 3, but we have only $R_4' \subseteq \rho_1\left(R_3'\right)$ in general.

With one more prolongation, applying again the δ-map to the top symbol sequence, we get the following commutative diagram:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & S_6T^* \otimes T & \to & S_5T^* \otimes F_0 & \to & h_5 & \to & 0 \\
 & & \downarrow \delta & & \downarrow \delta & & \downarrow & & \\
0 & \to & T^* \otimes S_5T^* \otimes T & \to & T^* \otimes S_4T^* \otimes F_0 & \to & T^* \otimes h_4 & \to & 0 \\
 & & \downarrow \delta & & \downarrow \delta & & \downarrow & & \\
0 & \to & \wedge^2 T^* \otimes S_4T^* \otimes T & \to & \wedge^2 T^* \otimes S_3T^* \otimes F_0 & \to & \wedge^2 T^* \otimes h_3 & \to & 0 \\
 & & \downarrow \delta & & \downarrow \delta & & \downarrow & & \\
0 & \to & \wedge^3 T^* \otimes S_3T^* \otimes T & \to & \wedge^3 T^* \otimes S_2T^* \otimes F_0 & \to & \wedge^3 T^* \otimes h_2 & \to & 0 \\
 & & \downarrow \delta & & \downarrow \delta & & \downarrow & & \\
0 & \to & \wedge^4 T^* \otimes S_2T^* \otimes T & \to & \wedge^4 T^* \otimes T^* \otimes F_0 & \to & 0 & & \\
 & & \downarrow & & \downarrow & & & & \\
 & & 0 & & 0 & & & &
\end{array}$$

where the right exact vertical column is $0 \to 224 \to 504 \to 360 \to 80 \to 0$. It just remains to replace in the two upper right epimorphisms h_5 by $T^* \otimes Q_4$ and h_4 by Q_4 along with the following commutative diagram where we have chosen $F_1 = Q_4$:

$$\begin{array}{ccccc}
0 & & 0 & & \\
 & \searrow & \downarrow & & \\
 & & h_5 & & \\
 & & \downarrow & \searrow & \\
0 & \to & T^* \otimes h_4 & \to & T^* \otimes Q_4
\end{array}$$

in order to obtain the long exact sequence $0 \to S_6T^* \otimes T \to S_5T^* \otimes F_0 \to T^* \otimes Q_4$.

Finally, chasing in the following commutative and exact introductory diagram:

$$\begin{array}{ccccccccccc}
 & & & & 0 & & 0 & & 0 & & 0 & \\
 & & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
 & & 0 & \to & S_6T^* \otimes T & \overset{\sigma_5(\Phi)}{\to} & S_5T^* \otimes F_0 & \to & T^* \otimes Q_4 & \to & Q_1' & \to 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \| & \\
0 & \to & R_6 & \to & J_6\left(T\right) & \overset{\rho_5(\Phi)}{\to} & J_5\left(F_0\right) & \to & J_1\left(Q_4\right) & \to & Q_1' & \to 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0 & \to & R_5 & \to & J_5\left(T\right) & \overset{\rho_4(\Phi)}{\to} & J_4\left(F_0\right) & \to & Q_4 & \to & 0 & \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & & \\
 & & 0 & & 0 & & 0 & & 0 & & &
\end{array}$$

we deduce that $R'_5 = \rho_1(R'_4)$ is involutive with $dim(R'_5) = 840 - 4 = 836$ and symbol $g'_5 \simeq S_6T^* \otimes T$.

Unhappily, the reader will check at once that *a similar procedure cannot be applied* in order to prove that $R'_4 = \rho_1(R'_3)$. Indeed, *if we still have a monomorphism* $0 \to h_4 \to Q_4$ *we do not have a monomorphism* $h_3 \to Q_3$ *because now this map has a kernel of dimension equal to* $dim\left(R_3/R_3^{(1)}\right) = 5 - 4 = 1$ according to the corresponding long exact connecting sequence.

IT IS THUS NOT POSSIBLE TO PROVE THAT THERE ARE ONLY SECOND AND THIRD ORDER GENERATING CC IN A SIMPLE INTRINSIC WAY.

However, like in the first motivating example in which we *should* be waiting for third order CC but a direct computation was proving that only second order ones *could* be used, we have:

THEOREM 4.B.4: The CC of the first order operator $D: T \to F_0$ are generated by a third order operator $D_1: F_0 \to F_1 = Q_3$ and we have thus $R'_4 = \rho_1(R'_3)$.

Proof: With $F_0 = S_2T^*$ and $F_1 = Q_3$ while applying the Spencer operator, we obtain the following commutative diagram in which the two central vertical columns are locally exact ([1] [5]):

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow, & & \downarrow & & \\
0 \to & & \Theta & \to & T & \overset{\mathcal{D}}{\to} & \boxed{F_0} & \overset{\mathcal{D}_1}{\to} & F_1 \\
 & & \downarrow j_4 & & \downarrow j_4 & & \downarrow j_3 & \searrow & \| \\
0 \to & & R_4 & \to & J_4(T) & \to & J_3(F_0) & \to & F_1 \to 0 \\
 & & \downarrow d & & \downarrow d & & \downarrow d & & \\
0 \to & & \boxed{T^* \otimes R_3} & \to & T^* \otimes J_3(T) & \to & T^* \otimes J_2(F_0) & & \\
 & & \downarrow d & & \downarrow d & & & & \\
0 \to & & \wedge^2 T^* \otimes R_2 & \to & \wedge^2 T^* \otimes J_2(T) & & & &
\end{array}$$

Chasing in this diagram by using the Snake lemma of the second section, we discover that the local exactness at F_0 of the top row is equivalent to the local exactness at $T^* \otimes R_3$ of the left column.

Now, we have the commutative diagram:

$$\begin{array}{ccccccccc}
 & & & & 0 & & 0 & & 0 \\
 & & & & \downarrow & & \downarrow & & \downarrow \\
0 \to & & \Theta & \overset{j_5}{\to} & R_5 & \overset{d}{\to} & T^* \otimes R_4 & \overset{d}{\to} & \wedge^2 T^* \otimes R_3 \\
 & & \| & & \downarrow & & \downarrow & & \downarrow \\
0 \to & & \Theta & \overset{j_4}{\to} & R_4 & \overset{d}{\to} & \boxed{T^* \otimes R_3} & \overset{d}{\to} & \wedge^2 T^* \otimes R_2 \\
 & & & & \downarrow & & \downarrow & & \downarrow \\
 & & & & 0 & \to & T^* \otimes \left(R_3/R_3^{(1)}\right) & \overset{d}{\to} & \wedge^2 T^* \otimes \left(R_2/R_2^{(1)}\right) \\
 & & & & & & \downarrow & & \downarrow \\
 & & & & & & 0 & & 0
\end{array}$$

The top row is known to be locally exact as it is isomorphic to a part of the Poincaré sequence according to the commutative diagram with $R_4 \simeq R_5 \simeq R_6$:

$$\begin{array}{ccccccccc} & & & & 0 & & 0 & & 0 \\ & & & & \downarrow & & \downarrow & & \downarrow \\ 0 \to & \Theta & \overset{j_6}{\to} & R_6 & \overset{d}{\to} & T^* \otimes R_5 & \overset{d}{\to} & \wedge^2 T^* \otimes R_4 & \\ & \| & & \downarrow & & \downarrow & & \downarrow & \\ 0 \to & \Theta & \overset{j_5}{\to} & R_5 & \overset{d}{\to} & T^* \otimes R_4 & \overset{d}{\to} & \wedge^2 T^* \otimes R_3 & \\ & & & \downarrow & & \downarrow & & & \\ & & & 0 & & 0 & & & \end{array}$$

