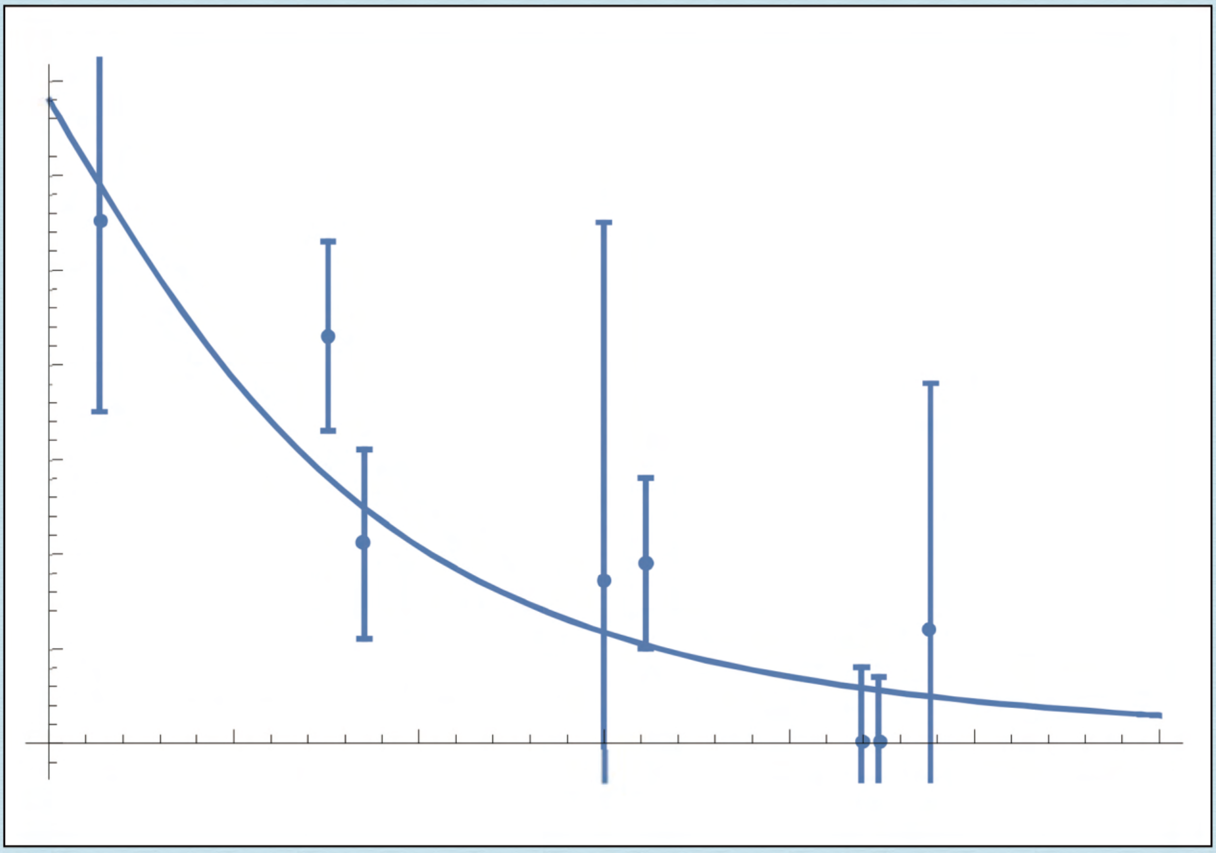
ISSN: 2153-1196 Volume 12, Number 13, November 2021

Scientific Research Publishing

Journal of Modern Physics



ISSN: 2153-1196

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

Journal Editorial Board

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

Journal of Modern Physics, 2021, 12, 1729-1929
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Table of Contents

Volume 12 Number 13 November 2021

The figure on the front cover is from the article published in Journal of Modern Physics, 2021, Vol. 12, No. 13, pp. 1729-1748 by Miguel Portilla.

Journal of Modern Physics (JMP)
Journal Information

SUBSCRIPTIONS

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

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

SERVICES

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

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

COPYRIGHT

PRODUCTION INFORMATION

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

Journal of Modern Physics, 2021, 12, 1729-1748
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Inertial and Gravitational Mass in General Relativity and Their Cosmological Consequences

Miguel Portilla

Departament d'Astronomia i Astrofísica, Universitat de València, València, Spain
Email: miguel.portilla@uv.es

How to cite this paper: Portilla, M. (2021) Inertial and Gravitational Mass in General Relativity and Their Cosmological Consequences. *Journal of Modern Physics*, **12**, 1729-1748.
https://doi.org/10.4236/jmp.2021.1213101

Received: July 16, 2021
Accepted: November 1, 2021
Published: November 4, 2021

Abstract

We revise the concept of mass of a particle in general relativity initiated by Einstein, Brans, and Rosen in the fifties, using the results of P. Havas and J.N. Goldberg on the equations of motion for point-like particles. We show how one can define a constant inertial mass, and a variable gravitational mass dependent on their gravitational interaction with the rest of particles. The introduced gravitational mass allows us to construct a cosmological model that satisfactorily accounts for the observed deficit of mass, the dark energy and the cosmological constant, without the assumption of new forms of matter or energy: dark matter and dark energy can be explained as a gravitational effect in the framework of the standard general theory of relativity.

Keywords

Dark Matter, Dark Energy, Cosmological Constant: Gravitational Mass

1. Introduction

We interpret in this paper the cosmological dark matter and energy present in the epoch of galactic dominance as a pure gravitational effect, using primarily baryonic matter and standard general relativity.

In a Minkowski spacetime, a fluid is the result of a statistical average of the dynamical properties of a system of particles; in a curved space, as it is the case when the particles are self gravitating, one needs also to revise the concept of mass. The distinction between inertial mass and active and passive gravitational mass and its relation with the weak equivalence principle was thoroughly compiled in Jammer's book [1], of which we give a few details in the next section. The revision in the framework of general relativity will permit to explain the dark components present in the galactic dominant epoch. We shall use the metric generated by a finite system of particles, *i.e.*, a solution of the Einstein's equations, having as source a distribution with support on the world lines of the particles. In the sixties, P. Havas and J.N. Goldberg [2] developed a method of successive approximations to find it. They used harmonic coordinates, unique up to Lorentz transformations [3] [4], and an auxiliary Minkowski tensor η, defined as the one that in harmonic coordinates verifies $\eta(a,b)=\eta_{\mu\nu}a^{\mu}b^{\nu}$, with $\eta_{\mu\nu}=\mathrm{diag}(1,-1,-1,-1)$, for any pair of vectors. Serious problems of convergence over the particle world lines were easily overcome in the first approximation [2], and needed of more efforts in the second one [5]. These results were useful to interpret in the eighties the first indirect proves of gravitational radiation by compact bodies [6]. Fortunately, we shall have enough with the first approximation.

There is some rejection to consider point-like particles (the use of unidimensional Dirac distribution as source), due to the problems of convergence, predicted by a theorem due to Geroch-Traschen [7] that states a necessary condition for the existence of regular metrics (locally bounded, with locally square integrable weak first derivative), by requiring the support of the distributions to be tridimensional. However, Katanaev [8] solved exactly the problem in the case of only one point-like particle, obtaining, as hoped, a non regular metric in the sense of Geroch-Traschen, with an additional coordinate singularity that makes necessary a metric extension. Recently, we have obtained [9] a maximal extension of this metric that is regular and has, as hoped, a distribution with tridimensional support as source, in accordance to the Geroch-Traschen theorem. Unfortunately, these results has not been extended to more than one particle, and we must use the approximated solutions [2] [5].

Following Havas-Goldberg (H-G), we consider the spacetime as a manifold provided with two tensors fields (M, g, η): the first is the true metric, generated by point-like particles, and the second is an auxiliary tensor used to obtain the true metric. In the next section, we show that the paper by H-G [2] tacitly contains valuable definitions for a constant inertial and a time dependent passive gravitational mass, though the last one, inexplicably, was not presented as such. We shall show that their quotient, though being the same for all the particles, is not a constant, because the passive gravitational increases with the particle's proper time, impeding the equality of inertial and gravitational mass and suggesting that their difference might account for both dark components. Section 3 summarizes the dynamical properties of a cosmological fluid made of self gravitating particles obtained recently by us [10], and Section 4 shows how the gravitational mass introduced in this paper explains satisfactorily both dark matter and dark energy.

2. Inertial and Passive Gravitational Mass of Point-Like Particles

Classical physics distinguishes three kinds of mass: inertial and, active and passive gravitational mass; but these concepts change when we move from Newtonian mechanics to special and general relativity.

Newtonian mechanics identifies the three masses, and links them to the law of motion as proportionality between force and acceleration or, alternatively, to the law of conservation of the total tri-momentum: given an isolated system of N interacting particles there exist N real numbers such that the sum $\sum_{a=1}^{N} m_a \frac{\mathrm{d}x_a}{\mathrm{d}t}$ is constant [1] [11].

Special relativity distinguishes the proper inertial mass (or rest mass), related to the conservation of the total-four-momentum (where the absolute Newtonian time t is substituted by the particle's proper time τ_a), and the relativistic mass, that in a general inertial frame has the familiar velocity dependence $m_a \big/ \sqrt{1-v_a^2}$.

General relativity, as reported in Jammer's book [1], has not yet achieved general agreement with the distinction between inertial mass and active and passive gravitational mass:

Einstein in 1950 [12], using a weak field approximation to describe the metric due to a finite distribution of matter, concluded that the inertial mass of a body, identified with the gravitational mass by principle, is not constant because depends on its gravitational interaction with all the other present masses.

C.H. Brans in 1962 [13] by studying the motion of a test particle in the field of a massive particle, at rest at the centre of a shell with mass M_S simulating the effect of the universe, contrarily argued the equality of inertial and active gravitational mass, and its independence of the gravitational interaction, at variance with the Mach's principle. N. Rosen in 1965 [14] replied analyzing the equation of motion of a test particle in the field produced by a more massive particle, immersed now in an true expanding universe, obtaining a variable inertial mass (equal to the passive gravitational mass) which increases as the universe expands, and a constant active gravitational mass. So, in short: Rosen partially extended the Einstein's result to a cosmological setting, obtaining a variable inertial mass equal to the passive gravitational mass; and coinciding with Brans about the constancy of the active gravitational mass.

H.C. Ohanian in 2013, [15] defined the inertial mass of an object in an asymptotically flat spacetime as the volume integral of an energy density determined by the canonical energy-momentum tensor, and its gravitational mass by the asymptotically Newtonian potential at large distance from it; and, he also demonstrated the equality of them. But asymptotic flat conditions are not satisfied in cosmology.

The concept of mass varies when we move from Newtonian mechanics to special relativity and to general relativity.

Not all the authors have considered that in general relativity might exist different definitions of mass. Rosen clearly stated that his definition of inertial mass, based on an interpretation of the equation of the geodesics, is not the only possible [14]. The aforementioned Einstein, and Rosen models, were developed under the tacit assumption of the Mach's principle: *the inertial mass of a body is due to its interaction with all the other masses in the universe*. Brans abandoned this principle in his critic of these models.

In this paper we also contradict the Mach's principle by considering a constant inertial mass, and identical active and passive mass depending on the gravitational interaction. A satisfactory extension to general relativity of the inertial mass of a point-like particle can be proposed, in the framework of the H-G work [2], by linking the concept directly to the Einstein's field equations (in this way it is linked also to the equations of motion). To follow the process it will be convenient to recall the two world line parametrizations used by H-G, namely, the physical proper time $\mathrm{d}s=\sqrt{g_{\alpha\beta}\mathrm{d}z^{\alpha}\mathrm{d}z^{\beta}}$, and the Minkowski proper time $\mathrm{d}\tau=\sqrt{\eta_{\alpha\beta}\mathrm{d}z^{\alpha}\mathrm{d}z^{\beta}}$. We shall denote by $u_a^{\mu}=\frac{\mathrm{d}z^{\mu}}{\mathrm{d}s_a}$ and $v_a^{\mu}=\frac{\mathrm{d}z^{\mu}}{\mathrm{d}\tau_a}$ the tangent vectors to the world line of the particle "*a*", corresponding to both parametrizations. Consequently, $g(u_a,u_a)=1$, $\eta(v_a,v_a)=1$. From the first equation one gets

$$\frac{\mathrm{d}\tau_a}{\mathrm{d}s_a}=\frac{1}{\sqrt{g(v_a,v_a)}},\quad g(v_a,v_a)=g_{\mu\nu}\left(x(\tau_a)\right)v_a^{\mu}v_a^{\nu} \tag{1}$$

We propose to define the particle's inertial mass m_a and enunciate at once the field equations for a system of point-like particles:

Given a system of N self-gravitating particles there exist N real numbers $\{m_a, a=1,\cdots,n\}$, *such that the metric g of the spacetime is a solution of the Einstein's equations, with a distribution with support on the world lines of the particles as energy-momentum tensor*:

$$G^{\mu\nu}(g)=8\pi G\sum_{a=1}^{N}\int m_a u_a^{\mu}u_a^{\nu}\delta^{(4)}\left(x-z_a(s_a)\right)\mathrm{d}s_a,\quad u_a^{\mu}=\frac{\mathrm{d}z_a^{\mu}}{\mathrm{d}s_a}. \tag{2}$$

We shall justify bellow to denote inertial mass to these constants. In order to define the passive gravitational mass we consider the particle's four-momentum $p_a=m_a u_a, a=1,2,\cdots,N$. By a simple calculation:

$p_a^{\mu}=m_a u_a^{\mu}=m_a\frac{\mathrm{d}\tau_a}{\mathrm{d}s_a}\frac{\mathrm{d}z^{\mu}}{\mathrm{d}\tau_a}=m_a\frac{\mathrm{d}\tau_a}{\mathrm{d}s_a}v_a^{\mu}$, and substituting (1) we obtain

$p_a^{\mu}=M_a(\tau_a)v_a^{\mu}$, with $M_a(\tau_a)$ related to m_a:

$$M_a(\tau_a)=\frac{m_a}{\sqrt{g(v_a,v_a)}} \tag{3}$$

coinciding with equation (16) in [2]. One can now interpret physically the constants m_a and the functions $M_a(\tau_a)$ using the equations of motion derived by Havas-Goldberg from the null divergence of the energy momentum tensor (Equation (14) in [2]):

$$\frac{\mathrm{d}p_{a\mu}}{\mathrm{d}\tau_a}=F_{a\mu},\quad F_{a\mu}=\frac{1}{2}M_a(\tau_a)\frac{\partial\left(g(v_a,v_a)\right)}{\partial x_a^{\mu}} \tag{4}$$

that can be rewritten as:

$$m_a\frac{\mathrm{d}u_{a\mu}}{\mathrm{d}\tau_a}=M_a(\tau_a)\frac{\partial U_a}{\partial x_a^{\mu}},\quad U_a=\frac{1}{2}g(v_a,v_a) \tag{5}$$

to the equation of motion of a particle in special relativistic mechanics, justifying in this way to consider the constants m_a as the inertial masses, and prompting us to interpret the functions $M_a(\tau_a)$ given in (3) as the passive gravitational masses of the particles. The aim of this paper is to show that this gravitational mass can explain the dark components and the cosmological constant (Inexplicably, H-G [2] did not used this concept).

Let us outline that the equations of motion (4) or (5) (suppressing subindex a for the sake of simplicity) are equivalent to the geodesic equations. The geodesics equations generated by the Lagrangian $L = g_{\alpha\beta}\dot{x}^\alpha \dot{x}^\beta$, where dots denote derivatives respect the proper time are $\frac{\mathrm{d}}{\mathrm{d}s}\left(g_{\mu\nu}\dot{x}^\nu\right) = \frac{1}{2}\frac{\partial g_{\alpha\beta}}{\partial x^\mu}\dot{x}^\alpha \dot{x}^\beta$. Now, inserting the con-vector $p_\mu = g_{\mu\nu}p^\nu$ we can write: $\frac{\mathrm{d}p_\mu}{\mathrm{d}s} = \frac{m}{2}\frac{\partial g_{\alpha\beta}}{\partial x^\mu}\dot{x}^\alpha \dot{x}^\beta$, and changing the physical proper time to the Minkowski proper time we get the Equation (4)
$\frac{\mathrm{d}p_\mu}{\mathrm{d}\tau} = \frac{1}{2}m\frac{\mathrm{d}\tau}{\mathrm{d}s}\frac{\partial g_{\alpha\beta}}{\partial x^\mu}v^\alpha v^\beta = \frac{1}{2}M(\tau)\frac{\partial g_{\alpha\beta}}{\partial x^\mu}v^\alpha v^\beta$.

The question arises if the definitions of mass given in Equations (2), (3) satisfy the weak equivalence principle (WEP). The answer is affirmative. In terms of the Hammer's book [1], pg. 103, the WEP for a point-like particle is stated as follows: "the world line of a particle, released at an initial space-time event with a given velocity is independent of the weight" (it means the passive mass).The WEP defines thus a preferred set of geodetic curves, as we have proven above.

Let us add a comment to the Jammer's distinction of two versions of WEP: the kinematic $\mathrm{WEP}_{\mathrm{kin}}$, states that at a given location all bodies fall with the same acceleration (principle of universality of free fall), and the dynamic $\mathrm{WEP}_{\mathrm{dyn}}$ states that the ratio m_p/m_i is the same for all the particles, or in appropriate units $m_i = m_p$.

When considering the torsion balance used in Eötvös-like experiments, the quotient $M_a/m_a = 1/\sqrt{g_{00}}$ is a constant, and the same for any of the two small masses of the balance, because g_{00} corresponds to the metric produced by a static earth. So, in this particular case the $\mathrm{WEP}_{\mathrm{dyn}}$, in the sense of identity of inertial and gravitational mass, is fulfilled.

However, in subsection IIIC, where we study the Milne's universe with mass and prove the equality active and passive gravitational mass, we shall find that the quotient $M_a(\tau_a)/m_a$ is the same for all the particles, and therefore all them with the same initial conditions move equal in the gravitational field in accord with the $\mathrm{WEP}_{\mathrm{kin}}$, but the quotient of masses is a monotonous increasing function of the expansion factor (see Equation (23)) and the $\mathrm{WEP}_{\mathrm{dyn}}$ does not imply the equality of inertial and gravitational mass by choosing appropriated units. In section IV we shall develop, using the gravitational mass concept, far reaching cosmological consequences.

3. A Cosmological Model Built with Point-Like Particles

In this section we study two properties of the gravitational mass: the equivalence between active and passive gravitational mass, and finally their time dependence presented in Equation (18); but, the main objective is to construct a cosmological model composed of self-gravitating point-like particles. A cosmological model is a spacetime locally isotropic everywhere respect a cosmological observer [16], whose main characteristics we summarize in the next subsection with notations that will be useful to describe both, the standard model with dark components, and our model based uniquely on the gravitational mass defined in the previous section. It will be useful to start recalling the Milne's universe without mass, described in [16], and then to construct a Milne's universe with mass in section IIIC.

1) The FRLW cosmological models

A cosmological model may be characterized by an index of curvature ($k = 1, 0, -1$), and giving the energy density $\rho_F(a)$ as function of the expansion factor, that will be convenient to write in the form

$$\rho_F(a) = \frac{3H_0^2}{8\pi G}\left(\frac{\Omega_M}{a^3} + f(a)\right), \tag{6}$$

with $H_0 = \dot{a}/a\big|_{a=1}$ and $\Omega_M + f(1) = 1$. In the current cosmological model the first and second summands describe the unknown pressure-less dark matter and dark energy densities respectively; the cosmological constant Λ may be considered as an important particular case of dark energy, with $f(a) = \frac{\Lambda}{3H_0^2}$, because a successful cosmological model, the ΛCDM model, has been developed with this assumption [17]. The model presented in this paper with a non constant $f(a)$ is equally successful as we show in IV; but, it has the advantage that the dark components may be accounted for with the gravitational mass concept. The Friedmann equation:

$$\frac{\mathrm{d}a}{\mathrm{d}s} = H_0 a\sqrt{\frac{\Omega_M}{a^3} + \frac{\Omega_k}{a^2} + f(a)},\ \Omega_k = -\frac{k}{H_0^2 R_0^2}, \tag{7}$$

determines the expansion factor as function of the cosmological time s. Henceforth we shall take $k = 0$. Taking into account the equation of continuity $\dot{\rho}_F + 3H(\rho_F + p) = 0$, and well known calculations, one obtains the pressure and acceleration of the model:

$$p_F(a) = \frac{3H_0^2}{8\pi G}\left(-f(a) + \frac{1}{3}a\frac{\mathrm{d}f}{\mathrm{d}a}\right) \tag{8}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho_F + 3p_F) = -H_0^2\left(\frac{\Omega_M}{2a^3} - f(a) + \frac{1}{2}a\frac{\mathrm{d}f}{\mathrm{d}a}\right) \tag{9}$$

In the ΛCDM model, Ω_M is the sum of the baryonic and dark matter density fractions: $\Omega_M = \Omega_{ba} + \Omega_{dm}$; in our model $\Omega_M = \Omega_{ba}(1+\alpha)$ and, accordingly with the interpretation of α we shall give bellow in Equation (13), $\frac{3H_0^2}{8\pi G}\frac{\Omega_{ba}\alpha}{a_i^3}$ is the gravitational mass density present at the beginning of the galactic dominance epoch, due to the gravitational interaction of the particles that collapsed to form the galaxies in the precedent epoch with $a < a_i$.

2) The Milne's universe

It is a portion of the Minkowski spacetime formed by a future half-light cone filled with a set of straight time like lines concurrent on its vertex. The world lines are the integral curves of the cosmological observer, and their points of intersection with any hyperboloid of constant proper time τ define an homogeneous Poisson process, with constant number density equal to N/τ^3, being N a constant; however, the number density of the intersections with the hypersurface of inertial coordinate t = constant is not uniform: $Nt\left(t^2-r^2\right)^{-2}$. The change of coordinates $t=\tau\cosh\Psi$, $r=\sinh\Psi$ produces the metric for the Milne's universe $\mathrm{d}s^2=-\mathrm{d}\tau^2+\frac{1}{\tau^2}\left(\mathrm{d}\psi^2+\sinh^2\psi\left(\mathrm{d}\theta^2+r^2\sin^2\theta\mathrm{d}\psi^2\right)\right)$. That is a cosmological model without mass and therefore with null energy-momentum tensor.

3) The Milne's universe with mass

We consider now the concurrent straight lines of the Milne's universe as world lines of particles (galaxies) with equal inertial mass m, and use the metric obtained by H-G [2] at first order in G in harmonic coordinates. To describe the recent epoch of galactic dominance we must take into account the evolution of the universe from small perturbations to bigger condensations of matter. We shall do that, with extreme simplification, by assuming that the galaxies began to dominate at some initial Minkowski proper time τ_i. For times $\tau<\tau_i$ the matter content was dominated by other particles ("first stars") that finally did aggregate to form the galaxies of the more recent universe. To complete the Einstein's equations the energy momentum tensor is written in the form of a distribution with support over the world lines of the particles, parametrized with the Minkowskian proper time, as given by H-G in Equations (29) and (13) of [2]

$$\sqrt{-g}T^{\mu\upsilon}=\sum_{a=1}^{N}\int_{\tau_i}^{\tau_0}M_a\left(\tau_a\right)v_a^\mu v_a^\nu\delta^{(4)}\left(x-z_a\left(\tau_a\right)\right)\mathrm{d}\tau_a,\quad v_a^\mu=\frac{\mathrm{d}z_a^\mu}{\mathrm{d}\tau_a}. \tag{10}$$

We have denoted with a subindex the different particles, though all the inertial masses are equal: $m_a=m$. The quantity $M_a(\tau_a)$ is the passive gravitational mass introduced in the previous section by the Equation (3).

The energy density $\rho(x)$ of the fluid formed by the system of interacting particles can be obtained as result of the action $\left(\rho_{pp},\phi_x\right)$ of the point particle distribution $\rho_{pp}=\sqrt{-g}T^{\mu\nu}u_\mu u_\nu$ over a convenient test function ϕ_x defined as follow: let A be the neighbourhood of the point x represented in **Figure 1**, defined by two neighbour hypersurfaces: $\tau=t+\Delta t/2$ (that are also hypersurfaces of constant proper time: $s=s(t)+\Delta s/2$, $s=s(t)-\Delta s/2$) and a thin time like cone; and let S be the intersection of the hypersurface $\tau=t$ with the neighbourhood A. We choose as test function $\phi(x)$ the characteristic function of the set A: $\phi_x(u)=1$ if $u\in A$, $\phi_x(u)=0$ otherwise. One can prove [16], by averaging over the ref line L, that

$$\rho(x)=\lim_{\Delta t\to 0}\frac{\left(\sqrt{-g}T^{00},\varphi_x\right)}{Vol^{(3)}(S)\Delta t}=n(t)M(t) \tag{11}$$

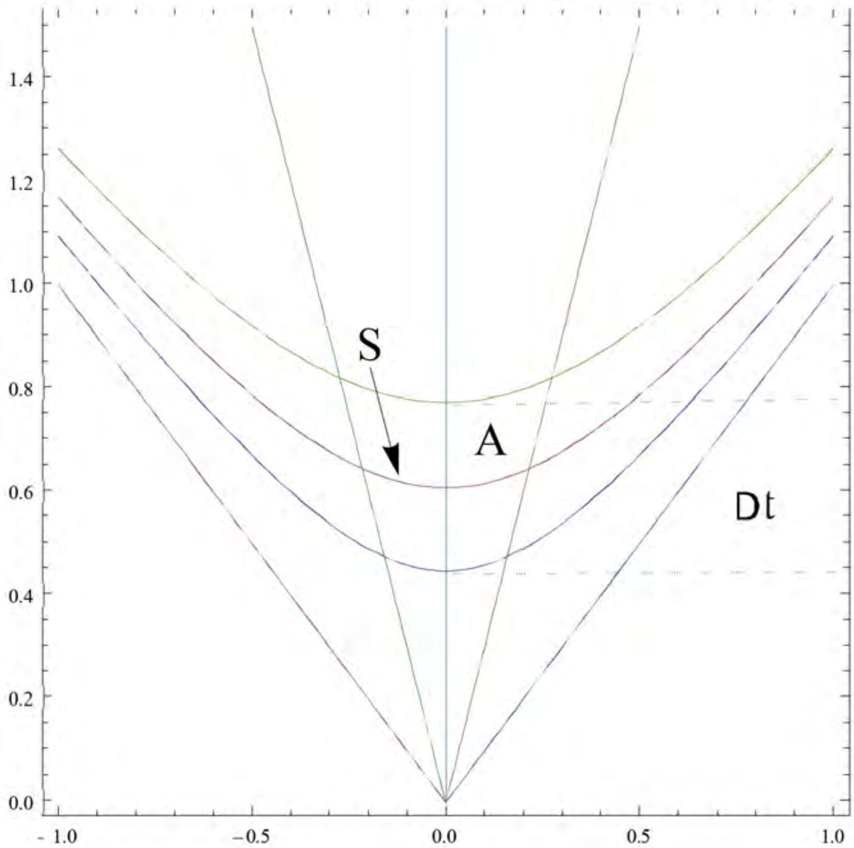


Figure 1. Element of volume A centered at a point over a reference world line, limited by two surfaces $\tau = const.$ in a Milne's universe. One shows the tridimensional surface S necessary to estimate the mass density.

Therefore $M(t)$ is also the active gravitational mass. This result proves that passive and active gravitational mass coincide, because according to equation (5), $M(t)$ was the passive gravitational mass. Henceforth we shall refer to $M(t)$ as the gravitational mass.

The iterative method involving a power series in G developed in [2] allows to determine the metric and the gravitational mass in the form of two functional series: $g_{\alpha\beta} = \eta_{\alpha\beta} + g^{(1)}_{\alpha\beta} + \cdots$, and $M_a(\tau_a) = m_a + M^{(1)}_a(\tau_a) + \cdots$. A regularization is necessary at each order of approximation, because the metric diverges over the world lines of the particles. In our case, a great simplification is that in harmonic coordinates the world lines of the particles are still straight lines, despite the fact of having mass. The solution at first order in G obtained by H-G (eq. 62 and 46) is the following:

$$g^{(1)}_{\mu\nu}(z_a(\tau)) = -4G \sum_{b \neq a} \frac{m_b \left(\eta_{\mu\alpha} \eta_{\upsilon\beta} v^{\alpha}_b v^{\beta} - \frac{1}{2} \eta_{\mu\nu} \right)_{ret}}{\eta(x - z_b, v_b)_{ret}} \tag{12}$$

$$M^{(1)}_a(\tau_a) = -\frac{1}{2} m_a g^{(1)}_{\alpha\beta} v^{\alpha} v^{\beta} + m_a \alpha \tag{13}$$

where in Equation (13) we have written as $m_a \alpha$ the constant of integration noted as ${}_2C_i$ by H-G in their Equation (46).

We interpret it as the gravitational mass present at time τ_i: $m_a\alpha = M_a^{(1)}(\tau_i)$, *i.e.*, as the gravitational mass acquired by the particles that collapsed during the anterior phase to form a galaxy in the present epoch; and bellow, in subsection IIIC1, this quantity will be identified with the detected dark matter.

As all the world lines of the Milne's universe are equivalent, by the Lorentz invariance of the equations, we shall obtain the physical metric and the gravitational mass over a world-line of reference *L*: $x^\alpha(t) = (t,0,0,0)$, with null three-velocities $v^k = 0$. Over this line the initial Minkowskian proper time τ_i ($\tau_i^2 = t_i^2 - |x_i|^2$) coincides with the initial harmonic time t_i. The gravitational mass at a point $p = (t > t_i, 0) \in L$ is $M(t) = m\alpha + m\left(1 - \frac{1}{2} g_{00}^{(1)}(p)\right)$, with

$$g_{00}^{(1)}(p) = -G\sum_{b\neq a} \frac{m_b\left(\gamma_b^2 - 1/2\right)}{r_b\gamma_b\left(1+\bar{v}_b\right)},\ \bar{v}_b = \frac{r_b}{t},\ \gamma_b = \frac{1}{\sqrt{1-\left(\bar{v}_b\right)^2}}. \tag{14}$$

The number of summands is finite even if the universe is made of infinite particles because only intervene the world lines intersecting the part of the past light cone of the point *p* limited by the hyperboloid Σ_{τ_i}, as shown in **Figure 2**. The

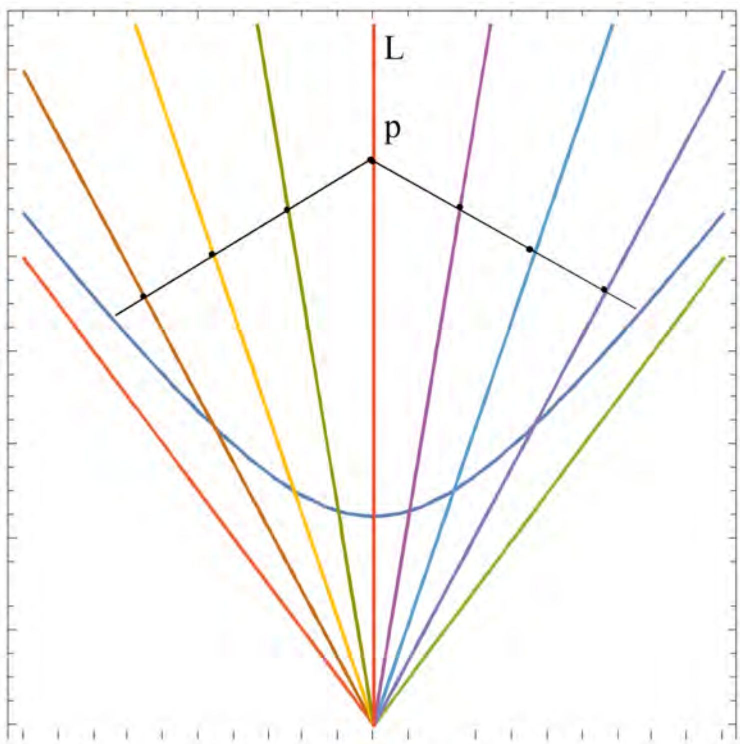


Figure 2. The past light cone at the point $p = (t, 0), t > \tau_i$ on the reference world-line *L* intersects a finite number of world lines, limited by the hyperboloid $\tau = \tau_i$. When *t* increases the number of particles contributing to its gravitational mass augments, and as a result thereof, an increment of their gravitational mass. The cosmological expansion increments the volume enclosing a given number of particles in a way (described in section IVA) that the gravitational mass density rapidly tends to a constant value, that we shall identify with the cosmological constant.

sum in Equation (14) was estimated by integration in [10], hence we can obtain the metric over the reference line L for $t > t_i$, and the ratio $\frac{\mathrm{d}t}{\mathrm{d}s} = \frac{1}{g(v,v)}$ up to first order in G:

$$\frac{\mathrm{d}t}{\mathrm{d}s} = 1 - \frac{1}{2}\left\langle g_{00}^{(1)}(t)\right\rangle = 1 + \frac{\pi GNm}{3t_i^3} t^2 g(t,t_i) \tag{15}$$

$$\left\langle g_{00}^{(1)}(t)\right\rangle = -\frac{2\pi GNm}{3t_i^3} t^2 g(t,t_i), \quad g(t,t_i) = \left(1 + \frac{t_i^2}{t^2}\right) - 4\frac{t_i^2}{t^2}\left(\frac{3}{2}\left(1 + \frac{t_i^2}{t^2}\right) - \frac{t_i}{t}\right) \tag{16}$$

where $\frac{N}{t_i^3}$ is the number density introduced in IIIB. It is evident from the last equation that $g(t_i,t_i)$. Let us identify in Equation (15) the two first terms of a series: $\frac{\mathrm{d}t}{\mathrm{d}s} = \sum_{l=0}^{\infty} F^{(l)}(t)$ with $F^{(0)}(t) = 1$ and

$$F^{(1)}(t) = \frac{\pi GNm}{3t_i^3} t^2 g(t,t_i). \tag{17}$$

The time evolution of the passive gravitational mass is $M(t) = m\alpha + m\frac{\mathrm{d}t}{\mathrm{d}s}$; and, taking into account (15) can be expressed in the form:

$$M(t) = m\left(1 + \alpha + \frac{\pi GNm}{3t_i^3} t^2 g(t,t_i) + O\left(G^2\right)\right), \tag{18}$$

that verifies $M(t_i) = m + m\alpha$. Let us recall that the constant $m\alpha$ introduced in Equation (13) was interpreted as the passive gravitational mass acquired during the epoch $a < a_i$. Multiplying (18) by the number density of particles we get the energy density valid for the recent universe, at redshift $z < z_i$:

$$\rho(t) = n(t)m\left(1 + \alpha + \frac{\pi GNm}{3t_i^3} t^2 g(t,t_i) + O\left(G^2\right)\right). \tag{19}$$

a) *Construction of a cosmological model with gravitational mass*

As we are considering a model made exclusively of baryons, the term baryonic mass should refer to the constant inertial mass m introduced in (2) and $\Omega_{ba} = \frac{8\pi G}{3H_0^2} n_o m$, be interpreted as the dimensionless baryonic density parameter.

The baryonic matter, as shown in **Figure 2**, is continuously acquiring gravitational mass (3) due to the gravitational interaction. To obtain the cosmological model, denoted above as Milne's universe with mass, we need to transform the harmonic coordinates used in section II into standard cosmological coordinates, but that is not so simple as was in the Milne's universe without mass treated in section IIIB. We shall assume now that $t(a)$ may be expressed as a series $t(a) = \sum_{k=0}^{\infty} t^{(k)}(a)$ (though we shall need only the first term $t^{(0)}$) and consider that:

1.It is verified $n(t(a)) = \frac{n_o}{a^3}$, with n_o equal to the number density of baryons at the present epoch. Let us recall that $\frac{N}{t_i^3}$ is the constant number density of particles over the hyperboloid $\tau_i = t_i$, introduced in IIIB, therefore we can write $\frac{N}{t_i^3} = \frac{n_o}{a_i^3}$.

b) The Equations (15) and (7), with $k = 0$, determine the functions $t = t(s)$ and $s = s(a)$ up to first order in G.

First we identify in (15) the two first terms of the series $\frac{\mathrm{d}t}{\mathrm{d}s} = \sum_{l=0}^{\infty} F^{(l)}(t)$, namely: $F^{(0)}(t) = 1$, and $F^{(1)}(t) = \frac{\pi G N m}{3t_i^3} t^2 g(t, t_i)$. Then, from Equation (7) we have $\frac{\mathrm{d}s}{\mathrm{d}a} = \frac{1}{H_0 a}\left(\frac{\Omega_M}{a^3} + f(a)\right)^{-1/2}$ and using $\frac{\mathrm{d}t}{\mathrm{d}a} = \frac{\mathrm{d}s}{\mathrm{d}a}\frac{\mathrm{d}t}{\mathrm{d}s}$ we obtain the differential equation that determines the series $t(a) = \sum_{l=0}^{\infty} t^{(l)}(a)$

$$\frac{\mathrm{d}t}{\mathrm{d}a} = \frac{1}{H_0 a}\left(\frac{\Omega_M}{a^3} + \frac{\Omega_{ba}}{a^3}\sum_{l=1}^{\infty} F^{(l)}(t(a))\right)^{-1/2} \sum_{l=0}^{\infty} F^{(l)}(t(a)). \tag{20}$$

To order zero one obtains $\frac{\mathrm{d}t^{(0)}}{\mathrm{d}a} = \frac{1}{aH_0}\sqrt{\frac{a^3}{\Omega_M}}$, whose solution is

$$H_0 t^{(0)}(a) = \frac{2a^{3/2}}{3\Omega_M^{1/2}}, \tag{21}$$

and substituting (21) into (16) we have $g\left(t^{(0)}(a), t_i\right) = \tilde{g}(a, a_i)$ with

$$\tilde{g}(a, a_i) = \left(1 + \frac{a_i^3}{a^3}\right)^3 - 4\frac{a_i^3}{a^3}\left(\frac{3}{2}\left(1 + \frac{a_i^3}{a^3}\right) - \frac{a_i^{3/2}}{a^{3/2}}\right). \tag{22}$$

Using the prescription 1 given above we get $n(t(a))m = \frac{3H_0^2}{8\pi G}\frac{\Omega_{ba}}{a^3}$, and taking into account (18) and (21) we obtain the dependence of the gravitational mass of a galaxy on the expansion factor, valid for the galactic dominance epoch $a > a_i$:

$$M(a) = m\left(1 + \alpha + \frac{G\Omega_{ba}}{18\Omega_M}\frac{a^3}{a_i^3}\tilde{g}(a, a_i)\right). \tag{23}$$

It is manifest that the quotient $M(a)/m$ is independent of m, but not a constant. This fact prevents from identifying inertial and gravitational mass as quoted at the end of section II, but just this un-equality, far of been a drawback, is the clue to explain in the next section the origin of the cosmological dark components. Now, we can construct a cosmological model with dominant gravitational mass. By considering Equations (19) and (21) we obtain $\rho_F(a) := \rho(t(a))$, with

$$\rho_F(a) = \frac{3H_0^2}{8\pi G}\frac{\Omega_{ba}}{a^3}\left(1 + \alpha + \frac{\Omega_{ba}}{\Omega_M}\frac{a^3}{18a_i^3}\tilde{g}(a, a_i)\right) + O\left(G^2\right) \tag{24}$$

that can be written in the form (6) to identify Ω_M and $f(a)$ as follow:

$$\rho_F(a) = \frac{3H_0^2}{8\pi G}\left(\frac{\Omega_M}{a^3} + f(a)\right) \tag{25}$$

$$\Omega_M = \Omega_{ba}(1+\alpha),\quad f(a) = \frac{\Omega_{ba}^2 \tilde{g}(a,a_i)}{18\Omega_M a_i^3}. \tag{26}$$

Our model has introduced three parameters: Ω_{ba}, α, a_i, but the identity $\Omega_M + f(1) = 1$ gives a relation between them:

$$\Omega_{ba}(1+\alpha) + \frac{\Omega_{ba}\tilde{g}(1,a_i)}{18(1+\alpha)a_i^3} = 1. \tag{27}$$

The function $\tilde{g}(a,a_i)$ vanishes at $a = a_i$ and rapidly tends to unity for increasing a. From $\tilde{g}(a_i,a_i) = 0$ we obtain, using Equations (23) and (25), that $M(a_i) = m(1+\alpha)$ and $\rho_F(a_i) = \frac{n_o}{a_i^3} m(1+\alpha)$; hence, we conclude that: at the beginning of the galactic epoch, $m\alpha$ is the gravitational mass contained in a galaxy, and $\frac{n_o m\alpha}{a_i^3}$ the gravitational mass density. Both were generated during the anterior epoch, with $a < a_i$, by the gravitational interaction of the particles that collapsed to form a galaxy.