The bottom row is purely algebraic as it is induced by the exact sequence obtained by applying the Spencer operator to the long exact connecting sequence and chasing along the south west diagonal:

$$0 \to h_4 \overset{\delta}{\to} T^* \otimes h_3 \overset{\delta}{\to} \wedge^2 T^* \otimes h_2 \to \wedge^4 T^* \otimes g_1 \to 0$$

$$0 \to 126 \to 240 \to 120 \to 6 \to 0$$

Changing the confusing notations used in ([24]), we prove that the bottom Spencer operator is injective. Indeed, we have the following representative parametric jets for the various Lie equations:

$$dim(R_2) = 10 \Rightarrow \left\{\xi^0, \xi^1, \xi^2, \xi^3, \xi^1_0, \xi^0_2, \xi^0_3, \xi^1_2, \xi^1_3, \xi^2_3\right\}$$

$$dim(R_3) = 5 \Rightarrow \left\{\xi^0, \xi^2, \xi^3, \xi^1_0, \xi^2_3\right\}, \quad dim(R_4) = 4 \Rightarrow \left\{\xi^0, \xi^2, \xi^3, \xi^2_3\right\}$$

$$dim\left(R_3 / R_3^{(1)}\right) = 1 \Rightarrow \left\{\xi^1_0\right\}, \quad dim\left(R_2 / R_2^{(1)}\right) = 5 \Rightarrow \left\{\xi^1, \xi^1_2, \xi^1_3, \xi^0_2, \xi^0_3\right\}$$

We also recall the definition of the Spencer operator $d: T^* \otimes J_{q+1}(T) \to \wedge^2 T^* \otimes J_q(T)$:

$$\left(\xi^k_{\nu,i}\right) \to \xi^k_{\mu,ij} = \left(\partial_i \xi^k_{\mu,j} - \partial_j \xi^k_{\mu,i} + \xi^k_{\mu+1_j,i} - \xi^k_{\mu+1_i,j}\right)$$

Accordingly, we may choose local coordinates $\left(\xi^1_{0,i}\right)$ for a representative and a representative of the image by d is for example $\left(\xi^1_{,0i} = \xi^1_{0,i} - \xi^1_{i,0}\right)$. Now, as $dim\left(R_3 / R_3^{(1)}\right) = 4 - 3 = 1$, we may introduce the four local coordinates $\xi^1_{0,i}$ and $\xi^0_{1,i}$ such that $\xi^1_{0,i} - A^2 \xi^0_{1,i} = 0$, $\forall i = 0,1,2,3$. We may also use the $6 \times 5 = 30$ local coordinates $\left(\xi^1_{,ij}, \xi^1_{2,ij}, \xi^1_{3,ij}, \xi^0_{2,ij}, \xi^0_{3,ij}\right)$ in order to describe $\wedge^2 T^* \otimes \left(R_2 / R_2^{(1)}\right)$. In the kernel of d, we have in particular

$\xi^1_{,0i} = \xi^1_{0,i} - \xi^1_{i,0} = 0 \Rightarrow \xi^1_{0,i} = \xi^1_{i,0} = 0, \forall i = 1,2,3$ because $\xi^1_1 = \dfrac{m}{2Ar^2}\xi^1$ in R_1 but also $\xi^0_{,01} = \xi^0_{0,1} - \xi^0_{1,0} = 0 \Rightarrow \xi^0_{1,0} = 0 \Rightarrow \xi^1_{0,0} = 0$ because $\left\{\xi^0\right\}$ is among the parametric jets of R_3 and thus $\xi^1_{0,i} = 0, \forall i = 0,1,2,3$. The bottom Spencer operator is thus injective and the bottom sequence is thus exact. A circular chase ends the proof: If $b \in T^* \otimes R_3$ is killed by d, then its projection $c \in T^* \otimes \left(R_3 / R_3^{(1)}\right)$ is also killed by d and is such that $c = 0$. Accordingly, $\exists a \in T^* \otimes R_4$ with image b under the monomorphism $T^* \otimes R_4 \to T^* \otimes R_3$ and such that $da = 0$. We may thus find $e \in R_5$ and $f \in R_4$ because $R_5 \simeq R_4$ with $a = de \Rightarrow b = df$.

□

Like in the second motivating example, the sequence constructed in the previous theorem may have "*jumps*" in the order of the successive operators and we have therefore (Compare to [1]):

COROLLARY 4.B.5: The symbol of $\mathcal{D}_1$ is *not* 2-acyclic and the CC operator $\mathcal{D}_2$ is thus of order 2. Accordingly, if one does want a formally exact canonical Janet sequence, *the only possibility* is to use the involutive operator $\mathcal{D}_1$ of order 4 defined by $R'_4 = \rho_1(R'_3)$.

Proof: Recapitulating the results so far obtained, we have successively $R'_{r+1} \subseteq \rho_1(R'_r)$ with:

$$R'_0 = F_0, dim(R'_1) = dim(J_2(T)) - dim(R_2) = 60 - 10 = 50$$

$$\Rightarrow R'_1 = J_1(F_0), \rho_1(R'_1) = J_2(F_0)$$

$$dim(R'_2) = dim(J_3(T)) - dim(R_3) = 140 - 5 = 135,$$

$$dim(\rho_1(R'_1)) - dim(R'_2) = 150 - 135 = 15$$

$$dim(R'_3) = dim(J_4(T)) - dim(R_4) = 280 - 4 = 276,$$

$$dim(\rho_1(R'_2)) - dim(R'_3) \geq 3,$$

$$dim(R'_4) = dim(J_5(T)) - dim(R_5) = 504 - 4 = 500, \quad R'_4 = \rho_1(R'_3),$$

$$dim(R'_5) = dim(J_6(T)) - dim(R_6) = 840 - 4 = 836, \quad R'_5 = \rho_1(R'_4).$$

the long exact sequence: $0 \to \rho_1(R'_2) \to J_3(F_0) \to J_1(Q_2) \to Q'_1 \to 0$ with $dim(Q'_1) = x \geq 4$ because of the divergence CC condition for Einstein equations implied by the Bianchi identities.

It also follows that:

$$g'_1 \simeq T^* \otimes F_0 \Rightarrow dim(g'_1) = 40,$$

$$S_3T^* \otimes T \subset g'_2 \Leftrightarrow 80 < dim(g'_2) = 135 - 50 = 85,$$

$$S_4T^* \otimes T \subset g'_3 \Leftrightarrow 140 < 141 = 276 - 135, \quad S_{5+r}T^* \otimes T \simeq g'_{4+r}, \forall r \geq 0$$

and we have the basic commutative and exact "*defining diagram*" of the system $R_2 \subset J_2(T)$:

$$\begin{array}{ccccccccc}
 & & & & 0 & & 0 & & \\
 & & & & \downarrow & & \downarrow & & \\
 & & 0 & \to & S_2T^* \otimes T & \to & T^* \otimes F_0 & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & R_2 & \to & J_2(T) & \to & J_1(F_0) & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & R_1 & \to & J_1(T) & \to & F_0 & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

allowing to obtain the central vertical short exact sequence $0 \to g'_1 \to R'_1 \to F_9 \to 0$.

Now, it is known that a symbol g_q of finite type is involutive if an only if it is vanishing ([1] [5] [10]). Using a similar proof, let us consider the commutative

diagram of δ-sequences:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
\cdots\rightarrow & T^*\otimes S_6T^* \otimes T & \rightarrow & \wedge^3T^*\otimes S_5T^*\otimes T & \rightarrow & \wedge^4T^*\otimes S_4T^*\otimes T & \rightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
\cdots\rightarrow & T^*\otimes g_5' & \rightarrow & \wedge^3T^*\otimes g_4' & \rightarrow & \boxed{\wedge^4T^*\otimes g_3'} & \rightarrow & 0 \\
 & & \downarrow & & \downarrow & & & & \\
 & & 0 & & 0 & & & &
\end{array}$$

Using the fact that the upper sequence is known to be exact as a δ-sequence and that we have $dim\left(S_4T^*\otimes T\right)=140<141=dim\left(g_3'\right)$, an easy chase proves that the lower sequence *cannot* be exact and thus g_3' *cannot* be involutive after counting the dimensions. The corollary follows from the fact that $g_4'=\rho_1\left(g_3'\right)\simeq S_5T^*\otimes T$ is indeed 3-acyclic *one step ahead* by chasing and even involutive.

Finally, with vector bundles A,B such that $dim\left(A\right)=1$, $dim\left(B\right)=5$, we have the commutative diagram of δ-sequences in which we recall that $g_1'\simeq T^*\otimes F_0$:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
T^*\otimes S_5T^*\otimes T & \rightarrow & \wedge^2T^*\otimes S_4T^*\otimes T & \rightarrow & \wedge^3T^*\otimes S_3T^*\otimes T & \rightarrow & \wedge^4T^*\otimes S_2T^*\otimes T & \rightarrow & 0 \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
T^*\otimes g_4' & \rightarrow & \boxed{\wedge^2T^*\otimes g_3'} & \rightarrow & \wedge^3T^*\otimes g_2' & \rightarrow & \wedge^4T^*\otimes g_1' & \rightarrow & 0 \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \rightarrow & \wedge^2T^*\otimes A & \rightarrow & \wedge^3T^*\otimes B & \rightarrow & 0 & & \\
 & & \downarrow & & \downarrow & & & & \\
 & & 0 & & 0 & & & &
\end{array}$$

Taking into account that the top row is exact and proceeding as in the last theorem with similar local coordinates, we get:

$$\xi^1_{,123}=\xi^1_{1,23}+\xi^1_{2,31}+\xi^1_{3,12}=0+0+0=0$$

always.

$$\xi^1_{,012}=\xi^1_{0,12}+\xi^1_{1,20}+\xi^1_{2,01}=\xi^1_{0,12}+0+0=0\Rightarrow\xi^1_{0,12}=0\Rightarrow\xi^1_{0,ij}=0,\forall i,j=1,2,3$$

$$\xi^0_{,10i}=\xi^0_{1,0i}+\xi^0_{0,i1}+\xi^0_{i,01}=\xi^0_{1,0i}+0+0=0\Rightarrow\xi^0_{1,0i}=0\Rightarrow\xi^1_{0,0i}=0,\forall i=2,3.$$

We are thus only left with $\xi^1_{0,01}$ that may not vanish though $\xi^1_{0,01}+\xi^1_{0,10}+\xi^1_{1,00}=0$ *in any case* and the bottom map δ is not injective.

Let us prove that g_3' is not 2-acyclic because the central δ-sequence *cannot* be exact at $\wedge^2T^*\otimes g_3'$. Indeed, if it were, let $c\in\wedge^2T^*\otimes A$ be killed by δ. Then, we may lift c to $b\in\wedge^2T^*\otimes g_3'$ such that $\delta b=f\in\wedge^3T^*\otimes S_3T^*\otimes T$ and obtain by commutativity $\delta f=0$ because the last vertical downarrow on the right is an isomorphism, thus a monomorphism. As the upper row is an exact sequence, we may thus find $a\in\wedge^2T^*\otimes S_4T^*\otimes T$ such that $f=\delta a$. Chasing

circularly, it follows from the exactness assumption at $\wedge^2T^*\otimes g_3'$ that we can find $e\in T^*\otimes g_4'\simeq T^*\otimes S_5T^*\otimes T$ such that $b=a+\delta e=a'\in\wedge^2T^*\otimes S_4T^*\otimes T$. It should follow that $c=0$ and a contradiction, that is g_3' cannot be 2-acyclic.

As we know from ([1] [5] [10]) that the order of the generating CC for $\mathcal{D}_1$ is equal to $s+1$ if one needs s prolongations in such a way that $\rho_s\left(g_3'\right)=g_{3+s}'$ becomes 2-acyclic. As we already know that $g_4'=\rho_1\left(g_3'\right)\simeq T^*\otimes S_5T^*\otimes T$ is involutive, we get $s=1$ and the generating CC $\mathcal{D}_2$ of $\mathcal{D}_1$ are of order 2. We have just a "*jump*" in the order and, for the details, refer the reader to the quite delicate Example 3.14 of ([9], p 119-125) in which it is already difficult to discover how many new second order CC should be introduced though the initial system is trivially FI with coefficients in $\mathbb{Q}$. Such a result could not even be imagined while using the methods of ([6] [7] [8]).

□

There are "natural" reason for which we do not believe that these results could be useful in physics. Indeed, considering like in the previous reference the long exact sequence of jet bundles allowing to define F_2 when $F_1=Q_3$, namely:

$$0\to R_6\to J_6\left(T\right)\underset{1}{\to}J_5\left(F_0\right)\underset{3}{\to}J_2\left(F_1\right)\underset{2}{\to}F_2\to 0$$

$$0\to 4\to 840\underset{1}{\to}1260\underset{3}{\to}1110\underset{2}{\to}686\to 0$$

and the large values of these dimensions need no comment for *any* application.

□

C) KERR METRIC:

We now write the Kerr metric in Boyer-Lindquist coordinates:

$$\mathrm{d}s^2=\frac{\rho^2-mr}{\rho^2}\mathrm{d}t^2-\frac{\rho^2}{\Delta}\mathrm{d}r^2-\rho^2\mathrm{d}\theta^2-\frac{2amr\sin^2\left(\theta\right)}{\rho^2}\mathrm{d}t\mathrm{d}\phi$$
$$-\left(r^2+a^2+\frac{mra^2\sin^2\left(\theta\right)}{\rho^2}\right)\sin^2\left(\theta\right)\mathrm{d}\phi^2$$

where we have set $\Delta=r^2-mr+a^2$, $\rho^2=r^2+a^2\cos^2\left(\theta\right)$ as usual and we check that we recover the Schwarschild metric when $a=0$. We notice that t or ϕ do not appear in the coefficients of the metric. We shall change the coordinate system in order to confirm theses results by using computer algebra and the idea is to use the so-called "*rational polynomial*" coefficients as follows:

$$\left(x^0=t,x^1=r,x^2=c=\cos\left(\theta\right),x^3=\phi\right)\Rightarrow\mathrm{d}x^2=-\sin\left(\theta\right)\mathrm{d}\theta$$
$$\Rightarrow\left(\mathrm{d}x^2\right)^2=\left(1-c^2\right)\mathrm{d}\theta^2$$