4) Comparison of the model with gravitational mass with the ΛCDM cosmological model

The success of the ΛCDM cosmological model has been corroborated by the Planck 2013 results [17]. This model considers null pressure and introduces the cosmological constant Λ, by substituting $f(a) = \Lambda/3H_0^2 \equiv \Omega_\Lambda$ into Equation (6), to get $\rho_F = \frac{3H_0^2}{8\pi G}\left(\frac{\Omega_M}{a^3} + \Omega_\Lambda\right)$. The Equations (8) and (9) imply a constant negative pressure and a variable acceleration: $p = -\frac{3H_0^2}{8\pi G}\Omega_\Lambda$, $\frac{\ddot{a}}{a} = -\frac{H_0^2}{2}\left(\frac{\Omega_M}{a^3} - 2\Omega_\Lambda\right)$. The density parameter Ω_M contains baryon and dark matter contributions: $\Omega_M = \Omega_{ba} + \Omega_{dm}$. The model works well, fitting the observed supernovae moduli-distance redshift relation [18], and the unexpected recent transition from decelerated to accelerated universe at redshift $z = 0.6$.

However, the problematic physical interpretation of the cosmological constant as vacuum energy density, caused the introduction of the new physical field dubbed dark energy. A comprehensive review of the new dark components can be found in the books [19] [20]. Assuming $h = 0.67$ for the reduced Hubble constant, one gets estimations for the density fractions: $(\Omega_{ba}, \Omega_{dm}, \Omega_\Lambda) = (0.049, 0.268, 0.683)$.

Let us give our interpretation of the dark matter parameter Ω_{dm} in the ΛCDM model: our model use the parameter $\Omega_M = \Omega_{ba}(1+\alpha)$, where we have substituted $\Omega_{ba}\alpha$ in place of Ω_{dm}. Accordingly to the interpretation of $m\alpha$ at the end of the anterior section, $\frac{n_o}{a_i^3} m\alpha \equiv \frac{3H_0^2}{8\pi G}\frac{\Omega_{ba}\alpha}{a_i^3}$ is the gravitational mass density at the beginning of the galactic dominance epoch, acquired by the particles dominant in the precedent epoch with $a < a_i$.

Finally, although we will not use the cosmological constant as parameter, in the next section we shall give a physical interpretation of this constant as the asymptotic value of the gravitational mass density.

4. Gravitational Mass Density in the Galactic Dominance Epoch: An Interpretation of Dark Matter and Dark Energy

The objective of this section is to show that the notion of gravitational mass, introduced in this paper to construct the Milne's universe with mass, affords a satisfactory explanation, at cosmological scales, of both dark components.

1) Gravitational mass instead of dark matter, cosmological constant, and dark energy

The gravitational mass fraction $f(a)$ obtained in (26) depends of three parameters (Ω_{ba}, α, a_i) constrained by the Equation (27). We can obtain good values for them identifying our parameter $\Omega_M = \Omega_{ba}(1+\alpha)$ with the equivalent in the ΛCDM model $\Omega_M = \Omega_{ba} + \Omega_{dm}$ described in IIID, obtaining the relations $\Omega_{ba} = 0.049$, and $\Omega_{ba}\alpha = \Omega_{dm} = 0.268$ and from the last one we get $\alpha = \frac{\Omega_{dm}}{\Omega_{ba}} = 5.47$. To determine the remaining parameter a_i we observe that, as it is shown in **Figure 3**, $f(a)$ is a monotonous increasing function of the expansion factor, verifying $f(a_i) = 0$, that rapidly tends to a constant value $f(1)$. If we identify $f(1) = \Lambda/3H_0^2 = \Omega_\Lambda = 0.683$ we obtain the equation $\frac{g(1,a_i)}{a_i^3} = \frac{18\Omega_M\Omega_\Lambda}{\Omega_{ba}^2}$, that determines the beginning of the galactic epoch at $a_i = 0.085$, that corresponds to redshift $z_i = 10.76$. This is our explanation of the cosmological constant Λ as a limit value of the gravitational mass density introduced in this paper, that makes unnecessary to surmise the existence of a new physical field, so-called dark-energy. This is a physical interpretation of the cosmological constant Λ as the limit of the gravitational mass density. This interpretation of the constant Ω_Λ has no problem, unlike the dark energy assumption, with the coincidence of the densities $\frac{\Omega_m}{a^3}$ and $f(a)$ at a so recent epoch as $z_i = 0.3$ [20], because in our

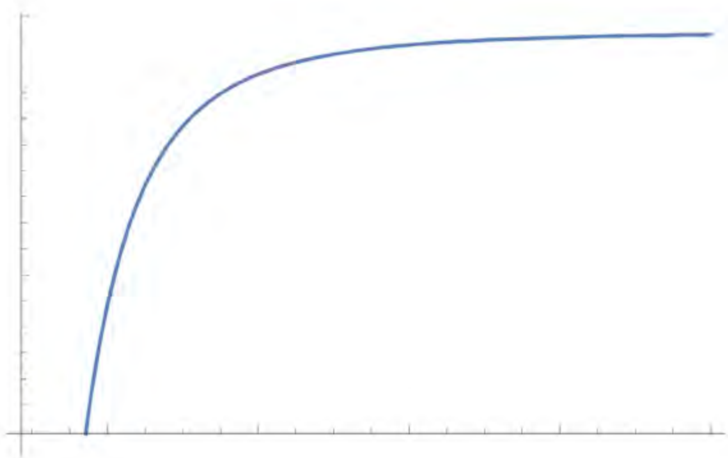


Figure 3. The gravitational mass fraction as function of the expansion factor. The gravitational mass fraction ($f(a)$ = gravitational mass density/ρ_{cri}) of our model is a function of the expansion factor a. In a neighbourhood of $a = 1$ the function $f(a)$ is practically constant.

case, $\frac{\Omega_M}{a^3}$ and $f(a)$ should not be considered as two unrelated magnitudes, rather they are two components of a sole gravitational mass density. Though $f(a_i = 0)$, $f(a)$ increases rapidly, the $\frac{\Omega_m}{a^3}$ is a decreasing function, and the graphs of both components cross for $a_c = \frac{1}{1.3}$.

In the next subsections we show how using the values $(\Omega_{ba}, \alpha, a_i) = (0.049, 5.47, 0.085)$ obtained in this section, one can reproduce with great exactitude the main predictions of the ΛCDM model.

2) The moduli-distance redshift relation

With the gravitational mass density $f(a)$ we can explain the luminosity distance $d_L(z)$, or the equivalent logarithmic moduli-distance $\mu(z) = 5\log d_L(z) + 25$, to a source with redshift *z*. The difference between our prediction $\mu_f(z)$ and the ΛCDM prediction $\mu_\Lambda(z)$ for the supernovae observations accounted in [18] is of the order of 10^{-5}. The discrepancy, $\frac{\mu_f(z) - \mu_\Lambda(z)}{\mu_\Lambda(z)}$, between both models is represented in Figure 4.

3) The accelerated universe

With the Equation (9) without cosmological constant it is impossible to explain an accelerated universe, unless the pressure of the cosmological fluid be negative, but an important consequence of the gravitational interaction between the particles is that the gravitational mass density, $f(a)\frac{3H_0^2}{8\pi G}$, given in (26), when substituted into the Equations (8) and (9), gives the necessary negative pressure. We can conclude that the gravitational pressure is the cause of the

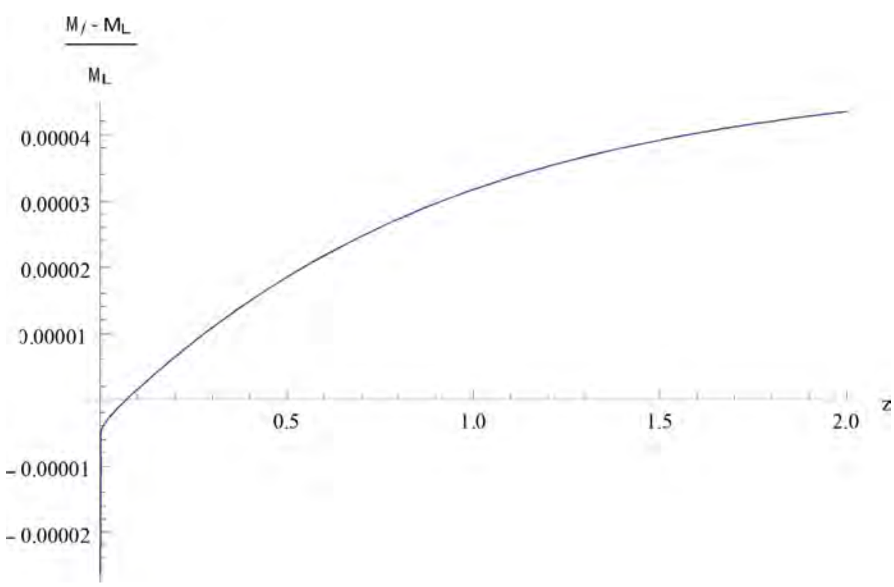


Figure 4. Our prediction of the moduli-distance redshift relation about the supernovae observations [18], is practically indistinguishable from the one based on the ΛCDM cosmological model.

acceleration of the universe. As it is shown in **Figure 5**, the deceleration factor $q=-\ddot{a}a/\dot{a}^2$ becomes negative recently because $f(a)$ corresponds to the galactic dominance epoch with expansion factor $a_i < a < 1$.

4) The pressure to density ratio $w(a)$

Let us decompose in two summands the energy density given in (25), (26): $\rho_F(a)=\rho_M(a)+\rho_f(a)$, with $\rho_M(a)=\frac{3H_0^2}{8\pi G}\frac{\Omega_M}{a^3}$ and $\rho_f(a)=\frac{3H_0^2}{8\pi G}f(a)$. As only the component $\rho_f(a)$ contributes to the pressure given in (8), it is worthwhile to know our prediction for the ratio $w_f(a)=\frac{p(a)}{\rho_f(a)}$. Using (8) one gets:

$$w(a)=\frac{a}{3}\frac{\mathrm{d}\ln f}{\mathrm{d}a}-1. \tag{28}$$

In **Figure 6** we have represented the function $w(a)$ predicted by our model. It is manifest a linear dependence for values close to unity (very low redshifts) and a rapid increase when a decreases (for high redshifts).

The variability of the ratio $w(a)$ has been tested in [21] under a linear dependence hypothesis: $w(a)=w_0+(1-a)w(a)$, obtaining the constraint $-1.33 < w_0 < -0.79$ for the present value $w_0=w(1)$ of the pressure density ratio. With equation (28) we get $w(1)=-0.998$ that satisfies the constraint, but the linear dependence in our prediction, showed in **Figure 6**, clearly fails for $a<0.4$ ($z>1.5$). Let us remark that in the today vast literature on dark energy one has introduced, besides the scalar field language with the quintessences, some barotropic fluid models as origin of the acceleration [22], whose $w(a)$ ratio are qualitatively similar to our prediction. The pressure given in (8) has dynamic

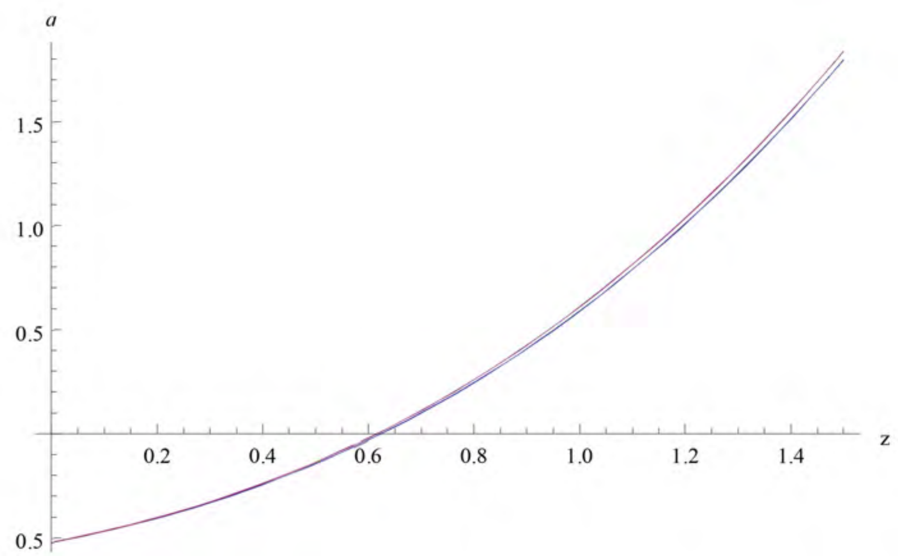


Figure 5. The deceleration factor q becomes negative recently at redshift 0.6. The upper curb corresponds to our model, as a consequence of the gravitational energy density $f(a)$, the lower one is the prediction made by the ΛCDM model based on the assumption of a cosmological constant.

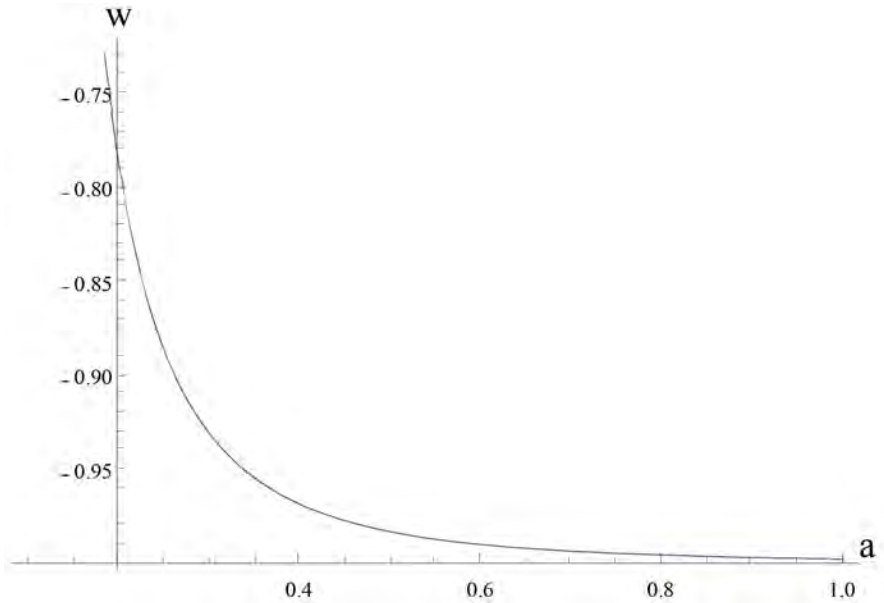


Figure 6. This figure shows the pressure to density ratio $w(a) = p(a)/\rho_t(a)$ predicted by our model for the galactic dominance epoch, $\frac{1}{11.7} \le a \le 1$.

character rather than kinetic, because it is linked to the gravitational energy density $f(a)$ described in section IVA. The potential energy contribution to a dynamic pressure is well known in classical mechanics, see Chp.17.2 in [23] and Chp. 7.4 in [24], but little is known about that in general relativity.

5) Time evolution of the gravitational mass fraction

A very interesting result comes from Equation (23), that describes the dependence of the gravitational mass of a point-like particle on the expansion factor. We can estimate the evolution of the gravitational mass fraction of a generic galaxy with the expression

$$f_{GM}(a) = \frac{M(a) - m(1+\alpha)}{M(a)}. \tag{29}$$

Our interpretation of dark matter as the gravitational mass acquired by the particles that collapsed to form a galaxy implies that we must compare $f_{GM}(a)$ with dark matter observations at different redshifts. The observed redshift dependence of the dark matter fraction $f_{DM}(z)$ can be accounted for with the function $f_{GM}(a)$, after substituting $a = \frac{1}{1+z}$, showed in **Figure 7** as a continuous curve. The two first fractions were observed by Dutton [25] in 2011 and Suyu [26] in 2012, and the last six, correspond to more recent observations by Gencel *et al.* [27].

5. Conclusions

We have used P. Havas and J.N. Goldberg results on dynamics of a finite number of gravitating point-like particles to revise the concepts of inertial and gravitational

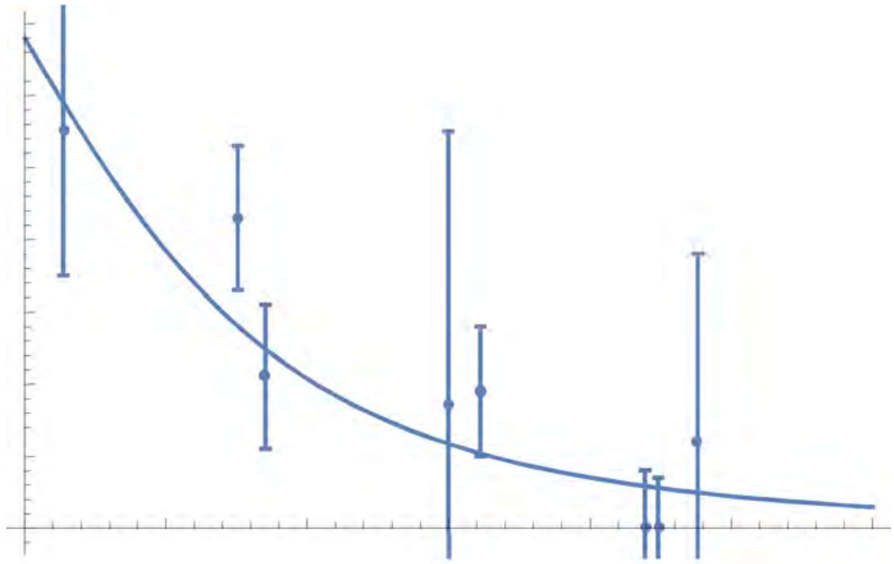


Figure 7. The gravitational mass fraction $f_{GM}(z)$ as function of the redshift calculated with our model for point-like particles fits well to a sample of observed values in different galaxies up to redshifts of the order 2.

mass in general relativity and derive their cosmological consequences:

a) *Inertial and gravitational mass*. The law of motion of a particle (5) suggests definitions for a constant inertial mass, m, and a passive gravitational mass, $M = m\big/\sqrt{g(v,v)}$, depending on the metric generated by the particles. In section IIIC, we proved the equivalence of passive and active gravitational mass, and obtained its monotonous increasing dependence on the expansion factor $M(a)$. In Figure 2, we show how the excess of gravitational mass, $M(a)-m$, at an event P on a world line L is due to the gravitational interaction with the world lines intersecting its past light cone. When time passes, the number of intersected lines, and $M(a)$, increases.

b) *Dark matter, dark energy and cosmological constant as different aspects of the gravitational mass*. In section IVA, we have estimated the beginning of the galactic dominance at $a_i = 0.085$ ($z_i = 10.76$). The gravitational mass fraction $f(a)$, showed in Figure 3, verifies $f(a_i) = 0$ and is practically a constant in the interval $0.4 < a \leq 1$. The value $f(1)$ is identified with the cosmological constant. We have reinterpreted the ΛCDM density parameters $(\Omega_{dm}, \Omega_\Lambda) = (0.268, 0.683)$ in terms of gravitational mass: Ω_{dm} is now the gravitational mass density, $\frac{8\pi G}{3H_0^2} n_o m\alpha$, acquired by the dominant particles before the galactic era; Ω_Λ is the present value, $f(1)$, of the time dependent gravitational mass density. Therefore, in our model, the universe is roughly formed by 5% of baryonic and 95% of gravitational mass: a 27%, usually referred as dark matter, was acquired during the epoch $a < a_i$ previous to the galactic dominance, and a 68%, usually described as dark energy, is the gravitational mass generated in the galactic dominance era $a > a_i$. This interpretation is free of the coincidence problem [20]: the value of Ω_Λ is close to $\Omega_M = \Omega_{ba} + \Omega_{dm}$, because having a common origin, they are not unrelated magnitudes.

c) *Relation distance-redshift*. As showed in **Figure 4**, we have obtained the moduli-distance as function of the redshift that explains the supernova observations. It is indistinguishable from the one obtained with the ΛCDM model using a cosmological constant.

d) *The accelerated universe*. The pressure of the non ideal gas filling the universe given in (8) is not kinetic. It is a functional of the gravitational mass density $\rho_f(a) = f(a)\frac{3H_0^2}{8\pi G}$. This pressure is liable of the recent acceleration of the universe: substituted in (9) produces the deceleration factor q showed in **Figure 5**.

e) *The equation of state of the cosmological fluid*. We have obtained the equation of state of the cosmological fluid by the pressure to density ratio $w(a) = \frac{p(a)}{\rho_f(a)}$ given in (28) and shown in **Figure 6**, satisfying the observational constraint $-1.33 < w(1) < -0.79$.

f) *Time evolution of the gravitational mass fraction*. We have given the time evolution of the gravitational mass of a point-like particle in Equation (23) and the consequent gravitational mass fraction (29). They are in good agreement, as shown in **Figure 7**, with the mass measurements of galaxies up to redshift $z \sim 2$, usually reported as dark matter fractions.

We conclude that the gravitational energy density introduced in this paper explains the large scale cosmological observations in the galactic dominance epoch, making unnecessary either the cosmological constant or the dark matter and energy assumptions. To extend the analysis of the gravitational mass to an inhomogeneous universe is more complicated, but necessary to predict its distribution inside the galaxies.

Acknowledgements

This work has been supported by the Spanish "Ministerio de Economía y Competitividad" and the "Fondo Europeo de Desarrollo Regional" MINECO-FEDER Project No. PGC2018-095251-B-100, and by the Spanish Ministerio de Ciencia, Innovación y Universidades and the Fondo Europeo de Desarrollo Regional, Projects PID2019-109753GB-C21 and PID2019-109753GB-C22, the Generalitat Valenciana Project AICO/2020/125 and the University of Valencia Special Action Project UV-INVAE19-1197312.

Conflicts of Interest

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

References

[1] Jammer, M. (2000) Concepts of Mass in Contemporary Physics and Philosophy. Princeton University Press, Princeton. https://doi.org/10.1515/9781400823789

[2] Havas, P. and Goldberg, J.N. (1962) *Physical Review*, **128**, 398. https://doi.org/10.1103/PhysRev.128.398

[3] Fock, V. (1959) The Theory of Space Time and Gravitation. Pergamon Press, Oxford, 342-352.

[4] Belifante, F.J. (1962) *Physical Review*, **125**, 1124. https://doi.org/10.1103/PhysRev.125.1124

[5] Bel, L., Damour, T., Deruelle, N., *et al.* (1981) *General Relativity and Gravitation*, **13**, 963-1004. https://doi.org/10.1007/BF00756073

[6] Damour, T. (1982) Gravitational Radiation and the Motion of Compact Bodies in Gravitational Radiation. North-Holland, Amsterdam.

[7] Geroch, R. and Traschen, J. (1962) *Physical Review*, **36**, 1124.

[8] Katanaev, M.O. (2013) *General Relativity and Gravitation*, **45**, 1861-1875. https://doi.org/10.1007/s10714-013-1564-3

[9] Beltran, P. and Portilla, M. (2021) The Extended Hyperbolic Einstein-Rosen Bridge.

[10] Portilla, M. (2015) *Journal of Physics*: *Conference Series*, **600**, Article ID: 012007. https://doi.org/10.1088/1742-6596/600/1/012007

[11] Landau, L.D., Akhiezer, A.I. and Lifschitz, E.M. (1967) General Physics, Mechanics and Molecular Physics. Pergamon Press, Oxford. https://doi.org/10.1016/B978-0-08-009106-8.50004-2

[12] Einstein, A. (1950) The Meaning of Relativity. 4th Edition, Methuen, London.

[13] Brans, C.H. (1962) *Physical Review*, **125**, 388. https://doi.org/10.1103/PhysRev.125.388

[14] Rosen, N. (1965) *Annals of Physics*, **35**, 426-436. https://doi.org/10.1016/0003-4916(65)90248-4

[15] Ohanian, H.C. (2013) The Energy-Momentum Tensor in General Relativity and in Alternative Theories of Gravitation, and the Gravitational vs. Inertial Mass. https://arxiv.org/abs/1010.5557

[16] Rindler, W. (2006) Relativity, Special, General, and Cosmological. 2th Edition, Oxford University Press, Oxford.

[17] Planck Collaboration (2014) *Astronomy and Astrophysics*, **571**, A16.

[18] Riess, A.G., *et al.* (2007) *The Astrophysical Journal*, **659**, 98. https://doi.org/10.3139/146.070802

[19] Peebles, P.J.E. (1993) Principles of Physical Cosmology. Princeton University Press, Princeton.

[20] Amendola, L. and Tsujikawa, S. (2010) Dark Energy. Theory and Observations. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511750823

[21] Komatsu, E., *et al.* (2009) *The Astrophysical Journal*, **180**, 330. https://doi.org/10.1088/0067-0049/180/2/330

[22] Goodstein, D. (2012) Adventures i Cosmology. World Scientific, Singapore. https://doi.org/10.1142/7820

[23] Hill, T.L. (1986) An Introduction to Statistical Thermodynamics. Dover Publications Inc., Mineola.

[24] Kreuzer, H.J. (1981) Nonequilibrium Thermodynamics and Its Statistical Foundations. Oxford Science Publications, Oxford.

[25] Dutton, A.A., *et al.* (2011) *Monthly Notices of the Royal Astronomical Society*, **417**,

1621.

[26] Suyu, S.H., *et al*. (2012) *The Astrophysical Journal*, **750**, 10. https://doi.org/10.1088/0004-637X/750/1/10

[27] Genzel, R., *et al*. (2017) *Nature*, **543**, 397.

Journal of Modern Physics, 2021, 12, 1749-1757
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Origin and Evolution of the Universe

Walter Petry

Mathematical Institute of the University Düsseldorf, Düsseldorf, Germany
Email: wpetry@meduse.de

How to cite this paper: Petry, W. (2021) Origin and Evolution of the Universe. *Journal of Modern Physics*, **12**, 1749-1757.
https://doi.org/10.4236/jmp.2021.1213102

Received: September 26, 2021
Accepted: November 7, 2021
Published: November 10, 2021

Abstract

A theory of gravitation in flat space-time is shortly summarized and applied to cosmological models. These models start with a gravitational field and without matter. Gravitational energy is converted to matter and the total energy is conserved. The arising universe has no singularity (no big bang) and is not expanding. The redshift is a gravitational effect. It follows by converting gravitational energy to matter changing the gravitational field. This is another gravitation theory different from general relativity which also gives the presently most accepted results of general relativity for weak gravitational fields but has not the problems of general relativity with a big bang.

Keywords

Gravitation, Flat Space-Time, Field Theory, Gravitational Energy, Matter Arises, No Big Bang, Conservation of Total Energy, Non-Expanding Space, Redshift Is a Gravitation Effect

1. Introduction

The presently most accepted universe is based on Einstein's general theory of relativity (GR). The application of GR to flat cosmological models gives an expanding universe with a singularity in the beginning, the so-called big bang. The universe must have cosmic inflation in the beginning to explain our big universe.

In this article, a theory of gravitation in flat space-time (GFST) which has till now not received attention by cosmologists is applied to cosmological models. The universe starts with gravitational energy and not with matter. In the course of time, gravitational energy is converted to matter implying the observed redshift of distant objects by virtue of the change of the gravitational field. The total energy is always conserved. The universe is not singular and it doesn't expand. Spherically, symmetric perturbations in the universe can arise in the beginning of the universe and they grow quickly in the matter dominated universe. This may explain the big galaxies in the universe whereas the big bang gives only small galaxies in contradiction to observations.

2. The Universe

GR is a theory of gravitation giving for weak gravitational fields agreement to experimental accuracy of theory and experiment. This may be the high acceptance of GR by many scientists. GR is also applied to homogeneous, cosmological models to get origin and development of the universe. The universe begins with a big bang (singularity) and it must expand very quickly (cosmic inflation) by virtue of the observed big universe and the many galaxies.

More than forty years ago, in the year 1979, I published a theory of gravitation in flat space-time. This theory can be described by a gravitational field in the pseudo-Euclidean geometry. Gravitation is a field theory and not a geometry as by GR. GFST gives for weak gravitational fields to measurable accuracy the same results as GR. Hence, we get the same acceptance of GFST as GR for weak gravitational fields. But the results of cosmological models are very different by these two theories. The universe starts with uniformly distributed gravitational field by GFST (no big bang). Gravitation is attracting. The densified gravitational field contracts to matter and a part of the gravitational field surrounds the originated object. Every object is surrounded by a gravitational field. Hence, objects attract one another by the surrounded gravitational field and not by their masses. In the course of time matter and other types of energy arise at coasts of gravitational energy. The sum of all types of arising energies is conserved. The universe is not singular and doesn't expand. The redshift follows by converting gravitational field to matter.

3. Gravitation in Flat Space-Time

Nearly all cosmologists use GR and the results of this theory to describe and study the universe. This theory is well known and gives a singularity, the big bang in the beginning of the universe. In addition cosmic inflation is needed to explain the observed big universe. Therefore, we will use another theory of gravitation, namely gravitation in flat space-time, to study the universe (compare the book [1]).

This theory is not a geometry as GR but a field theory of gravitation in flat space-time. The metric is given by

$$(\mathrm{d}s)^2 = -\eta_{ij}\mathrm{d}x^i\mathrm{d}x^j \tag{1}$$

where (η_{ij}) is a symmetric tensor. A special case is the pseudo-Euclidean geometry with

$$\eta_{ij} = \delta^i_j\,(i, j = 1,2,3), \eta_{i4} = \eta_{4i} = 0\,(i = 1,2,3), \eta_{44} = -1\,. \tag{2}$$

Here, $(x^i) = (x^1, x^2, x^3)$ are the Cartesian coordinates and $x^4 ct$. Put

$$\eta = \det(\eta_{ij})\,. \tag{3}$$

The gravitational field is described by a symmetric tensor (g_{ij}). Let (g^{ij}) be defined by

$$g_{ik}g^{kj} = \delta_i^j. \tag{4}$$

Put

$$G = \det(g_{ij}). \tag{5}$$

The proper-time τ is defined by

$$(c\mathrm{d}\tau)^2 = -g_{ij}\mathrm{d}x^i\mathrm{d}x^j. \tag{6}$$

The Lagrangean of the gravitational field is given by

$$L(G) = -\left(\frac{-G}{-\eta}\right)^{1/2} g_{ij}g_{kl}g^{mn}\left(g^{ik}_{/m}g^{jl}_{/n} - \frac{1}{2}g^{ij}_{/m}g^{kl}_{/n}\right). \tag{7}$$

Here, the bar/denotes the covariant derivative relative to the flat space-time metric (1).

The Lagrangean of dark energy (given by the cosmological constant Λ) has the form

$$L(\Lambda) = -8\Lambda\left(\frac{-G}{-\eta}\right)^{1/2}. \tag{8}$$

Let

$$\kappa = 4\pi k/c^4 \tag{9}$$

(k: gravitational constant) and assume that matter is a perfect fluid. Then, the mixed energy-momentum tensors of gravitation, of dark energy and of matter for a perfect fluid are

$$T_j^i(G) = \frac{1}{8\kappa}\left[\left(\frac{-G}{-\eta}\right)^{\frac{1}{2}} g_{kl}g_{mn}g^{i\nu}\left(g^{km}_{/j}g^{ln}_{/\nu} - \frac{1}{2}g^{kl}_{/j}g^{mn}_{/\nu}\right) + \frac{1}{2}\delta_j^i L(G)\right] \tag{10a}$$

$$T_j^i(\Lambda) = \frac{1}{16\kappa}\delta_j^i L(\Lambda) \tag{10b}$$

$$T_j^i(M) = (\rho + p)g_{jk}u^k u^i + \delta_j^i pc^2. \tag{10c}$$

where ρ, p and (u^i) denote density, pressure and four-velocity of matter. It follows by (6)

$$c^2 = g_{ij}u^i u^j. \tag{11}$$

Let us define the covariant differential operator

$$D_j^i = \left[\left(\frac{-G}{-\eta}\right)^{1/2} g^{kl}g_{jm}g^{mi}_{/l}\right]_{/k}. \tag{12}$$

Then the field equations for the gravitational potentials (g_{ij}) have the form

$$D_j^i - \frac{1}{2}\delta_j^i D_k^k = 4\kappa T_j^i. \tag{13}$$

Here, T_j^i is the sum of the energy-momentum tensors of gravitation, of matter and of dark energy

$$T_j^i = T_j^i(G) + T_j^i(M) + T_j^i(\Lambda). \quad (14)$$

Define the symmetric energy-momentum tensor

$$T^{ij}(M) = g^{ik} T_k^j(M), \quad (15)$$

then the equations of motion are (in covariant form)

$$T_{i/k}^{k}(M) = \frac{1}{2} g_{kl/i} T^{kl}(M). \quad (16)$$

In addition to the field Equation (13) and the equations of motion (16) the conservation of the total energy-momentum holds, *i.e.*

$$T_{i/k}^{k} = 0. \quad (17)$$

The field Equation (13) are formally similar to those of GR where D_j^i corresponds to the Ricci tensor and T_j^i is the total energy-momentum without that of gravitation which is no tensor by GR. These results can be found in the book [1] and in the article [2].

4. Homogeneous, Isotropic Universe

We follow along the lines of article [2].