We obtain over the differential field $K=\mathbb{Q}\left(a,m\right)\left(t,r,c,\phi\right)=\mathbb{Q}\left(a,m\right)\left(x\right)$:

$$\mathrm{d}s^2=\frac{\rho^2-mx^1}{\rho^2}\left(\mathrm{d}x^0\right)^2-\frac{\rho^2}{\Delta}\left(\mathrm{d}x^1\right)^2-\frac{\rho^2}{1-\left(x^2\right)^2}\left(\mathrm{d}x^2\right)^2-\frac{2amx^1\left(1-\left(x^2\right)^2\right)}{\rho^2}\mathrm{d}x^0\mathrm{d}x^3$$
$$-\left(1-\left(x^2\right)^2\right)\left(\left(x^1\right)^2+a^2+\frac{ma^2x^1\left(1-\left(x^2\right)^2\right)}{\rho^2}\right)\left(\mathrm{d}x^3\right)^2$$

with now $\Delta=\left(x^1\right)^2-mx^1+a^2=r^2-mr+a^2$ and $\rho^2=\left(x^1\right)^2+a^2\left(x^2\right)^2=r^2+a^2c^2$. For a later use, it is also possible to set $\omega_{33}=-\left(1-c^2\right)\left(\left(r^2+a^2\right)^2-a^2\left(1-c^2\right)\left(a^2-mr+r^2\right)\right)\Big/\left(r^2+a^2c^2\right)$ and we have $det(\omega)=-\left(r^2+a^2c^2\right)^2$. Framing the leading derivatives, we obtain:

$$R_1\subset J_1(T)\left\{\begin{array}{rcl}
\Omega_{33} & \equiv & 2\left(\omega_{33}\boxed{\xi^3_3}+\omega_{03}\xi^0_3\right)+\xi\partial\omega_{33}=0\\
\Omega_{23} & \equiv & \omega_{33}\boxed{\xi^3_2}+\omega_{03}\xi^0_2+\omega_{22}\xi^2_3=0\\
\Omega_{22} & \equiv & 2\omega_{22}\boxed{\xi^2_2}+\xi\partial\omega_{22}=0\\
\Omega_{13} & \equiv & \omega_{33}\boxed{\xi^3_1}+\omega_{03}\xi^0_1+\omega_{11}\xi^1_3=0\\
\Omega_{12} & \equiv & \omega_{22}\boxed{\xi^2_1}+\omega_{11}\xi^1_2=0\\
\Omega_{11} & \equiv & 2\omega_{11}\boxed{\xi^1_1}+\xi\partial\omega_{11}=0\\
\Omega_{03} & \equiv & \omega_{33}\boxed{\xi^3_0}+\omega_{03}\left(\xi^0_0+\xi^3_3\right)+\omega_{00}\xi^0_3+\xi\partial\omega_{03}=0\\
\Omega_{02} & \equiv & \omega_{22}\boxed{\xi^2_0}+\omega_{00}\xi^0_2+\omega_{03}\xi^3_2=0\\
\Omega_{01} & \equiv & \omega_{11}\boxed{\xi^1_0}+\omega_{00}\xi^0_1+\omega_{03}\xi^3_1=0\\
\Omega_{00} & \equiv & 2\left(\omega_{00}\boxed{\xi^0_0}+\omega_{03}\xi^3_0\right)+\xi\partial\omega_{00}=0
\end{array}\right.$$

Now, we know that if $R_q\subset J_q(T)$ is a system of infinitesimal Lie equations, then we have the algebroid bracket and its link with the *prolongation/projection* (PP) procedure ([1] [2] [5] [10]):

$$\left[R_q,R_q\right]\subset R_q\Rightarrow\left[R^{(s)}_{q+r},R^{(s)}_{q+r}\right]\subset R^{(s)}_{q+r},\forall q,r,s\geq 0$$

As $R_1^{(1)}=\pi_1^2\left(R_2\right)=R_1$, it follows that $R_1^{(2)}=\pi_1^3\left(R_3\right)$ is such that $\left[R_1^{(2)},R_1^{(2)}\right]\subset R_1^{(2)}$ with $dim\left(R_1^{(2)}\right)=20-16=4$ because we have obtained a total of 6 *new different* first order equations. Using the first general diagram of the Introduction, we discover that the operator defining R_1 has $10+4=14$ CC of order 2, a result obtained *totally independently of any specific GR technical object* like the *Teukolski scalars* or the *Killing-Yano tensors* introduced in ([6] [7] [8]).

Like in the case of the S metric, two prolongations allow to obtain 6 additional equations (instead of 5) that we set on the left side in the following list obtained $mod\left(j_2(\Omega)\right)$:

We have *on sections* (*care*) the 16 (linear) equations $mod\left(j_2(\Omega)\right)$ of $R_1^{(2)}$ as follows ([23]):

$$R_1^{(2)}\subset R_1\subset J_1(T)\left\{\begin{array}{lcl}
\xi^1=0,\xi^2=0 & \Rightarrow & \boxed{\omega_{00}\xi^0_1+\omega_{03}\xi^3_1}+\omega_{11}\xi^1_0=0,\xi^1_1=0,\xi^2_2=0\\
\xi^1_2=0 & \Rightarrow & \xi^2_1=0\\
\xi^1_3+lin\left(\xi^1_0,\xi^2_0\right)=0 & \Rightarrow & \boxed{\omega_{03}\xi^0_1+\omega_{33}\xi^3_1}+\omega_{11}\xi^1_3=0\\
\xi^2_3+lin\left(\xi^1_0,\xi^2_0\right)=0 & \Rightarrow & \boxed{\omega_{00}\xi^0_2+\omega_{03}\xi^3_2}+\omega_{22}\xi^2_0=0,\\
 & & \boxed{\omega_{03}\xi^0_2+\omega_{33}\xi^3_2}+\omega_{22}\xi^2_3=0\\
\xi^0_3=0 & \Rightarrow & \xi^3_0=0,\xi^0_0=0,\xi^3_3=0
\end{array}\right.$$

The coefficients of the linear equations lin involved depend on the Riemann tensor as in ([23]). Accordingly, we may choose only the 2 parametric jets $\left(\xi_0^1,\xi_0^2\right)$ among $\left(\xi_0^1,\xi_3^1,\xi_0^2,\xi_3^2\right)$ to which we must add $\left(\xi^0,\xi^3\right)$ *in any case* as they are not appearing in the Killing equations.

The system is *not* involutive because its symbol is finite type but non-zero.

Using one more prolongation, all the *sections* (*care again*) vanish but ξ^0 and ξ^3, a result leading to $dim\left(R_1^{(3)}\right)=2$ in a coherent way with the only nonzero Killing vectors $\left\{\partial_t,\partial_\phi\right\}$. We have indeed:

$$\boxed{\xi_0^1=0},\boxed{\xi_0^2=0}\Rightarrow\xi_3^1=0,\xi_3^2=0\Rightarrow\xi_1^0=0,\xi_1^0=0,\xi_2^0=0,\xi_2^3=0$$

Taking therefore into account that the metric only depends on $\left(x^1=r,x^2=\cos(\theta)\right)$ we obtain *after three prolongations* the first order system:

$$R_1^{(3)}\subset R_1^{(2)}\subset R_1^{(1)}=R_1\subset J_1(T)$$

$\xi_3^3=0$	0	1	2	3
$\xi_3^2=0$	0	1	2	3
$\xi_3^1=0$	0	1	2	3
$\xi_3^0=0$	0	1	2	3
$\xi_2^3=0$	0	1	2	•
$\xi_2^2=0$	0	1	2	•
$\xi_2^1=0$	0	1	2	•
$\xi_2^0=0$	0	1	2	•
$\xi_1^3=0$	0	1	•	•
$\xi_1^2=0$	0	1	•	•
$\xi_1^1=0$	0	1	•	•
$\xi_1^0=0$	0	1	•	•
$\xi_0^3=0$	0	•	•	•
$\xi_0^2=0$	0	•	•	•
$\xi_0^1=0$	0	•	•	•
$\xi_0^0=0$	0	•	•	•
$\xi^2=0$	•	•	•	•
$\xi^1=0$	•	•	•	•

Surprisingly and contrary to the situation found for the S metric, we have now an involutive first order system with only solutions $\left(\xi^0=cst,\xi^1=0,\xi^2=0,\xi^3=cst\right)$ and notice that $R_1^{(3)}$ does not depend any longer on the parameters $(m,a)\in K$. The difficulty is to know what second members must be used along the procedure met for all the motivating examples. In particular, we have again identities to zero like $d_0\xi^1-\xi_0^1=0$, $d_0\xi^2-\xi_0^2=0$ and thus *at least* 6 third order CC coming from the 6 following components of the Spencer operator, namely:

$$\begin{array}{l}d_1\xi^1-\xi_1^1=0,\ d_2\xi^1-\xi_2^1=0,\ d_3\xi^1-\xi_3^1=0,\\ d_1\xi^2-\xi_1^2=0,\ d_2\xi^2-\xi_2^2=0,\ d_3\xi^2-\xi_3^2=0\end{array}$$

a result that cannot be even imagined from ([6] [7] [8]). Of course, proceeding like in the motivating examples, we must substitute in the right members the

values obtained from $j_2(\Omega)$ and set for example $\xi_1^1 = -\frac{1}{2\omega_{11}}\xi\partial\omega_{11}$ while replacing ξ^1 and ξ^2 by the corresponding linear combinations of the Riemann tensor already obtained for the right members of the two zero order equations.

We have the fundamental diagram I *no longer depending on* (m,a) with fiber dimensions:

$$\begin{array}{ccccccccccccccccc}
 & & & & & & 0 & & 0 & & 0 & & 0 & & 0 & & \\
 & & & & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
 & 0 & \to & \Theta & \xrightarrow{j_1} & & 2 & \xrightarrow{D_1} & 8 & \xrightarrow{D_2} & 12 & \xrightarrow{D_3} & 8 & \xrightarrow{D_4} & 2 & \to & 0 \\
 & & & & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
 & 0 & \to & 4 & \xrightarrow{j_1} & & 20 & \xrightarrow{D_1} & 40 & \xrightarrow{D_2} & 40 & \xrightarrow{D_3} & 20 & \xrightarrow{D_4} & 4 & \to & 0 \\
 & & & \| & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & \Theta & \to & 4 & \xrightarrow{\mathcal{D}} & 18 & \xrightarrow{\mathcal{D}_1} & 32 & \xrightarrow{\mathcal{D}_2} & 28 & \xrightarrow{\mathcal{D}_3} & 12 & \xrightarrow{\mathcal{D}_4} & 2 & \to & 0 \\
 & & & & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
 & & & & & & 0 & & 0 & & 0 & & 0 & & 0 & &
\end{array}$$

providing the Euler-Poincaré characteristic $4-18+32-28+12-2=0$. However, the only intrinsic concepts associated with a differential sequence are the "*extension modules*" that only depend on the Kerr differential module but *not* on the differential sequence and it follows that ([23]):

THE ONLY IMPORTANT CONCEPT IS THE GROUP INVOLVED, NOT THE SEQUENCE.

In an equivalent way as $g_2 = 0 \Rightarrow g_{r+2} = 0, \forall r \geq 0$, we obtain successively:

$$dim(R_1) = dim(J_1(T)) - dim(S_2T^*) = 20 - 10 = 10,$$

$$dim(R_2) = dim(J_2(T)) - dim(J_1(S_2T^*)) = 60 - 50 = 10,$$

$$dim(R_3) = dim(R_2) - 6 = 4$$

$$\Rightarrow dim(R_{r+4}) = dim(R_4) = dim(R_3) - 2 = 4 - 2 = 2$$

and shall use these results from now on.

According to a cut of the preliminary diagram with now $m = n = 4$, $q = 1$, $K = \mathbb{Q}(m,a)$, we obtain the following commutative and exact diagrams:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & 0 & \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0 & \to & \rho_1(g_2') & \to & S_3T^* \otimes F_0 & \to & T^* \otimes Q_2 & \to & h_1' & \to 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0 & \to & \rho_1(R_2') & \to & J_3(F_0) & \to & J_1(Q_2) & \to & Q_1' & \to 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0 & \to & R_2' & \to & J_2(F_0) & \to & Q_2 & \to & 0 & \\
 & & \downarrow & & \downarrow & & \downarrow & & & \\
 & & 0 & & 0 & & 0 & & &
\end{array}$$

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & 0 & \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0 & \to & 144+x & \to & 200 & \to & 56 & \to & x & \to 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \| & \\
0 & \to & 280+x & \to & 350 & \to & 70 & \to & x & \to 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
0 & \to & 136 & \to & 150 & \to & 14 & \to & 0 & \\
 & & \downarrow & & \downarrow & & \downarrow & & & \\
 & & 0 & & 0 & & 0 & & &
\end{array}$$

Denoting as before by y the number of additional CC of strict order 3 and by x the number $dim(h_1') = dim(Q_1')$, we discover from the above diagram that the sum of the number of second order CC (that is 14) and the number of differentially independent third order CC obtained by one prolongation of these second order CC is equal to $70-x$. As now $dim(Q_3) = 72$, we obtain therefore $72-y = 70-x$ and thus $y = x+2$. However, as $x \geq 4$ because of the 4 divergence conditions implied on the Einstein tensor by the 20 Bianchi identities, we *must* have $y \geq 6$. As we have already found *effectively* only 6 CC of order 3, we must have indeed $x = 4$ *effectively* and, *in any case*, we cannot have $y = 4$ as claimed in ([6] [7] [8]).

From the short exact sequence:

$$0 \to R_4 \to J_4(T) \to J_3(F_0) \to Q_3 \to 0$$

$$0 \to 2 \to 280 \to 350 \to 72 \to 0$$

we obtain the commutative and exact diagrams:

$$\begin{array}{ccccccccccccccccc}
 & & 0 & & & & & & & & 0 & & & & & & \\
 & & \downarrow & & & & & & & & \downarrow & & & & & & \\
0 & \to & R_4 & \to & J_4(T) & \to & R_3' & \to 0 & 0 & \to & 2 & \to & 280 & \to & 278 & \to 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & R_3 & \to & J_3(T) & \to & R_2' & \to 0 & 0 & \to & 4 & \to & 140 & \to & 136 & \to 0 \\
 & & & & \downarrow & & \downarrow & & & & & & \downarrow & & \downarrow & & \\
 & & & & 0 & & 0 & & & & & & 0 & & 0 & &
\end{array}$$

As a byproduct, we have the commutative and exact diagrams:

$$\begin{array}{ccccccccccccccc}
 & & 0 & & 0 & & & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & g_3' & \to & \rho_1(g_2') & & 0 & \to & 142 & \to & 144+x & \to & y & \to 0 \\
 & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow & & \| & & \\
0 & \to & R_3' & \to & \rho_1(R_2') & & 0 & \to & 278 & \to & 280+x & \to & y & \to 0 \\
 & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & R_2' & = & R_2' & \to 0 & 0 & \to & 136 & = & 136 & \to & 0 & & \\
 & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow & & & & \\
 & & 0 & & 0 & & & & 0 & & 0 & & & &
\end{array}$$

leading thus to the strict inclusions $g_3' \subset \rho_1(g_2') \Leftrightarrow R_3' \subset \rho_1(R_2')$ and to the formula $y = x+2$.