Let us use the pseudo-Euclidean geometry (2), (3) as metric. The matter tensor is given by perfect fluid with

$$u^i = 0 \ \ (i = 1,2,3) \quad (18)$$

and pressure p and density ρ with

$$p = p_m + p_r, \quad \rho = \rho_m + \rho_r. \quad (19)$$

Here, the indices m and r denote matter and radiation. The equations of state for matter (dust) and radiation are

$$p_m = 0, \quad p_r = \frac{1}{3}\rho_r. \quad (20)$$

The potentials are given by virtue of (18), homogeneity and isotropy

$$g_{ij} = \begin{cases} a^2(t) & (i = j = 1,2,3) \\ -\dfrac{1}{h(t)} & (i = j = 4) \\ 0 & (i \neq j) \end{cases} \quad (21)$$

The four-velocity is by Equation (18)

$$(u^i) = \left(0,0,0,c\sqrt{h}\right). \quad (22)$$

Let $t_0 = 0$ be the present time and assume as initial condition at present

$$a(0) = h(0) = 1, \quad \dot{a}(0) = H_0, \quad \dot{h}(0) = \dot{h}_0, \quad \rho(0) = \rho_{m0}, \quad \rho_r(0) = \rho_{r0}. \quad (23)$$

Here, the dot denotes the time-derivative, H_0 is the Hubble constant and $\dot{h}_0$ is a further constant, ρ_{m0} and ρ_{r0} are the present densities of matter and radiation. It follows by (16) under the assumption that matter and radiation do not interact

$$\rho_m = \rho_{m0}/\sqrt{h}\rho_r = \rho_{r0}/\left(a\sqrt{h}\right). \tag{24}$$

The field Equation (13) implies by the use of (21) the two non-linear differential equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(a^3\sqrt{h}\frac{\dot{a}}{a}\right) = 2\kappa c^4\left(\frac{1}{2}\rho_m + \frac{1}{3}\rho_r + \frac{\Lambda}{2\kappa c^2}\frac{a^3}{\sqrt{h}}\right), \tag{25a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(a^3\sqrt{h}\frac{\dot{h}}{h}\right) = 4\kappa c^4\left(\frac{1}{2}\rho_m + \rho_r + \frac{1}{8\kappa c^2}L(G) - \frac{\Lambda}{2\kappa c^2}\frac{a^3}{\sqrt{h}}\right) \tag{25b}$$

where

$$L(G) = \frac{1}{c^2}a^3\sqrt{h}\left(-6\left(\frac{\dot{a}}{a}\right)^2 + 6\frac{\dot{a}}{a}\frac{\dot{h}}{h} + \frac{1}{2}\left(\frac{\dot{h}}{h}\right)^2\right). \tag{26}$$

The expression $\frac{1}{16\kappa}L(G)$ is the density of the gravitational field. The conservation of the total energy gives

$$\left(\rho_m + \rho_r\right)c^2 + \frac{1}{16\kappa}L(G) + \frac{\Lambda}{2\kappa}\frac{a^3}{\sqrt{h}} = \lambda c^2 \tag{27}$$

where λ is a constant of integration. The Equations (25), (26) and (27) give by the use of the initial conditions (23)

$$\frac{\dot{h}}{h} = -6\frac{\dot{a}}{a} + 2\frac{4\kappa c^4\lambda t + \varphi_0}{2\kappa c^4\lambda t^2 + \varphi_0 + 1} \tag{28}$$

with

$$\varphi_0 = 3H_0\left(1 + \frac{1}{6}\frac{\dot{h}_0}{h_0}\right). \tag{29}$$

Integration of (28) yields

$$a^3\sqrt{h} = 2\kappa c^4\lambda t^2 + \varphi_0 t + 1. \tag{30}$$

Equation (27) gives at present time $t_0 = 0$ by using the initial conditions (23)

$$\frac{1}{3}\left(8\kappa c^4\lambda - \varphi_0^2\right) = 4\left(\frac{8}{3}\pi k\left(\rho_{m0} + \rho_{r0} + \frac{\Lambda c^2}{8\pi k}\right) - H_0^2\right). \tag{31}$$

Let us define as usually the density parameters

$$\Omega_m = \frac{8\pi k\rho_{m0}}{3H_0^2},\quad \Omega_r = \frac{8\pi k\rho_{r0}}{3H_0^2},\quad \Omega_\Lambda = \frac{\Lambda c^2}{3H_0^2}$$

and put

$$K_0 = \left(\Omega_m + \Omega_r + \Omega_\Lambda - 1\right)/\Omega_m, \tag{32}$$

then relation (31) can be rewritten

$$\frac{8kc^4\lambda}{H_0^2} - \left(\frac{\varphi_0}{H_0}\right)^2 = 12\Omega_m K_0. \tag{33}$$

It follows from (27) by the use of (28), (30) and elimination of h and $\dot{h}$ the differential equation

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{H_0^2}{\left(2\kappa c^4\lambda t^2 + \varphi_0 t + 1\right)^2}\left[-\Omega_m K_0 + \Omega_r a^2 + \Omega_m a^3 + \Omega_\Lambda a^6\right] \quad (34a)$$

with the initial condition

$$a(0) = 1. \quad (34b)$$

A necessary and sufficient condition to avoid singular solutions of (34) is the condition

$$K_0 > 0. \quad (35)$$

This yields

$$2\kappa c^4\lambda t^2 + \varphi_0 t + 1 > 0 \quad (36)$$

for all $t \in \mathfrak{R}$. Hence, we get a non-singular solution, *i.e.* we receive a non-singular universe for all $t \in \mathfrak{R}$. It exists $t_1 < t_0 = 0$ such that

$$\dot{a}(t_1) = 0. \quad (37)$$

Put $a_1 = a(t_1)$ then it follows from (34a) with $t = t_1$

$$\Omega_r a_1^2 + \Omega_m a_1^3 + \Omega_\Lambda a_1^6 = \Omega_m K_0. \quad (38)$$

It follows for all $t \in \mathbb{R}$

$$a(t) \geq a_1 > 0. \quad (39)$$

Subsequently let us assume

$$a_1 \ll a(0) = 1, \quad (40)$$

then we get by (38)

$$K_0 \ll 1. \quad (41)$$

It follows from (32) by the use of (41)

$$\Omega_r + \Omega_m + \Omega_\Lambda = 1 + \Omega_m K_0. \quad (42)$$

That is the sum of the density parameters is a little bit greater than 1. The solution of (34) is by longer calculations and under the assumption $p_{r0} = 0$

$$a^3(t) = 2a_1^3\left(1 + \frac{\Omega_\Lambda}{\Omega_m}a_1^3\right)\Big/\left(1 + \left(1 + 2\frac{\Omega_\Lambda}{\Omega_m}a_1^3\right)\cos\left(\sqrt{3}\beta(t)\right)\right), \quad (43a)$$

$$\beta(t) = \operatorname{arctg}\left(\frac{\sqrt{3\Omega_m K_0}H_0(t - t_1)}{1 + \frac{1}{2}\frac{\varphi_0}{H_0}t_1 + (B)H_0 t}\right), \quad (B) = \left(\left(\left(\frac{1}{2}\frac{\varphi_0}{H_0}\right)^2 + 3\Omega_m K_0\right)H_0 t_1 + \frac{1}{2}\frac{\varphi_0}{H_0}\right). \quad (43b)$$

We get as $t \to -\infty$

$$a(-\infty) \approx \left(\frac{2}{\left(1 - \cos\left(\sqrt{3}\pi\right)\right)}\right)^{\frac{1}{3}} a_1 \approx 1.81 a_1.$$

Hence, $a(t)$ starts at $t=-\infty$ with $a(-\infty)=1.81a_1$, decreases to $a_1>0$ and then increases for all *t*. It follows from (30) by the use of (43b) for *t* sufficiently large

$$\sqrt{h(t)} \approx \left(\frac{1}{2}\frac{\varphi_0}{H_0}t+1\right)^2 \Big/ a^3(-\infty). \tag{44}$$

Hence, $a(t)$ starts from a small positive value decreases to a small positive value and then increases for all $t \in \mathcal{R}$. The function $h(t)$ can then be calculated by relation (30).

The proper-time from the beginning of the universe till time *t* is

$$\tilde{\tau}(t) = \int_{-\infty}^{t} 1\Big/\sqrt{h(t)}\,\mathrm{d}t, \tag{45}$$

i.e. the age of the universe is finite by GFST analogues to that of GR.

The metric (1) is not expanding, *i.e.* the universe is not expanding and not singular. In the beginning of the universe it consists only of gravitational energy and no matter and no radiation exist. They arise in the course of time at coasts of gravitational energy and the sum of the total energy is always conserved.

Introducing the proper-time $\tilde{\tau}$ into the differential Equation (34a) we get by the use of (30) the differential equation (see [1])

$$\left(\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}\tilde{\tau}}\right)^2 = H_0^2\left(-\frac{\Omega_m K_0}{a^6}+\frac{\Omega_r}{a^4}+\frac{\Omega_m}{a^3}+\Omega_\Lambda\right). \tag{46}$$

This differential equation is for $a(t)$ sufficiently large (that is: away from the beginning of the universe) by virtue of (41) identical with that of GR. Hence, the function $a(t)$ approximates the scaling factor of GR. In the beginning of the universe the function $a(t)$ is positive whereas the scaling factor of GR is zero what implies by GR the singularity of the expanding universe in the beginning. It is worth to mention that the resulting universes of GR and of GFST are very different. The results of GFST are contained in the book [1] and in article [2]. Additional results about the universe by GFST are found in the articles [3] [4] [5] [6].

A non-expanding universe is experimentally stated by Lerner [7] [8]. The redshift in a non-expanding space is given in [9] and Siegel [10] also asked the question of an expanding space. I must also mention the book of Fahr [11] where a universe without big bang is propagated. In article [12], the redshift of distant objects is derived from the frequency shift of the gravitational field which is changed by converting gravitation to matter.

5. Redshift of Distant Objects

It is useful to introduce the absolute time to simplify the calculations of the redshift in the universe. The absolute time t' of the universe is given by

$$\mathrm{d}t' = \frac{1}{a(t)\sqrt{h(t)}}\mathrm{d}t = \frac{1}{a(t)}\mathrm{d}\tilde{\tau}. \tag{47}$$

The proper-time (6) with (21) by the universe is

$$\left(c\mathrm{d}\tau\right)^2 = -a\left(t\right)^2\left(\sum_{i=1}^{3}\left(\mathrm{d}x^i\right)^2 - \mathrm{d}ct'^2\right). \tag{48}$$

The energy of a photon emitted at a distant object at time t'_e is given by $E\left(t'_e\right) \sim -g_{44}\dfrac{\mathrm{d}t'}{\mathrm{d}\tilde{\tau}} \to a\left(t'_e\right)E_0$, which means for the frequencies

$$\nu_e = a\left(t'_e\right)\nu_0 . \tag{49}$$

Here, ν_0 is the frequency of the same atom emitted at present. This gives for the frequency moving in the universe by virtue of constant light velocity (see (48))

$$\nu\left(t\right) = \nu_e = a\left(t'_e\right)\nu_0 .$$

The redshift is

$$z = \frac{\nu_0}{\nu\left(t'_e\right)} - 1 = \frac{1}{a\left(t'_e\right)} - 1 . \tag{50}$$

Light emitted at distance r at time t'_e and received at $r = 0$ at time t'_0 holds by the constant velocity of light

$$r = c\left(t'_0 - t'_e\right).$$

This gives by Taylor expansion of $a\left(t'_e\right)$ and relation (50)

$$z = H_0\frac{r}{c} + \left(1 - \frac{1}{2}\frac{1}{H_0^2}\frac{\mathrm{d}^2 a\left(t'_e\right)}{\mathrm{d}t'^2_e}\right)\left(H_0\frac{r}{c}\right)^2$$

The differential Equation (46) is by introducing the absolute time t' rewritten

$$\left(\frac{\mathrm{d}a}{\mathrm{d}t'}\right)^2 = \frac{H_0^2}{a^2}\left(-\Omega_m K_0 + \Omega_r a^2 + \Omega_m a^3 + \Omega_\Lambda a^6\right).$$

Differentiation of this relation implies by neglecting small expressions

$$\frac{\mathrm{d}^2 a}{\mathrm{d}t'^2_e} = H_0^2\left(1 - \frac{1}{2}\Omega_m + \Omega_\Lambda\right).$$

Hence, we get by introducing this expression in the relation for z

$$z \cong H_0\frac{r}{c} + \frac{3}{4}\Omega_m\left(H_0\frac{r}{c}\right)^2 . \tag{51}$$

This expression for the redshift is derived by the use of the frequency caused by the change of gravitation and not by a Doppler effect. It was already derived in previous articles (see e.g. the book [1]).

Some additional remarks to GFST:

Gravitational field is attracting. Spherically symmetric perturbations of the gravitational field in the universe attract and are partly converted to matter, an object arises. The rest of this gravitational field surrounds the object and may attract other objects. This seems that matter attracts matter but it is the gravitational field which is attracting. The process of arising objects is fast and implies big objects and may explain the galaxies in our universe. This result can be found in [13] and in the book [1]. This gives also an explanation of gravitation. Gravitational field is attracting and not the mass. The densified gravitational field is converted to matter. Matter is the result of the attracting gravitation. This result can be found in article [14].

Conflicts of Interest

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

References

[1] Petry, W. (2014) A Theory of Gravitation in Flat Space-Time. Science PG. https://doi.org/10.4236/jmp.2014.516156

[2] Petry, W. (2015) *Journal Modern Physics*, **6**, 1085-1094. https://doi.org/10.4236/jmp.2015.68113

[3] Petry, W. (2014) *Journal of Applied Mathematics and Physics*, **2**, 50-54. https://doi.org/10.4236/jamp.2014.25007

[4] Petry, W. (2013) *Journal Modern Physics*, **4**, 20-25. https://doi.org/10.4236/jmp.2013.47A1003

[5] Petry, W. (2018) *Journal Modern Physics*, **9**, 1441-1447. https://doi.org/10.4236/jmp.2018.97088

[6] Petry, W. (2017) *Journal Applied Mathematics and Physics*, **5**, 862-872.

[7] Lerner, E.J. (1991) The Big Bang Never Happened. Amazon.de. https://www.amazon.de/dp/0812918533/ref=cm_sw_r_awdo_navT_a_VWQF7EENJ5T1CBEKQBG3

[8] Lerner, E.J. (2014) Universe Is Not Expanding at All. Science News, May.

[9] Petry, W. (2007) Is the Universe really Expanding? arXiv: 0705.4859

[10] Siegel, E. (2018) Scientists Can't Agree on the Expanding Universe. https://www.forbes.com/sites/startswithabang/2018/12/06/scientists-cant-agree-on-the-expanding-universe/?sh=345f692d5e2c

[11] Fahr, H.-J. (1995) Universum ohne Urknall. Spektrum.

[12] Petry, W. (2016) *Journal Modern Physics*, **7**, 1492-1499. https://doi.org/10.4236/jmp.2016.712135

[13] Petry, W. (1996) *Astrophysics and Space-Science*, **235**, 41-57. https://doi.org/10.1007/BF00643589

[14] Petry, W. (2019) *Journal Modern Physics*, **10**, 157-162. https://doi.org/10.4236/jmp.2019.102012

Journal of Modern Physics, 2021, 12, 1758-1782
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

A New Black Hole Solution in Conformal Dilaton Gravity on a Warped Spacetime

Reinoud J. Slagter

Asfyon, Astronomisch Fysisch Onderzoek Nederland, Bussum, The Netherlands
Email: info@asfyon.com

How to cite this paper: Slagter, R.J. (2021) A New Black Hole Solution in Conformal Dilaton Gravity on a Warped Spacetime. *Journal of Modern Physics*, **12**, 1758-1782.
https://doi.org/10.4236/jmp.2021.1213103

Received: September 30, 2021
Accepted: November 13, 2021
Published: November 16, 2021

Abstract

An exact time-dependent solution of a black hole is found in a conformally invariant gravity model on a warped Randall-Sundrum spacetime, by writing the metric $g_{\mu\nu} = \omega^{\frac{4}{n-2}} \tilde{g}_{\mu\nu}$. Here, $\tilde{g}_{\mu\nu}$ represents the "un-physical" spacetime and ω the dilaton field, which will be treated on equal footing as any renormalizable scalar field. In the case of a five-dimensional warped spacetime, we thereafter write ${}^{(4)}\tilde{g}_{\mu\nu} = \bar{\omega}^2\, {}^{(4)}\bar{g}_{\mu\nu}$. The dilaton field $\bar{\omega}$ can be used to describe the different notion the in-going and outside observers have of the Hawking radiation by using different conformal gauge freedom. The disagreement about the interior of the black hole is explained by the antipodal map of points on the horizon. The free parameters of the solution can be chosen in such a way that $\bar{g}_{\mu\nu}$ is singular-free and topologically regular, even for $\omega \to 0$. It is remarkable that the 5D and 4D effective field equations for the metric components and dilaton fields can be written in general dimension $n = 4, 5$. From the exact energy-momentum tensor in Eddington-Finkelstein coordinates, we are able to determine the gravitational wave contribution in the process of evaporation of the black hole. It is conjectured that, in context of quantization procedures in the vicinity of the horizon, unitarity problems only occur in the bulk at large extra-dimension scale. The subtraction point in an effective theory will be in the UV only in the bulk, because the use of a large extra dimension results in a fundamental Planck scale comparable with the electroweak scale.

Keywords

Conformal Invariance, Dilaton Field, Black Holes, Brane World Models, Antipodal Map

1. Introduction

It is believed that the understanding of the quantum mechanical property of a black hole is one of the major challenges of modern physics. The quantum features of a black hole were investigated, decades ago, by Hawking in his epic work on radiation effects of a black hole [1] [2]. The thermal emission from the black hole can be described by a temperature $T = \frac{1}{8\pi M}$, where M is the mass of the black hole. Hawking considered the collapsing body on a background spacetime, which is time dependent and not symmetrical with respect to time reversal.

Vacuum pair-production at the horizon causes the Hawking radiation, which is thermal and contains no information. The anti-particle falls into the black hole. So when the black hole evaporates completely, it seems that information is lost, which is against Quantum Mechanics (QM): it dictates that the initial and final stage of a system is related to a unitary S-matrix. This is a first indication that there is a problem with QM when applied to a black hole spacetime. This is the information paradox.

Related to this issue, is the holographic principle [3], which states that the interior volume of spacetime of a black hole containing the information of the in-going particles is dual to the surface of the horizon. Could it be that the information is still at the horizon? The idea was extended to the well-known Anti-de Sitter/Conformal Field Theory (AdS/CFT) correspondence: in some way, the information must be present in the Hawking radiation. This model relies heavily on string theory, but would solve the information paradox, by introducing the notion of complementarity of the in- and out-side of the black hole. The in-going and out-going particles are entangled and the information of the in-going particle is also reflected back. However, this viewpoint conflicts with causality [4]. The previously emitted Hawking radiation and the corresponding in-going particles are independent systems and at the same time indirect entangled.

Another solution for the information paradox, which doesn't rely on string theory, is the introduction of a firewall [5]. The entanglement between the in-going and out-going particles is broken by a high energetic shield. The freely in-falling observer encounters high-energy particles at the horizon. This viewpoint conflicts with general relativity, *i.e.*, violation of the equivalence principle. Free falling observers, when falling through the horizon, perceive spacetime as Minkowski, so will not notice the horizon at all.

A fundamental issue which is omitted in all the treatments as described above, is the time-dependency of the spacetime structure near the horizon. The emitted Hawking particle will have a back-reaction effect on the spacetime [6] [7]. Could it be possible, that the topology of the black hole must be revised? It is well known that quantum field theory on a curved spacetime opens the possibility that a field theory can have different vacuum states. It can have intrinsic statistical features from a change in topology and not from a priori statistical description of the matter fields.

A spacetime with a given local geometry admits in principle, different possible global topologies. One can consider the modification of the spacetime topology of the form $\hat{\mathcal{M}}/\Gamma$, where Γ is a discrete subgroup of isometries of $\mathcal{M}$ [8]

[9] [10] [11], without fixed points. $\hat{\mathcal{M}}$ is non-singular and is obtained from its universal covering $\mathcal{M}$ by identifying points equivalent under Γ. A particular interesting case is obtained, when Γ is the antipodal transformation on $\mathcal{M}$

$$J : P(X) \to \hat{P}(\hat{X}). \tag{1}$$

where the light-cone of the antipode of $P(X)$ intersects the light-cone of $P(X)$ only in two point (at the boundary of the spacetime). This is the so-called "elliptic interpretation" [12] of spacetime, where antipodal points represents in fact the same world-point or event. The future and past event horizon intersect each other as a projected cylinder $\mathbb{R}_1 \times \$_1/\mathbb{Z}_2$ [1]. At the intersection one then identifies antipodal points. One must realize that the antipodal map is a boundary condition at the horizon, only observable by the outside observer. On a black hole spacetime, the inside is removed. So nothing can escape the interior, since there is no interior. The field theories formulated on $\mathcal{M}$ and $\hat{\mathcal{M}}$ are globally different, while locally $\mathcal{M}$ and $\hat{\mathcal{M}}$ are identically. The emitted radiation is only locally thermal. Antipodal identification, however, destroys the thermal features in the Fock space construction. In the construction, one needs unitary evolution operators for the in-going and out-going particles [13].

In order to avoid wormhole constellations or demanding "an other universe" in the construction of the Penrose diagram, it is essential that the asymptotic domain of $\mathcal{M}$ maps one-to-one onto the ordinary spacetime in order to preserve the metric. In fact, one deals with one black hole. A consequence is that time-inversion takes place in region II of the Penrose diagram, so interchange of the creation and annihilation operators and entangling positive energy particles at the horizon with positive energy antiparticles at the antipodes. So the antipodal identification is not in conflict with the general CPT invariance of our world. Further, for the outside observer, the thermodynamically mixed state is replaced by a pure state. So the Hawking particles at opposite sides of the black hole are entangled.

The former representation that observers have no access to the inside of the black hole is no longer valid. One arrives by this new geometrical description at pure quantum states for the black hole. It will solve, moreover, the information paradox and firewall problem as well[2].

The gravitational back-reaction as proposed by't Hooft [14] [15], suggests a cut-off of high momenta, which avoids the firewall. The in-going particle has a back-reaction on the other particles, leading to a unitary S-matrix. The gravitational interaction between the in-going and out-going particles will be strong, because we are dealing here with a strongly curved spacetime near the horizon. Using a "cut-and-paste" procedure, one replaces the high-energy particles ("hard"), *i.e.*, mass or momentum of the order of the Planck mass, by low-energy

[1]We work here in polar coordinates, because the spinning black hole we will consider, has a preferred spin axis. The antipodal identification is then $(U,V,z,\varphi) \to (-U,-V,-z,\pi+\varphi)$.

[2]The technical aspects in constructing the unitary S-matrix can be found in the literature, as provided by the references.

("soft") particles far away. These hard particles just caused the firewall problem. Hard particles will also influence the local spacetime (to become non-Schwarzschild) and causes the Shapiro effect. The interaction with the soft particles is described by the Shapiro delay. Effectively, all hard particles are quantum clones of all soft particles. By this "firewall-transformation", we look only at the soft particle clones. They define the Hilbert space and leads to a unitary scattering matrix. The net result is that the black hole is actually in a pure state, invalidating the entanglement arguments in the firewall paradox. The entanglement issue can be reformulated by considering the two regions I and II in the maximally extended Penrose diagram of the black hole, as representing two "hemispheres" of the same black hole. It turns out that the antipodal identification keeps the wave functions pure and the central $r = 0$ singularity has disappeared. This gravitational deformation will cause transitions from region I to II in the Penrose diagram. The fundamental construction then consists of the exchange of the position operator with the momentum operator of the in-going particles, which turn them into out-particles. Hereby, 't Hooft expands the moment distributions and position variables in partial waves in (θ, φ). So the Hawking particles emerging from I are entangled with the particles emerging from II. An important new aspect is the way particles transmit the information they carry across the horizon. In the new model, the Hawking particles emerging from I are maximally entangled with the particles emerging from II. The particles form a pure state, which solves the information paradox.

In order to describe the more realistic black holes, such as the axially symmetric Kerr black hole, it is not possible to ignore the dynamics of the horizon. Moreover, one must incorporate gravitation waves. There is another reason to consider axially symmetry. A spherical symmetric system cannot emit gravitation waves [16]. Astronomers conjecture that most of the black holes in the center of galaxies are of the Kerr type. A linear approximation is, of course, inadequate in high-curvature situations. In the linear approximation, the waves don't carry enough energy and momentum to affect their own propagation. The notion of the "classical" Hartle-Hawking vacuum thermal state, with a temperature $T \sim \frac{1}{M} \sim \kappa$ and the luminosity $\frac{\mathrm{d}M}{\mathrm{d}t} \sim -\frac{1}{M^2}$ must also be revised when the mass reaches the order of the Planck mass. On the Kerr black hole spacetime no analog of the Hartle-Hawking vacuum state exists. The Killing field ξ^μ generates a bifurcate Killing horizon ($\xi^\mu \xi_\mu = -1$ at infinity) and possesses space like orbits near infinity [17].

Another aspect of the huge curvature in the vicinity of the horizon, will be the problem of constructing a renormalizable (and maintaining unitarity) quantum gravity model of the Standard Model fields, which must be incorporated in the Lagrangian. Up till now, no convincing theory of quantum gravity is available. Many attempts were made in order to make a renormalizable and unitary quantum gravity model. One also can try to construct a renormalizable model, by adding fourth order derivative terms of the curvature tensor (Euler-term). However, one looses unitarity. Also the "old" effective field theory (EFT) has its problems. One ignores what is going on at high energy. In order to solve the anomalies one encounters in calculating the effective action, one can apply the so-called conformal dilaton gravity (CDG) model [6] [7] [18] [19]. CDG is a promising route to tackle the problems arising in quantum gravity model, such as the loss of unitarity close to the horizon. One assumes local conformal symmetry, which is spontaneously broken (for example by a quartic self-coupling of the Higgs field). Changing the symmetry of the action was also successful in the past, *i.e.*, in the SM of particle physics. A numerical investigation of a black hole solution of a non-vacuum CDG model, was recently performed [20]. The key feature in CDG, is the splitting of the metric tensor $g_{\mu\nu}=\omega^{\frac{4}{n-2}}\tilde{g}_{\mu\nu}$, with ω the dilaton field. Applying perturbation techniques (and renormalization/dimensional regularization), in order to find the effective action and its divergencies, one first integrate over ω (shifted to the complex contour), considered as a conventional renormalizable scalar field and afterwards over $\tilde{g}_{\mu\nu}$ and matter fields. The dilaton field is locally unobservable. It is fixed when we choose the global spacetime and coordinate system. If one applies this principle to a black hole spacetime, then the energy-momentum tensor of ω influences the Hawking radiation. When $\tilde{g}_{\mu\nu}$ is flat, then the handling of the anomalies simplifies considerably [15]. When $\tilde{g}_{\mu\nu}$ is non-flat, the problems are more deep-seated.

It is well known, that the antipodal transformation, or inversion, is part of the conformal group [21]. So conformal invariant gravity models could fit very well the models of antipodal mapping as described above. In this context, the modification of GRT by an additional spacetime dimension could be an alternative compromise, because Einstein gravity on the brane will be modified by the very embedding itself and opens up a possible new way to address the dark energy problem [22] [23] [24][3]. These models can be applied to the standard Friedmann-Lemaître-Robertson-Walker (FLRW) spacetime and the modification on the Friedmann equations can be investigated [25]. Recently, Maldacena, *et al.* [26], applies the RS model to two black hole spacetimes and could construct a traversable macroscopic wormhole solution by adding only a 5D U(1) gauge field (see also Maldacena [27]). However, an empty bulk would be preferable. Instead, one can investigate the contribution of the projected 5D Weyl tensor on the 4D brane. It carries information of the gravitational field outside the brane. If one writes the 5D Einstein equation in CDG setting, it could be possible that an effective theory can be constructed without an UV cutoff, because the fundamental scale M_5 can be much less than the effective scale M_{Pl} due to the warp factor. The physical scale is therefore not determined by M_{Pl}.

In this manuscript we will apply the antipodal map on a spinning black hole spacetime in conformal dilaton gravity applied to a warped 5D spacetime.

[3]There is another argument in favor of a (warped) 5D spacetime. It turns out, as we shall see, that a surface in 4D can be immersed to 5D, like a Klein bottle.

2. Conformal Transformations and Antipodal Mapping

2.1. The Origin of the Antipodal Mapping

The antipodal map originates from the so-called "elliptic" interpretation [12]. If one considers the hyperboloid H, $-t^2+x^2+y^2+z^2+w^2=R^2$, then the space-like sections through the origin are ellipses and the time-like sections are hyperbola branches.

Since the de Sitter spacetime can be isometrically embedded as a hyperboloid in $\mathbb{R}^5$, one can take $R^2=-\frac{3}{\Lambda}$. If one suppresses the coordinates (z,w), we have the $\mathbb{R}^3$ Minkowski metric. Lorentz transformations (LT's) around the origin transform H into itself. Circles on H represent space at different epochs. The bottle-neck parallel is a spatial geodesic, while the others are not. Further, the circumferences contract from $z=-\infty$ to $z=0$ and then expand. A LT of $\mathbb{R}^3$ turns the bottle-neck into an ellipse, cut out of H with an angle < 45° with the (x,y)-plane. See Figure 1. All the ellipses are equivalent space-like geodesics since each of them is transferred by a suitable automorphism into the bottle-neck, which is one of them. One defines the antipodal map

$$J:P(t,x,y)\to\hat{P}(-t,-x,-y), \tag{2}$$

on H. The antipodicity is Lorentz invariant. When the angle approaches 45°, then the ellipses degenerate into a couple of parallel generators (g_1,g_2) (null geodesics). The other plane of 45° delivers the set (g_3,g_4). The sets (g_1,g_4) and (g_2,g_3) form, for example, the light-cones at the points M and $\hat{M}$. If one

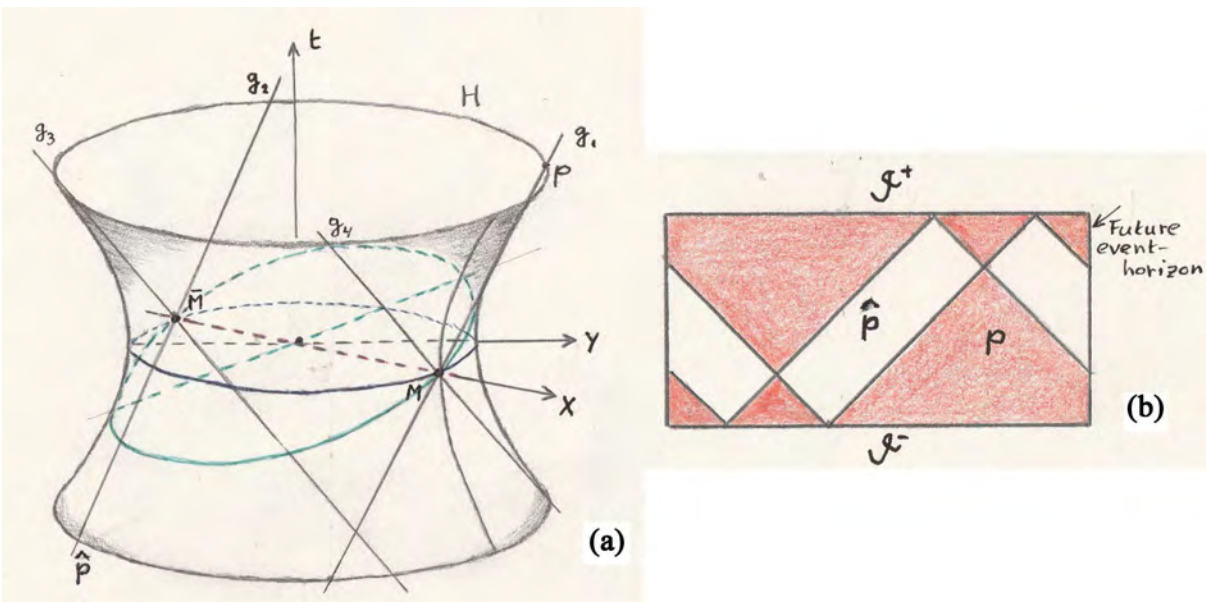


Figure 1. Hyperboloid H representing the $\mathbb{R}^3$ spacetime of the compactified de Sitter universe $\mathbb{R}^5$ ((z,w) suppressed) (a). In the Penrose diagram, the antipodal points P and $J(P)$ are spacelike separated. An observer moving in de Sitter spacetime cannot meet both P and $J(P)$ (b). He cannot receive a message from $P, J(P)$. Moreover, he cannot receive a message from P and send a message to $J(P)$.

moves upwards along t, the inner angles of the light-cones decrease. Note that the light-cones at P and $\hat{P}$ has no point in common and the antipodes are joined by a space-like geodesic. Now Schrödinger proposed to identify P and $\hat{P}$ with the same physical world-point or event. One half of H, containing no antipodal points, represents the "whole world". Thereafter, Schrödinger argues in a clever way that the total potential of experiences of any observer is complete and embraces the same events for any two observers, whatever their world lines be. But there is a price we have to pay for[4]. The direction of the arrow of time is lost (or the distinction between the "fore-cone" and "after-cone" is lost). The allotment of past and future is undecidable. The elliptic model is time-reversible. This can open perspective to the general CPT invariance of our world. The real problems arise, when one considers thermodynamical systems, as is the case for the Hawking effect in the vicinity of the horizon of a black hole. Then the entropy comes into play. Note, quoting Schrödinger, "the irreversible laws of thermodynamics can only be based on the statistical microscopically reversible systems on condition that statistical theory be autonomous in defining the arrow of time. If any other law of nature determines this arrow, the statistical theory collapses."

In a pseudo-polar frame $(\chi, T, \theta, \varphi)$ we can write the line element

$$\mathrm{d}s^2 = -R^2\mathrm{d}T^2 + R^2\cosh^2 T\left[\mathrm{d}\chi^2 + \sin^2\chi\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\varphi^2\right)\right], \tag{3}$$

where $0 < \chi < 2\pi$. The antipodal map becomes now

$$J : (T, \chi, \theta, \varphi) \to (-T, \pi + \chi, \pi - \theta, \pi + \varphi). \tag{4}$$

We already mentioned that de Sitter can be embedded as a hyperboloid in 5D Minkowski. We then say that $J : X^\mu \to -X^\mu$ is an inversion[5]. There exist another coordinate system (introduced by de Sitter himself) in which the line element is written as

$$\mathrm{d}s^2 = -\left(1 - \frac{\rho^2}{R^2}\right)\mathrm{d}T'^2 + \frac{1}{1 - \frac{\rho^2}{R^2}}\mathrm{d}\rho^2 + \rho^2\left(\mathrm{d}\theta'^2 + \sin^2\theta'\mathrm{d}\varphi'^2\right), \tag{5}$$

where we have taken the velocity of the LT $\tanh T = \frac{t}{y}$. This is the static de Sitter and the spaces of constant time are all equivalent. There are singularities for $x = \pm R$ ($\chi = \pm 90^\circ$), *i.e.*, the points $(M, \hat{M})$. However, as also observed by Schrödinger, this static model is not adequate for applying the antipodal map. In order to apply the antipodal map on a black hole spacetime in a more general setting, one needs a time dependent spacetime.

[4]This price is worth paying in the black hole situation, when the information paradox will be solved by the antipodal map. The antipodal half is not time orientable. There is a breakdown of the global distinction between past and future in the interior of the black hole.

[5]The inversion $X_\mu \to -\frac{X_\mu}{X^2}$ (as well as the dilatations) is part of the conformal group [21]. We shall see in the next sections that in general the conformal group is a projective group from 5D. The fifth "degree of freedom" is a sort of gauge space.

2.2. The "Classical" Hawking Effect and Its Problems

The famous result of Hawking states, that a black hole will radiate at "sufficiently" late times like a black body at a temperature

$$kT \sim \frac{\kappa}{2\pi} = \frac{\hbar c^3}{8\pi GM}, \tag{6}$$

with κ the surface gravity and M the mass. The entropy should then be $S_{bh} = \frac{kc^3}{4\hbar} A$, with A the area of the horizon. However, one runs into problems by the back-reaction effect of the particle creation, which will alter the area. It is questionable if the ordinary laws of thermodynamics can be applied to a black hole. It is clear that these laws must be constrained to form quantum states with orthonormality and unitarity conditions. Suppose that an isolated black hole completely evaporate within a finite time. Loss of quantum coherence should then occur *i.e.*, an initially pure quantum state should evolve to a mixed state. In general, in the classical picture, a black hole cannot causally influence its exterior, so it is hard to understand the mechanism by which thermal equilibrium could be achieved. Observe that the state of the field at late times in the region I of the Penrose diagram (and so the particles flux reaching infinity) is described by a density matrix by the S-matrix analysis. The particles present in region I are strongly correlated with the particles which entered the black hole at earlier times.

Consider now in **Figure 2** the evolution of two Cauchy surfaces ("time" Σ_1

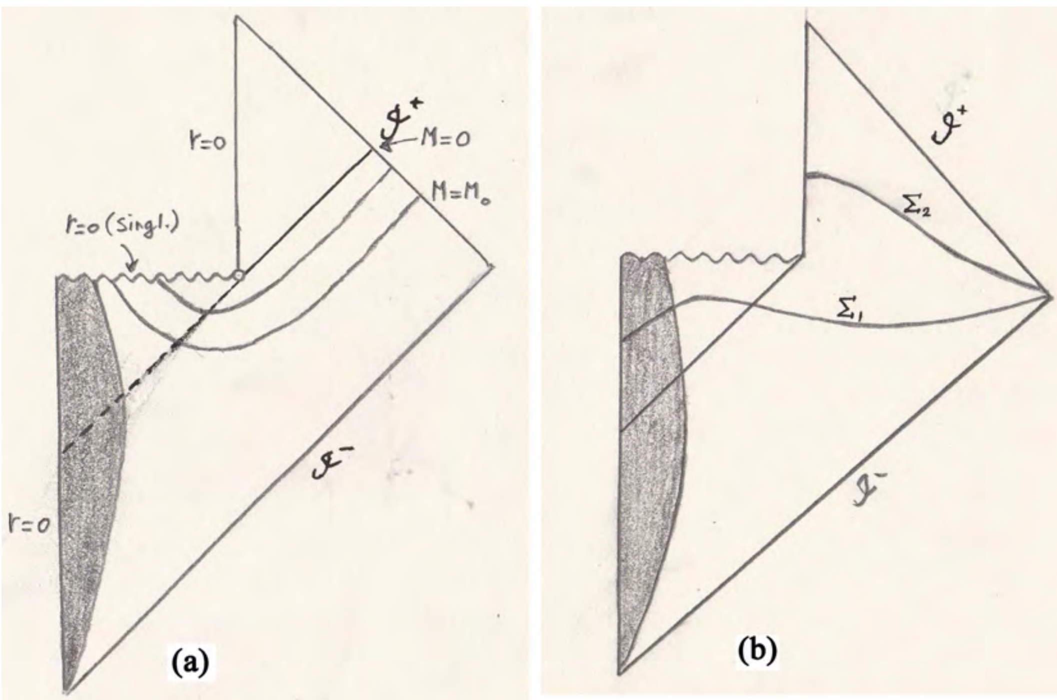


Figure 2. The formation and evaporation of the Schwarschild black hole. The contour $M = 0$ lies at the retarded time corresponding to the final evaporation (a). The geometry is flat above this contour. It turns out that there will be loss of quantum coherence, *i.e.*, an evolution from a pure state to a mixed state [17], as can be observed by the two Cauchy surfaces Σ_1 and Σ_2 (b).

to "time" Σ_2). When the black hole disappears from the spacetime, then at late times, the entire state of the field is mixed. If one takes the "out" Hilbert space to be the Fock space of the particles propagating out to infinity at late times, one cannot describe particle creation and scattering by an ordinary S-matrix. The initial pure state will evolve to a final density matrix. So we have a breakdown of quantum theory. The antipodal model, however, could "repair" this breakdown.