We obtain therefore the following most useful diagram with symbolic notations:

$$\begin{array}{ccccccccc}
 & & & & & & 0 & & \\
 & & & & & & \downarrow & & \\
 & & 0 & & 0 & & y & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 \to & & R_3' & \to & J_3(F_0) & \to & Q_3 & \to & 0 \\
 & & \downarrow & & \| & & \downarrow & & \\
0 \to & & \rho_1(R_2') & \to & J_3(F_0) & \to & J_1(Q_2) & \to & Q_1' \to 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & \rho_1(R_2')/R_3' & & 0 & & x & & \\
 & & \downarrow & & & & \downarrow & & \\
 & & 0 & & & & 0 & &
\end{array}$$

finally showing that $x = dim(Q_1')$,
$y = dim(\rho_1(R_2')/R_3') = dim(\rho_1(R_2')) - dim(R_3')$, a result leading to the long exact connecting sequence of vector bundles:

$$0 \to R_3' \to \rho_1(R_2') \to Q_3 \to J_1(Q_2) \to Q_1' \to 0.$$

in agreement with the main theorem of Section 2. We have the following dimensions:

$$\begin{array}{ccccccccc}
 & & & & & & 0 & & \\
 & & & & & & \downarrow & & \\
 & & 0 & & 0 & & 6 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 \to & & 278 & \to & 350 & \to & 72 & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 \to & & 284 & \to & 350 & \to & 70 & \to & 4 \to 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 6 & & 0 & & 4 & & \\
 & & \downarrow & & & & \downarrow & & \\
 & & 0 & & & & 0 & &
\end{array}$$

Prolonging once while taking into account that $R_5 \simeq R_4$ with common dimension 2, namely the dimension of the Kerr algebra generated by $\{\partial_t, \partial_\phi\}$, we obtain the following commutative and exact diagram in which Q_2 and Q_3 are replaced by Q_3 and Q_4:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 \to & & g_4' & \to & S_4T^* \otimes F_0 & \to & h_4 & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 \to & & R_4' & \to & J_4(F_0) & \to & Q_4 & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 \to & & R_3' & \to & J_3(F_0) & \to & Q_3 & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0. & &
\end{array}$$

showing that $g_4' \simeq S_5T^* \otimes T$ with $dim(R_4') = 504 - 2 = 502$ and $dim(Q_4) = 700 - 502 = 198$. It follows that $R_4' = \rho_1(R_3')$ is an involutive fourth order system allowing to construct a formally exact Janet sequence following the Killing operator as in ([22]), namely (exercise!!):

$$0 \to \Theta \to 4 \underset{1}{\to} 10 \underset{4}{\to} 198 \underset{1}{\to} 568 \underset{1}{\to} 652 \underset{1}{\to} 348 \underset{1}{\to} 72 \to 0$$

Of course, *such a sequence is quite far from being minimum*. However, as the Killing operator for the Kerr metric is not formally integrable as we saw, the corresponding free resolution of the Kerr differential module, namely:

$$0 \to D^{72} \underset{1}{\to} D^{348} \underset{1}{\to} D^{652} \underset{1}{\to} D^{568} \underset{1}{\to} D^{198} \underset{4}{\to} D^{10} \underset{1}{\to} D^4 \overset{p}{\to} M \to 0$$

is *not* strictly exact though we have indeed:

$$rk_D(M) = 4 - 10 + 198 - 568 + 652 - 348 + 72 = 0$$

As the maximum size of the matrices involved is $dim(J_4(198)) \times dim(J_3(568))$, that is 13860×19880, we hope to have convinced the reader that there is no hope for using computer algebra.

As $R_3' \subset \rho_1(R_2')$ with a strict inclusion, the only possibility to escape from the above difficulty is to use only R_3' and third order CC. However, as we have the strict inclusion $S_4T^* \otimes T \subset g_3'$ with a strict inclusion because $140 < 142$. As for the S metric, we have the crucial theorem:

THEOREM 4.C.1: The operator $\mathcal{D}_1 : F_0 \overset{j_3}{\to} J_3(F_0) \to F_1 = Q_3$ generates the CC of $\mathcal{D} : T \overset{j_1}{\to} J_1(T) \to F_0$.

Proof: First of all, as R_3' is *strictly* contained into $\rho_1(R_2')$, we have *at least* one third order generating CC but we already know that we have the six $(d_i\xi^1 - \xi_i^1 = 0, d_i\xi^2 - \xi_i^2 = 0)$ for $i = 1,2,3$.

Collecting the previous results and applying the Spencer operator, we obtain the following commutative diagram in which the two central columns are known to be locally exact ([1] [5]):

$$\begin{array}{ccccccccccc}
 & & 0 & & 0 & & 0 & & & & \\
 & & \downarrow & & \downarrow & & \downarrow & & & & \\
0 \to & & \Theta & \to & T & \overset{\mathcal{D}}{\to} & F_0 & & & & \\
 & & \downarrow j_4 & & \downarrow j_4 & & \downarrow j_3 & \searrow \mathcal{D}_1 & & & \\
0 \to & & R_4 & \to & J_4(T) & \to & J_3(F_0) & \to & Q_3 & \to & 0 \\
 & & \downarrow d & & \downarrow d & & \downarrow d & & \downarrow & & \\
0 \to & & \boxed{T^* \otimes R_3} & \to & T^* \otimes J_3(T) & \to & T^* \otimes J_2(F_0) & \to & T^* \otimes Q_2 & \to & 0 \\
 & & \downarrow d & & \downarrow d & & \downarrow d & & \downarrow & & \\
0 \to & & \wedge^2 T^* \otimes R_2 & \to & \wedge^2 T^* \otimes J_2(T) & \to & \wedge^2 T^* \otimes J_1(F_0) & \to & 0 & &
\end{array}$$

Chasing around the right upper commutative triangle, it follows from Theorem 2.A.3 that a section $\Omega \in F_0$ with $\mathcal{D}_1\Omega = 0$ is such that there exists $\xi \in T$ with $\mathcal{D}\xi = \Omega$ if and only if the left vertical Spencer sequence is locally exact at $T^* \otimes R_3$.

However, one has isomorphisms $R_4 \simeq R_5 \simeq R_6$ because they have the same dimension equal to 2 and the map $R_4 \to R_3$ is a monomorphism because $dim(R_3) = 4$ with parametric jets $(\xi^0, \xi^3, \xi^1_0, \xi^2_0)$ and $g_4 = 0$. Accordingly, as the Spencer sequence for the Killing algebroid is locally exact as it is isomorphic to the tensor product of the Poincaré sequence by a Lie algebra of dimension 2, it is locally exact. We have therefore the commutative diagram with exact columns:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
0 & \to & \Theta & \xrightarrow{j_5} & R_5 & \xrightarrow{d} & T^* \otimes R_4 & \xrightarrow{d} & \wedge^2 T^* \otimes R_3 \\
 & & \| & & \downarrow & & \downarrow & & \downarrow \\
0 & \to & \Theta & \xrightarrow{j_4} & R_4 & \xrightarrow{d} & \boxed{T^* \otimes R_3} & \xrightarrow{d} & \wedge^2 T^* \otimes R_2 \\
 & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
 & & 0 & & 0 & \to & T^* \otimes \left(R_3 / R_3^{(1)}\right) & \xrightarrow{d} & \wedge^2 T \otimes \left(R_2 / R_2^{(1)}\right) \\
 & & & & & & \downarrow & & \downarrow \\
 & & & & & & 0 & & 0
\end{array}$$

in which the upper row is locally exact. Chasing in this diagram, we discover that the central row is locally exact at $T^* \otimes R_3$ if the lower Spencer operator d is injective.

Indeed, we have the following representative parametric jets:

$$dim(R_2) = 10 \Rightarrow \left\{\xi^0, \xi^1, \xi^2, \xi^3, \xi^1_0, \xi^2_0, \xi^0_3, \xi^1_2, \xi^1_3, \xi^2_3\right\}$$

$$dim(R_3) = 4 \Rightarrow \left\{\xi^0, \xi^3, \xi^1_0, \xi^2_0\right\}, \quad dim(R_4) = 2 \Rightarrow \left\{\xi^0, \xi^3\right\}$$

$$dim\left(R_3 / R_3^{(1)}\right) = 2 \Rightarrow \left\{\xi^1_0, \xi^2_0\right\}, \quad dim\left(R_2 / R_2^{(1)}\right) = 6 \Rightarrow \left\{\xi^1, \xi^2, \xi^0_3, \xi^1_2, \xi^1_3, \xi^2_3\right\}$$

Accordingly, we may choose local coordinates $\left(\xi^1_{0,i}, \xi^2_{0,i}\right)$ for a representative and a representative of the image by d is for example $\left(\xi^1_{,0i} = \xi^1_{0,i} - \xi^1_{i,0}, \xi^2_{,0i} = \xi^2_{0,i} - \xi^2_{i,0}\right)$. In the kernel of d, we have $\left(\xi^1_{0,i} = \xi^1_{i,0}, \xi^2_{0,i} = \xi^2_{i,0}\right), \forall i = 1,2,3$ but the situation is more tricky than for the S metric.

For $i = 1$, we have $\xi^1_1 = 0, \xi^1_2 = 0, \xi^2_1 = 0, \xi^2_2 = 0 \Rightarrow \xi^1_{0,i} = 0$ and similarly $\xi^2_{0,i} = 0, \forall i = 1,2$.

Then, as $\left\{\xi^0, \xi^3\right\}$ are among the parametric jets of R_3, we have $\xi^0_{,13} = \xi^0_{1,3} - \xi^0_{3,1} = 0 \Rightarrow \xi^0_{1,3} = \xi^0_{3,1} = 0$ and similarly $\xi^3_{1,3} = \xi^3_{3,1} = 0$. Using the Lie Equations of $R_1^{(2)}$ we obtain successively:

$$\omega_{00}\xi^0_{1,3} + \omega_{03}\xi^3_{1,3} + \omega_{11}\xi^1_{0,3} = 0 \Rightarrow \xi^1_{0,3} = 0,$$
$$\omega_{03}\xi^0_{1,3} + \omega_{33}\xi^3_{1,3} + \omega_{11}\xi^1_{3,3} = 0 \Rightarrow \xi^1_{3,3} = 0$$

Exchanging 1 and 2, we obtain similarly with the two other framed Lie equations the two relations $\xi^2_{0,3} = 0, \xi^2_{3,3} = 0$. Hence, when $i = 3$, it follows that $\xi^1_{0,3} = \xi^1_{3,0} = lin\left(\xi^1_{0,0}, \xi^2_{0,0}\right) = 0$ and $\xi^2_{0,3} = \xi^2_{3,0} = lin\left(\xi^1_{0,0}, \xi^2_{0,0}\right) = 0$, a result leading to $\xi^1_{0,0} = 0, \xi^2_{0,0} = 0$ on one side and $\xi^1_{0,3} = 0, \xi^2_{0,3} = 0$ on the other side because

ξ_0^1 and ξ_0^2 are linearly independent jet coordinates, that is finally $\xi_{0,i}^1 = 0, \xi_{0,i}^2 = 0, \forall i = 0,1,2,3$.

□

Like in the example of the S metric, the sequence constructed in the previous theorem for the K metric have the same "*jumps*" in the order of the successive operators. Indeed, using the same proof but now with $dim(g_3') = dim(S_4T^* \otimes T) + 2 = 142$, $dim(A) = 2$, $dim(B) = 6$, we have:

COROLLARY 4.C.2: The symbol of $\mathcal{D}_1$ is *not* 2-acyclic and the CC operator $\mathcal{D}_2$ is of order 2.

Following closely the procedure used for all the motivating examples, we have thus transformed the search for the generating CC of the Killing operator into *a purely mathematical problem of formal integrability and diagram chasing*, quite far away from any physical background ([28]).

5. Conclusions

To end this paper with a rather personal story, let me come back 60 years ago when I was preparing the competition for the French "Grandes Ecoles" at the State College Louis le Grand in Paris which is famous for one of his former student Evariste Galois. To give a few statistics, let us say that, for the one I had in mind, 30,000 students were trying, 3000 were selected after the written exam and 300 were only elected after oral exam! This college was known to have the maximum number of success in France and the teachers were carefully selected for that purpose, in particular in the best class room where I was. Once, this teacher was writing on the board the text of the problem we had to solve for the next day about what is now called "Desargues theorem". Roughly, if you consider in a plane two triangles $(ABC),(A'B'C')$ that are not flat and such that the 3 straight lines AA', BB', CC' have a common origin O (*Center of perspective*), then the intersection P of BC and $B'C'$, Q of AC and $A'C'$, R of AB and $A'B'$ are on a straight line (*Axis of perspective*). Though I knew nothing about this result at that time, I suddenly "saw" the figure as a volume in space and shouted "P, Q, R are on a straight line", even before the teacher had been asking the question in front of the astonished students. Surprisingly, and I will never forget, the teacher said "Pommaret, this is true but how did you find it". When I said "Well, Sir, I have seen in space that the common line is the intersection of the two planes containing the triangles" (the reader may draw the picture for fun), his only comment has been "Better don't do that on the day of the competition". I replied "Sir, a result is important but the way you find it may even be more important". As a byproduct he never asked me any question during the full academic year and became a "private enemy" in my scholar life during 10 years till he retired.

In a similar way, I point out the fact that during a visit for lecturing at the Albert Einstein Institute (AEI, Berlin/Postdam) in October 23-27, 2017 ([18], arXiv: 1802.02430), I discovered that the members of the inviting research team

were not interested about the new tools I had developed in the many books or papers already quoted, in particular the link existing between the Spencer operator and the bracket of Lie algebroids. I also claim that the few references they quote for defining involutive systems are not the best ones as it happens that I have been regularly lecturing in Aachen during more than fifteen years and I know that the authors involved are only using Janet, Gröbner or Pommaret bases for computer algebra packages but are unable to deal with acyclicity in general. The situation met in the case of the Lie pseudogroup of conformal transformations is a good example (See [4] and [9]). As a byproduct, it became a personal challenge to clarify the CC for the Killing operators over the Schwarzschild and Kerr metrics without using any of their tedious computations. The surprise is that, if I found again the 15 second order CC for the S metric and the 14 second order CC for the K metric, I also found explicitly 3 third order CC for the S metric and 6 third order CC for the K metric. All the formulas can be written and framed within less than one line provided one uses these new methods from differential homological algebra that have never been introduced in GR up to now. The main reason is that they also prove that Einstein equations *cannot* be parametrized by a potential like Maxwell equations ([4] [19]) as such a result deeply questions the existence of gravitational waves… but this is surely another story!

Conflicts of Interest

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

References

[1] Pommaret, J.-F. (1994) Partial Differential Equations and Group Theory. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-2539-2

[2] Pommaret, J.-F. (2005) 5 Algebraic Analysis of Control Systems Defined by Partial Differential Equations. In: Lamnabhi-Lagarrigue, F., Loría, A. and Panteley, E., Eds., *Advanced Topics in Control Systems Theory*, Lecture Notes in Control and Information Sciences Vol. 311, Springer, London, 155-223. https://doi.org/10.1007/11334774_5

[3] Pommaret, J.-F. (2019) Generating Compatibility Conditions and General Relativity. *Journal of Modern Physics*, **10**, 371-401. (arXiv: 1811.12186). https://doi.org/10.4236/jmp.2019.103025

[4] Pommaret, J.-F. (2021) The Conformal Group Revisited. *Journal of Modern Physics*, **12**, 1822-1842. (arXiv:2006.03449). https://doi.org/10.4236/jmp.2021.1213106

[5] Pommaret, J.-F. (1978/1983) Systems of Partial Differential Equations and Lie Pseudogroups. Gordon and Breach, New York; Russian translation, MIR, Moscow.