2.3. Conformal Map between Manifolds and the Antipodal Map

Let us consider a smooth regular map $f: M^n \curvearrowright N^n$, with metrics g_1 and g_2 [21]. This represents a local isometry if it preserves the metric, *i.e.*, $f * g_2 = g_1$ and a global isometry if it is a diffeomorphism too. It is a conformal map if it rescales the metric, *i.e.*, $f * g_2 = \Omega^2(x) g_1$, with Ω^2 a positive scalar field. Moreover, it preserves the light-cone structure. Further, an isometry maps geodesics into geodesics and preserves the affine parameter. Conformal maps preserve null geodesics. In many physical applications, it is preferable to consider global isometries: they constitute a group of the manifold. On Minkowski spacetime, the diffeomorphism is of the form $y^\alpha = A^\alpha_\beta + b^\alpha$, with A^α_β a Lorentz matrix. In this context, one must not confuse this transformation with the Poincare transformations, which are of the same form. They connect two inertial frames. They are the basic of special relativity. They are coordinate transformations and are linear. Conformal maps in Minkowski spacetime do not act as linear transformations. Nevertheless, one can generate them from linear transformations in a higher-dimensional spacetime. Now the antipodal map can be represented as a conformal transformation generated from pseude-othogonal matrices of $O(3)$, *i.e.*, the conformal group. Each conformal transformation in this group can be presented by a pair of antipodal matrices. In language of group theory, the map of a pair of antipodal points into a pair of antipodal points can be considered as a conformal transformation on $M\left(\mathbf{R}^1 \otimes \mathbf{R}^1\right)$ and is represented by the pseudo-orthogonal group of matrices $O(2,2)$. The matrix $-\mathbb{I}$ will interchange antipodal points. Details can be found in Felsager [21]. A very illuminating presentation of conformal transformations, in particular the inversions, can be given by the stereographic projection ($SP: S^2 \to \mathbb{C}_\infty$) by using complex numbers $z \in \mathbb{C}$. If one extend the complex plane, $\mathbb{C}_\infty = \mathbb{C} \cup \infty$, then one has a bijection between $\mathbb{C}_\infty$ and S^2. This is the Riemann sphere and one says that $\mathbb{C}_\infty$ is a one-point compactification. Moreover, the map is a homeomorphism. Further, SP^{-1} are conformal maps. The inversion map $T(z) = \frac{1}{z} = \frac{\bar{z}}{|z|^2}$ is a conformal map in $\mathbb{C}_\infty \to \mathbb{C}_\infty$. One can proof that the Möbius transformations $M(\mathbb{C})$ $f: \mathbb{C}_\infty \to \mathbb{C}_\infty$ with $f(z) = az + b/cz + b$, are the conformal maps of $\mathbb{C}_\infty$. The set $M(\mathbb{C})$ is a surjective group with a homomorphism $\Gamma: GL_2(\mathbb{C}) \to M(\mathbb{C}_\infty)$ and kernel the diagonal matrices. The group $GL_2(\mathbb{C})/kI$ is the $PGL_2(\mathbb{C})_\infty$, with k a constant. If $SL_2(\mathbb{C})$ represents the complex matices with determinant 1, then $\Gamma: SL_2(\mathbb{C}) \to M(\mathbb{C}_\infty)$ is onto and has kernel $\pm I$. One then has an isomorphism $\Gamma: SL_2(\mathbb{C})/\pm I \rightarrow M(\mathbb{C}_\infty)$. The class of the Möbius transformations where a,b,c,d are $\in \mathbb{R}$, are interesting, because they apply to hyperbolic geometry. The group $PSL_2(\mathbb{C})$ can then be defined, in order to define conjugate classes and to classify the fixed points, that means in our situation, no fixed points. If an element $f \in (\mathbb{C})$ has period m with $f^M(z) = z$ for the smallest *m*, then *f* has no fixed points. Rotations in $(\mathbb{C})_\infty$ are also Möbius transformations. A rotation of S^2 is a linear map with positive determinant that maps S^2 onto itself. Because there is a fixed axis, one can represent the rotation in $\mathbb{R}^3$ (the $SO(3)$, the orthogonal matrices with determinant 1) by

$$A = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

They are conformal maps of $\mathbb{R}^3$. A map $f: \mathbb{C} \rightarrow \mathbb{C}$ is a rotation in $\mathbb{C}$ if $SP^{-1} \circ f \circ SP: S^2 \rightarrow S^2$. So *f* is conformal too. Suppose $P = (u,v,w) \in S^2$, and $\bar{P} = (-u,-v,-w) \in S^2$ is the antipodal point in S^2. Then, if $z = SP((u,v,w)) \in \mathbb{C}_\infty$, the antipodal point of $z, \bar{z} \in \mathbb{C}_\infty$ is given by $\frac{-1}{\bar{z}}$. So if $f \in Rot(\mathbb{C}_\infty)$, then the antipodal pair $(z, \bar{z})$ is mapped to an antipodal pair $(f(z), f(\bar{z}))$. Further, one proofs that $f\left(\frac{-1}{\bar{z}}\right) = \frac{-1}{\overline{f(z)}}$ and $Rot(\mathbb{C}_\infty) = PSU_2(\mathbb{C})$ is isomorphic with $SO(3)$. So $SO(3)$ will generate the conformal group. which can be used in our 4D spacetime, specially the inversion (by defining self-dual and anti-self-dual forms). One then can formulate the Cauchy-Riemann equations. In physics, they play an important role, because the solution of these equations is automatically a harmonic function of the Laplace equation. Moreover, the equations are conformally invariant.

There is another interesting application of the Möbius presentation: define a complex manifold in 4D. This is the Ernst formulation [28]. If one introduces two complex metric components, one reformulates the Kerr spacetime in a very transparent way. Non-vacuum models can then be generated from the vacuum situation. Just as the holomorphic smooth mappings on the complex manifold of the Riemann sphere $f: S^2 \rightarrow S^2$. These mapping are conformal if they are holomorphic. It could be well possible to extend this approach to 4D. A holomorphic map has interesting properties. It can be represented by an algebraic function $f(z) = P(z)/Q(z)$, with (P,Q) polynomials. So smooth function can be replaced by a holomorphic one. Further, the polynomials can have zero's or singular points, real or complex. Compare this with the conformal maps on the Riemann sphere (generated by the inverse stereographic projection), where the north and south poles causes poles. Some notes must be made about the antipodal map when one uses polar coordinates (θ,φ) on S^2 of $(\theta_0,\varphi) \rightarrow (\theta_0, n\varphi)$ (rotation over the azimuthal angle $n\varphi$, with *n* the winding number). It is singular at the poles, unless we take $\cos^2 n\varphi + \sin^2 n\varphi = 1$, which is true for $n = \pm 1$. For

$n=-1$ we have the antipodal map!

Remember, when adding a scalar gauge field to the Lagrangian (which becomes the axially symmetric Nielsen-Olesen vortex), n represents the number of magnetic flux quanta. It is conjectured that the antipodal map can be applied to our exact solution presented in the next sections.

3. The Black Hole Solution on a 5D Warped Spacetime in Conformal Dilaton Gravity

3.1. The 5D Warped Spacetime

Instead of considering the static metric of de Sitter, *i.e.*, Equation (5), we will now investigate the dynamical 5D spacetime warped spacetime [23] [24] [25] [29]

$$\mathrm{d}s^2=\omega(t,r,y)^2\left[-N(t,r)^2\,\mathrm{d}t^2+\frac{1}{N(t,r)^2}\mathrm{d}r^2+\mathrm{d}z^2+r^2\left(\mathrm{d}\varphi+N^{\varphi}(t,r)\,\mathrm{d}t\right)^2+\mathrm{d}y^2\right],\quad(7)$$

where y is the extra dimension and ω a warp factor in the formulation of Randall-Sundrum's (RS) 5D warped spacetime with one large extra dimension and negative bulk tension Λ_5. The Standard Model (SM) fields are confined to the 4D brane, while gravity acts also in the fifth dimension. Originally, the RS model was applied to a 5-dimensional anti-de Sitter (AdS) spacetime with a positive brane tension. This is the so-called RS-1 model, with one brane. The RS-2 model treats two branes with $\mathbb{Z}_2$ symmetry. However, the effective cosmological constant on the brane can be zero by fine tuning with the negative Λ_5. In the RS model there is a bound state of the graviton confined to the wall as well as a continuum of Kaluza-Klein (KK) states. Four dimensional gravity is then recovered on the brane and the hierarchy problem seems to be solved. Since the pioneering publication of RS, many investigation were done in diverge domains. In particular, Shiromizu *et al.* [30], extended the RS model to a fully covariant curvature formalism. See also the work of Maartens [31]. It this extended model, an effective Einstein equation is found on the brane, with on the right-hand side a contribution from the 5D Weyl tensor which carries information of the gravitational field outside the brane. So the brane world observer may be subject to influences from the bulk. The field equations are (were we took an empty bulk) [20]

$$^{(5)}G_{\mu\nu}=-\Lambda_5\,{}^{(5)}g_{\mu\nu},\quad(8)$$

$$^{(4)}G_{\mu\nu}=-\Lambda_{eff}\,{}^{(4)}g_{\mu\nu}+\kappa_4^2\,{}^{(4)}T_{\mu\nu}+\kappa_5^4\mathcal{S}_{\mu\nu}-\mathcal{E}_{\mu\nu},\quad(9)$$

where we have written

$$^{(5)}g_{\mu\nu}={}^{(4)}g_{\mu\nu}+n_\mu n_\nu,\quad(10)$$

with n^μ the unit normal to the brane. Here $^{(4)}T_{\mu\nu}$ is the energy-momentum tensor on the brane and $S_{\mu\nu}$ the quadratic contribution of the energy-momentum tensor $^{(4)}T_{\mu\nu}$ arising from the extrinsic curvature terms in the projected Einstein tensor. Further,

$$\mathcal{E}_{\mu\nu} = {}^{(5)}C^{\alpha}_{\beta\rho\sigma} n_\alpha n^\rho \, {}^{(4)}g^{\beta}_{\mu} \, {}^{(4)}g^{\sigma}_{\nu}, \tag{11}$$

represents the projection of the bulk Weyl tensor orthogonal to n^μ. The effective gravitational field equations on the brane are not closed. One must solve at the same time the 5D gravitation field in the bulk.

3.2. The Conformal Dilaton Gravity (CDG) Model on a 5D Warped Spacetime

One can distinguish several possible "routes" to the unification of GR and QFT. One can start, for example, with a given classical theory and applies heuristic quantization rules. One then can make a division in canonical and covariant approaches, *i.e.*, uses a Hamiltonial formalism or employs covariance at some stage. The CDG model we consider here, is part of the covariant approach to quantum gravity. The key feature in CDG, is the splitting of the metric tensor

$$g_{\mu\nu} = \omega^{\frac{4}{n-2}} \tilde{g}_{\mu\nu}, \tag{12}$$

with ω the dilaton field and $\tilde{g}_{\mu\nu}$ the ""un-physical" spacetime. At high energy, ω will be treated as a (renormalizable) quantum field. One can prove that the action (without matter terms for the time being)

$$S = \int \mathrm{d}^n x \sqrt{-\tilde{g}} \left[\frac{1}{2} \xi \omega^2 \tilde{R} + \frac{1}{2} \tilde{g}^{\mu\nu} \partial_\mu \omega \partial_\nu \omega + \Lambda \kappa^{\frac{4}{n-2}} \xi^{\frac{n}{n-2}} \omega^{\frac{2n}{n-2}} \right], \tag{13}$$

is conformal invariant under

$$\tilde{g}_{\mu\nu} \to \Omega^{\frac{4}{n-2}} \tilde{g}_{\mu\nu}, \quad \omega \to \Omega^{-\frac{n-2}{2}} \omega. \tag{14}$$

The covariant derivative is taken with respect to $\tilde{g}_{\mu\nu}$. For details, see Slagter [20]. Now we implement the 5D warped spacetime Equation (7). So

$$ {}^{(5)}g_{\mu\nu} = \omega^{4/3} \, {}^{(5)}\tilde{g}_{\mu\nu}, \quad {}^{(5)}\tilde{g}_{\mu\nu} = {}^{(4)}\tilde{g}_{\mu\nu} + n_\mu n_\nu, \tag{15}$$

and write again

$$ {}^{(4)}\tilde{g}_{\mu\nu} = \bar{\omega}^2 \bar{g}_{\mu\nu}. \tag{16}$$

Variation of the action leads to the field equations

$$\xi \omega \tilde{R} - \tilde{g}^{\mu\nu} \tilde{\nabla}_\mu \tilde{\nabla}_\nu \omega - \frac{2n}{n-2} \Lambda \kappa^{\frac{4}{n-2}} \xi^{\frac{n}{n-2}} \omega^{\frac{n+2}{n-2}} = 0 \tag{17}$$

and

$$\omega^2 \tilde{G}_{\mu\nu} = T^{\omega}_{\mu\nu} - \Lambda \tilde{g}_{\mu\nu} \kappa^{\frac{4}{n-2}} \xi^{\frac{2}{n-2}} \omega^{\frac{2n}{n-2}}, \tag{18}$$

with

$$T^{\omega}_{\mu\nu} = \tilde{\nabla}_\mu \tilde{\nabla}_\nu \omega^2 - \tilde{g}_{\mu\nu} \tilde{\nabla}^2 \omega^2 + \frac{1}{\xi} \left(\frac{1}{2} \tilde{g}_{\alpha\beta} \tilde{g}_{\mu\nu} - \tilde{g}_{\mu\alpha} \tilde{g}_{\nu\beta} \right) \partial^\alpha \omega \partial^\beta \omega. \tag{19}$$

From the 5D Einstein equations Equation (8) one obtains

$\omega(t,r,y)=\omega_1(t,r)\omega_2(y)$, with $\omega_2(y)=l=\text{constant}$ (the length scale of the extra dimension). The dilaton equations Equation (17) is superfluous. Note that the effective Einstein equations Equation (9) contains the $\mathcal{E}_{\mu\nu}$, while $T_{\mu\nu}$ and $\mathcal{S}_{\mu\nu}$ are taken zero in our case. The dilaton equation is again superfluous.

It turns out that one can write the field equations for ω and N in the form (n = 4, 5)

$$\ddot{\omega}=-N^4\omega''+\frac{n}{\omega(n-2)}\left(N^4\omega'^2+\dot{\omega}^2\right), \tag{20}$$

$$\begin{aligned}\ddot{N}&=\frac{3\dot{N}^2}{N}-N^4\left(N''+\frac{3N'}{r}+\frac{N'^2}{N}\right)\\&\quad-\frac{n-1}{(n-3)\omega}\left[N^5\left(\omega''+\frac{\omega'}{r}+\frac{n}{2-n}\frac{\omega'^2}{\omega}\right)+N^4\omega'N'+\dot{\omega}\dot{N}\right].\end{aligned} \tag{21}$$

One can solve these equations exact (we took $\Lambda_{eff}=0$):

$$\omega=\left(\frac{a_1}{(r+a_2)t+a_3r+a_4}\right)^{\frac{1}{2}n-1},N^2=\frac{1}{5r^2}\frac{10a_2^3r^2+20a_2^2r^3+15a_2r^4+4r^5+C_1}{C_2(a_3+t)^4+C_3}, \tag{22}$$

with a_i some constants. There is a constraint equation

$$\bar{\omega}''=-\frac{2n}{n-2}\frac{\Lambda l\kappa^{\frac{4}{(n-2}}\xi^{\frac{n-2}{4(n-1)}}\bar{\omega}^{\frac{n+2}{n-2}}}{N^2}-\frac{\omega'N'}{N}-\frac{\omega'}{2r}+\frac{4}{n-2}\frac{\dot{\bar{\omega}}^2}{\bar{\omega}N^4}-\frac{\dot{\bar{\omega}}\dot{N}}{N^5}, \tag{23}$$

which l the dimension of y. The solution for the two dilaton fields ω and $\bar{\omega}$ differs only by the different exponent $\frac{3}{2}$ and 1 respectively. The solution for the metric component is the same (apart from the constants). The solution for the angular momentum component is

$$N^{\varphi}=F_n(t)+\int\frac{1}{r^3\bar{\omega}^{\frac{n-1}{n-3}}}\mathrm{d}r. \tag{24}$$

The Ricci scalar for $\bar{g}_{\mu\nu}$ ($\Lambda=0$) is given by

$$\bar{R}=\frac{12}{N^2}\left[\dot{\bar{\omega}}^2-N^4\bar{\omega}'^2\right], \tag{25}$$

with is consistent with the null condition for the two-dimensional (t,r) line element, when $\bar{R}=0$. One can easily check that the trace of the Einstein equations is zero. Note that N^2 can be written as

$$N^2=\frac{4\int r(r+a_2)^3\,\mathrm{d}r}{r^2\left[C_2(a_3+t)^4+C_3\right]}. \tag{26}$$

So the spacetime seems to have two poles. However, the $r=0$ is questionable. The conservation equations become

$$\bar{\nabla}^{\mu}\mathcal{E}_{\mu\nu}=\bar{\nabla}^{\mu}\left[\frac{1}{\bar{\omega}^2}\left(-\Lambda\kappa^{2\,(4)}\bar{g}_{\mu\nu}\bar{\omega}^4+{}^{(4)}T_{\mu\nu}^{(\bar{\omega})}\right)\right], \tag{27}$$

which yields differential equations for $\ddot{N}'$ and $\dot{N}$ as boundary conditions at the brane. It can be described as the non-local conservation equation. In the high energy case close to the horizon, one must include the $\mathcal{S}_{\mu\nu}$ term. So the divergence of $\mathcal{E}_{\mu\nu}$ is constrained. In the non-conformal case, Equation (27) contains on the right hand side also the quadratic correction $\mathcal{S}_{\mu\nu}$ of the matter fields on the brane. The effective field equations, Equation (9), are then not a closed system. One needs the Bianchi equations. In fact, $\mathcal{E}_{\mu\nu}$ encodes corrections from the 5D graviton effects and are for the brane observer non-local. In our model under consideration, we have only the $T_{\mu\nu}^{(\omega)}$ term and no source terms (only the 5D Λ_5). But it still sources the KK modes. The dilaton ω plays the role of a "scalar field". But we don't need the 5D equations themselves, because the solution for N is the same! It is only the $\omega^{4/3}$ which represents the 5D contribution. There is no exchange of energy-momentum between the bulk and brane. If one applies the model to a FLRW model [31], then the evolution equations are very complicated. Inhomogeneous and anisotropic effects from the 4D matter radiation distribution on the brane are sources for the 5D Weyl tensor $\mathcal{E}_{\mu\nu}$ and cause non-local back-reaction on the brane. One needs an approximation scheme in order to find the missing evolution equation for $\mathcal{E}_{\mu\nu}$.

The locations of the horizon's and ergo-spheres are found by solving $N^2=0$ and $\bar{g}_{tt}=0$ respectively. N^2 becomes singular at coordinate time $t=t_H=-b_3+\sqrt[4]{-\frac{C_3}{C_2}}$. However, $\bar{g}_{\mu\nu}$ can be made regular everywhere and singular free by suitable choices of the parameters b_i, c_i and C_i. For $C_1=0$, $\bar{g}_{\mu\nu}$ has one real zero $r_H=\sim|1.606 b_2|$ and two complex zero's $\sim(0.178\pm 0.638I)b_2$. In Figure 3, we plotted the possible graphs. If one ignores the contribution from the bulk, then N^2 has for $C_1=0$ no real roots, so only naked singularities. The contribution from the bulk then generates at least one horizon.

3.3. Penrose Diagram

If we define the coordinates, $\mathrm{d}r^*\equiv\frac{1}{N_1(r)^2}\mathrm{d}r$ and $\mathrm{d}t^*\equiv N_2(t)^2\,\mathrm{d}t$, then our induced spacetime can be written as

$$\mathrm{d}s^2=\omega^{4/3}\bar{\omega}^2\left[\frac{N_1^2}{N_2^2}\left(-\mathrm{d}t^{*2}+\mathrm{d}r^{*2}\right)+\mathrm{d}z^2+r^2\left(\mathrm{d}\varphi+\frac{N^\varphi}{N_2^2}\mathrm{d}t^*\right)^2\right], \tag{28}$$

with

$$N_1^2=\frac{10b_2^3r^2+20b_2^2r^3+15b_2r^4+4r^5+C_1}{5r^2}, N_2^2=\frac{1}{C_2\left(t+b_3\right)^4+C_3} \tag{29}$$

and

$$r^*=\frac{1}{4}\sum_{r_i^H}\frac{r_i^H\log\left(r-r_i^H\right)}{\left(r_i^H+b_2\right)^3},\quad t^*=\frac{1}{4C_2}\sum_{t_i^H}\frac{\log\left(t-t_i^H\right)}{\left(t_i^H+b_3\right)^3}. \tag{30}$$

The sum it taken over the roots of $\left(10b_2^3r^2+20b_2^2r^3+15b_2r^4+4r^5+C_1\right)$ and

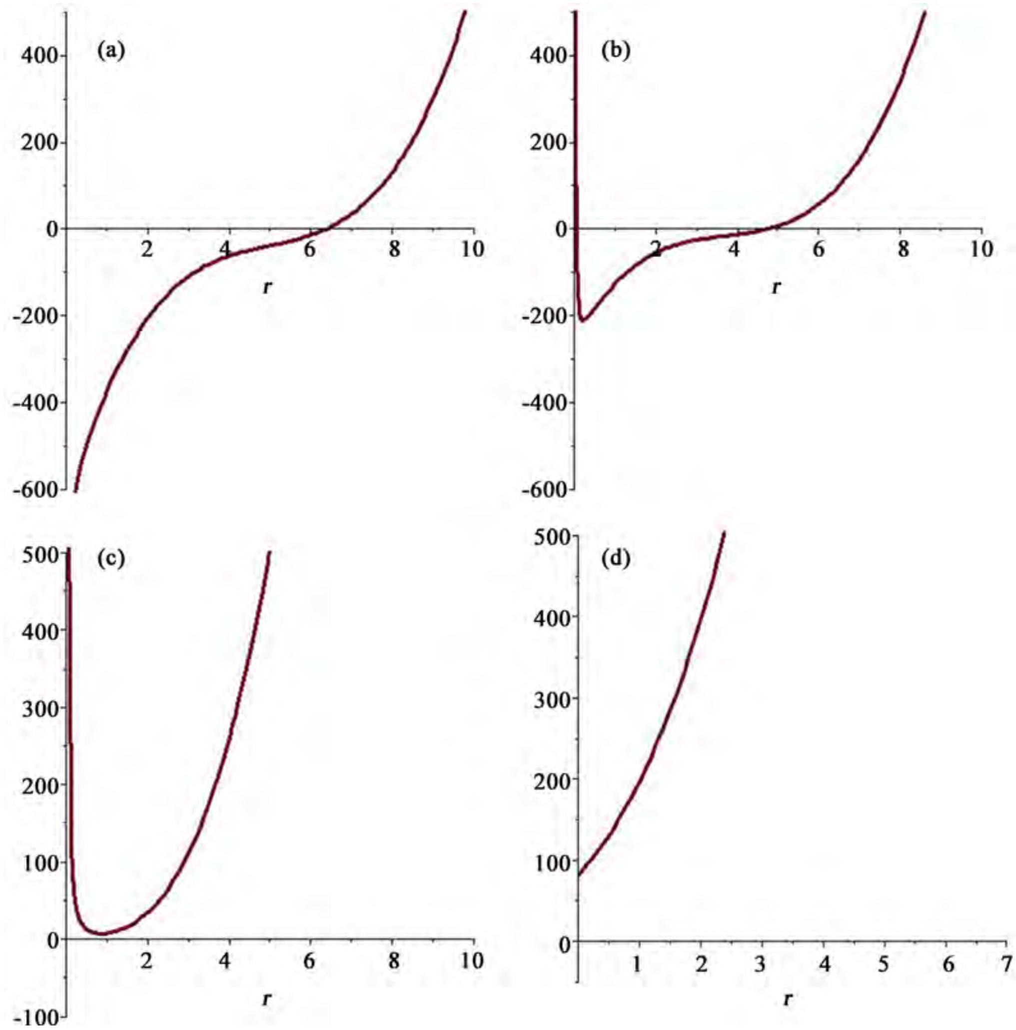


Figure 3. Four possible plots of N^2 as function of r.

$C_2(t+b_3)^4+C_3$, *i.e.*, r_i^H and t_i^H. This polynomial in r defining the roots of N_1^2, is a quintic equation, which has some interesting connection with Klein's icosahedral solution (see appendix). Further, one can define the azimuthal angular coordinate $\mathrm{d}\varphi^* \equiv \left(\mathrm{d}\varphi + \frac{N^\varphi}{N_2^2}\mathrm{d}t^*\right)$, which can be used when an incoming null geodesic falls into the event horizon. φ^* is the azimuthal angle in a coordinate system rotating about the z-axis relative to the Boyer-Lindquist coordinates. Next, we define the coordinates [32] (in the case of $C_1 = C_3 = 0$ and 1 horizon, for the time being)

$$\begin{aligned} U_+ &= \mathrm{e}^{\kappa(r^*-t^*)}, \quad V_+ = \mathrm{e}^{\kappa(r^*+t^*)} \quad r > r_H \\ U_- &= -\mathrm{e}^{\kappa(r^*-t^*)}, \quad V_- = -\mathrm{e}^{\kappa(r^*+t^*)} \quad r < r_H, \end{aligned} \tag{31}$$

with κ a constant. The spacetime becomes

$$\mathrm{d}s^2 = \omega^{4/3}\bar{\varpi}^2\left[\frac{N_1^2}{N_2^2}\log(UV)^{\frac{1}{2\kappa}}\mathrm{d}U\mathrm{d}V + \mathrm{d}z^2 + r^2\mathrm{d}\varphi^{*2}\right]. \tag{32}$$

In **Figure 4**, we plotted the Penrose diagram (a). The antipodal points $P(X)$ and $\bar{P}(\bar{X})$ are physically identified. If we compactify the coordinates,

$$\tilde{U} = \tanh U, \quad \tilde{V} = \tanh V, \tag{33}$$

then the spacetime can be written as

$$\mathrm{d}s^2 = \omega^{4/3}\bar{\omega}^2\left[H\left(\tilde{U},\tilde{V}\right)\mathrm{d}\tilde{U}\mathrm{d}\tilde{V} + \mathrm{d}z^2 + r^2\mathrm{d}\varphi^{*2}\right], \tag{34}$$

with

$$H = \frac{N_1^2}{N_2^2}\frac{1}{\kappa^2 \operatorname{arctanh}\tilde{U}\operatorname{arctanh}\tilde{V}\left(1-\tilde{U}^2\right)\left(1-\tilde{V}^2\right)}. \tag{35}$$

We can write r and t as

$$r = r_H + \left(\operatorname{arctanh}\tilde{U}\operatorname{arctanh}\tilde{V}\right)^{\frac{1}{2\kappa\alpha}}, \quad t = t_H + \left(\frac{\operatorname{arctanh}\tilde{V}}{\operatorname{arctanh}\tilde{U}}\right)^{\frac{1}{2\kappa\beta}}, \tag{36}$$

with

$$\alpha = \frac{r_H}{4\left(r_H + b_2\right)^3}, \quad \beta = \frac{1}{4C_2\left(t_H + b_3\right)^3}. \tag{37}$$

Observe that N_1 and N_2 can be expressed in $\left(\tilde{U},\tilde{V}\right)$. The Penrose diagram is drawn in **Figure 4(b)**. Note that $\mathrm{d}s^2$ and H are invariant under $\tilde{U} \to -\tilde{U}$ and $\tilde{U} \to -\tilde{U}$. $\tilde{g}_{\mu\nu}$ is regular everywhere and conformally flat. The "scale-term" H is consistent with the features of the Penrose diagram. Now we have still the φ dependency. We assume no z-dependency. It is expected that the differential equation for ω can be separated in a (U,V) part and a φ

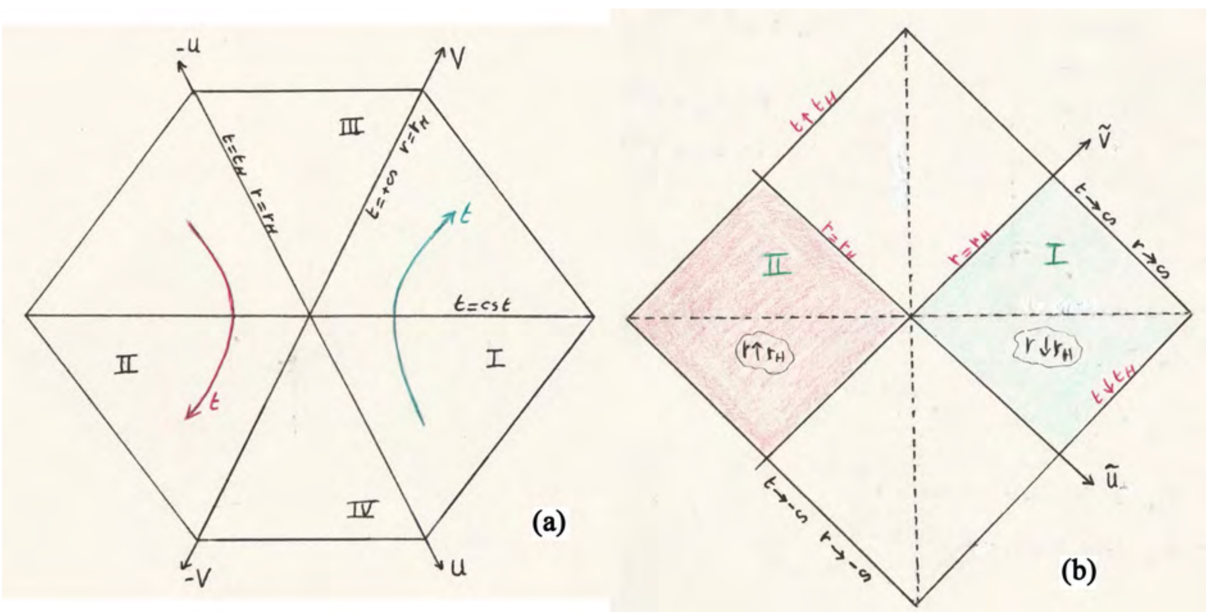


Figure 4. Plot of the Kruskal diagram for $\bar{g}_{\mu\nu}$ in (U,V) coordinates (a). The antipodal map between region I and II is quite clear here. If one approaches the horizon from the outside and passes the horizon, one approaches from the "other side" the horizon. One can also plot the Kruskal diagram for $\bar{g}_{\mu\nu}$ in $\left(\tilde{U},\tilde{V}\right)$ coordinates (b).

part. The method of 't Hooft can then be applied. In the next sections, we will briefly come back to this issue.

4. Related Issues of the New Black Hole Solution

4.1. Treatment of the Quantum Fields

The physical identification in the de Sitter spacetime of $P(X)$ and $\hat{P}(\hat{X})$ are considered as different representations in Kruskal space of one and the same Schwarzschild event. There is only one world with one singularity and one exterior region. Fields which are symmetric under J are identified as

$$\Psi_{JS} = \frac{1}{2}\left[\Psi(X) + \hat{\Psi}(\hat{X})\right]. \tag{38}$$

One then builds these fields from fields with arguments specified in [9]. Each of these fields, positive or negative frequency in I, can be extended to global spacetime surfaces. However, due to the time reversal, the inner product on the full Hilbert space has zero norm for the symmetric fields. One then defines negative frequency functions $\Psi^{\uparrow}_{(-)}(X) = \Psi^{\downarrow}_{(+)}(JX)$ and $\Psi^{\downarrow}_{(-)}(X) = \Psi^{\uparrow}_{(+)}(JX)$, where the arrows stands for the solutions on the future/past singularity. The symmetric (anti-) solutions ($\varepsilon = \pm 1$) are then

$$\Psi^{(1)}_{JS}(X) = \frac{1}{2}\left[\Psi^{\uparrow}_{(+)}(X) + \varepsilon\Psi^{\downarrow}_{(-)}(X)\right], \quad \Psi^{(2)}_{JS}(X) = \frac{1}{2}\left[\Psi^{\downarrow}_{(+)}(X) + \varepsilon\Psi^{\uparrow}_{(-)}(X)\right]. \tag{39}$$

Introducing then reflection and transmission coefficients, one can construct a wave function regular at the singularities, $\Psi^{(r)}_{JS} = \frac{\varepsilon K}{(K+\varepsilon)^2}\left[\Psi^{(1)}_{JS} + \varepsilon\Psi^{(2)}_{JS}\right]$, with $K = \mathrm{e}^{\pi\omega/\kappa}$, $\kappa = 1/4M$. Thereafter, one constructs hermitian field operators for the Fock space. Next, one needs the renormalized expectation value of the stress-energy tensor $\langle T_{\mu\nu}\rangle$ in the "semiclassical" equations of Einstein $G_{\mu\nu} = 8\pi G\langle T_{\mu\nu}\rangle$. If one assumes that there is a $r = 0$ singularity, then back-reaction will be small in the vicinity of the horizon (at least for massless fields). The spacetime can then be approximated by Schwarzschild geometry. The mass will decrease slowly with time and evaporates. In a flat spacetime, this is easily done, because the vacuum is well defined. One can calculate the zero-energy state and can construct finite quantum operators. In curved spacetime, the vacuum state is dependent of the boundary condition for the propagators (positive frequency modes). In principle, we can follow the method of Sanchez (for the de Sitter spacetime) for the dilaton field and our "un-physical" spacetime $\bar{g}_{\mu\nu}$ ($\Lambda = 0$),

$$\langle\bar{\omega}^2\rangle\bar{G}_{\mu\nu} = \left\langle T^{(\omega)}_{\mu\nu}(\bar{\omega}, \bar{g}_{\mu\nu})\right\rangle - \langle\bar{\omega}^2\rangle\mathcal{E}_{\mu\nu}, \tag{40}$$

where $T^{(\omega)}_{\mu\nu}$ depends on the geometry and boundary conditions (see Equation (19)). Further, $\langle T^{(\omega)}\rangle = -\langle\bar{\omega}^2\rangle\bar{R}$, because $\mathcal{E}_{\mu\nu}$ is traceless. We have now contributions from the antipode:

$$\left\langle T^{(\omega)}_{\mu\nu}\right\rangle \to \left\langle T^{(\omega)}_{\mu\nu}\right\rangle \pm \left\langle \hat{T}^{(\omega)}_{\mu\nu}\right\rangle, \quad \langle\bar{\omega}^2\rangle \to \langle\bar{\omega}^2\rangle \pm \left\langle\widehat{\bar{\omega}^2}\right\rangle. \tag{41}$$

In the simplified de Sitter space, one then easily construct Green functions [10]

$$G_{\alpha JS}(X,X') = \mathrm{e}^{2\alpha}\left[G(X,X') + G(X,JX')\right],$$
$$G_{\alpha JA}(X,X') = \mathrm{e}^{2\alpha}\left[G(X,X') - G(X,JX')\right], \tag{42}$$

with α labels the one parameter family of the de Sitter vacua. The expectation values for a scalar field and the energy momentum tensor can then be calculated. One obtain, for example [9],

$$\left\langle \widehat{\Phi^2_{JS,JA}} \right\rangle = \frac{1}{16\pi\cos\pi\nu}\left[m^2 + \left(\xi - \frac{1}{6}\right)R\right], \tag{43}$$

with $\nu = \left(9/4 - M^2/H^2\right)^{1/2}, M^2 = m^2 + \xi R$, m the mass of the field and $H \sim \Lambda$. Recently, a different analysis of perturbative quantum gravity on the de Sitter spacetime was done by Sofi, *et al.* [33].

In our case we have no scalar field, but instead ω. The expression for $T_{\tilde{U}\tilde{U}}$ becomes [34]

$$T^{(\omega)}_{UU} = \frac{c_1 \mathrm{e}^{-2c_1 U}}{c_3^2 c_4^2 \left(c_2\rho + c_3\right)^2}\left(c_2 c_3 \rho^2 F(U) - c_1^2 c_2^2 \rho^2 + c_1 c_3^2\right), \tag{44}$$

which can be used to evaluate the expectation value. In order to apply the full antipodal map, one includes the φ-dependency in the dilaton equation. The relevant operator (d'Alembertian) can be separated in the used coordinate system. The relevant φ contribution comes from periodic Mathieu functions (in variable φ). They converge uniformly on all compact sets in the *z*-plane. Next, one applies the method of 't Hooft, by expanding the position variables $u^\pm(z,\varphi)$ and momentum distributions $p^\pm(z,\varphi)$ in the partial waves of Mathieu functions[6]. Further, one then calculates the gravitational shift $\delta\tilde{U}(z,\varphi)$, in order to carry a particle over from I to II, or back [7], using the Shapiro delay.

4.2. The Surface Gravity and the Conformal Gauge

Since we have now the description of the antipodal map in our black hole spacetime, we will look more closely at the conformal invariance. First of all, one should rely in the dynamical situation on (conformal) Killing vectors in order to describe the spacetime symmetries. Our Lagrangian is conformal invariant under Equation (14), so we can use the freedom of the conformal factor Ω. Remember, different ω means different notion of the vacuum state for the in-going and outside observer, so they will use different conformal gauge freedom. It is desirable that for the out-going observer, the surface gravity of the horizon is conformal invariant. Further, conformal transformations must preserve affinely parameterized null geodesics. This will deliver Ω for the in-going observer. We can define out-going and in-going null normals [20] for $\bar{g}_{\mu\nu}$

$$\bar{l}^\mu = \left(1, N\sqrt{N^2 - r^2 N^{\varphi 2}}, 0, 0\right),$$
$$\bar{m}^\mu = \left(-\frac{1}{2r^2 N^{\varphi 2} - 2N^2}, -\frac{N}{2\sqrt{N^2 - r^2 N^{\varphi 2}}}, 0, 0\right). \tag{45}$$

[6]So the spherical harmonics are replaced by the Mathieu harmonics.

with $\bar{l}^{\mu}\bar{l}_{\mu}=\bar{m}^{\mu}\bar{m}_{\mu}=0$, $\bar{l}^{\mu}\bar{m}_{\mu}=-1$. The surface gravity then becomes

$$\kappa=2N\left(\partial_r\left(\sqrt{N^2-r^2N^{\varphi 2}}\right)+\partial_t\left(\frac{1}{N}\right)\right)=2N\left(\partial_t\sqrt{\bar{g}_{rr}}-\partial_r\sqrt{\bar{g}_{tt}}\right). \tag{46}$$

This is consistent with the metric definition of κ.

4.3. The Meaning of the Warped Spacetime

Let us now return to $g_{\mu\nu}=\omega^{4/3}\varpi^2\bar{g}_{\mu\nu}$. In the CDG setting, the evaporation of the black hole is also determined by the complementarity transformation of ω between the in-going and outside observer. Our spacetime is now ($b_4=b_3b_2$)

$$\begin{aligned} ds^2=\omega^{4/3}\varpi^2\Bigg[&\left(C_2\left(t+b_3\right)^4+C_3\right)\frac{10b_2^3r^2+20b_2^2r^3+15b_2r^4+4r^5+C_1}{5r^2}\\ &\cdot\left(-\mathrm{d}t^{*2}+\mathrm{d}r^{*2}\right)+\mathrm{d}z^2+r^2\left(\mathrm{d}\varphi+\frac{N^{\varphi}}{N_2^2}\mathrm{d}t^*\right)^2\Bigg],\end{aligned} \tag{47}$$

with

$$\omega^{4/3}\varpi^2=\frac{1}{\left(r+c_2\right)^2\left(t+c_3\right)^2\left(r+b_2\right)^2\left(t+b_3\right)^2}, \tag{48}$$

We observe that $\omega^{4/3}\varpi^2$ approaches zero for coordinate time $t\to\infty$, so $g_{\mu\nu}$ shrinks to zero, so the distant observer sees a gradually shrinking black hole when the metric time runs to infinity. Further, the only contribution from the 5D spacetime is the $\omega^{4/3}$. Remarkable, the projected Weyl component is necessary in order to obtain the same form of N^2 and to avoid naked singularities. So $\omega^{4/3}=\left[\left(r+c_2\right)^2\left(t+c_3\right)^2\right]^{-1}$ is the "scale" term from the 5D warped spacetime (the warpfactor in the RS model is the product of *y*-dependent part and ω part). Suppose one wants combine the conformal transformation with an internal symmetry transformation, *i.e.*, a spacetime transformation. In particular, the scale transformations. One can proof in that case, $\Box\log\Omega=0$ [21], which is consistent with our 2D null hypersurface of Equation (32). Further, in dimension $n\neq 4$ only the scale-invariant theories based upon scalar fields (so ω from 5D) are conformally invariant. Conclusion: ϖ of our $\bar{g}_{\mu\nu}$ can be used in non-vacuum models. An additional advantage of the warped spacetime in connection with cosmology and hierarchy problem was already mentioned in the introduction. A new aspect will be the embedding of the 5D in the 4D spacetime and the relation with the 3D BTZ blackhole solution.

4.4. The Relation with the 3D Bañados-Teitelboim-Zanelli Black Hole

In the spacetime under consideration, the $\mathrm{d}z^2$ term can be omitted. One obtains then the 3D Bañados-Teitelboim-Zanelli (BTZ) black hole spacetime. It solves the Einstein equations with a negative cosmological constant [35]. The BTZ solution is related to the AdS/CFT correspondence and intensively studies in connection with black hole entropy issues. However, we should like to take the cosmological constant zero. In a former study [34], an exact solution was found in a CDG setting in Eddington-Finkelstein retarded coordinates (U,ρ) (or advanced V) where the antipodal map $(U,V,\varphi)\to(-U,-V,\varphi+\pi)$ is applicable:

$$\mathrm{d}s^2=\frac{\mathrm{e}^{-2c_1U}}{(c_2\rho+c_3)^2}\left[\pm\frac{c_1\left(c_3^2-c_2^2\rho^2\right)}{c_2c_3}\mathrm{d}U^2-2\mathrm{d}U\mathrm{d}\rho+\mathrm{d}z^2+\rho^2\left(\mathrm{d}\xi+F(U)\mathrm{d}U\right)^2\right], \quad (49)$$

which is Ricci flat, while $\tilde{R}^{(4)}=\frac{6c_1c_2}{c_3}$. The function $F(U)$ will be fixed when matter terms are incorporated (*i.e.* for example, a scalar gauge field). The metric Equation (49) will then contain a term $b(U,\rho)^2\mathrm{d}\varphi^2$ and a relation like $\left(N^\xi\right)'=\frac{b}{\eta^2X^2+\omega^2}$ will be obtained. It has no curvature singularity. The location of the apparent horizon in U:

$$\rho_{AH}=\pm\frac{c_3}{\sqrt{c_2\left(c_2+\frac{c_3}{c_1}F(U)^2\right)}}, \quad (50)$$

with

$$\frac{\mathrm{d}\rho}{\mathrm{d}U}=\frac{1}{2}\mathrm{e}^{-2c_1U}\cdot\begin{cases}-\frac{c_1}{c_2c_3} & \rho\to 0\\ \frac{c_3F(U)^2+c_1c_2}{c_2^2c_3} & \rho\to\infty\\ 0 & \rho=\rho_{AH}\end{cases} \quad (51)$$

which is independent of ω. Here C_i are constants and $F(U)$ a function determined by the non-diagonal contribution. Further, we have

$$\lim_{\rho\to 0}g_{UU}\to\pm\frac{c_1}{c_2c_3\mathrm{e}^{2c_1U}}, \quad (52)$$

So when the evaporation speeds up, it approaches zero. We are dealing here with null-radiation in the (ρ,z)-plane. One could compare this solution with that found by Chan [36] in standard GR of a spinning black hole. They also find a solution for $F(U)$ which is determined by an energy-momentum tensor of null spinning dust. It is again curious that the "uplifted" BTZ has a solution, comparable with the "up-lifted" 5D solution.

5. Metric Fluctuation and Hawking Radiation

In the original deviation of the Hawking radiation, one uses the propagation of a linear quantized field in a classical background metric. However, near the horizon, high-frequencies metric fluctuations can contribute to the vacuum polarization and the impact of gravitational back reactions can be large. These zero-point fluctuations result in a modification of the Hawking radiation by gravitational waves [37]. One could question what the effect is of these waves in our

CDG model, where we have instead the dilaton field. Of course, one should need a quantum gravitational approach, which is not available yet. So need some approximation. However, effect of the scattering of these quanta at the horizon can be investigated in the context of the antipodal mapping considered here.[7] Without the contribution of the metric fluctuations, the mean number of quanta reaching $\mathcal{J}^+$ takes the form

$$\langle \bar{n}_\lambda \rangle_0 \sim \frac{1}{e^{2\pi E_+/\kappa} - 1}, \tag{53}$$

with E_+ the energy measured at $\mathcal{J}^+$ for the out modes. This is the Planck distribution with temperature $T = \kappa/2\pi = (8\pi M)^{-1}$. The correction terms can then be calculated by using the s-modes of a quantum massless scalar field and by using the fact that the in-going and out-going modes decouple [37]. One makes use of the mean energy flux, by calculating $\mathrm{d}E/\mathrm{d}U = 4\pi r^2 \langle T_{UU} \rangle_{ren}$, where the renormalized surface gravity is used. However, in this approximation, the reflection conditions are at r = 0, with in our antipodal map must be revised (we have no inside). We can use the (U,U) energy-momentum component of our model and can apply Equation (41) for the antipodal contribution.

Notice that the meaning of the local dilaton $\bar{\omega}$, is twofold. First, it determines the metric fluctuations (one also must incorporate in the dilaton equation the φ-dependency). Secondly, the in-going observer will use a different conformal gauge freedom Ω on $\bar{\omega}$ to describe the vacuum. Further, $\bar{\omega}$ is locally unobservable, unless we include metric fluctuations (gravitational waves. It will be necessary to compare this with the usual contribution using the Bunch-Davies method (and to taken into count the antipodal contribution). Note that the outside observer will use a different gauge and he/she experiences a mass $\sim \omega^2 N^2$ and Hawking radiation $\sim \partial_U (\omega^2 N^2)$, while for the in-going observer it is part of his vacuum. On the other hand, the outside observer is not aware of the antipodal identification. One could also say that they disagree about the observed scales. Or differently stated, they disagree about the back reaction from the Hawking radiation.

6. Conclusions

We investigated the conformal dilaton gravity model on a warped 5D spacetime, where the warp factor is interpreted as a dilaton field, to be treated as a renormalized quantum field. This approach is very suitable when one is dealing with a high curvature situation, for example, in the vicinity of the horizon of a black hole spacetime. It is a promising route to tackle the problems arising in quantum gravity models, such as the loss of unitarity when one investigates the Hawking radiation, emitted during the final stage of a black hole. Moreover, it could solve the information and firewall paradox. The basic concept behind the model is

[7]A suitable approximation is the high-frequency approximation applied to a Vaidya spacetime, where the not-flat background spacetime is distorted by the gravitational waves [38]. A recent application was provided by Slagter [39] [40].

conformal invariance, spontaneously broken when matter fields are incorporated in the Einstein-Hilbert action. The conformal symmetry group contains the antipodal map, so it is quite natural to apply the antipodal map on the black hole spacetime. It then turns out that the notion of the interior of the black hole changes dramatically, *i.e.*, there is no inside.

In this manuscript, we find an exact time dependent solution in the conformal dilaton gravity model on a warped 5D spacetime. The spacetime is written as ${}^{(5)}g_{\mu\nu} = \omega^{4/3}\,{}^{(5)}\tilde{g}_{\mu\nu}$ and ${}^{(4)}\tilde{g}_{\mu\nu} = \bar{\omega}^2\,{}^{(4)}\bar{g}_{\mu\nu}$. In our model, ω can be seen as the contribution from the bulk, while $\bar{\omega}$ is the brane component. It is conjectured that the different conformal gauge freedom, Ω, the in-going and outside observers possess, can be calculated by demanding a conformal invariant surface gravity and the preservation of affinely parameterized null geodesics. This means that the complementarity is expressed by the different notion of the vacuum state. The solution guarantees regularity of the action when $\omega \to 0$. We don't need a Weyl term in the action (generates negative metric states). Instead, we have a contribution from the bulk, i.e., the electric part of the 5D Weyl tensor. It is remarkable that the 5D field equations and the effective 4D equations can be written for general dimension *n*, with $n = 4, 5$. The energy-momentum tensor of the time-dependent dilaton, determining also the Hawking radiation, can be calculated exactly. By suitable choice of the parameters, the spacetime $\bar{g}_{\mu\nu}$ can be regular and singular free. In context of quantization procedures, counter terms in an effective action will cause problems, only in the bulk spacetime of the "large" extra dimension and not for the brane spacetime. When the extra dimensional volume is significantly above the Planck scale, then the true fundamental scale can be much less than the effective scale 10^{19} GeV. This means that no UV cut-off is necessary on the brane. This exact solution, nonetheless without mass terms, can be used to tackle the deep-seated problem of the black hole complementarity: the infalling and outside observer experience different ω by the choice of Ω. The solution fits also very well in the antipodal mapping, when crossing the horizon. The Penrose diagram for $\bar{g}_{\mu\nu}$, in suitable Kruskal coordinates, shows the features of the antipodal map of region I on region II: the inside of the black hole is removed. The in-going observer, when crossing the horizon, turns up at "the other side" of the horizon. The next task is to incorporate mass into our model and investigate the dilaton-scalar field interaction. The conformal invariance will then spontaneously be broken.

Acknowledgements

This research was supported by Asfyon, Astronomisch Fysisch Onderzoek Nederland.

Conflicts of Interest

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

References

[1] Hawking, S. (1974) *Nature*, **248**, 30-31. https://doi.org/10.1038/248030a0

[2] Hawking, S. (1975) *Communications in Mathematical Physics*, **43**, 199-220. https://doi.org/10.1007/BF02345020

[3] ‘t Hooft, G. (1993) Dimensional Reduction in Quantum Gravity. *Conference on Highlights of Particle and Condensed Matter Physics* (*SALAMFEST*), Trieste, 8-12 March 1993, 184.

[4] Page, D.N. (1993) *Physical Review Letters*, **71**, 3743-3746. https://doi.org/10.1103/PhysRevLett.71.3743

[5] Almheiri, A., Marolf, D., Polchinski, J. and Sully, J. (2013) *JHEP*, **2**, 62. https://doi.org/10.1007/JHEP02(2013)062

[6] ‘t Hooft, G. (2016) *Foundations of Physics*, **46**, 1185-1198. https://doi.org/10.1007/s10701-016-0014-y

[7] ‘t Hooft, G. (2018) The Firewall Transformation for Black Holes and Some of Its Implications.

[8] ‘t Hooft, G. (1984) *Journal of Geometry and Physics*, **1**, 45-52. https://doi.org/10.1016/0393-0440(84)90013-5

[9] Sanchez, N. and Whiting, B.F. (1987) *Nuclear Physics B*, **283**, 605-623. https://doi.org/10.1016/0550-3213(87)90289-6

[10] Sanchez, N. (1986) Two- and Four-Dimensional Semi-Classical Gravity and Conformal Mappings.

[11] Folacci, A. and Sanchez, N. (1987) Quantum Field Theory and the Antipodal Identification of de Sitter Space. Elliptic Inflation.

[12] Schrödinger, E. (1957) Expanding Universe. Cambridge Univ. Press, Cambridge.

[13] ‘t Hooft, G. (2018) Discreteness of Black Hole Microstates.

[14] ‘t Hooft, G. (2018) *Foundations of Physics*, **48**, 1134-1149. https://doi.org/10.1007/s10701-017-0133-0

[15] ‘t Hooft, G. (2015) Diagonalizing the Black Hole Information Retrieval Process.

[16] Zakharov, V.D. (1973) Gravitational Waves in Einstein’s Theory. John Wiley & Sons, Inc., New York.

[17] Wald, R.M. (1994) Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics. The Univ. of Chicago Press, Chicago.

[18] Codello, A., D’Odorico, G., Pagani, G. and Percacci, R. (2013) *Classical and Quantum Gravity*, **30**, Article ID: 115015. https://doi.org/10.1088/0264-9381/30/11/115015

[19] Alvarez, E., Herrero-Valea, M. and Martin, C.P. (2014) *JHEP*, **10**, 214. https://doi.org/10.1007/JHEP10(2014)115

[20] Slagter, R.J. (2021) Conformal Dilaton Gravity and Warped Spacetimes in 5D.

[21] Felsager, B. (1998) Geometry, Particles and Fields. Springer, New York. https://doi.org/10.1007/978-1-4612-0631-6

[22] Mannheim, P.D. (2005) *Progress in Particle and Nuclear Physics*, **56**, 340-445. https://doi.org/10.1016/j.ppnp.2005.08.001

[23] Randall, L. and Sundrum, R. (1999) *Physical Review Letters*, **83**, 3370-3373. https://doi.org/10.1103/PhysRevLett.83.3370

[24] Randall, L. and Sundrum, R. (1999) *Physical Review Letters*, **83**, 4690-4693.

https://doi.org/10.1103/PhysRevLett.83.4690

[25] Slagter, R.J. and Pan, S. (2016) *Foundations of Physics*, **46**, 1075. https://doi.org/10.1007/s10701-016-0002-2

[26] Maldacena, J. and Milekhin, A. (2021) *Physical Review D*, **103**, Article ID: 066007. https://doi.org/10.1103/PhysRevD.103.066007

[27] Maldacena, J. (2011) Einstein Gravity from Conformal Gravity.

[28] Islam, J.N. (1985) Rotating Fields in General Relativity. Cambridge Univ. Press, Cambridge. https://doi.org/10.1017/CBO9780511735738

[29] Shiromizu, T., Maeda, K. and Sasaki, M. (2000) *Physical Review D*, **62**, Article ID: 024012. https://doi.org/10.1103/PhysRevD.62.024012

[30] Shiromizu, T., Maeda, K. and Sasaki, M. (2003) *Physical Review D*, **7**, Article ID: 084022. https://doi.org/10.1103/PhysRevD.67.084022

[31] Maartens, R. and Koyama, K. (2010) *Living Reviews in Relativity*, **13**, 5. https://doi.org/10.12942/lrr-2010-5

[32] Strauss, N.A., Whiting, B.F. and Franzen, A.T. (2020) Classical Tools for Antipodal Identification in Reissner-Nordstrom Spacetime. https://doi.org/10.1088/1361-6382/ab9a9d

[33] Sofi, A.H., Akhoon, S.A., Rather, A.A. and Maini, A. (2015) *Open Access Library Journal*, **2**, 1. https://doi.org/10.1088/1361-6382/ab9a9d

[34] Slagter, R.J. (2019) On the Dynamical BTZ Black Hole in Conformal Gravity. In Spacetime 1909-2019. *Proceedings Second H. Minkowski Meeting*, Albena, May 2019, 17-37.

[35] Compère, G. (2019) Advanced Lectures on General Relativity. Lecture Notes in Physics 952, Springer, Heidelberg. https://doi.org/10.1007/978-3-030-04260-8

[36] Chan, J.S.F., Chan, K.C.H. and Mann, R.B. (1996) *Physical Review D*, **54**, 1535-1539. https://doi.org/10.1103/PhysRevD.54.1535

[37] Barrabès, C. and Hogan, P.A. (2013) Advanced General Relativity. Oxford Univ. Press, Oxford. https://doi.org/10.1093/acprof:oso/9780199680696.001.0001

[38] Choquet-Bruhat, Y. and Geroch, R.P. (1969) *Communications in Mathematical Physics*, **14**, 329-335. https://doi.org/10.1007/BF01645389

[39] Slagter, R.J. and Miedema, P.G. (2020) *Monthly Notices of the Royal Astronomical Society*, **459**, 3054. https://doi.org/10.1093/mnras/staa3840

[40] Slagter, R.J. (2021) New Evidence of the Azimuthal Alignment of Quasars Spin Vector in the LQG U1.28, U1.27, U1.11, Cosmologically Explained. https://doi.org/10.20944/preprints202103.0560.v1

[41] Toth, G. (2002) Finite Möbius Groups, Minimal Immersions of Spheres and Moduli. Springer, Heidelberg. https://doi.org/10.1007/978-1-4613-0061-8

[42] Klein, F. (1888) Lectures on the Icosahedron and the Solution of Equations of the Fifth Degree. Trübner & Co, London.

Appendix: The Quintic Horizon Equation and Related Issues

Our quintic polynomial, determining the horizons,

$$f = r^5 + \frac{15}{4} b_2 r^4 + 5 b_2^2 r^3 + \frac{5}{2} b_2^3 r^2 + \frac{C_1}{4} = 0, \quad (54)$$

can be written by a, so-called Tschirnhaus transformation, in the form

$$r^5 - \frac{15 b_2}{16}\left(C_1 + b_2^5\right) r^2 - \frac{125 b_2^3}{256}\left(C_1 + b_2^5\right) r - \frac{1}{16}\left(C_1 + b_2^5\right)^2 = 0 \quad (55)$$

By scaling, this form can be reduced to the Bring-Jerrard form $r^5 + r - c$, with c a function of b_2 and C_1 [41]. There is an interesting relation between the symmetry group of the icosahedron and our quintic equation. The symmetry group is isomorphic with the Galois group A_5 (of an irreducible quintic polynomial). The icosahedron is dual to the dodecahedron, *i.e.*, their symmetries are isomorphic. The A_5 is interesting in physics, because it is a simple group having no invariant subgroups. It has three orbits, which are invariant under the antipodal map. So the connection with the Möbius group is clear (see section 2.3). For details, we refer to Toth [41]. It was Klein [42], who first became aware of the relation between the solutions of the quintic equation and the icosahedron.

It is conjectured that our quintic polynomial (Equation (54)) has a deep-seated relation with the 5D spacetime solution. Further, it is remarkable that the resulting quintic equation is independent of the dimension of our manifold ($n = 4,5$). Moreover, the nice fitting of the antipodal map in our model cannot be a coincident. From Equation (26) we observe that the derivative of f is $5r\left(r + b_2\right)^3$. So it is expected that our quintic equation results from a immersion[8] of a closed surface S in $\mathbb{R}^3$ into $\mathbb{R}^4$.

This is currently under investigation by the author.

[8]An immersion is a differentiable function between differentiable manifold whose derivative is everywhere injective. It is also a topological embedding.

Journal of Modern Physics, 2021, 12, 1783-1794
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Lunar Eclipses and Allais Effect

Russell Bagdoo

Saint-Bruno-de-Montarville, Quebec, Canada
Email: rbagdoo@gmail.com

How to cite this paper: Bagdoo, R. (2021) Lunar Eclipses and Allais Effect. *Journal of Modern Physics*, **12**, 1783-1794.
https://doi.org/10.4236/jmp.2021.1213104

Received: September 8, 2021
Accepted: November 16, 2021
Published: November 19, 2021

Abstract

Two anomalies observed during lunar eclipses, the enlargement of the Earth's shadow and the excessive clarity of the penumbra, possibly attributed to insufficient causes if not doubtful, would refute the assertion of certain experimenters according to which the lunar Allais eclipse effect would be almost impossible to detect. The Earth's umbra seems to be 2% larger than what is expected from geometrical considerations and it is believed that the Earth's atmosphere is responsible for the extent of the enlargement, but it is realized that the atmospheric absorption cannot explain light absorption at a height as high as 90 km above the Earth, as required by this hypothesis. It was also argued that the irradiation of the Moon in the Earth's shadow during the eclipse is caused by the refraction of sunlight in the upper regions of the Earth's atmosphere. However, the shade toward the center is too bright to be accounted for by refraction of visible sunlight. Although these assumptions are not trifling, we attribute the majority of these abnormalities to the Allais eclipse effect. This effect would cause a slight decrease of gravity during the eclipse: the geodesics would be displaced a small amount outwards; the ray of light coming from the Sun, passing close by the Moon would be less attracted, which would expand the shadow cone of the Moon. On the other hand, the rays emanating from the Moon would have a shorter wavelength and therefore the luminescence would increase by anti-Stoke Raman effect: the scattered photon has more energy than the absorbed photon.

Keywords

Lunar Eclipses, Enlargement of the Earth's Shadow, Luminescence, Allais Eclipse Effect, Anti-Stokes Raman Effect

1. Introduction

We know that astronomical data give us accurate values of the radii of the Sun, the Earth and the Moon. Furthermore, the knowledge of their relative distances predicts quite accurately the instant when the umbra-penumbra limit sweeps some specific craters on the Moon during lunar eclipses. Since the 1830s, crater timing has been used during lunar eclipses to measure the length of the Earth's shadow. The method is simple: one takes the timing of lunar features (craters, limbs, ridges, peaks, bright spots) as they enter and exit the umbra. The Sun-Earth-Moon geometry being known quite precisely is then possible to calculate the size and shape of the Earth's umbra at the Moon. Measurements that vary from one eclipse to the next can now be made with low-power telescopes or a clock synchronized with radio time signals. However, it has systematically been found that the shadow of the Earth seems to be 2% larger than what is expected from geometrical predictions.

Even, if it is believed that the thickness of the Earth atmosphere is responsible for that displacement [1], it was realized that the atmospheric absorption cannot explain the absorption of light at a height of up to 90 km above the Earth, as required by this hypothesis. It may be noted in particular that Link [2] has firmly established a relationship between the enlargement of the Earth's shadow during lunar eclipses and the presence of meteors, which have the ability to distort the optical properties of the atmosphere when they are braked at high altitudes [3].

It has been said that the pronounced red colour in the inner portions of the umbra during an eclipse of the Moon is caused by refraction of sunlight through the upper regions of the Earth's atmosphere, but the umbral shadow towards the centre is too bright to be accounted for by refraction of visible sunlight.

In Sections 2 and 3, we give a brief history of the enlargement of the Earth's umbra and the excess of light into the Earth's shadow onto the Moon during lunar eclipses. We present some accepted interpretations and we show how the Allais effect, which occurs at the time when problems arise related to these anomalies, leads us to reject these interpretations. In Section 4, it emerges from a discussion that, failing to have an answer that would explain the two coexisting anomalies, the Allais eclipse effect currently remains the only viable option. Experiments are proposed as much to corroborate the observations of the two anomalies as to test the Allais eclipse effect. We conclude that both anomalies during lunar eclipses are caused by a lunar Allais effect.

2. Enlargement of the Earth's Umbra

2.1. Brief History of the Enlargement of the Earth's Umbra on the Moon during Lunar Eclipses

In the early 1700s, Philippe de La Hire made a curious observation about Earth's umbra. The predicted radius of the shadow needed to be enlarged by about 1/41 in order to fit timings made during an eclipse of the Moon (La Hire 1707). Additional observations over the next two centuries revealed that the shadow enlargement was somewhat variable from one eclipse to the next [4].

Chauvenet (1891) adopted a value of 1/50, which has become the standard enlargement factor for lunar eclipse predictions published by many national institutes worldwide. Some authorities dispute Chauvenet's shadow enlargement convention [5]. Danjon (1951) notes that the only reasonable way of accounting for a layer of opaque air surrounding Earth is to increase the planet's radius by the altitude of the layer [6]. This can be accomplished by proportionally increasing the parallax of the Moon. The radii of the umbral and penumbral shadows are then subject to the same absolute correction and not the same relative correction employed in the traditional Chauvenet 1/50 convention. Danjon estimates the thickness of the occulting layer to be 75 km and this results in an enlargement of Earth's radius and the Moon's parallax of about 1/85. Since 1951, the French almanac *Connaissance des Temps* has adopted Danjon's method for the enlargement Earth's shadows in their eclipse predictions

Danjon's method correctly models the geometric relationship between an enlargement of Earth's radius and the corresponding increase in the size of its shadows. Meeus and Mucke (1979), and Espenak (2006), both use Danjons method. However, the resulting umbral and penumbral eclipse magnitudes are smaller by approximately 0.006 and 0.026 respectively as compared to predictions using the traditional Chauvenet convention of 1/50.

For his part, in an analysis of 57 eclipses over a period of 150 years, Link (1969) found an enlargement of the shadow of 2.3% on average. Furthermore, schedules inputs and outputs of the crater through the umbra for four lunar eclipses from 1972 to 1982 strongly support the Chauvenet value of 2%. Of course, the small magnitude difference between the two methods is difficult to observe because the edge of the umbral shadow is diffuse. From a physical point of view, there is no well defined border between the umbra and the penumbra. The shadow density actually varies continuously as a function of radial distance from the central axis out to the extreme limit of the penumbral shadow. However, the density variation is most rapid near the theoretical edge of the umbra. Kuhl's (1928) contrast theory demonstrates that the verge of the umbra is perceived at the point of inflexion in the shadow density. This point appears to be equivalent to a layer in Earth's atmosphere at an altitude of about 120 to 150 km. The net enlargement of Earth's radius of 1.9% to 2.4% corresponds to an extension of the umbra of 1.5%, to 1.9%, in reasonably good agreement with the conventional value.

It seems that the increase of the Earth's umbral shadow during eclipses of the Moon is the classical value of 2% (the rule of the fiftieth) used in most calculations of lunar eclipses [7].

2.2. Accepted Interpretation of the Enlargement of the Umbra

Numerous reports show that the umbra-penumbra limit appears significantly displaced on the moon during an eclipse. It is believed that the thickness of the Earth atmosphere is responsible for that displacement [8] [9] [10]. In order to study more deeply the phenomenon showing that the umbra-penumbra limit appears significantly displaced on the Moon during an eclipse, it is important to evaluate if the reported increase of 2% of the Earth's shadow on the Moon corresponds to a reasonable value of the height at which the atmosphere is opaque. Calculations indicate that this enlargement corresponds to a terrestrial altitude of 92 km.

This usual interpretation of the umbral enlargement forces us to believe that the atmosphere is normally opaque up to 92 km or so. But how is this possible when, at this altitude, the air is extremely rarefied? It is the altitude close to the orbit on which a satellite travels around the Earth.

In fact, according to data [11], the atmospheric pressure at 90 km above sea level is about half a million times smaller than that at sea level. Above 15 km, the atmosphere becomes relatively transparent to light, since 90% of the air and almost all the humidity and pollution are below that level. That makes an enlarged obscuration due to the opacity of the atmosphere of only 0.3% which is much smaller than the 2.0% reported.

Furthermore, the eruption of volcanos cannot explain the larger shadow. According to some, the altitude reached by some material ejected from volcano El Chichon is in the stratosphere, some 26 kilometers (16 miles) above Earth's surface – roughly 50% higher than material from the famous Mount St. Helens [12]. Since the atmosphere does not appear to be responsible for the umbra-penumbra limit displacement of 2% on the Moon, then what is the cause?

F. Link argues that the meteoric dust in the upper atmosphere of the Earth is at the origin of the additional weakening of the light and the expansion of the Earth's darkness [13] [14]. We might point out, in particular, that Link has actually established a concomitance between the enlargement of the Earth's umbra during lunar eclipses and the presence of meteors, which are capable of distorting the optical properties of the atmosphere when they are decelerated at high elevations [3] [15].

Paul Marmet and Christine Couture [1], for their part, believe that the actual umbra of the Earth projected on the Moon is not as big as observed, that the sensitivity of the eyes is a factor leading necessarily to an umbral enlargement and that almost the totality of the reported umbra-penumbra limit displacement is an optical effect that has nothing to do with the thickness of the Earth atmosphere.

For our part, we believe that the observed times to browse the path of the Moon through the Earth's obscurity deviate from the predicted times and that some variations in colour, size and shape of the umbra occur in the darkness. We attribute this deviations and variations to the Allais eclipse effect.

2.3. Umbral Enlargement and the Allais Eclipse Effect

During an eclipse of the Moon, it is predicted geometrically that the photons from the Sun describe a rectilinear trajectory as if they were little deflected and pass at a minimum approach distance R'_E from the centre of the Earth (slightly larger than the radius R_E of the Earth), before moving to the Moon. A ray of sunlight passes close to the Moon at point R'_M of minimum approach to arrive at the point P at the end of the shadow cone of the Earth. The trajectory followed by the solar photons shapes the curvature of minimum approach of the Earth and the Moon.

However, we have serious reasons to believe that during a lunar eclipse, with the Earth interposed between the Moon and the Sun, there would be a kind of anti-gravity on the Moon which would be manifested by a deviation of light. This is precisely the Allais effect [16] [17] [18].

The orbital radius seems longer, which means that the curvature of minimum approach of the Earth and the Moon is shifted outward. Since the curvature is the inverse of the square of the radius, the curvature is even smaller than the radius is large. This is grounded in the Newtonian logic stating that gravity identified to the curvature is all the more weak as the orbital radius is large. The deviated photons will pass at a distance of minimum approach $R''_E\left(R''_E = R_E + \Delta R_E\right)$ from the centre of the Earth and at the point $R''_M\left(R''_M = R_M + \Delta R_M\right)$ of minimum approach of the Moon to end at the point $P''\left(P'' = P + \Delta P\right)$, casting an enlarged umbra cone. It matches with the observed cone of the enlargement of the Earth's umbra.

During a lunar eclipse, it is predicted geometrically that the photons from the Sun describe a rectilinear trajectory as if they were a little deflected (**Figure 1**).

Inevitably, sunlight, observed on the eclipsed Moon, will tend to move away. The photon has an "inertial mass" equal to $h\nu/c^2$ equivalent to a "gravitational mass" also equal to $h\nu/c^2$. The energy of the "fallen" photon from the Sun will be $m(g - g')H$ instead of mgH (m = inertial and gravitational mass of the photon; g = acceleration due to gravity, H = height). The gravitational mass of photons, lessened, takes distance, so increasing the darkness.

We are witnessing an abnormal gravitational frequency shift.

Suppose that in normal times the light is emitted by the Sun at the height H (distance Sun-moon) [19]. The total energy of a photon of frequency ν and

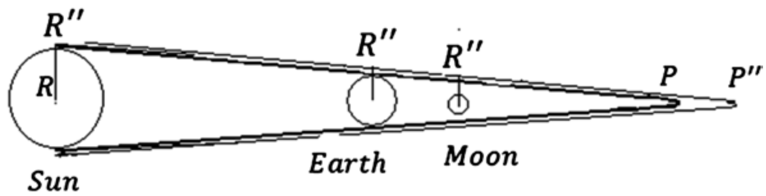


Figure 1. Exaggerated diagram (for comprehension) of the cones of the umbra. The illumination of the Earth by the Sun projects into space a converging cone of umbra and a divergent cone of penumbra whose conical generators are tangent to the Earth and the Sun. We represent here that the converging cone of umbra to illustrate the enlargement of the earth's shadow. The umbra cone does not completely obscure the Moon and, as early as the 18th century, astronomers knew that the shadow limit was a little further (P'') than according to the geometric path of the rays (P). The quasi-parallel interior tangents to the Sun and the Earth give the two interior cones, predicted and observed, of the umbra. The tangent $R''_S R''_E R''_M P''$ gives the observed cone of the umbra while the tangent $R_S R_E R_M P$ gives the calculated cone of the umbra.

energy $h\nu$, reaching the lunar surface has become

$$h\nu' = h\nu + h\nu gH/c^2 \quad (1)$$

The receiver, placed on the lunar ground, detects a frequency ν' greater than ν of the solar source (g designates the lunar gravitational field):

$$\nu' = \nu\left[1 + gH/c^2\right]. \quad (2)$$

During a lunar eclipse, due to a potential loss of attraction, the lunar gravitational field g amounts to $g - g'$. When a photon emitted by the Sun reaches the surface of the Moon, he lost potential energy $\left(h\nu/c^2\right)(g - g')H$ and won the kinetic energy $\left(h\nu/c^2\right)(g - g')H$. Its total energy has become

$$h(\nu' - \nu'') = h\nu + \left(h\nu/c^2\right)(g - g')H. \quad (3)$$

The frequency $\nu' - \nu''$ of the photon at its arrival at the surface of the eclipsed Moon is less red-shifted relative to its initial frequency, according to the relation

$$\nu' - \nu'' = \nu\left[1 + (g - g')H/c^2\right]. \quad (4)$$

The receiver detects a frequency $\nu' - \nu''$ slightly smaller than ν' of not eclipsed Moon. This means a small blue shift for the Sun during the eclipse.

If this hypothesis is correct which consists to declare that the Allais effect causes a kind of repulsion between the three celestial bodies involved, there should be a variation of the gravitational potential. This means that the gravitation will influence the geometry of space-time: the time of clocks and the length measured by a rule will be affected depending on whether there is more or less gravity. Einstein's general theory of relativity predicts that a clock in the presence of weak gravity runs more rapidly than one located where gravity is stronger. Consequently, the frequencies of radiation emitted by atoms in the presence of a weak gravitational field are shifted to higher frequencies when compared with the same emissions in the presence of a strong field. The light of the Sun observed on an eclipsed Moon should be blue shifted; a fraction of the solar gravitational redshift ($g'R/c^2$) which is about two parts in a million [20].

The widening of the Earth's shadow on the Moon would not be due to a greater density of the upper atmosphere of the Earth, which would make it as opaque as the lower atmosphere, it would be caused by a gravitational potential temporarily decreased. An "alleviated" matter would dictate to space-time a smaller degree of curvature; the space-time would in turn impose to matter to move on a larger orbital radius.

3. Excess of Light into the Earth's Shadow

3.1. Brief History of the Excess of Light into the Earth's Shadow on the Moon during Lunar Eclipses

The first work on the variation in brightness of eclipses was executed by André-Louis Danjon in 1920. He devised a scale of brightness for total lunar eclipses, from 0 for invisible to 4 for very brilliant. He used it to analyze data on eclipses extending back over three and a half centuries, and showed a correlation between the eclipse brightness and the solar activity. But a series of three-color photometric observations of Moon eclipse, made by him and his associates between 1932 and 1957, appears to show a clear correlation between the eclipse brightness and the geomagnetic planetary index K_p [21]. So the new data seems to contradict the first Danjon's conclusion.

An attempt to interpret the relation with K_p in terms of lunar luminescence indicates that the change in the eclipse brightness is in accord with the rate of increase of the plasma energy, as predicted by the experiment of Snyder *et al.* (1963) [22]. However, there is some difficulty in the required proton density being greater than the observed value by an order of magnitude. The same calculations show that luminescence to be visible in ordinary moonlight requires a plasma energy at least three orders of magnitude greater than the maximum value predicted by Snyder *et al.* They also show that the reported dates of these observations fall on geophysically quiet days, as well as on dates of high K_p. The above conclusion agrees with the result of calculations by Ney *et al.* in 1966.

On the other hand, J. Dubois and F. Link in 1969 found a correlation between the brightness of the eclipsed Moon and the solar activity, as had been suggested by Danjon on the basis of its first observations. They demonstrated that the brightness of eclipse was related not only to the sunspots number but also to the height of the latitude. They showed an annual correlation between the heliographic latitude of the apparent centre of the Sun's disk and eclipse brightness [23] [24].

It was suggested that the brightness anomaly of the umbral region during an eclipse of the Moon would be caused by refraction of sunlight through the upper regions of the Earth's atmosphere. The red coloring arises because, they say, sunlight reaching the Moon must pass through a long and dense layer of the Earth's atmosphere, where it is scattered. Shorter wavelengths are more likely to be scattered by the small particles and so, by the time the light has passed through the atmosphere, the longer wavelengths dominate. This resulting light we perceive as red. The amount of refracted light depends on the amount of dust or clouds in the atmosphere; this also controls how much light is scattered. In general, the dustier the atmosphere, the more that other wavelengths of light will be removed (compared to red light), leaving the resulting light a deeper red color [25].

Despite this reasoning, it has been found that towards the centre the umbra is too bright to be accounted for by refraction of visible sunlight. F. Link proposed that this excess be interpreted as luminescence [26]. He concluded that about 10 percent of the Moon's optical radiation is caused by luminescence. Observations seem to confirm the existence of lunar luminescence. The term luminescence can be applied to any object that emits light in addition to the usual reflected light [27]. The main characteristic of luminescence is that the emitted light is an attribute of the object itself, and the light emission is stimulated by some internal or external process. As external process, Link suggested the luminescence of the lunar surface by X-ray bombardment from the uneclipsed regions of the solar corona, as suggested by Link. This theory is supported by the variation of a factor of 100, between solar maximum and minimum, of the intensity of certain wavelengths of X-rays [28].

Another possible mechanism by which the eclipsed Moon shines results from the fact that the Moon is covered by a fine layer of meteoric dust, and would therefore contain quantities of the *achondritic enstatites*. This type of stony meteorite produces the luminescence when protons and electrons of the solar wind are deflected and impinge on the lunar surface during a total eclipse. An experiment performed by Zdenek Kopal, C. J. Derham and J. E. Geakel in 1963 showed that certain meteorite specimens glowed with a strange red light, same colour as the umbra in eclipse, when bombarded by high energy protons in the laboratory [24].

It appears to us that the excess of irradiation of the Moon in the shadow of the Earth during the eclipse is partially caused by refraction in the atmosphere, but that it prevailingly depends of the light emission stimulated by an internal process linked to the Allais eclipse effect.

3.2. Luminescence of the Eclipsed Moon and Allais Effect

The time of vibration (T) of atoms and molecules of luminescent gases in the fieldless region of space is $T = \tau$ (T is the time of vibration in the atom at rest; τ is the modified time of vibration) [29]. In a region of space with the gravitational field, the time of vibration is altered to

$$\tau = T/(1-\gamma/2) = T/\left(1-v^2/c^2\right) = T/\left(1-\phi/c^2\right) \tag{5}$$

[γ contains the gravitational potential ϕ ($\gamma = 2GM/Rc^2 = 2v^2/c^2 = 2\phi/c^2$)].

Our assumption is that at the surface of an eclipsed Moon there is a weaker gravitational field than in a frame of reference without eclipse ($\gamma' < \gamma$)

$$\underset{\text{with eclipse}}{\left[\tau = T/(1-\gamma'/2)\right]} < \underset{\text{without eclipse}}{\left[\tau = T/(1-\gamma/2)\right]} \tag{6}$$

During the lunar eclipse time, the gravitational potential ϕ of the Moon is conjecturally diminished; the time τ of the vibration of the atom is shorter. The metric of the obscured Moon is affected and the ticking of time accelerates relative to the system without eclipse in which the atom is considered at rest.

Consequently, the red shift of the spectral lines of light that comes from the layer of particles on the ground will have a small additional blue "Allais" shift which reduces the "Einstein" redshift [19]. As the Einstein effect (*i.e.* the tiny frequency shift of spectral lines in a gravitational field) is directed towards the blue, there is thus more internal electromagnetic energy. It appears that this blueshift by variation in the reduction of the mass could be a form of atomic excitement at the level of electrons, as Brownian motion. More specifically, we would say that the Allais eclipse effect would have engendered a significant change of wavelength within the molecules of matter [30]. The excess of luminescence would be the imprint left on the light by the intramolecular oscillation of the atoms constituting the molecules of the lunar soil which spreads it. A Raman effect caused by an Allais effect, in some way.