[6] Aksteiner, S., Andersson L., Backdahl, T., Khavkine, I. and Whiting, B. (2021) Compatibility Complex for Black Hole Spacetimes. *Communications in Mathematical Physics*, **384**, 1585-1614. (arXiv: 1910.08756). https://doi.org/10.1007/s00220-021-04078-y

[7] Aksteiner, S. and Backdahl, T. (2019) New Identities for Linearized Gravity on the

Kerr Spacetime. *Physical Review D*, **99**, Article ID: 044043. (arXiv: 1601.06084). https://doi.org/10.1103/PhysRevD.99.044043

[8] Aksteiner, S. and Backdahl, T. (2018) All Local Gauge Invariants for Perturbations of the Kerr Spacetime. *Physical Review Letters*, **121**, Article ID: 051104. (arXiv: 1803.05341). https://doi.org/10.1103/PhysRevLett.121.051104

[9] Pommaret, J.-F. (2016) Deformation Theory of Algebraic and Geometric Structures. Lambert Academic Publisher (LAP), Saarbrucken, Germany. In: Dubreil, S.P. and Malliavin, M.-P. (Eds), *Topics in Invariant Theory*, Springer Lecture Notes in Mathematics, 244-254. https://arxiv.org/abs/1207.1964

[10] Pommaret, J.-F. (2001) Partial Differential Control Theory. Kluwer, Dordrecht. (Zbl 1079.93001).

[11] Goldschmidt, H. (1968) Prolongations of Linear Partial Differential Equations: I Inhomogeneous equations. *Annales scientifiques de l'École Normale Supérieure*, **1**, 617-625. https://doi.org/10.24033/asens.1173

[12] Pommaret, J.-F. (1983) Differential Galois Theory. Gordon and Breach, New York.

[13] Pommaret, J.-F. (1988) Lie Pseudogroups and Mechanics. Gordon and Breach, New York.

[14] Pommaret, J.-F. (2018) New Mathematical Methods for Physics. Mathematical Physics Books, Nova Science Publishers, New York, 150 p.

[15] Rotman, J.J. (1979) An Introduction to Homological Algebra, (Pure and Applied Mathematics). Academic Press, Cambridge.

[16] Pommaret, J.-F. (2016) Airy, Beltrami, Maxwell, Einstein and Lanczos Potentials revisited. *Journal of Modern Physics*, **7**, 699-728. (arXiv: 1512.05982). https://doi.org/10.4236/jmp.2016.77068

[17] Pommaret, J.-F. (2010) Parametrization of Cosserat Equations. *Acta Mechanica*, **215**, 43-55. https://doi.org/10.1007/s00707-010-0292-y

[18] Pommaret, J.-F. (2019) The Mathematical Foundations of Elasticity and Electromagnetism Revisited. *Journal of Modern Physics*, **10**, 1566-1595. (arXiv: 1802.02430). https://doi.org/10.4236/jmp.2019.1013104

[19] Pommaret, J.-F. (2021) Minimum Parametrization of the Cauchy Stress Operator. *Journal of Modern Physics*, **12**, 453-482. (arXiv: 2101.03959). https://doi.org/10.4236/jmp.2021.124032

[20] Pommaret, J.-F. (2021) Homological Solution of the Lanczos Problems in Arbitrary Dimensions. *Journal of Modern Physics*, **12**, 829-858. (arXiv: 1807.09122). https://doi.org/10.4236/jmp.2021.126053

[21] Khavkine, I. (2017) The Calabi Complex and Killing Sheaf Cohomology. *Journal of Geometry and Physics*, **113**, 131-169. https://doi.org/10.1016/j.geomphys.2016.06.009

[22] Pommaret, J.-F. (2018) Minkowski, Schwarzschild and Kerr Metrics Revisited. *Journal of Modern Physics*, **9**, 1970-2007. (arXiv: 1805.11958v2). https://doi.org/10.4236/jmp.2018.910125

[23] Pommaret, J.-F. (2020) A Mathematical Comparison of the Schwarzschild and Kerr Metrics, Journal of Modern Physics. *Journal of Modern Physics*, **11**, 1672-1710. (arXiv: 2010.07001). https://doi.org/10.4236/jmp.2020.1110104

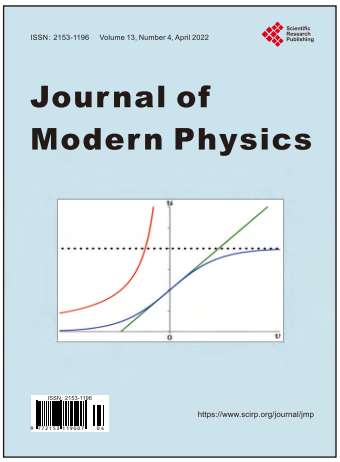
[24] Pommaret, J.-F. (2021) Differential Correspondences and Control Theory. *Advances in Pure Mathematics*, **11**, 835-882. (arXiv: 2107.08797). https://doi.org/10.4236/apm.2021.1111056

[25] Pommaret, J.-F. (2015) Relative Parametrization of Linear Multidimensional Systems. *Multidimensional Systems and Signal Processing*, **26**, 405-437. https://doi.org/10.1007/s11045-013-0265-0

[26] Macaulay, F.S. (1916) The Algebraic Theory of Modular Systems, Cambridge Tract. Cambridge University Press, Cambridge, 19. https://doi.org/10.3792/chmm/1263317740

[27] Vessiot, E. (1903) Théorie des Groupes Continus. *Annales scientifiques de l'École Normale Supérieure*, **20**, 411-451 (Can Be Found in Numdam). https://doi.org/10.24033/asens.529

[28] Pommaret, J.-F. (2013) The Mathematical Foundations of General Relativity Revisited. *Journal of Modern Physics*, **4**, 223-239. (arXiv: 1306.2818). https://doi.org/10.4236/jmp.2013.48A022



Call for Papers

Journal of Modern Physics

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

Journal of Modern Physics **(JMP)** is an international journal dedicated to the latest advancement of modern physics. The goal of this journal is to provide a platform for scientists and academicians all over the world to promote, share, and discuss various new issues and developments in different areas of modern physics.

Editor-in-Chief

Prof. Yang-Hui He City University, UK

Subject Coverage

Journal of Modern Physics publishes original papers including but not limited to the following fields:

Biophysics and Medical Physics
Complex Systems Physics
Computational Physics
Condensed Matter Physics
Cosmology and Early Universe
Earth and Planetary Sciences
General Relativity
High Energy Astrophysics
High Energy/Accelerator Physics
Instrumentation and Measurement
Interdisciplinary Physics
Materials Sciences and Technology
Mathematical Physics
Mechanical Response of Solids and Structures
New Materials: Micro and Nano-Mechanics and Homogeneization
Non-Equilibrium Thermodynamics and Statistical Mechanics
Nuclear Science and Engineering
Optics
Physics of Nanostructures
Plasma Physics
Quantum Mechanical Developments
Quantum Theory
Relativistic Astrophysics
String Theory
Superconducting Physics
Theoretical High Energy Physics
Thermology

We are also interested in: 1) Short Reports—2-5 page papers where an author can either present an idea with theoretical background but has not yet completed the research needed for a complete paper or preliminary data; 2) Book Reviews—Comments and critiques.

Notes for Intending Authors

Submitted papers should not have been previously published nor be currently under consideration for publication elsewhere. Paper submission will be handled electronically through the website. All papers are refereed through a peer review process. For more details about the submissions, please access the website.

Website and E-Mail

https://www.scirp.org/journal/jmp **E-mail: jmp@scirp.org**

What is SCIRP?

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

What is Open Access?

All original research papers published by SCIRP are made freely and permanently accessible online immediately upon publication. To be able to provide open access journals, SCIRP defrays operation costs from authors and subscription charges only for its printed version. Open access publishing allows an immediate, worldwide, barrier-free, open access to the full text of research papers, which is in the best interests of the scientific community.

- High visibility for maximum global exposure with open access publishing model
- Rigorous peer review of research papers
- Prompt faster publication with less cost
- Guaranteed targeted, multidisciplinary audience

Scientific Research Publishing

Website: https://www.scirp.org

Subscription: sub@scirp.org

Advertisement: service@scirp.org