3.3. An Anti-Stoke Raman Effect Induced by an Allais Effect

We assume that the lunar gravitational potential ϕ can be reduced in times of eclipse, what would accelerate the vibration of atoms.

A molecule can be excited to a very high energy state. The amount of energy necessary to reach this excited state is $h\nu_o$. Therefore, the relaxation of the molecule to the ground-state vibrational energy level $v=0$ results in the emission of a photon of energy $h\nu_o$. This emission is usually observed in the visible spectral region and is called Rayleigh scattering. We think that the rapid oscillation of a light wave passing by the intramolecular level of atoms which constitute the molecules diffused by the lunar soil could be similar to an effect Raman anti-Stoke [31].

The scattering of light on the optical modes is designated Raman effect. It is different from the Rayleigh scattering because the scattered light changes the frequency of the spectrum active vibration. Historically, the effect was first observed with molecules. Molecules vibrate, and each molecular oscillation corresponds to a certain amount of energy. In the scattering process, this energy is added or subtracted from the incident light. An anti-Stokes Raman effect occurs when the molecule absorbs an incident light of frequency ν_o and reemits it at a higher frequency.

Thus, during the eclipse of the Moon, the excited molecule would oscillate from a superior vibrational energy level, say $v=1$. The energy absorbed in this process is still $h\nu_o$. The molecule can relax to the original $v=1$ vibration energy level and emits a photon $h\nu_o$; however, the relaxation can be to the ground state. The return to the state $v=0$ results in the emission of a photon which is $h\nu_1$ greater than the exciting energy $h\nu_o$ from level 1. The photon energy emitted is $h(\nu_o+\nu_1)$. Spectral lines with frequencies higher than ν_o are labeled anti-Stokes lines [32]. This Raman shift induces a brighter electromagnetic radiation.

4. Discussion and Conclusion

In 2009, NASA's *Lunar Reconnaissance Orbiter* (LRO) was launched with the *Lunar Crater Observation and Sensing Satellite* (LCROSS) on the first U.S. mission to the Moon in over 10 years. LRO gathered information on day-night temperature maps, contributed data for a global geodetic grid, and conducted high-resolution imaging. During lunar eclipses, the solar-powered orbiter also falls in Earth's shadow, cutting it off from the source of its power. The mission controllers can then use an instrument—called *Diviner*—that can watch how the lunar surface responds to the rapid change in temperature caused by a lunar eclipse. These data which help scientists better understand the composition and properties of the surface could be a scientific boon for understanding both anomalies [33] [34] [35].

We believe that almost the totality of the reported umbra-penumbra boundary shift and the excessive clarity of the penumbra reveal a lunar Allais effect on the Moon's shadow that has nothing to do with the thickness of the Earth's atmosphere. Both phenomena were reported during each and every lunar eclipse recorded for the past 180 years. They occur during lunar eclipses and are correlated. Dr Marmet demonstrated that the Earth's atmosphere cannot be the cause of the enlargement of the Earth's shadow. He concludes that it is an optical illusion, but neither addresses nor explains the second offset: the excessive brightness of the penumbra [1]. If he is right to say that the Earth's atmosphere is not responsible for the 2% umbra-penumbra limit shift on the Moon, he is wrong to evoke the optical illusion. NASA records anomalies without providing an explanation, the priorities being elsewhere. The door is open to researchers to probe the reasons and suggest fields to explore. The list of our references shows that they cannot be explained by current science, which leaves only one option: the lunar Allais effect.

But the bottleneck, which means that this aspect of science remains speculative even as Professor Allais' experiments have validated the solar eclipse effect, is the question of a lack of willingness to experiment. How can interest be aroused in experimenters for whom the scientific value of precise experience is dependent on their theoretical interpretation? Classical conservative physical thought cannot tolerate the defeat of the current theory of gravitation when applied to the case of the influence of the attraction of the Sun and the Moon on the motion of the paraconic pendulum, whether these are the amplitudes of the lunisolar periodic components or the anomalies observed during the total eclipses of the Sun [16].

However, the most accommodating physicists say they do not rely on the more or less contradictory experiments carried out so far; they would like experiments operated with a paraconic pendulum at any point similar to the pendulum used by M. Allais, or they would like to turn to more radical experiments, like those intended for modern theories. For example, the atomic clock cooled by cesium laser (PHARAO) [36] placed by the European Space Agency (ESA) on the International Space Station (ISS) could have been used. In default of confirm doubtful fashionable theories, this high technology could test the Allais effect and supply the way to tie, by a new theoretical link, the facts observed during the eclipses to the physical laws having received the sanction of a rigorous experimental control.

In conclusion, it seems that two noticed anomalies during lunar eclipses, the enlargement of the Earth's shadow delineated onto its satellite and an excessive illumination of the penumbra, adjusted *ad hoc* to the Earth's atmosphere, would rather be caused by an Allais eclipse effect, *i.e.*, a repulsion that occurs when the

Moon passes directly behind the Earth into its umbra, when Sun, Earth and Moon are closely aligned in space. This is consistent with our knowledge of the solar eclipse, with the calculation of the abnormal spontaneous acceleration of the Moon during the solar eclipse in June 1954 (paraconical pendulum of Maurice Allais) and the result recorded by a gravimeter during the solar eclipse of 1997.

(To know more about lunar eclipses by pictures, references [37] [38] [39] [40] have been added).

Conflicts of Interest

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

References

[1] Marmet, P. and Couture, C. (1997-99) Enlargement of the Earth's Shadow on the Moon: An Optical Illusion. Physics Department, University of Ottawa, Ottawa.

[2] Link, F. (1970) La Lune, Collection Science d'Aujourd'hui. Albin Michel, Paris, 107.

[3] Link, F. and Linkova, Z. (1954) Agrandissement de l'ombre terrestre pendant les éclipses de Lune. Institut astronomique, Ondrerov.

[4] Espenak, F. and Meeus, J. (2009) Enlargement of Earth's Shadows. NASA, Washington DC.

[5] Chauvenet, W.A. (1891) Manual of Spherical and Practical Astronomy, Vol. 1. (1960) Dover Reprint, New York.

[6] Danjon, A. (1951) *L'Astronomie*, **65**, 51-53.

[7] Meeus, J. (1972) *Ciel et Terre*, **88**, 491.

[8] Sinnott, R.W. (1983) Sky & Telescope, Readers Gauge the Umbra Again. 387.

[9] Sinnott, R.W. (1992) Sky & Telescope, a Tale of Two Eclipses. 678.

[10] Soulsby, B.W. (1984) *Journal of the British Astronomical Association*, *Lunar Eclipse Crater Timing Programme*, **95**, 16-21.

[11] Lang, K.R. (1992) Astrophysical Data: Planets and Stars. Springer-Verlag, New York, 36.

[12] McCormick, P. (1982) Sky & Telescope. 390.

[13] Link, F. (1953) Sur la grandeur des poussières météoriques dans l'atmosphère terrestre. Publishing House of the Czechoslovak Academy of Sciences. NASA Astrophysics Data System.

[14] Link, F. and Linková, Z. (1954) Agrandissement de l'ombre terrestre pendant les éclipses de Lune; influences météoriques. Publishing House of the Czechoslovak Academy of Sciences. Provided by the NASA Astrophysics Data System.

[15] Fesenkov, V.G. (1970) *Astronomicheskii Zhurnal*, **47**, 237-245.

[16] Allais, M. (1997) L'Anisotropie de l'Espace. Edition Clément Juglar, Paris, 200.

[17] Bagdoo, R. (2009) *Journal of Modern Physics*, **11**, 1620-1638. https://doi.org/10.4236/jmp.2020.1110101

[18] Bagdoo, R. (2010) Is the Growth of the Astronomical Unit Caused by the Allais Eclipse Effect? ViXra, General Science Journal, Internet Archive, Issuu, Scribd.

[19] Mavridès, S. (1988) La Relativité, Que sais-je? Presses universitaires de France,

Paris, 105-106, 112, 119.

[20] Schwinger, J. (2002) Einstein's Legacy. Dover Publication, Inc., Mineola, 146.

[21] Matsushima, S. (1966) *The Astronomical Journal*, **71**, 699-705. https://doi.org/10.1086/110174

[22] Snyder, C.W., Neugebauer, M. and Rao, U.R. (1963) *Journal of Geophysical Research*, **68**, 6361-6370. https://doi.org/10.1029/JZ068i024p06361

[23] Dubois, J. and Link, F. (1970) Analyse photométrique de l'ombre intérieure pendant les éclipses de Lune. Publishing House of the Czechoslovak Academy of Sciences. Provided by the NASA Astrophysics Data System.

[24] Shepherd, J.S. (1982) *Journal of the British Astronomical Association*, **92**, 66-67.

[25] Wikipedia. Lunar Eclipse. https://en.wikipedia.org/wiki/Lunar_eclipse

[26] Link, F. (1959) Densité de l'ombre pendant les éclipses de Lune. Publishing House of the Czechoslovak Academy of Sciences. Provided by the NASA Astrophysics Data System.

[27] Sandulak, N. (1964) Jurgen Stock, Indication of Luminescence Found in the December 1964 Lunar Eclipse. Cerro Tololo Inter-American Observatory, La Serena, 237. (1965) Provided by the NASA Astrophysics Data System.

[28] Jamieson, D.N. (1995) The July Lectures in Physics-1995 Light without Heat: Luminescence in Moonlight. School of Physics, University of Melbourne, Melbourne.

[29] Born, M. (1962) Einstein's Theory of Relativity. Dover Publications, Inc., Mineola, 352-353.

[30] Radounskaïa, I. (1972) Idées folles. Édition MIR, Moscou, 68-69, 78-82.

[31] Lerner, R.G. and Trigg, G.L. (1990) Raman Spectroscopy, Encyclopaedia of Physics. VCH Publishers, Inc., Hoboken, 1034.

[32] Möller, K.D. (1988) Optics. University Science Books, Sausalito, 574-5, 621.

[33] Earth's Moon. https://solarsystem.nasa.gov/moons/earths-moon/lunar-phases-and-eclipses

[34] Lunar Reconnaissance Orbiter. https://solarsystem.nasa.gov/missions/lro/in-depth

[35] NASA (2011) Need to Know: Lunar Eclipse and LRO. YouTube, NASA Goddard.

[36] Bagdoo, R. (2012) The PHARAO/ACES Mission and the Allais Effect. General Science Journal, ViXra, Issuu, Scribd. http://gsjournal.net/Science-Journals/Research%20Papers/View/4292

[37] Jones, T. (2011) Total Lunar Eclipse 10th December 2011. https://communicatescience.com/zoonomian/2011/12/10/total-lunar-eclipse-10th-december-2011

[38] Google, 8 477 Lunar Eclipse Premium High Res Photos, Getty Images.

[39] Google, Images correspondant à Lunar eclipses.

[40] NASA (2014) Understanding Lunar Eclipses. YouTube.

Journal of Modern Physics, 2021, 12, 1795-1821
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

The Staggered Fermion for the Gross-Neveu Model at Non-Zero Temperature and Density

Daming Li

School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai, China
Email: lidaming@sjtu.edu.cn

How to cite this paper: Li, D.M. (2021) The Staggered Fermion for the Gross-Neveu Model at Non-Zero Temperature and Density. *Journal of Modern Physics*, **12**, 1795-1821.
https://doi.org/10.4236/jmp.2021.1213105

Received: October 9, 2021
Accepted: November 22, 2021
Published: November 25, 2021

Abstract

The 2 + 1d Gross-Neveu model with finite density and finite temperature is studied by the staggered fermion discretization. The kinetic part of this staggered fermion in momentum space is used to build the relation between the staggered fermion and Wilson-like fermion. In the large N_f limit (the number N_f of staggered fermion flavors), the chiral condensate and fermion density are solved from the gap equation in momentum space, and thus the phase diagram of fermion coupling, temperature and chemical potential is obtained. Moreover, an analytic formula for the inverse of the staggered fermion matrix is given explicitly, which can be calculated easily by parallelization. The generalization to the 1 + 1d and 3 + 1d cases is also considered.

Keywords

Gross-Neveu Model, Phase Diagram, Staggered Fermion, Gap Equation

1. Introduction

The chiral phase transition in quantum chromodynamics (QCD) from the hadronic phase at low temperature T (low density μ_B) to the quark-gluon plasma phase at high temperature (high density) has been studied intensively in the last decade. Although the relative firm statements for the phase structure can be made in two limit cases: finite T with small baryon density $\mu_B \ll T$ and asymptotically high density $\mu_B \gg \Lambda_{\mathrm{QCD}}$, the phase structures at the intermediate baryon density are not clear. For a recent and review and related work of QCD with finite density, see Ref. [1]-[9].

Since the chiral symmetry breaking and restoration are intrinsically nonperturbative, the number of techniques is limited and most results come from the lattice QCD. Unfortunately, the lattice QCD at finite density suffers from the notorious sign problem, especially for the intermediate or large baryon density. For some simpler quantum field models, e.g., the dense two-color QCD [10], the sign problem can be avoided. The recent progress of the sign problem in lattice field models can refer to [11] and references therein. In the last decades, the tensor network becomes very popular in condensed matter physics and high energy physics, especial for lower dimension models, since probability is not used and thus it is free of sign problem [12] [13] [14] [15].

This paper addresses a simplest four-fermion model with Z_2 symmetry: Gross-Neveu model at non-zero temperature and density [16] [17] [18] [19] [20]. The 2 + 1d Gross-Neveu model has an interesting continuum limit and there is a critical coupling indicating the threshold for the symmetry breaking at zero temperature and density. Although the 2 + 1d Gross-Neveu model is not renormalisable in the weak coupling expansion, it is renormalisable in $1/N_f$ expansion [16], where N_f is the number of flavors of fermions.

The symmetry breaking of Gross-Neveu model for the 1 + 1d case has been studied extensively [21]-[29]. Recently, 2 + 1d Gross-Neveu model is used to study the inhomogeneous phases [30] and the symmetry breaking [31].

Compared with the Wilson fermion, the staggered fermion is more adequate for studying spontaneous chiral symmetry breaking. Another advantage of the staggered fermion is due to the reduced computational cost since the Dirac matrices have been replaced by the staggered phase factor. The reconstruction of the Wilson-like fermion from the staggered fermion is rather technique, thus needing a careful explanation of the physical fermions for lattice QCD [32] and for Gross-Neveu model [18].

In this paper, we revisit the staggered fermion for the 1 + 1d, 2 + 1d and 3 + 1d Gross-Neveu model at non-zero temperature and finite density. The gap equation, which is based on the large N_f limit, is solved in the momentum space. Moreover, we derive an explicit formula for the inverse matrix of the staggered fermion matrix, which is easy to be implemented by parallelization and thus make the large scale calculation of the gap equation feasible.

The arrangement of the paper is as follows. The continuum 2 + 1d Gross-Neveu model at finite density and non-zero temperature is introduced in Section 2. In Section 3, the 2 + 1d staggered fermion is shown and non-dimensional quantities are introduced. The kinetic part of staggered fermion in the momentum space is given in Section 4, where the trace of the inverse matrix and elements of inverse matrix are given explicitly in momentum space. In Section 5, the results in Section 4 are generalized to the 1 + 1d and 3 + 1d staggered fermion. The gap equation is given in Section 6, where the chiral condensate and fermion density are calculated. The simulation results in the large N_f limit are obtained in Section 7. Finally, the conclusion is given in Section 8.

2. The Gross-Neveu Model

The Gross-Neveu model for interacting fermions in 2 + 1d is defined by the continuum Euclidiean Lagrangian density at finite density

$$\mathcal{L}=\bar{\psi}\left(\partial\!\!\!/+\tilde{\mu}\gamma_0+\tilde{m}\right)\psi-\frac{\tilde{g}^2}{2N_f}\left(\bar{\psi}\psi\right)^2 \tag{1}$$

where $\partial\!\!\!/=\sum_{\nu=0}^{2}\gamma_\nu\partial_\nu$, $\tilde{\mu}$ is the chemical potential, $\tilde{m}$ the bare mass, ψ and $\bar{\psi}$ are an N_f-flavor 4 component spinor fields. Here we choose the Gamma matrices

$$\gamma_\nu=\begin{pmatrix}\sigma_{\nu+1} & 0\\ 0 & -\sigma_{\nu+1}\end{pmatrix},\quad \nu=0,1,2 \tag{2}$$

$$\gamma_3=\begin{pmatrix} & -i\mathbb{I}_2\\ i\mathbb{I}_2 & \end{pmatrix},\quad \gamma_5=\gamma_0\gamma_1\gamma_2\gamma_3=\begin{pmatrix} & \mathbb{I}_2\\ \mathbb{I}_2 & \end{pmatrix} \tag{3}$$

where $\sigma_i\left(i=1,2,3\right)$ are the Pauli matrices. The Gamma matrices satisfies

$$\gamma_\mu\gamma_\nu+\gamma_\nu\gamma_\mu=\delta_{\mu\nu}2\mathbb{I}_4,\quad \mu,\nu=0,1,2,3,5$$

There is a discrete Z_2 symmetry $\psi\to\gamma_5\psi$, $\bar{\psi}\to-\bar{\psi}\gamma_5$, which is broken by the mass term but not the interaction. Introducing the bosonic field σ, the interaction between fermions is decoupled with the Lagrangian density,

$$L=\bar{\psi}\left(\partial\!\!\!/+\tilde{\mu}\gamma_0+\tilde{m}+\sigma\right)\psi+\frac{N_f}{2\tilde{g}^2}\sigma^2 \tag{4}$$

The dimension of quantities for the 2 + 1d Gross-Neveu model is as follows

$$\left[\bar{\psi}\right]=\left[\psi\right]=\left[\tilde{\mu}\right]=\left[\tilde{m}\right]=\left[\sigma\right]=\text{length}^{-1},\quad \left[\tilde{g}\right]=\text{length}^{1/2} \tag{5}$$

The partition function for this model is

$$\begin{aligned}Z&=\int\mathrm{d}\bar{\psi}\mathrm{d}\psi\mathrm{d}\sigma\mathrm{e}^{-\int\mathcal{L}}\\&=\int\mathrm{d}\sigma\mathrm{e}^{-\int\frac{N_f}{2\tilde{g}^2}\sigma^2}\left[\det\left(\partial\!\!\!/+\tilde{\mu}\gamma_0+\tilde{m}+\sigma\right)\right]^{N_f}\\&=\int\mathrm{d}\sigma\exp\left(-\int\frac{N_f}{2\tilde{g}^2}\sigma^2+N_f\ln\left[\det\left(\partial\!\!\!/+\tilde{\mu}\gamma_0+\tilde{m}+\sigma\right)\right]\right)\end{aligned} \tag{6}$$

where $\int\equiv\int_0^\beta\mathrm{d}x_0\int_0^L\mathrm{d}x_1\mathrm{d}x_2$ with the inverse temperature $\beta=1/T$ and the space size *L*. $\bar{\psi}$ and ψ are antiperiodic in x_0 direction, and are periodic in x_1 and x_2 directions. We want to calculate the chiral condensate for one flavor

$$\frac{1}{N_fV}\frac{\partial\ln Z}{\partial\tilde{m}}=\left\langle-\frac{1}{V}\int\bar{\psi}_i\psi_i\right\rangle=\frac{1}{\tilde{g}^2}\left\langle\frac{1}{V}\int\sigma\right\rangle\equiv\frac{1}{\tilde{g}^2}\Sigma \tag{7}$$

where $V=\beta L^2$ is the volume of 2 + 1d system. In the second equality we used

$$0=\int\mathrm{d}\bar{\psi}\mathrm{d}\psi\mathrm{d}\sigma\frac{\delta}{\delta\sigma(x)}\mathrm{e}^{-\int\mathcal{L}}=\int\mathrm{d}\bar{\psi}\mathrm{d}\psi\mathrm{d}\sigma\mathrm{e}^{-\int\mathcal{L}}\left(-1\right)\left(\bar{\psi}\psi+\frac{N}{\tilde{g}^2}\sigma\right)(x)$$

Since the Lagrangian density is translation invariant, $\left\langle\bar{\psi}(x)\psi(x)\right\rangle$ and $\left\langle\sigma(x)\right\rangle$ does not depend on *x*. This model in the large N_f limit can be solved exactly [18] in the chiral limit $\tilde{m}=0$, which is based on the saddle approximation (gap equation) in (6)

$$\begin{aligned}0&=-\frac{V}{\tilde{g}^2}\Sigma+\frac{\mathrm{d}}{\mathrm{d}\Sigma}\ln\left[\det\left(\partial\!\!\!/+\tilde{\mu}\gamma_0+\tilde{m}+\Sigma\right)\right]\\&=-\frac{V}{\tilde{g}^2}\Sigma+\mathrm{Tr}\left(\partial\!\!\!/+\tilde{\mu}\gamma_0+\tilde{m}+\Sigma\right)^{-1}\\&=-\frac{V}{\tilde{g}^2}\Sigma+\sum_k\mathrm{tr}\left(ik\!\!\!/+\tilde{\mu}\gamma_0+\tilde{m}+\Sigma\right)^{-1}\\&=-\frac{V}{\tilde{g}^2}\Sigma+4\left(\tilde{m}+\Sigma\right)\sum_k\left(\left(k_0-i\tilde{\mu}\right)^2+\sum_{\nu=1,2}k_\nu^2+\left(\tilde{m}+\Sigma\right)^2\right)^{-1}\end{aligned}\tag{8}$$

where in the third equality we write the trace of operator in momentum space and the summation over $k=\left(k_0,k_1,k_2\right)$

$$k_0=\left(2n-1\right)\pi T,\quad k_\nu=2n_\nu\pi/L,\quad n,n_\nu\in\mathbf{Z},\quad \nu=1,2$$

3. The Staggered Fermion

The staggered fermion discretization of the action $\int\mathcal{L}$ is

$$\begin{aligned}S&=a^2a_t\sum_{x,y}\bar{\psi}\left(x\right)\left(\sum_{\alpha=1,2}\frac{\eta_{x,\alpha}}{2a}\left(\delta_{x+\hat{\alpha},y}-\delta_{x,y+\hat{\alpha}}\right)\right)\psi\left(y\right)\\&+a^2a_t\sum_{x,y}\bar{\psi}\left(x\right)\left(\frac{\eta_{x,0}}{2a_t}\left(\mathrm{e}^{a_t\tilde{\mu}}s_x^1\delta_{x+\hat{0},y}-\mathrm{e}^{-a_t\tilde{\mu}}s_x^2\delta_{x,y+\hat{0}}\right)\right)\psi\left(y\right)\\&+a^2a_t\sum_x\left(\tilde{m}+\phi\left(x\right)\right)\bar{\psi}\left(x\right)\psi\left(x\right)+a^2a_t\frac{N_f}{2\tilde{g}^2}\sum_{\tilde{x}}\sigma\left(\tilde{x}\right)^2\end{aligned}\tag{9}$$

with staggered phase factor $\eta_{x,0}=1$, $\eta_{x,1}=\left(-1\right)^{x_0/a}$, $\eta_{x,2}=\left(-1\right)^{\left(x_0+x_1\right)/a}$. $aN_x=L$, $a_tN_t=\beta=1/T$. The boundary condition for ψ and $\bar{\psi}$ are accounted for by the sign s^1 and s^2

$$s_x^1=\begin{cases}-1 & \text{if } x_0=N_t-1\\1 & \text{Otherwise}\end{cases},\quad s_x^2=\begin{cases}-1 & \text{if } x_0=0\\1 & \text{Otherwise}\end{cases}\tag{10}$$

Here ϕ is defined on lattice x by $\sigma\left(\tilde{x}\right)$

$$\phi\left(x\right)=\frac{1}{8}\sum_{\left[x,\tilde{x}\right]}\sigma\left(\tilde{x}\right)\quad\Leftrightarrow\quad\sigma\left(\tilde{x}\right)=\frac{1}{8}\sum_{\left[x,\tilde{x}\right]}\phi\left(x\right)\tag{11}$$

where $\left[x,\tilde{x}\right]$ denotes 8 dual lattices $\tilde{x}$ which is neighbour to x. The auxiliary field on dual lattice for two dimensional Gross-Neveu model was first studied in Ref. [33].

According to (5), the non-dimensional quantities are introduced by

$$a\sigma\to\sigma,\quad a\phi\to\phi,\quad a\bar{\psi}\to\bar{\psi},\quad a\psi\to\psi\tag{12}$$

$$a\tilde{\mu}=\mu,\quad a\tilde{m}=m,\quad a^{-1/2}\tilde{g}=g,\quad x/a\to x,\quad a_1=a_t/a\tag{13}$$

and thus the action in (9) can be rewritten as

$$\begin{aligned}S&=a_1\sum_{x,y}\bar{\psi}\left(x\right)\left(\sum_{\alpha=1,2}\frac{\eta_{x,\alpha}}{2}\left(\delta_{x+\hat{\alpha},y}-\delta_{x,y+\hat{\alpha}}\right)\right)\psi\left(y\right)\\&+\sum_{x,y}\bar{\psi}\left(x\right)\left(\frac{\eta_{x,0}}{2}\left(\mathrm{e}^{a_1\mu}s_x^1\delta_{x+\hat{0},y}-\mathrm{e}^{-a_1\mu}s_x^2\delta_{x,y+\hat{0}}\right)\right)\psi\left(y\right)\\&+a_1\sum_x\left(m+\phi\left(x\right)\right)\bar{\psi}\left(x\right)\psi\left(x\right)+a_1\frac{N_f}{2g^2}\sum_{\tilde{x}}\sigma\left(\tilde{x}\right)^2\end{aligned}$$

The partition function for the Gross-Neveu model with N_f flavors is:

$$Z = \int \prod_i \mathrm{d}\bar{\psi}_i \mathrm{d}\psi_i \mathrm{d}\sigma \mathrm{e}^{-S} \tag{14}$$

where ψ_i and $\bar{\psi}_i$ denote the Grassmann fields of flavors $i = 0, \cdots, N_f - 1$ at the sites x, σ is the real field defined at the dual lattice sites $\tilde{x}$. The action is

$$S = \sum_{i,x,y} \bar{\psi}_i(x) D_{x,y} \psi_i(y) + \sum_{i,x} a_1 \phi(x) \bar{\psi}_i(x) \psi_i(x) + \frac{a_1 N_f}{2g^2} \sum_{\tilde{x}} \sigma(\tilde{x})^2 \tag{15}$$

where

$$D_{x,y} = \begin{cases} a_1 \dfrac{\eta_{x,\alpha}}{2} & \text{if } y = x + \hat{\alpha}, \quad \alpha = 1,2 \\ -a_1 \dfrac{\eta_{x,\alpha}}{2} & \text{if } y = x - \hat{\alpha}, \quad \alpha = 1,2 \\ \dfrac{\eta_{x,0}}{2} \mathrm{e}^{a_1 \mu} s_x^1 & \text{if } y = x + \hat{0} \\ -\dfrac{\eta_{x,0}}{2} \mathrm{e}^{-a_1 \mu} s_x^2 & \text{if } y = x - \hat{0} \\ a_1 m & \text{if } y = x \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

The derivative of this matrix D with respect to the chemical potential and bare mass are rather simple

$$\frac{\partial D_{x,y}}{\partial(a_1 \mu)} = \frac{\mathrm{e}^{a_1 \mu}}{2} s_x^1 \delta_{x+\hat{0},y} + \frac{\mathrm{e}^{-a_1 \mu}}{2} s_x^2 \delta_{x,y+\hat{0}}, \quad \frac{\partial D_{x,y}}{\partial(a_1 m)} = \delta_{x,y}$$

The real matrix $D(\mu, m)$ satisfies the following symmetry

$$D(\mu, m)_{x,y} = -D(-\mu, -m)_{y,x}$$

$$\varepsilon_x D(\mu, m)_{x,y} \varepsilon_y = -D(\mu, -m)_{x,y} = D(-\mu, m)_{y,x}$$

where $\varepsilon_x = (-1)^{x_0 + x_1 + x_2}$ is the parity of site x.

By integrating the Grassmann fields, the partition function in (14) can be rewritten as

$$Z = \int \prod_{\tilde{x}} \mathrm{d}\sigma(\tilde{x}) \mathrm{e}^{-S_{\text{eff}}} \tag{17}$$

with the effective action

$$S_{\text{eff}} = \frac{a_1 N_f}{2g^2} \sum_{\tilde{x}} \sigma^2(\tilde{x}) - N_f \ln \det D[\phi] \tag{18}$$

and

$$\left(D[\phi]\right)_{x,y} = D_{x,y} + a_1 \phi(x) \delta_{x,y} \tag{19}$$

The computational results, e.g., non-dimensional chiral condensate and fermion density, depend on the non-dimensional quantities

$$\left(N_f, g, \mu, m, N_x, N_t\right)$$

The physical dimensional quantities can be recovered from the non-dimensional ones by introducing lattice size a according to (12), (13). For notation simplicity, we set $a_t = a$ and thus $a_1 = 1$ in the following discussion.

4. Staggered Fermion in Momentum Space

The kinetic part in (15) in one flavor is $\sum_{x,y} \bar{\chi}(x) D_{x,y} \chi(y)$ where

$$D_{x,y} = \begin{cases} \frac{\eta_{x,\alpha}}{2} & \text{if } y = x + \hat{\alpha}, \quad \alpha = 1,2 \\ -\frac{\eta_{x,\alpha}}{2} & \text{if } y = x - \hat{\alpha}, \quad \alpha = 1,2 \\ \frac{\eta_{x,0}}{2} \mathrm{e}^{\mu} s_x^1 & \text{if } y = x + \hat{0} \\ -\frac{\eta_{x,0}}{2} \mathrm{e}^{-\mu} s_x^2 & \text{if } y = x - \hat{0} \\ m & \text{if } y = x \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

$\bar{\chi}$ and χ are the Grassmann fields defined on lattices. A Wilson-like fermion can be obtained from the stagger fermion $\sum_{x,y} \bar{\chi}(x) D_{x,y} \chi(y)$ [18].

Assume that N_x and N_t are even integers. Let $Y = (Y_0, Y_1, Y_2)$ denotes a site on a lattice of twice the spacing of the original, and $A = (A_0, A_1, A_2), A_i = 0,1$ is a lattice vector, which ranges over the corners of the elementary cube associated with Y, so that each site on the original lattice x uniquely corresponds to A and Y: $x = 2Y + A$. Introducing notation

$$\chi(x) = \chi(2Y + A) = \chi(A, Y)$$

A shift along μ direction can be represented by

$$\begin{aligned} \chi(x + \hat{\mu}) &= \chi(2Y + A + \hat{\mu}) = \chi(2(Y + \hat{\mu}) + A - \hat{\mu}) \\ &= \sum_{A'} \left(\delta_{A+\hat{\mu},A'} \chi(A', Y) + \delta_{A-\hat{\mu},A'} \chi(A', Y + \hat{\mu}) \right) \end{aligned} \tag{21}$$

Similarly,

$$\chi(x - \hat{\mu}) = \sum_{A'} \left(\delta_{A-\hat{\mu},A'} \chi(A', Y) + \delta_{A+\hat{\mu},A'} \chi(A', Y - \hat{\mu}) \right) \tag{22}$$

$\chi(x)$ is defined on the fine lattice sites x with lattice size $a = 1$

$$\{ x = (x_0, x_1, x_2), 0 \le x_0 < N_t, 0 \le x_1, x_2 < N_x \} \tag{23}$$

while $\chi(A, \cdot)$ on the coarse lattice sites Y with lattice size $2a = 2$

$$\{ 2Y = 2(Y_0, Y_1, Y_2), 0 \le Y_0 < N_t/2, 0 \le Y_1, Y_2 < N_x/2 \} \tag{24}$$

A unitary transformation of $\chi(A, \cdot)$ is defined by [34]

$$u^{\alpha a}(Y) = \frac{1}{4\sqrt{2}} \sum_A \Gamma_A^{\alpha a} \chi(A, Y), \quad d^{\alpha a}(Y) = \frac{1}{4\sqrt{2}} \sum_A B_A^{\alpha a} \chi(A, Y) \tag{25}$$

$$\bar{u}^{\alpha a}(Y) = \frac{1}{4\sqrt{2}} \sum_A \bar{\chi}(A, Y) \Gamma_A^{*\alpha a}, \quad \bar{d}^{\alpha a}(Y) = \frac{1}{4\sqrt{2}} \sum_A \bar{\chi}(A, Y) B_A^{*\alpha a} \tag{26}$$

where 2×2 matrices Γ_A and B_A is given by

$$\Gamma_A = \sigma_1^{A_0}\sigma_2^{A_1}\sigma_3^{A_2}, \quad B_A = \left(-\sigma_1\right)^{A_0}\left(-\sigma_2\right)^{A_1}\left(-\sigma_3\right)^{A_2} \tag{27}$$

Γ_A and B_A satisfies the following properties (The indices $\alpha,\alpha',\beta,a,a'$ and b always run from 1 to 2)

$$\Gamma_{A\pm\hat{\mu}} = \eta_\mu(A)\sigma_{\mu+1}\Gamma_A, \quad B_{A\pm\hat{\mu}} = \eta_\mu(A)\left(-\sigma_{\mu+1}\right)B_A, \quad \mu = 0,1,2 \tag{28}$$

$$\mathrm{Tr}\left(\Gamma_A^\dagger\Gamma_{A'} + B_A^\dagger B_{A'}\right) = 4\delta_{AA'} \tag{29}$$

$$\begin{aligned}&\sum_A \Gamma_A^{\alpha a}\Gamma_A^{*\beta b} = \sum_A B_A^{\alpha a}B_A^{*\beta b} = 4\delta_{\alpha\beta}\delta_{ab},\\ &\sum_A \Gamma_A^{\alpha a}B_A^{*\beta b} = \sum_A B_A^{\alpha a}\Gamma_A^{*\beta b} = 0\end{aligned} \tag{30}$$

$$\sum_{A,A_\mu=1}\Gamma_A^{\alpha a}\left(\Gamma_A^*\right)^{\alpha' a'} = \sum_{A,A_\mu=0}\Gamma_A^{\alpha a}\left(\Gamma_A^*\right)^{\alpha' a'}, \quad \mu = 0,1,2 \tag{31}$$

Equation (31) is also valid if Γ is replaced by B.

$$\sum_{A,A_\mu=1}\Gamma_A^{\alpha a}\left(\sigma_{\mu+1}^* B_A^*\right)^{\alpha' a'} = -2\sigma_{\mu+1}^{*aa'}\delta_{\alpha\alpha'} \tag{32}$$

$$\sum_{A,A_\mu=0}\Gamma_A^{\alpha a}\left(\sigma_{\mu+1}^* B_A^*\right)^{\alpha' a'} = 2\sigma_{\mu+1}^{*aa'}\delta_{\alpha\alpha'} \tag{33}$$

See **Appendix A** for these properties.

Using (29), the inverse transformation of (25) and (26) are

$$\chi(A,Y) = \sqrt{2}\sum_{\alpha,a}\left[\Gamma_A^{*\alpha a}u^{\alpha a}(Y) + B_A^{*\alpha a}d^{\alpha a}(Y)\right] \tag{34}$$

$$\bar{\chi}(A,Y) = \sqrt{2}\sum_{\alpha,a}\left[\bar{u}^{\alpha a}(Y)\Gamma_A^{\alpha a} + \bar{d}^{\alpha a}(Y)B_A^{\alpha a}\right] \tag{35}$$

Let us introduce the two Dirac fields with 4 components ($a = 1,2$)

$$q^a(Y) = \begin{pmatrix} q_1^a(Y) \\ q_2^a(Y)\end{pmatrix} = \begin{pmatrix} u^{\alpha a}(Y) \\ d^{\alpha a}(Y)\end{pmatrix}, \quad \bar{q}^a(Y) = \left(\bar{q}_1^a(Y), \bar{q}_2^a(Y)\right) = \left(\bar{u}^{\alpha a}(Y), \bar{d}^{\alpha a}(Y)\right)$$

From the properties (30), it is easy to show that

$$\begin{aligned}&\sum_x \bar{\chi}(x)\chi(x)\\ &= \sum_{A,Y}\sqrt{2}\sum_{\alpha,a}\left(\bar{u}^{\alpha a}(Y)\Gamma_A^{\alpha a} + \bar{d}^{\alpha a}(Y)B_A^{\alpha a}\right)\sqrt{2}\sum_{\alpha',a'}\left[\Gamma_A^{*\alpha' a'}u^{\alpha' a'}(Y) + B_A^{*\alpha' a'}d^{\alpha' a'}(Y)\right]\\ &= 8\sum_Y\sum_{\alpha,a}\left(\bar{u}^{\alpha a}(Y)u^{\alpha a}(Y) + \bar{d}^{\alpha a}(Y)d^{\alpha a}(Y)\right)\\ &= 8\sum_{Y,a}\bar{q}^a(Y)q^a(Y) = 8\sum_Y \bar{q}(Y)q(Y) = 8\sum_k \bar{q}(k)q(k)\end{aligned}$$

where in the last equality the inner produce between $\bar{q}$ and q is given in momentum space corresponding to the coarse lattice with lattice size 2

$$k = 2\pi\left(\frac{m_0 + \frac{1}{2}}{N_t}, \frac{m_1}{N_x}, \frac{m_2}{N_x}\right), \quad 0 \le m_0 < N_t/2, 0 \le m_1, m_2 < N_x/2 \tag{36}$$

For any fixed $\mu = 0,1,2$,

$$
\begin{aligned}
&\frac{1}{2}\sum_{x}\eta_{\mu}(x)\bar{\chi}(x)\big(\chi(x+\hat{\mu})-\chi(x-\hat{\mu})\big)\\
&=\frac{1}{2}\sum_{A,A',Y}\eta_{\mu}(A)\bar{\chi}(A,Y)\big(\delta_{A+\hat{\mu},A'}\big(\chi(A',Y)-\chi(A',Y-\hat{\mu})\big)\\
&\quad+\delta_{A-\hat{\mu},A'}\big(\chi(A',Y+\hat{\mu})-\chi(A',Y)\big)\big)\\
&=\frac{1}{2}\sum_{A,A',Y}\eta_{\mu}(A)\bar{\chi}(A,Y)\Bigg(\frac{\delta_{A+\hat{\mu},A'}+\delta_{A-\hat{\mu},A'}}{2}\partial_{\mu}\chi(A',Y)\\
&\quad+\frac{\delta_{A-\hat{\mu},A'}-\delta_{A+\hat{\mu},A'}}{2}\partial_{\mu}^{2}\chi(A',Y)\Bigg)\\
&=\frac{1}{2}\sum_{A,A',Y}\eta_{\mu}(A)\sqrt{2}\sum_{\alpha,a}\big(\bar{u}^{\alpha a}(Y)\Gamma_{A}^{\alpha a}+\bar{d}^{\alpha a}(Y)B_{A}^{\alpha a}\big)\\
&\quad\times\Bigg\{\frac{\delta_{A+\hat{\mu},A'}+\delta_{A-\hat{\mu},A'}}{2}\sqrt{2}\sum_{\alpha',a'}\big(\Gamma_{A'}^{*\alpha'a'}\partial_{\mu}u^{\alpha'a'}(Y)+B_{A'}^{*\alpha'a'}\partial_{\mu}d^{\alpha'a'}(Y)\big)\\
&\quad+\frac{\delta_{A-\hat{\mu},A'}-\delta_{A+\hat{\mu},A'}}{2}\sqrt{2}\sum_{\alpha',a'}\big(\Gamma_{A'}^{*\alpha'a'}\partial_{\mu}^{2}u^{\alpha'a'}(Y)+B_{A'}^{*\alpha'a'}\partial_{\mu}^{2}d^{\alpha'a'}(Y)\big)\Bigg\}
\end{aligned}
$$

where in the second equality (21) and (22) are used. According to the properties of Γ_A and B_A in (30) (31) (32) and (33)

$$
\begin{aligned}
&\frac{1}{2}\sum_{x}\eta_{\mu}(x)\bar{\chi}(x)\big(\chi(x+\hat{\mu})-\chi(x-\hat{\mu})\big)\\
&=2\sum_{Y}\Big(\bar{u}^{\alpha a}(Y)\big(\sigma_{\mu+1}\big)^{\alpha\alpha'}\delta_{aa'}\partial_{\mu}u^{\alpha'a'}(Y)+\bar{d}^{\alpha a}(Y)\big(-\sigma_{\mu+1}\big)^{\alpha\alpha'}\delta_{aa'}\partial_{\mu}d^{\alpha'a'}(Y)\\
&\quad+\bar{u}^{\alpha a}(Y)\big(\sigma_{\mu+1}^{*}\big)^{aa'}\delta_{\alpha\alpha'}\partial_{\mu}^{2}d^{\alpha'a'}(Y)+\bar{d}^{\alpha a}(Y)\big(-\sigma_{\mu+1}^{*}\big)^{aa'}\delta_{\alpha\alpha'}\partial_{\mu}^{2}u^{\alpha'a'}(Y)\Big)\\
&=2\sum_{Y}\Big[\bar{q}(Y)\big(\gamma_{\mu}\otimes\mathbb{I}_{2}\big)\partial_{\mu}q(Y)+\bar{q}(Y)\big(i\gamma_{3}\otimes\sigma_{\mu+1}^{*}\big)\partial_{\mu}^{2}q(Y)\Big]\\
&=8\sum_{Y}\left[\bar{q}(Y)\big(\gamma_{\mu}\otimes\mathbb{I}_{2}\big)\frac{\partial_{\mu}q(Y)}{4}+\bar{q}(Y)\big(i\gamma_{3}\otimes\sigma_{\mu+1}^{*}\big)\frac{\partial_{\mu}^{2}q(Y)}{4}\right]\\
&=8\sum_{k}\left[\bar{q}(k)\big(\gamma_{\mu}\otimes\mathbb{I}_{2}\big)\frac{i}{2}\sin(2k_{\mu})q(k)+\bar{q}(k)\big(i\gamma_{3}\otimes\sigma_{\mu+1}^{*}\big)\frac{1}{2}\big[\cos(2k_{\mu})-1\big]q(k)\right]
\end{aligned}
\tag{37}
$$

where we used the notations

$$\partial_{\mu}q(Y)=q(Y+\hat{\mu})-q(Y-\hat{\mu})$$

$$\partial_{\mu}^{2}q(Y)=q(Y+\hat{\mu})-2q(Y)+q(Y-\hat{\mu})$$

and the summation over k is taken for all modes in (36). Similarly, we have (see **Appendix B**)

$$
\begin{aligned}
&\frac{1}{2}\sum_{x}\bar{\chi}(x)\big(\chi(x+\hat{0})+\chi(x-\hat{0})\big)\\
&=8\sum_{k}\Big[\bar{q}(k)\big(i\gamma_{3}\otimes\sigma_{1}^{*}\big)i2^{-1}\sin(2k_{0})q(k)\\
&\quad+\bar{q}(k)\big(\gamma_{0}\otimes\mathbb{I}_{2}\big)2^{-1}\big[\cos(2k_{0})+1\big]q(k)\Big]\\
&\equiv 8\sum_{k}\bar{q}(k)A_{+}(k)q(k)
\end{aligned}
\tag{38}
$$

Using

$$\begin{aligned}&\frac{1}{2}\sum_x \bar{\chi}(x)\left(\mathrm{e}^{\mu}\chi(x+\hat{0})-\mathrm{e}^{-\mu}\chi(x-\hat{0})\right)\\&=\cosh\mu\left[\frac{1}{2}\sum_x \bar{\chi}(x)\left(\chi(x+\hat{0})-\chi(x-\hat{0})\right)\right]\\&\quad+\sinh\mu\left[\frac{1}{2}\sum_x \bar{\chi}(x)\left(\chi(x+\hat{0})+\chi(x-\hat{0})\right)\right]\end{aligned}$$

and (37) (38), the kinetic part $\sum_{x,y}\bar{\chi}(x)D_{x,y}\chi(y)$ can be rewritten as in the momentum space

$$\sum_{x,y}\bar{\chi}(x)D_{x,y}\chi(y)=8\sum_k \bar{q}(k)D(k)q(k) \tag{39}$$

where the summation over k is taken for all momentum mode of coarse lattice according to (36), and the staggered matrix in the momentum space is diagonal

$$\begin{aligned}D(k)&=m+\sum_{\mu=1,2}\frac{i}{2}\left\{(\gamma_\mu\otimes\mathbb{I}_2)\sin(2k_\mu)+(\gamma_3\otimes\sigma^*_{\mu+1})\left[\cos(2k_\mu)-1\right]\right\}\\&\quad+\frac{1}{2}\left\{(\gamma_0\otimes\mathbb{I}_2)\left(i\cosh\mu\sin(2k_0)+\sinh\mu\left[\cos(2k_0)+1\right]\right)\right.\\&\quad\left.+(\gamma_3\otimes\sigma^*_1)\left(i\cosh\mu\left[\cos(2k_0)-1\right]-\sinh\mu\sin(2k_0)\right)\right\}\\&\equiv m+\sum_{\mu=0,1,2}(\gamma_\mu\otimes\mathbb{I}_2)a_\mu+\sum_{c=1,2,3}(\gamma_3\otimes\sigma^*_c)b_c\end{aligned} \tag{40}$$

where a_μ and b_c depends on k. The inverse matrix of $D(k)$ is

$$D(k)^{-1}=\frac{1}{N(k)}\left[m-\sum_{\mu=0,1,2}(\gamma_\mu\otimes\mathbb{I}_2)a_\mu-\sum_{c=1,2,3}(\gamma_3\otimes\sigma^*_c)b_c\right] \tag{41}$$

where

$$\begin{aligned}N(k)&=m^2+\frac{1}{4}\sum_{\mu=0,1,2}(\sin 2k_\mu)^2+\frac{1}{4}\sum_{\mu=0,1,2}(1-\cos 2k_\mu)^2\\&\quad-\sinh^2\mu\cos 2k_0-i\cosh\mu\sinh\mu\sin 2k_0\end{aligned} \tag{42}$$

We can calculate the trace of inverse matrix D in (20) from (39)

$$\begin{aligned}\sum_x D^{-1}_{x,x}&=-\frac{\int\mathrm{e}^{-\bar{\chi}D\chi}\sum_x\bar{\chi}(x)\chi(x)}{\int\mathrm{e}^{-\bar{\chi}D\chi}}\\&=-\frac{\int\mathrm{e}^{-\sum_k\bar{q}(k)8D(k)q(k)}8\sum_k\bar{q}(k)q(k)}{\int\mathrm{e}^{-\sum_k\bar{q}(k)8D(k)q(k)}}\\&=-8\sum_k\frac{\int\mathrm{e}^{-\bar{q}(k)8D(k)q(k)}\bar{q}(k)q(k)}{\int\mathrm{e}^{-\bar{q}(k)8D(k)q(k)}}\\&=8\sum_k\mathrm{tr}\left[(8D(k))^{-1}\right]\\&=\sum_k\mathrm{tr}\left[D(k)^{-1}\right]=\sum_k\frac{8m}{N(k)}\end{aligned} \tag{43}$$

where the summation over k is given by (36). Note that the right hand side of (43) is real since $\sum_{k_0}\sin 2k_0/|N(k)|^2=0$ for any k_1 and k_2 modes in (36). Similarly,

$$\sum_x \left(D^{-1}_{x+\hat{0},x} s^1_x + D^{-1}_{x-\hat{0},x} s^2_x \right) = 8 \sum_k \frac{b_1 \sin 2k_0 - a_0 \left(\cos 2k_0 + 1 \right)}{N(k)} \tag{44}$$

and

$$\sum_x \left(D^{-1}_{x+\hat{0},x} s^1_x - D^{-1}_{x-\hat{0},x} s^2_x \right) = (-8i) \sum_k \frac{a_0 \sin 2k_0 + b_1 \left(\cos 2k_0 - 1 \right)}{N(k)} \tag{45}$$

The inverse matrix of D in (20) is

$$\begin{aligned} D^{-1}_{x',x} = \frac{1}{4} \sum_{\alpha,a,\alpha',a'} \Big[& \Gamma^{\alpha a}_A \Gamma^{*\alpha' a'}_{A'} D^{-1}_{(Y'a'\alpha'1;Ya\alpha1)} + \Gamma^{\alpha a}_A B^{*\alpha' a'}_{A'} D^{-1}_{(Y'a'\alpha'2;Ya\alpha1)} \\ & + B^{\alpha a}_A \Gamma^{*\alpha' a'}_{A'} D^{-1}_{(Y'a'\alpha'1;Ya\alpha2)} + B^{\alpha a}_A B^{*\alpha' a'}_{A'} D^{-1}_{(Y'a'\alpha'2;Ya\alpha2)} \Big] \end{aligned} \tag{46}$$

See **Appendix C** for the derivation of (44)-(46).

Since D is diagonal in momentum space, the inverse matrix in the $\bar{q}q$ basis is

$$\begin{aligned} D^{-1}_{Y';Y} &= \frac{1}{N_t/2 \left(N_x/2 \right)^2} \sum_k \mathrm{e}^{ik\cdot 2(Y'-Y)} D^{-1}(k) \\ &= \frac{1}{N_t/2 \left(N_x/2 \right)^2} \sum_k \mathrm{e}^{ik\cdot 2(Y'-Y)} \frac{1}{N(k)} \left[m - \sum_{\mu=0,1,2} \left(\gamma_\mu \otimes \mathbb{I}_2 \right) a_\mu - \sum_{c=1,2,3} \left(\gamma_3 \otimes \sigma^*_c \right) b_c \right] \\ &\equiv m \left(\mathbb{I}_4 \otimes \mathbb{I}_2 \right) \tilde{1}(Y'-Y) - \sum_{\mu=0,1,2} \left(\gamma_\mu \otimes \mathbb{I}_2 \right) \tilde{a}_\mu (Y'-Y) - \sum_{c=1,2,3} \left(\gamma_3 \otimes \sigma^*_c \right) \tilde{b}_c (Y'-Y) \end{aligned}$$

where the notation with tilde denotes the inverse Fourier transformation, e.g.,

$$\begin{aligned} \tilde{a}_\mu (Y) &= \frac{1}{\frac{N_t}{2} \left(\frac{N_x}{2} \right)^2} \sum_k \mathrm{e}^{ik\cdot 2Y} \frac{a_\mu(k)}{N(k)} \\ &= \mathrm{e}^{i\frac{2\pi Y_0}{N_t}} \frac{1}{\frac{N_t}{2} \left(\frac{N_x}{2} \right)^2} \sum_{m_0=0}^{\frac{N_t}{2}-1} \sum_{m_1=0}^{\frac{N_x}{2}-1} \sum_{m_2=0}^{\frac{N_x}{2}-1} \mathrm{e}^{i2\pi\left(\frac{m_0 Y_0}{N_t/2} + \frac{m_1 Y_1}{N_x/2} + \frac{m_2 Y_2}{N_x/2} \right)} \frac{a_\mu(m_0,m_1,m_2)}{N(m_0,m_1,m_2)} \end{aligned}$$

for $|Y_0| \le \frac{N_t}{2} - 1$, $|Y_1|, |Y_2| \le \frac{N_x}{2} - 1$. We first use the fast Fourier transformation to calculate $\tilde{a}_\mu(Y) \exp\left(-i \frac{2\pi Y_0}{N_t} \right)$ and thus $\tilde{a}_\mu(Y)$ for $0 \le Y_0 \le \frac{N_t}{2} - 1$, $0 \le Y_1, Y_2 \le \frac{N_x}{2} - 1$. Then $\tilde{a}_\mu(Y)$ for $|Y_0| \le \frac{N_t}{2} - 1$, $|Y_1|, |Y_2| \le \frac{N_x}{2} - 1$ can be obtained since it is anti-periodic in Y_0 direction and periodic in Y_1 and Y_2 direction.

Each term in $D^{-1}_{Y';Y}$ has a tensor product $A \otimes B$ between 4×4 matrix $A = \left(A_{ij} \right)_{i,j=1,2}$ with 2×2 matrix A_{ij} and 2×2 matrix *B*. The indices of $D^{-1}_{(Y'a'\alpha'i;Ya\alpha j)}$ of the inverse matrix $D^{-1}_{Y';Y}$ in (46) is related to $\left(A_{ij} \right)_{\alpha'\alpha} B_{a'a}$. The analytic formula for the inverse matrix of the staggered fermion is the main contribution of this paper. Compared to the computational complexity $O\left(\left(N_t N_x^2 \right)^3 \right)$

of the usual inverse matrix, the computational cost is $O\left(16\left(N_t N_x^2\right)^2\right)$ since each element of the inverse matrix needs the summation over $\alpha, a, \alpha', a' = 1,2$. Moreover a parallel implementation can be realized easily for the formula (46).

The trace of the inverse matrix in (43) can be derived from (46)

$$\sum_x D_{x,x}^{-1} = \sum_{\alpha,a}\left[D_{(Ya\alpha 1;Ya\alpha 1)}^{-1} + D_{(Ya\alpha 2;Ya\alpha 2)}^{-1}\right] = \sum_k \frac{8m}{N(k)}$$

5. The 1 + 1d and 3 + 1d Staggered Fermion

The staggered fermion matrix in (20) can be generalized to the 1 + 1d and 3 + 1d case, where α is 1 for the 1 + 1d case and α run from 1 to 3 for the 3 + 1d case.

For the 1 + 1d case, the 2×2 matrices γ_μ are defined to be

$$\gamma_\mu = \sigma_\mu, \quad \mu = 1,2, \quad \gamma_5 = i\gamma_1\gamma_2, \quad \gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = \delta_{\mu\nu} 2\mathbb{I}_2, \quad \mu,\nu = 1,2,5$$

The unitary transformation in (25) and (26) are modified to be

$$\psi^{\alpha a}(Y) = \frac{1}{2}\sum_A \Gamma_A^{\alpha a}\chi(A,Y), \quad \bar{\psi}^{\alpha a}(Y) = \frac{1}{2}\sum_A \bar{\chi}(A,Y)\Gamma_A^{*\alpha a}$$

The kinetic part $\sum_{x,y}\bar{\chi}(x)D_{x,y}\chi(y)$ can be written as

$$\sum_{x,y}\bar{\chi}(x)D_{x,y}\chi(y) = \sum_k \bar{\psi}(k)D(k)\psi(k) \tag{47}$$

where the summation is taken over all modes

$$k = 2\pi\left(\frac{m_0+\frac{1}{2}}{N_t}, \frac{m_1}{N_x}\right), \quad 0 \le m_0 < N_t/2, 0 \le m_1 < N_x/2 \tag{48}$$

The fermion matrix in momentum space is diagonal

$$\begin{aligned} D(k) &= 2m + \sum_{\mu=1}\left\{\left(\gamma_{\mu+1}\otimes\mathbb{I}_4\right)i\sin\left(2k_\mu\right) + \left(\gamma_5\otimes\gamma_{\mu+1}^*\gamma_5^*\right)\left[\cos\left(2k_\mu\right)-1\right]\right\} \\ &\quad + \left\{\left(\gamma_1\otimes\mathbb{I}_4\right)\left(i\cosh\mu\sin\left(2k_0\right) + \sinh\mu\left[\cos\left(2k_0\right)+1\right]\right)\right. \\ &\quad \left. + \left(\gamma_5\otimes\gamma_1^*\gamma_5^*\right)\left(\cosh\mu\left[\cos\left(2k_0\right)-1\right] + i\sinh\mu\sin\left(2k_0\right)\right)\right\} \\ &\equiv 2m + \sum_{\mu=0,1}\left(\gamma_{\mu+1}\otimes\mathbb{I}_4\right)a_\mu + \sum_{\mu=0,1}\left(\gamma_5\otimes\gamma_{\mu+1}^*\gamma_5^*\right)b_\mu \end{aligned} \tag{49}$$

with its inverse

$$D(k)^{-1} = \frac{1}{N(k)}\left[2m - \sum_{\mu=0,1}\left(\gamma_{\mu+1}\otimes\mathbb{I}_4\right)a_\mu - \sum_{\mu=0,1}\left(\gamma_5\otimes\gamma_{\mu+1}^*\gamma_5^*\right)b_\mu\right] \tag{50}$$

where

$$\begin{aligned} N(k) &= 4m^2 + \sum_{\mu=1}\left(\sin 2k_\mu\right)^2 - \left(i\cosh\mu\sin 2k_0 + \sinh\mu\left(\cos 2k_0+1\right)\right)^2 \\ &\quad + \sum_{\mu=1}\left(1-\cos 2k_\mu\right)^2 + \left(\cosh\mu\left(\cos 2k_0 - 1\right) + i\sinh\mu\sin 2k_0\right)^2 \end{aligned} \tag{51}$$

The trace of the inverse matrix is

$$\sum_x D_{x,x}^{-1} = \sum_k \frac{16m}{N(k)} \tag{52}$$

The inverse matrix of D can be calculated

$$D_{x',x}^{-1} = \sum_{\alpha,a,\alpha',a'} \Gamma_A^{\alpha a} \Gamma_{A'}^{*\alpha' a'} D_{(Y'a'\alpha';Ya\alpha)}^{-1} \tag{53}$$

where

$$D_{Y';Y}^{-1} = \frac{1}{N_t/2(N_x/2)} \sum_k \mathrm{e}^{ik\cdot 2(Y'-Y)} D^{-1}(k) \tag{54}$$

For the 3 + 1d case, the 4×4 matrices γ_μ are defined to be

$$\Gamma_A = \gamma_1^{A_0}\gamma_2^{A_1}\gamma_3^{A_2}\gamma_4^{A_3}, \quad \mu = 1,2,3,4, \quad \gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_4$$

$$\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = \delta_{\mu\nu} 2\mathbb{I}_2, \quad \mu,\nu = 1,2,3,4,5$$

The unitary transformation in (25) and (26) are modified to be

$$\psi^{\alpha a}(Y) = \frac{1}{2\sqrt{2}} \sum_A \Gamma_A^{\alpha a} \chi(A,Y), \quad \bar{\psi}^{\alpha a}(Y) = \frac{1}{2\sqrt{2}} \sum_A \bar{\chi}(A,Y) \Gamma_A^{*\alpha a}$$

The kinetic part can also be written as (47) where the summation is taken for all modes

$$k = 2\pi \left(\frac{m_0 + \frac{1}{2}}{N_t}, \frac{m_1}{N_x}, \frac{m_2}{N_x}, \frac{m_3}{N_x} \right), \quad 0 \le m_0 < N_t/2, 0 \le m_1, m_2, m_3 < N_x/2$$

Equations (49) - (51) are still valid except that μ runs from 1 to 3. Equations (52) - (54) are modified to be

$$\sum_x D_{x,x}^{-1} = \sum_k \frac{64m}{N(k)} \tag{55}$$

$$D_{x',x}^{-1} = \frac{1}{2} \sum_{\alpha,a,\alpha',a'} \Gamma_A^{\alpha a} \Gamma_{A'}^{*\alpha' a'} D_{(Y'a'\alpha';Ya\alpha)}^{-1} \tag{56}$$

$$D_{Y';Y}^{-1} = \frac{1}{N_t/2(N_x/2)^3} \sum_k \mathrm{e}^{ik\cdot 2(Y'-Y)} D^{-1}(k) \tag{57}$$

respectively. We have checked the formula (46), (53), (56) for the inverse matrices by Matlab.

6. The Gap Equation

The main contribution of the effective action (18) to the partition function can be obtained by the gap equation if $N_f \to \infty$,

$$\frac{\Sigma}{g^2} = \frac{1}{N_t N_x^2} \sum_x D_{x,x}^{-1} \tag{58}$$

Here D is defined in (20) where m is replaced by $m+\Sigma$. The right hand side of (58) can be calculated from (42), (43) where m is replaced by $m+\Sigma$. The first derivative of Σ^2 with respect to μ can be computed from the gap equation

(For simplicity, we assume that $m=0$)

$$\frac{\partial \Sigma^2}{\partial \mu} = \frac{\sum_k \left(\sinh 2\mu \cos 2k_0 + i\cosh 2\mu \sin 2k_0\right) N(k)^{-2}}{\sum_k N(k)^{-2}} \tag{59}$$

If the average Σ of σ has been calculated from the gap equation, the free energy density in the large N_f limit is

$$\ln Z = -N_t N_x^2 \frac{\Sigma^2}{2g^2} + \ln\det D$$

where $\ln\det D = \prod_k \det D(k)$ up to a constant. The other thermodynamic quantities can be calculated. For example, the fermion density can be analytically calculated

$$\begin{aligned}\frac{1}{N_t N_x^2}\frac{\partial \ln Z}{\partial \mu} &= -\frac{1}{2g^2}\frac{\partial \Sigma^2}{\partial \mu} + \frac{1}{N_t N_x^2}\left(\mathrm{e}^{\mu}\sum_x D^{-1}_{x+\hat{0},x} s^1_x + \mathrm{e}^{-\mu}\sum_x D^{-1}_{x-\hat{0},x} s^2_x\right) \\ &= -\frac{1}{2g^2}\frac{\partial \Sigma^2}{\partial \mu} + \frac{1}{N_t N_x^2}\left(\cosh\mu \sum_x \left(D^{-1}_{x+\hat{0},x} s^1_x + D^{-1}_{x-\hat{0},x} s^2_x\right)\right. \\ &\quad \left. + \sinh\mu \sum_x \left(D^{-1}_{x+\hat{0},x} s^1_x - D^{-1}_{x-\hat{0},x} s^2_x\right)\right)\end{aligned} \tag{60}$$

where $\frac{\partial \Sigma^2}{\partial \mu}$, and two sums over x in (60) are given in (59), (44) and (45), respectively. The $N(k)$ for each mode k in (44), (45), (59) is given by (42) with the replacement of m by $m+\Sigma$ (Here for simplicity we assume that $m=0$) and Σ is solved from the gap equation (58).

7. Simulation Results

7.1. Large Volume Limit

Let us consider the large volume limit for the non-interacting 2 + 1d Gross-Neveu model. The partition function $Z = \int \mathrm{d}\bar{\chi}\mathrm{d}\chi \mathrm{e}^{-\bar{\chi} D \chi} = \det D$, where the stagger fermion matrix D is given by (20). The ratio of the non-dimensional chiral condensate $a^2\langle\bar{\psi}\psi\rangle$ and non-dimensional mass $m = a\tilde{m}$ is

$$\frac{a^2\langle\bar{\psi}\psi\rangle}{a\tilde{m}} = \frac{\langle\bar{\chi}\chi\rangle}{a\tilde{m}} = \frac{\left\langle \sum_x \bar{\chi}(x)\chi(x)\right\rangle}{a\tilde{m}\left(N_t N_x^2\right)} = \frac{\sum_x D^{-1}_{x,x}}{a\tilde{m}\left(N_t N_x^2\right)} = \frac{8}{N_t N_x^2}\sum_k \frac{1}{N(k)} \tag{61}$$

where in the last equality we used Equation (43) where $N(k)$, depending on m and μ, is given by (61). Note that there are $N_t N_x^2/8$ modes k in (61). The ratio of the non-dimensional fermion density $a^3\rho$ and $(a\tilde{\mu})^3$

$$\begin{aligned}\frac{a^3\rho}{(a\tilde{\mu})^3} &= \frac{1}{\tilde{\mu}^3}\left(\frac{1/\beta}{\beta L^2}\right)\frac{\partial \ln Z}{\partial \tilde{\mu}} = \frac{1}{N_t \beta L^2 \tilde{\mu}^3}\frac{\partial \ln Z}{\partial \mu} \\ &= \frac{1}{N_t \beta L^2 \tilde{\mu}^3}\left[\frac{\cosh\mu}{2} 8\sum_k \frac{b_1 \sin 2k_0 - a_0\left(\cos 2k_0 + 1\right)}{N(k)}\right. \\ &\quad \left. + \frac{\sinh\mu}{2}(-8i)\sum_k \frac{a_0 \sin 2k_0 + b_1\left(\cos 2k_0 - 1\right)}{N(k)}\right]\end{aligned} \tag{62}$$

where in the last equality we used (44) and (45).

We consider the case $L=\beta$, $a=a_t$ and thus $N_x=N_t\equiv N$. We fix $\tilde{\mu}L$ and $\tilde{m}L$ and then calculate $\frac{a^2\langle\bar{\psi}\psi\rangle}{a\tilde{m}}$ and $\frac{a^3\rho}{(a\tilde{\mu})^3}$ in the large N limit for fixed lattice size a. In fact $\frac{a^2\langle\bar{\psi}\psi\rangle}{a\tilde{m}}$ and $\frac{a^3\rho}{(a\tilde{\mu})^3}$ does not depend on the lattice size a since the non-dimensional mass $m=a\tilde{m}=\frac{\tilde{m}L}{N}$ and non-dimensional chemical potential $\mu=a\tilde{\mu}=\frac{\tilde{\mu}L}{N}$ does not depends on lattice size a. Figure 1 shows the dependence of $\frac{a^2\langle\bar{\psi}\psi\rangle}{a\tilde{m}}$ on N with fixed $\tilde{\mu}L,\tilde{m}L=0,1$. The linear fitting with respect to $1/N$ shows that the large N limit of $\frac{a^2\langle\bar{\psi}\psi\rangle}{a\tilde{m}}$ is close to 1.008 for all four cases, this is because $m=1/N$ and $\mu=1/N$ both vanish for large N limit. Figure 2 shows the dependence of $\frac{a^3\rho}{(a\tilde{\mu})^3}$ on N, where $\tilde{\mu}L=1$ and $\tilde{m}L=0,1$. The large N limit is close to 1.9271 for $m=0$ and 1.9234 for $m=0.1/N$, respectively.

7.2. Phase Diagram

The phase diagram of the 2 + 1d Gross-Neveu model in the large N_f limit is well known [16] [17] [18]. In this limit the phase diagram of (g^{-2},μ,T) is based on the calculation of Σ. Basically for $T=0$ and $\mu=0$, there is a critical coupling g_c^{-2} such that the chiral symmetry is broken $\Sigma>0$ if the coupling is

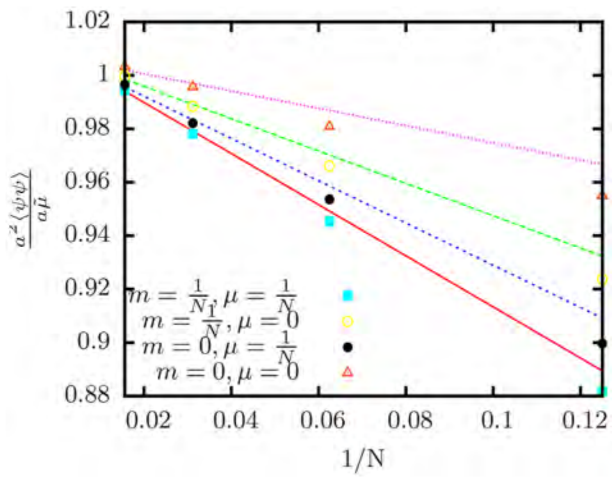


Figure 1. The dependence of $\frac{a^2\langle\bar{\psi}\psi\rangle}{a\tilde{m}}$ on N, $N=4,8,16,32,64,128,256,512$. (1) $m=1/N,\mu=1/N$ with fitting $-0.9563/N+1.009$, (2) $m=1/N,\mu=0$ with fitting $-0.6051/N+1.008$, (3) $m=0,\mu=1/N$ with fitting $-0.7904/N+1.008$, (4) $m=0,\mu=0$ with fitting $-0.3224/N+1.007$.

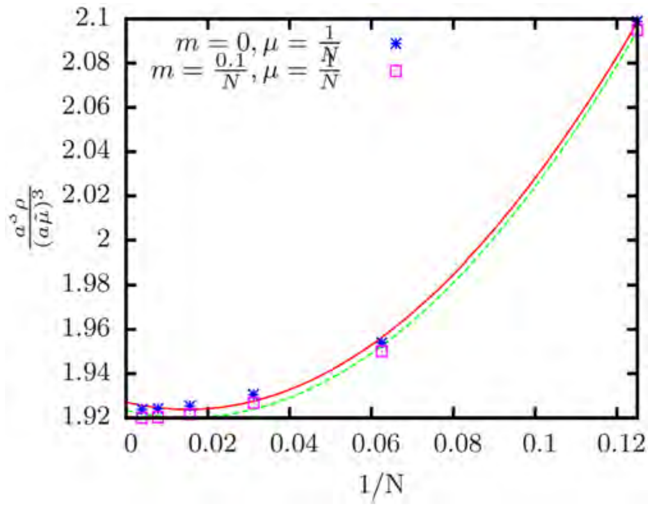


Figure 2. The dependence of $\frac{a^3\rho}{(a\tilde{\mu})^4}$ on N, $N = 8,16,32,64,128,256$. (1) $m = 0, \mu = 1/N$ with fitting $14.4370/N^2 - 0.4345/N + 1.9271$, (2) $m = 0.1/N, \mu = 0$ with fitting $14.4288/N^2 - 0.4343/N + 1.9234$.

strong enough $g^{-2} < g_c^{-2}$. This critical coupling depends on the regularization of the continuum model. For the lattice regularization in this paper, $g_c^{-2} \sim a^{-1}$ where a is the lattice size. For fixed coupling $g^{-2} < g_c^{-2}$ which is not far away from the critical coupling (Otherwise, the continuum limit $a \to 0$ cannot be taken), denote Σ_0 be the value of Σ at this coupling g^{-2} with vanishing temperature T and chemical potential μ. The gap Equation (8), which is solved exactly in the chiral limit in Ref. [18], shows that there exists a critical temperature $T_c = \frac{\Sigma_0}{2\ln 2}$ such that the chiral symmetry is broken if $T < T_c$ at this coupling g^{-2} and $\mu = 0$. Moreover, there is another critical chemical potential $\mu_c = \Sigma_0$ such that this symmetry is broken only if $\mu < \mu_c$ at this coupling g^{-2} and $T = 0$. The mean field results predict that the first order transition only exists at $T = 0$ and $\mu = \mu_c$ for this coupling g^{-2}.

For the 2 + 1d Gross-Neveu model, we first study the dependence of Σ on the coupling g and temperature $T = 1/N_t$ with vanishing chemical potential $\mu = 0$. Figure 3 is the phase diagram of $\left(N_t, 1/g^2\right)$ for $m = 0$ and $N_x = 36$. We always choose $N_x = 36$ to ensure the thermodynamic limit is achieved: the simulation results change very small for larger N_x. The marks + separate the symmetry phase $\Sigma = 0$ (above marks) and the chiral symmetry broken phase $\Sigma > 0$ (below marks). For fixed temperature T there is a critical coupling g_c^{-2} such that Σ decreases to zero if $1/g^2$ is increasing to $1/g_c^2$. Figure 3 shows that $1/g_c^2$ is a increasing function of $N_t = 1/T$ and it will close to 1 at very low temperature. On the other hand, if g^{-2} is fixed, there is a critical temperature $T_c = T_c(g)$ such that Σ is increasing from zero if T is decreasing from T_c.

Figure 4 shows the dependence of Σ on N_t for the different coupling $1/g^2$. For small $1/g^2$, e.g., $1/g^2 = 0.65$, Σ changes small with the temperature.

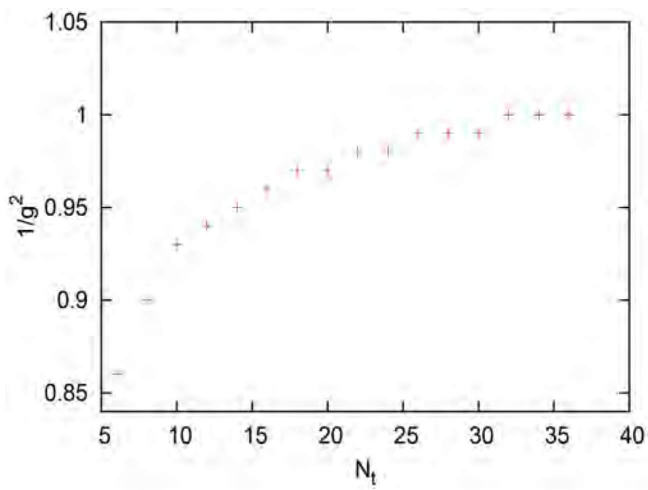


Figure 3. Phase diagram of $(N_t, 1/g^2)$ for $\mu = 0$, $m = 0$, $N_x = 36$. Below the marks + is the broken phase $\Sigma > 0$.

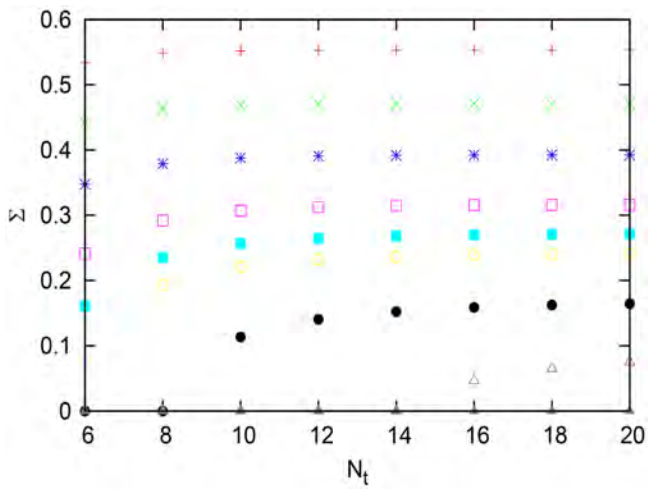


Figure 4. Σ versus N_t, $\mu = 0$, $m = 0$, $N_x = 36$. $1/g^2 = 0.65, 0.70, 0.75, 0.80, 0.83, 0.85, 0.90, 0.95, 1.00$ from top to bottom.

For these range of parameters, it is in the deep chiral symmetry broken phase and we cannot obtain the chiral symmetry phase $\Sigma = 0$ even at very high temperature. For a slightly larger $1/g^2$, for example, $1/g^2 = 0.90$ (black dots in Figure 4), we can find a transition point T_c, which is between $\frac{1}{8}$ and $\frac{1}{10}$ in lattice unit. The symmetry phase and broken phase are realized for $T > T_c(g)$ and $T < T_c(g)$, respectively.

Figure 5 shows the dependence of Σ on $1/g^2$ at different temperature. Σ drops continuously to 0 if $1/g^2$ is increasing to $1/g_c^2(T)$ from below, which show that the transition at the critical coupling constant $g_c(T)$ is second order. At very low temperature $T = 1/N_t = 1/36$, $g_c(T)$ is close to 1, which is consistent with those obtained in [19]. This is because in the limit of $N_t, N_x \to \infty$, the gap equation at $\Sigma = 0$ is reduced to

$$\frac{1}{g^2} = \frac{1}{N_t N_x^2} \sum_k \frac{8}{N(k)} \approx \frac{8}{\pi^3} \int_0^{\pi/2} \mathrm{d}k \frac{1}{\sum_{\mu=0,1,2} (\cos k_\mu)^2} = 1$$

The critical temperature $T_c = \frac{\Sigma_0}{2\ln 2}$ at the coupling g^{-2} and $\mu = 0$ can be verified numerically. Here we choose $N_x = 36$ and $g^{-2} = 0.95$ which is not too far away from the critical coupling $g_c^{-2} \approx 1$. We also choose $N_t = 36$ such that it is very close to zero temperature, the value of Σ at the zero temperature and vanishing chemical potential is $\Sigma_0 = 0.0944$. To calculate the critical temperature at this coupling, we calculate Σ at $N_t = 8, \cdots, 36$ and found that Σ is zero if N_t is between 14 and 16. Thus the critial temperature is between $1/16 = 0.0625$ and $1/14 = 0.0667$ which is very close to $T_c = \frac{\Sigma_0}{2\ln 2} = \frac{0.0944}{2\ln 2} = 0.0680$.

Now let us study the effect of chemical potential on the chiral condensate Σ. Figure 6 shows the dependence of Σ on the chemical potential at the different

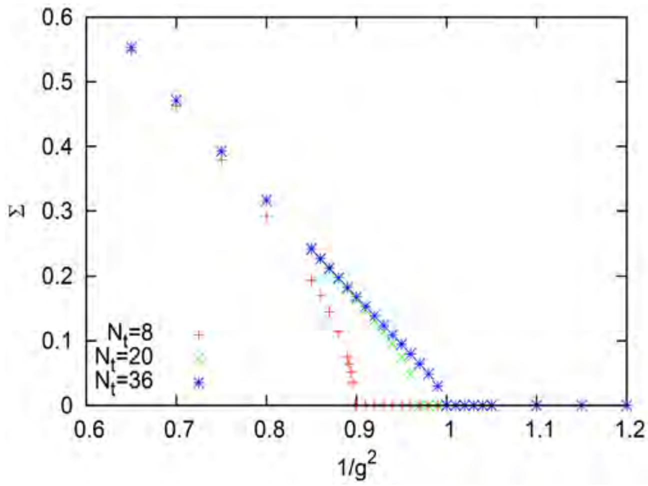


Figure 5. Σ versus $1/g^2$ for different N_t. $\mu = 0$, $m = 0$, $N_x = 36$.

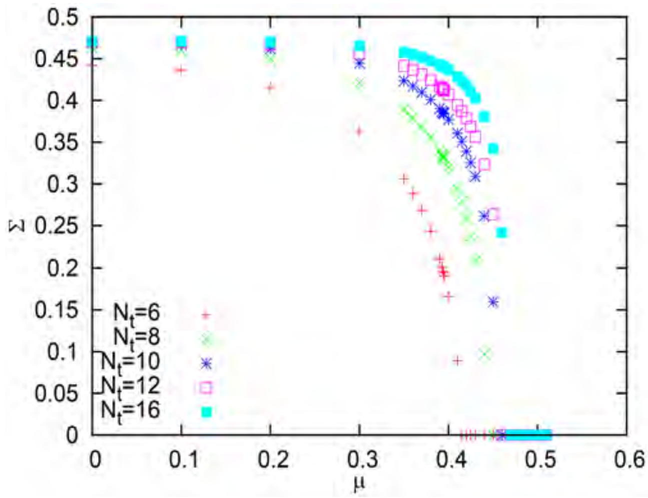


Figure 6. Σ versus μ, $m = 0$, $g = 1.19525$ ($1/g^2 = 0.70$), $N_x = 36$.

temperature $T = 1/N_t$. Σ drops sharply near $\mu_c \approx 0.45$ in the limit of zero temperature $N_t = 16$, *i.e.*, $T = 1/16$, which suggest a first order transition at the zero temperature. This first order transition at the zero temperature is verified by the analytical calculation, $\mu_c = \Sigma_0$ where Σ_0 is the Σ with $\mu = 0$ [18]. For the temperature $T = 1/16$, $\Sigma_0 \approx 0.47$ is slightly larger than $\mu_c \approx 0.45$. If the temperature is raised, e.g., $N_t = 6$, it is more difficult to find a critical chemical potential such that the chiral symmetry is restored. This is not caused by the smallness of $N_x = 36$, since the our results is always obtained for $N_x = 36$, which is very close to the thermodynamics limit, *i.e.*, the result changes very small if N_x is larger than 36. We also note that the transition at finite temperature is the second order, as explained in [18]. **Figure 7** shows the dependence of Σ on μ for a larger $1/g^2 = 0.80$. Compared with **Figure 6**, Σ at $\mu = 0$ and the critical chemical potential in **Figure 7** become smaller, and thus the figures in **Figure 7** is obtained by moving those figures of **Figure 6** in the left-down direction. For the same temperature, for example, $N_t = 16$, it is more difficult to find the critical chemical potential in **Figure 7** than those in **Figure 6**. Both **Figure 6** and **Figure 7** show that the critical chemical potential μ_c is decreased if the temperature is increased. At zero temperature, the mean field exact result show the critical chemical potential μ_c is just the value of Σ_0 at the vanishing chemical potential. This is exactly recovered in **Figure 7** where $\mu_c = 0.32$ for $g^{-2} = 0.80$ with $N_t = 16$.

Figure 8 shows the dependence of Σ and fermion density on the chemical potential at $1/g^2 = 0.7$. At low temperature $N_t = 16$, Σ drops rapidly near the critical chemical potential $\mu_c \approx 0.45$, and the fermion density increase very fast, which suggest Σ and fermion density are not continuous at μ_c at zero temperature and thus they can be regarded as the order parameters.

For the 3 + 1d Gross-Neveu model, we also calculate the dependence of Σ on the coupling and chemical potential at different temperature. **Figure 9** shows the value of Σ depending on the coupling for the vanishing chemical potential.

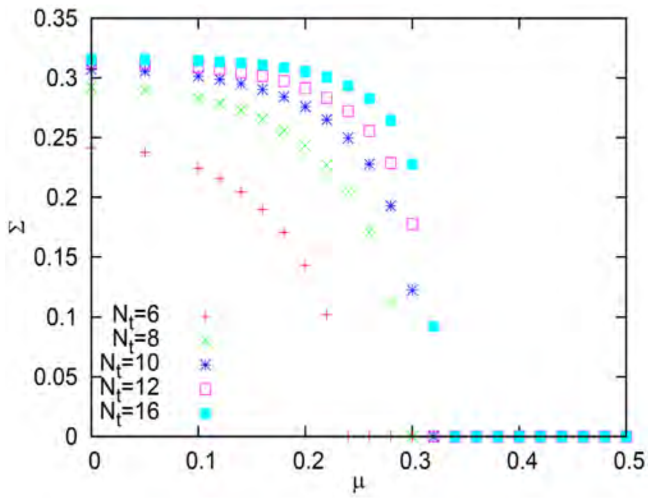


Figure 7. Σ versus μ, $m = 0$, $g = 1.1180$ ($1/g^2 = 0.80$), $N_x = 36$.

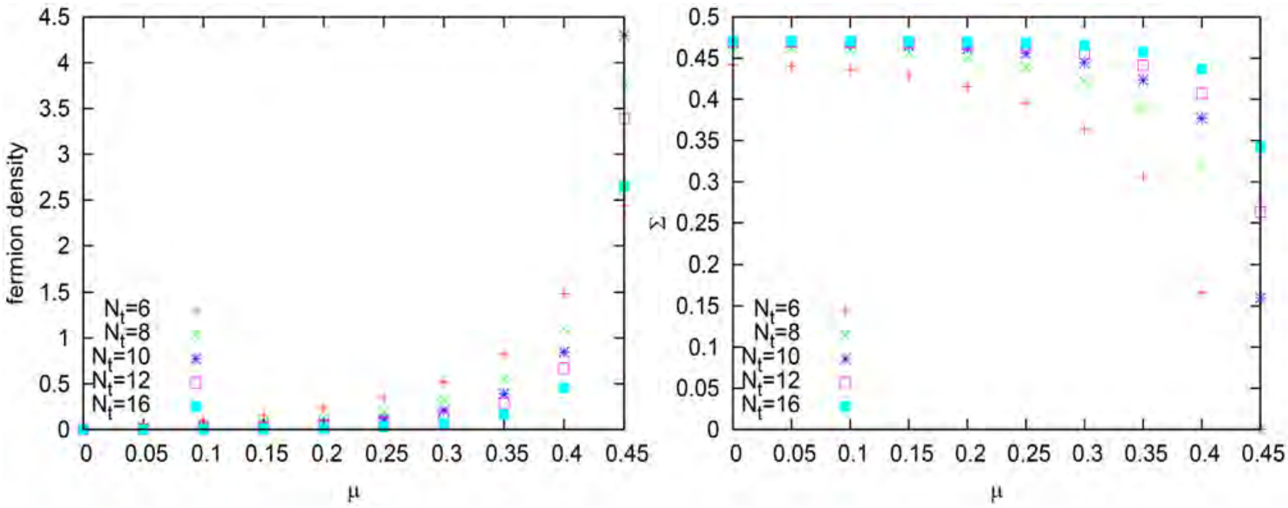


Figure 8. Σ and fermion density vs μ, $m = 0$, $g = 1.19525$ ($1/g^2 = 0.70$), $N_x = 36$.

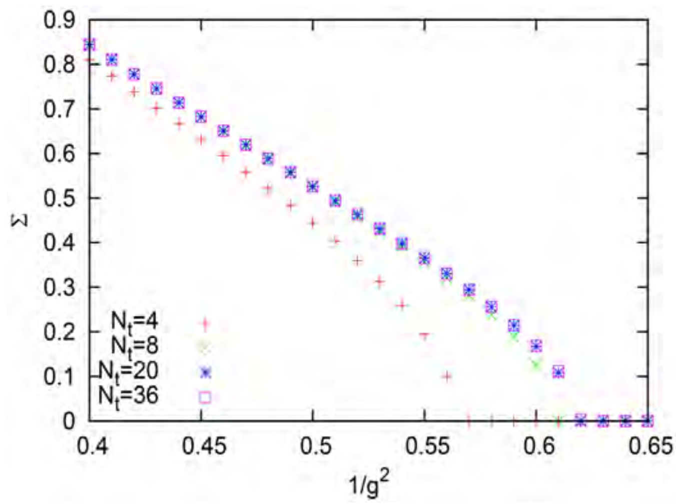


Figure 9. Σ versus $1/g^2$ for different N_t. $\mu = 0$, $m = 0$, $N_x = 36$.

Compared to **Figure 5** for the 2 + 1d model, the critical coupling becomes smaller. Moreover, the dependence of Σ on the temperature is less sensitive. **Figure 10** shows the dependence of Σ on the chemical potential at the coupling $1/g^2 = 0.58$ for the 2 + 1d and 3 + 1d Gross-Neveu model, the critical chemical potential is larger for the 2 + 1d model than those for the 3 + 1d model.

8. Conclusions

The staggered fermion for the Gross-Neveu model at finite density and temperature is revisited. In the large N_f limit, this model in 1 + 1d, 2 + 1d and 3 + 1d dimension can be easily solved in momentum space. Moreover, an explicit formula for the inverse matrix for the 1 + 1d, 2 + 1d and 3 + 1d staggered fermion matrix is found, which can be implemented by parallelization. This formula can also be generalized to the other space dimensions. For the odd space dimension,

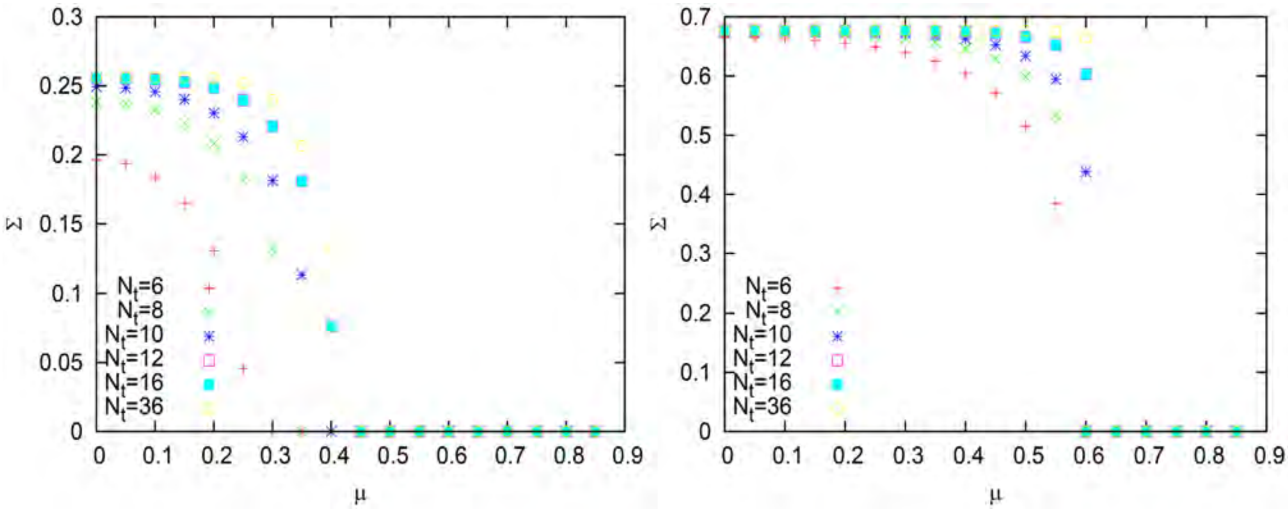


Figure 10. Σ versus μ, $m=0$, $(1/g^2=0.58)$, $N_x=36$. Left (3 + 1d), Right (2 + 1d).

the orthogonal transformation was found [33]. The key point to find the explicit formula for the inverse matrix is to use the properties of Γ_A and B_A as shown in Section 4. These properties for the even number of space dimension are simpler, as shown in the supplement material.

The dependence of chiral condensate and fermion density on the coupling, temperature and chemical potential are obtained by solving the gap equation. Our results for the 2 + 1d case reproduce the analytical results. We also compare the chiral condensate for the 2 + 1d and 3 + 1d case in the same range of parameters, showing that the reason for symmetry breaking and restoration can be explained by the suitable choice of the coupling, temperature and chemical potential.

Acknowledgements

Daming Li was supported by the National Science Foundation of China (No. 11271258, 11971309).

Conflicts of Interest

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

References

[1] de Forcrand, P. (2009) *PoS*, **LAT2009**, 10. https://doi.org/10.22323/1.091.0010

[2] Fukushima, K. and Hatsuda, T. (2011) *Reports on Progress in Physics*, **74**, Article ID: 014001. https://doi.org/10.1088/0034-4885/74/1/014001

[3] Bazavov, A., *et al.* (2012) *Physical Review D*, **85**, Article ID: 054503. https://doi.org/10.1103/PhysRevD.85.054503

[4] Bhattacharya, T., *et al.* (2014) *Physical Review Letters*, **113**, Article ID: 082001. https://doi.org/10.1103/PhysRevLett.113.082001

[5] Schmidt, C. and Sharma, S. (2017) *Journal of Physics G: Nuclear and Particle Phys-*

ics, **44**, Article ID: 104002. https://doi.org/10.1088/1361-6471/aa824a

[6] Bazavov, A., *et al.* (2019) *Physics Letters B*, **795**, 15. https://doi.org/10.1016/j.physletb.2019.05.013

[7] Fischer, C.S. (2019) *Progress in Particle and Nuclear Physics*, **105**, 1. https://doi.org/10.1016/j.ppnp.2019.01.002

[8] Fu, W.J., Pawlowski, J.M. and Rennecke, F. (2020) *Physical Review D*, **101**, Article ID: 054032. https://doi.org/10.1103/PhysRevD.101.054032

[9] Guenther, J.N. (2021) *The European Physical Journal A*, **57**, 136. https://doi.org/10.1140/epja/s10050-021-00354-6

[10] Cotter, S., Giudice, P., Hands, S. and Skullerud, J.-I. (2013) *Physical Review D*, **87**, Article ID: 034507. https://doi.org/10.1103/PhysRevD.87.034507

[11] Li, D.M. (2016) *Physical Review D*, **94**, Article ID: 114501. https://doi.org/10.1103/PhysRevD.94.114501

[12] Banuls, M.C. and Cichy, K. (2020) *Reports on Progress in Physics*, **83**, Article ID: 024401. https://doi.org/10.1088/1361-6633/ab6311

[13] Hauschild, J. and Pollmann, F. (2018) *SciPost Physics Lecture Notes*, **5**, 1.

[14] Verstraete, F. and Ignacio Cirac, J. (2010) *Physical Review Letters*, **104**, Article ID: 190405. https://doi.org/10.1103/PhysRevLett.104.190405

[15] Haegeman, J., Ignacio Cirac, J., Osborne, T.J. and Verstraete, F. (2013) *Physical Review B*, **88**, Article ID: 085118. https://doi.org/10.1103/PhysRevB.88.085118

[16] Rosenstein, B., Warr, B.J. and Park, S.H. (1991) *Physics Reports*, **205**, 59. https://doi.org/10.1016/0370-1573(91)90129-A

[17] Rosenstein, B., Warr, B.J. and Park, S.H. (1989) *Physical Review D*, **39**, 3088. https://doi.org/10.1103/PhysRevD.39.3088

[18] Hands, S., Kocic, A. and Kogut, J.B. (1993) *Nuclear Physics B*, **390**, 355. https://doi.org/10.1016/0550-3213(93)90460-7

[19] Hands, S., Kocić, A. and Kogut, J.B. (1993) *Annals of Physics*, **224**, 29. https://doi.org/10.1006/aphy.1993.1039

[20] Kogut, J.B. and Strouthos, C.G. (2001) *Physical Review D*, **63**, Article ID: 054502. https://doi.org/10.1103/PhysRevD.63.054502

[21] Wolff, U. (1985) *Physics Letters B*, **157**, 303. https://doi.org/10.1016/0370-2693(85)90671-9

[22] Thies, M. and Urlichs, K. (2003) *Physical Review D*, **67**, Article ID: 125015. https://doi.org/10.1103/PhysRevD.67.125015

[23] Schnetz, O., Thies, M. and Urlichs, K. (2004) *Annals of Physics* (*Amsterdam*), **314**, 425. https://doi.org/10.1016/j.aop.2004.06.009

[24] Thies, M. (2006) *Journal of Physics A*, **39**, 12707. https://doi.org/10.1088/0305-4470/39/41/S04

[25] de Forcrand, P. and Wenger, U. (2006) *PoS*, **LAT2009**, 152.

[26] Wagner, M. (2007) *Physical Review D*, **76**, Article ID: 076002. https://doi.org/10.1103/PhysRevD.76.076002

[27] Lenz, J., Pannullo, L., Wagner, M., Wellegehausen, B. and Wipf, A. (2020) *Physical Review D*, **101**, Article ID: 094512. https://doi.org/10.1103/PhysRevD.101.094512

[28] Lenz, J.J., Pannullo, L., Wagner, M., Wellegehausen, B. and Wipf, A. (2020) *Physical Review D*, **102**, Article ID: 114501. https://doi.org/10.1103/PhysRevD.102.114501

[29] Castorina, P., Mazza, M. and Zappala, D. (2003) *Physics Letters B*, **567**, 31-38.

https://doi.org/10.1016/j.physletb.2003.06.005

[30] Buballa, M., Kurth, L., Wagner, M. and Winstel, M. (2021) *Physical Review D*, **103**, Article ID: 034503. https://doi.org/10.1103/PhysRevD.103.034503

[31] Hands, S. (2016) *JHEP*, **11**, 15. https://doi.org/10.1007/JHEP11(2016)015

[32] Rothe, H.J. (2005) Lattice Gauge Theory, an Introduction. World Scientific Lecture Notes in Physics Vol. 74, World Scientific Publishing, Singapore. https://doi.org/10.1142/5674

[33] Cohen, Y., Elittzur, S. and Rabinovici, E. (1983) *Nuclear Physics B*, **220**, 102. https://doi.org/10.1016/0550-3213(83)90136-0

[34] Burden, C.J. and Burkitt, A.N. (1987) *Europhysics Letters*, **3**, 545. https://doi.org/10.1209/0295-5075/3/5/006

Appendix A. Proof of Properties of Γ_A and B_A

The notations for $\{A_i\}_{i=0}^{2}$ in (27) is a little awkward. I replace A_0, A_1 and A_2 in (27) by A_1, A_2 and A_3, respectively. Thus

$$\Gamma_A = \sigma_1^{A_1}\sigma_2^{A_2}\sigma_3^{A_3}, \quad B_A = (-\sigma_1)^{A_1}(-\sigma_2)^{A_2}(-\sigma_3)^{A_3} = (-1)^{A_1+A_2+A_3}\Gamma_A \tag{A1}$$

The three Pauli matrices

$$(\sigma_1)^{\alpha\beta} = (-1)^{\beta}\varepsilon_{\alpha\beta},$$
$$(\sigma_2)^{\alpha\beta} = (-i)\varepsilon_{\alpha\beta},$$
$$(\sigma_3)^{\alpha\beta} = (-1)^{\beta-1}\delta_{\alpha\beta}, \quad \alpha,\beta = 1,2$$

satisfies the completeness relation

$$\delta_{\alpha a}\delta_{\beta b} + \sum_{\mu=1,2,3}\sigma_\mu^{\alpha a}\sigma_\mu^{*\beta b} = 2\delta_{\alpha\beta}\delta_{ab} \tag{A2}$$

We first have

$$\begin{aligned}
\sum_{A,A_3=0}\Gamma_A^{\alpha a}\Gamma_A^{*\beta b} &= \sum_{A_1,A_2}\left(\sigma_1^{A_1}\sigma_2^{A_2}\right)^{\alpha a}\left(\sigma_1^{*A_1}\sigma_2^{*A_2}\right)^{\beta b} \\
&= \sum_{A_1,A_2}\left(\sigma_1^{A_1}\right)^{\alpha\gamma}\left(\sigma_2^{A_2}\right)^{\gamma a}\left(\sigma_1^{*A_1}\right)^{\beta\gamma'}\left(\sigma_2^{*A_2}\right)^{\gamma' b} \\
&= \sum_{A_1}\left(\sigma_1^{A_1}\right)^{\alpha\gamma}\left(\sigma_1^{*A_1}\right)^{\beta\gamma'}\sum_{A_2}\left(\sigma_2^{A_2}\right)^{\gamma a}(\sigma_2^{*A_2})^{\gamma' b} \\
&= \left(\delta_{\alpha\gamma}\delta_{\beta\gamma'} + \sigma_1^{\alpha\gamma}\sigma_1^{*\beta\gamma'}\right)\left(\delta_{\gamma a}\delta_{\gamma' b} + \sigma_2^{\gamma a}\sigma_2^{*\gamma' b}\right) \\
&= \delta_{\alpha a}\delta_{\beta b} + \sigma_2^{\alpha a}\sigma_2^{*\beta b} + \sigma_1^{\alpha a}\sigma_1^{*\beta b} + \left(\sigma_1\sigma_2\right)^{\alpha a}\left(\sigma_1^*\sigma_2^*\right)^{\beta b} \\
&= \delta_{\alpha a}\delta_{\beta b} + \sigma_2^{\alpha a}\sigma_2^{*\beta b} + \sigma_1^{\alpha a}\sigma_1^{*\beta b} + \sigma_3^{\alpha a}\sigma_3^{*\beta b} \\
&= 2\delta_{\alpha\beta}\delta_{ab} \quad \text{by (A2)}
\end{aligned} \tag{A3}$$

which is also valid if $(1,2)$ is replaced by $(1,3)$ or $(2,3)$. Secondly,

$$\begin{aligned}
\sum_A\Gamma_A^{\alpha a}\Gamma_A^{*\beta b} &= \sum_{A_1,A_2,A_3}\left(\sigma_1^{A_1}\sigma_2^{A_2}\sigma_3^{A_3}\right)^{\alpha a}\left(\sigma_1^{*A_1}\sigma_2^{*A_2}\sigma_3^{*A_3}\right)^{\beta b} \\
&= \sum_{A_1,A_2,A_3}\left(\sigma_1^{A_1}\sigma_2^{A_2}\right)^{\alpha t}\left(\sigma_3^{A_3}\right)^{ta}\left(\sigma_1^{*A_1}\sigma_2^{*A_2}\right)^{\beta t'}\left(\sigma_3^{*A_3}\right)^{t'b} \\
&= 2\delta_{\alpha\beta}\delta_{tt'}\left(\delta_{ta}\delta_{t'b} + (-1)^{a+b}\delta_{ta}\delta_{t'b}\right) \\
&= 4\delta_{\alpha\beta}\delta_{ab}
\end{aligned}$$

Inserting $B_A = (-1)^{A_1+A_2+A_3}\Gamma_A$ in the above equality, we have $\sum_A B_A^{\alpha a}B_A^{*\beta b} = 4\delta_{\alpha\beta}\delta_{ab}$.

$$\begin{aligned}
\sum_A\Gamma_A^{\alpha a}B_A^{*\beta b} &= \sum_{A_1,A_2,A_3}(-1)^{A_1+A_2+A_3}\left(\sigma_1^{A_1}\sigma_2^{A_2}\sigma_3^{A_3}\right)^{\alpha a}\left(\sigma_1^{*A_1}\sigma_2^{*A_2}\sigma_3^{*A_3}\right)^{\beta b} \\
&= \sum_{A_1,A_2,A_3}(-1)^{A_1+A_2}(-1)^{A_3}\left(\sigma_1^{A_1}\sigma_2^{A_2}\right)^{\alpha t}\left(\sigma_3^{A_3}\right)^{ta}\left(\sigma_1^{*A_1}\sigma_2^{*A_2}\right)^{\beta t'}\left(\sigma_3^{*A_3}\right)^{t'b} \\
&= \left(\delta_{\alpha t}\delta_{\beta t'} - \sigma_2^{\alpha t}\sigma_2^{*\beta t'} - \sigma_1^{\alpha t}\sigma_1^{*\beta t'} + \sigma_3^{\alpha t}\sigma_3^{*\beta t'}\right)\left(\delta_{ta}\delta_{t'b} - (-1)^{a+b}\delta_{ta}\delta_{t'b}\right) \\
&= \left(\delta_{\alpha a}\delta_{\beta b} - \sigma_2^{\alpha a}\sigma_2^{*\beta b} - \sigma_1^{\alpha a}\sigma_1^{*\beta b} + \sigma_3^{\alpha a}\sigma_3^{*\beta b}\right)\left(1-(-1)^{a+b}\right) = 0
\end{aligned}$$

where in the last equality we used

$$\begin{aligned}&\delta_{\alpha a}\delta_{\beta b}-\sigma_2^{\alpha a}\sigma_2^{*\beta b}-\sigma_1^{\alpha a}\sigma_1^{*\beta b}+\sigma_3^{\alpha a}\sigma_3^{*\beta b}\\&=\delta_{\alpha a}\delta_{\beta b}-\varepsilon_{\alpha a}\varepsilon_{\beta b}-(-1)^{a+b}\varepsilon_{\alpha a}\varepsilon_{\beta b}+(-1)^{a+b}\delta_{\alpha a}\delta_{\beta b}\\&=\left(1+(-1)^{a+b}\right)\left(\delta_{\alpha a}\delta_{\beta b}-\varepsilon_{\alpha a}\varepsilon_{\beta b}\right)=0,\quad \text{if } a\neq b\end{aligned}$$

To prove that

$$\sum_{A,A_\mu=1}\Gamma_A^{\alpha a}\left(\sigma_\mu^* B_A^*\right)^{\alpha' a'}=-2\left(\sigma_\mu^*\right)^{aa'}\delta_{\alpha\alpha'} \tag{A4}$$

we want to prove that

$$\sum_{A,A_\mu=1}\sigma_\mu^{*ba}\Gamma_A^{\alpha a}\left(\sigma_\mu^* B_A^*\right)^{\alpha' a'}=-2\sigma_\mu^{*ba}\left(\sigma_\mu^*\right)^{aa'}\delta_{\alpha\alpha'}$$

i.e.,

$$\sum_{A,A_\mu=1}\left(\Gamma_A\sigma_\mu\right)^{\alpha b}\left(\sigma_\mu^* B_A^*\right)^{\alpha' a'}=-2\delta_{ba'}\delta_{\alpha\alpha'}$$

This is obvious since the left hand side is

$$\begin{aligned}&\sum_{A,A_\mu=1}\left(\sigma_1^{A_1}\cdots\sigma_{\mu-1}^{A_{\mu-1}}\sigma_{\mu+1}^{A_{\mu+1}}\cdots\sigma_3^{A_3}\right)^{\alpha b}(-1)^{A_{\mu+1}+\cdots+A_3}\\&\times\left(\sigma_1^{*A_1}\cdots\sigma_{\mu-1}^{*A_{\mu-1}}\sigma_{\mu+1}^{*A_{\mu+1}}\cdots\sigma_3^{*A_3}\right)^{\alpha' a'}(-1)^{A_1+\cdots+A_{\mu-1}}(-1)^{A_1+A_2+A_3}\\&=\sum_{A,A_\mu=1}(-1)^{A_\mu}\left(\sigma_1^{A_1}\cdots\sigma_{\mu-1}^{A_{\mu-1}}\sigma_{\mu+1}^{A_{\mu+1}}\cdots\sigma_3^{A_3}\right)^{\alpha b}\left(\sigma_1^{*A_1}\cdots\sigma_{\mu-1}^{*A_{\mu-1}}\sigma_{\mu+1}^{*A_{\mu+1}}\cdots\sigma_3^{*A_3}\right)^{\alpha' a'}\\&=-2\delta_{ba'}\delta_{\alpha\alpha'},\ \text{by (A3) if}\ \ \mu=3\end{aligned} \tag{A5}$$

Similarly, (A4) is also valid if $A_\mu=1$ and -2 are replaced by $A_\mu=0$ and +2, respectively. This is because the sign $(-1)^{A_\mu}=-1$ in (A5) is replaced by $(-1)^{A_\mu}=+1$. Obviously,

$$\Gamma_{A\pm\hat{\mu}}=\eta_\mu(A)\sigma_\mu\Gamma_A,\quad \mu=1,2,3$$

For example, $\mu=2$,

$$\Gamma_{A\pm\hat{2}}=\sigma_1^{A_1}\sigma_2^{A_2\pm1}\sigma_3^{A_3}=\sigma_1^{A_1}\sigma_2^{A_2+1}\sigma_3^{A_3}=\eta_2(A)\sigma_2\Gamma_A,\quad \eta_2(A)=(-1)^{A_1}$$

Finally, we have

$$\frac{1}{4}\mathrm{Tr}\left(\Gamma_A^\dagger\Gamma_{A'}+B_A^\dagger B_{A'}\right)=\delta_{AA'}$$

since the left hand side is

$$\begin{aligned}&\frac{1}{4}\mathrm{Tr}\left[\begin{pmatrix}\Gamma_A & \\ & B_A\end{pmatrix}^\dagger\begin{pmatrix}\Gamma_{A'} & \\ & B_{A'}\end{pmatrix}\right]\\&=\frac{1}{4}\mathrm{Tr}\left(\left(\gamma_3^{A_3}\gamma_2^{A_2}\gamma_1^{A_1}\right)\left(\gamma_1^{A_1'}\gamma_2^{A_2'}\gamma_3^{A_3'}\right)\right)\\&=\frac{1}{4}(-1)^{(A_1+A_1')(A_2+A_3)+(A_2+A_2')A_3}\,\mathrm{Tr}\left(\gamma_1^{A_1+A_1'}\gamma_2^{A_2+A_2'}\gamma_3^{A_3+A_3'}\right)=\delta_{AA'}\end{aligned}$$

where we used

$$\gamma_\mu^i \gamma_\nu^j = (-1)^{ij} \gamma_\nu^j \gamma_\mu^i, \quad \mu \neq \nu, i, j = 0,1,2$$

$$\mathrm{Tr}(\gamma_\mu) = 0, \quad \mathrm{Tr}(\gamma_\mu \gamma_\nu) = 0, \quad \mu \neq \nu, \quad \mathrm{Tr}(\gamma_1 \gamma_2 \gamma_3) = 0$$

Here the we define $\gamma_\mu = \begin{pmatrix} \sigma_\mu & 0 \\ 0 & -\sigma_\mu \end{pmatrix} (\mu = 1,2,3)$.

Appendix B: The Derivation of (38)

The derivation of (38) is similar to the calculation of $\frac{1}{2}\sum_x \eta_\mu(x)\bar{\chi}(x)\left(\chi(x+\hat{\mu}) - \chi(x-\hat{\mu})\right)$.

$$\begin{aligned}
&\frac{1}{2}\sum_x \bar{\chi}(x)\left(\chi\left(x+\hat{0}\right)+\chi\left(x-\hat{0}\right)\right) \\
&= \frac{1}{2}\sum_{A,A',Y} \bar{\chi}(A,Y)\Big(\delta_{A+\hat{0},A'}\left(\chi(A',Y)+\chi\left(A',Y-\hat{0}\right)\right) \\
&\quad + \delta_{A-\hat{0},A'}\left(\chi\left(A',Y+\hat{0}\right)+\chi(A',Y)\right)\Big) \\
&= \frac{1}{2}\sum_{A,A',Y} \bar{\chi}(A,Y)\left(\frac{\delta_{A-\hat{0},A'}-\delta_{A+\hat{0},A'}}{2}\partial_0\chi(A',Y)+\frac{\delta_{A-\hat{0},A'}+\delta_{A+\hat{0},A'}}{2}\delta\chi(A',Y)\right) \\
&= \frac{1}{2}\sum_{A,A',Y}\sqrt{2}\sum_{\alpha,a}\left(\bar{u}^{\alpha a}(Y)\Gamma_A^{\alpha a}+\bar{d}^{\alpha a}(Y)B_A^{\alpha a}\right) \\
&\quad \times\Bigg\{\frac{\delta_{A-\hat{0},A'}-\delta_{A+\hat{0},A'}}{2}\sqrt{2}\sum_{\alpha',a'}\left(\Gamma_{A'}^{*\alpha'a'}\partial_0 u^{\alpha'a'}(Y)+B_{A'}^{*\alpha'a'}\partial_0 d^{\alpha'a'}(Y)\right) \\
&\quad +\frac{\delta_{A-\hat{0},A'}+\delta_{A+\hat{0},A'}}{2}\sqrt{2}\sum_{\alpha',a'}\left(\Gamma_{A'}^{*\alpha'a'}\delta u^{\alpha'a'}(Y)+B_{A'}^{*\alpha'a'}\delta d^{\alpha'a'}(Y)\right)\Bigg\} \\
&= 2\sum_Y\Big(\bar{u}^{\alpha a}(Y)\left(\sigma_1^*\right)^{aa'}\partial_0 d^{\alpha a'}(Y)+\bar{d}^{\alpha a}(Y)\left(-\sigma_1^*\right)^{aa'}\partial_0 u^{\alpha a'}(Y) \\
&\quad +\bar{u}^{\alpha a}(Y)(\sigma_1)^{\alpha\alpha'}\delta u^{\alpha' a}(Y)+\bar{d}^{\alpha a}(Y)(-\sigma_1)^{\alpha\alpha'}\delta d^{\alpha' a}(Y)\Big) \\
&= 8\sum_Y\left[\bar{q}(Y)\left(i\gamma_3\otimes\sigma_1^*\right)\frac{\partial_0 q(Y)}{4}+\bar{q}(Y)\left(\gamma_0\otimes\mathbb{I}_2\right)\frac{\delta q(Y)}{4}\right] \\
&= 8\sum_k\Big[\bar{q}(k)\left(i\gamma_3\otimes\sigma_1^*\right)i2^{-1}\sin(2k_0)q(k) \\
&\quad +\bar{q}(k)\left(\gamma_0\otimes\mathbb{I}_2\right)2^{-1}\left[\cos(2k_0)+1\right]q(k)\Big]
\end{aligned}$$

where

$$\delta q(Y) = q\left(Y+\hat{0}\right)+2q(Y)+q\left(Y-\hat{0}\right)$$

In the fourth equality, we used the formula like

$$\begin{aligned}
&\sum_{A,A'}\Gamma_A^{\alpha a}B_{A'}^{*\alpha'a'}\left(\delta_{A-\hat{0},A'}-\delta_{A+\hat{0},A'}\right) \\
&= \sum_{A,A_0=1}\Gamma_A^{\alpha a}B_{A-\hat{0}}^{*\alpha'a'}-\sum_{A,A_0=0}\Gamma_A^{\alpha a}B_{A+\hat{0}}^{*\alpha'a'} \\
&= \sum_{A,A_0=1}\Gamma_A^{\alpha a}\left(-\sigma_1 B_A\right)^{*\alpha'a'}-\sum_{A,A_0=0}\Gamma_A^{\alpha a}\left(-\sigma_1 B_A\right)^{*\alpha'a'} \\
&= 4\sigma_1^{*aa'}\delta_{\alpha\alpha'}
\end{aligned}$$

Appendix C. The Derivation of (44)-(46)

First,

$$\begin{aligned}
&\sum_x\left(D^{-1}_{x+\hat{0},x}s^1_x+D^{-1}_{x-\hat{0},x}s^2_x\right)\\
&=-\frac{\int \mathrm{e}^{-\bar{\chi}D\chi}\sum_x\bar{\chi}(x)\left[\chi\left(x+\hat{0}\right)+\chi\left(x-\hat{0}\right)\right]}{\int \mathrm{e}^{-\bar{\chi}D\chi}}\\
&=-\frac{\int \mathrm{e}^{-\sum_k\bar{q}(k)8D(k)q(k)}16\sum_k\bar{q}(k)A_+(k)q(k)}{\int \mathrm{e}^{-\sum_k\bar{q}(k)8D(k)q(k)}}\quad \text{by (38) (39)}\\
&=-16\sum_k\frac{\int \mathrm{e}^{-\bar{q}(k)8D(k)q(k)}\bar{q}(k)A_+(k)q(k)}{\int \mathrm{e}^{-\bar{q}(k)8D(k)q(k)}}\\
&=16\sum_k\mathrm{tr}\left[\left(8D(k)\right)^{-1}A_+(k)\right]\\
&=2\sum_k\mathrm{tr}\left[D(k)^{-1}A_+(k)\right]\\
&=\sum_k\frac{2}{N(k)}\mathrm{tr}\left\{\left[m-\sum_{\mu=0,1,2}\left(\gamma_\mu\otimes\mathbb{I}_2\right)a_\mu-\sum_{c=1,2,3}\left(\gamma_3\otimes\sigma^*_c\right)b_c\right]\right.\quad\text{by (41)}\\
&\quad\left.\times\left[\left(i\gamma_3\otimes\sigma^*_1\right)i2^{-1}\sin\left(2k_0\right)+\left(\gamma_0\otimes\mathbb{I}_2\right)2^{-1}\left[\cos\left(2k_0\right)+1\right]\right]\right\}\\
&=\sum_k\frac{2}{N(k)}\mathrm{tr}\left\{\left(\mathbb{I}_4\otimes\mathbb{I}_2\right)\left(b_12^{-1}\sin 2k_0-a_02^{-1}\left(\cos 2k_0+1\right)\right)\right\}\\
&=8\sum_k\frac{b_1\sin 2k_0-a_0\left(\cos 2k_0+1\right)}{N(k)}
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\sum_x\left(D^{-1}_{x+\hat{0},x}s^1_x-D^{-1}_{x-\hat{0},x}s^2_x\right)\\
&=\sum_k\frac{2}{N(k)}\mathrm{tr}\left\{\left[m-\sum_{\mu=0,1,2}\left(\gamma_\mu\otimes\mathbb{I}_2\right)a_\mu-\sum_{c=1,2,3}\left(\gamma_3\otimes\sigma^*_c\right)b_c\right]\right.\\
&\quad\left.\times\left[\left(\gamma_0\otimes\mathbb{I}_2\right)i2^{-1}\sin\left(2k_0\right)+\left(i\gamma_3\otimes\sigma^*_1\right)2^{-1}\left[\cos\left(2k_0\right)-1\right]\right]\right\}\\
&=\sum_k\frac{2}{N(k)}\mathrm{tr}\left\{\left(\mathbb{I}_4\otimes\mathbb{I}_2\right)\left(-a_0i2^{-1}\sin 2k_0-b_1i2^{-1}\left(\cos 2k_0-1\right)\right)\right\}\\
&=(-8i)\sum_k\frac{a_0\sin 2k_0+b_1\left(\cos 2k_0-1\right)}{N(k)}
\end{aligned}$$

The inverse matrix of D in (20) can be calculated as follows

$$\begin{aligned}
D^{-1}_{x',x}&=-\frac{\int \mathrm{e}^{-\bar{\chi}D\chi}\bar{\chi}(x)\chi(x')}{\int \mathrm{e}^{-\bar{\chi}D\chi}}\\
&=-2\sum_{\alpha,a,\alpha',a'}\frac{\int \mathrm{e}^{-\bar{q}8Dq}\left[\bar{u}^{\alpha a}(Y)\Gamma^{\alpha a}_A+\bar{d}^{\alpha a}(Y)B^{\alpha a}_A\right]\left[\Gamma^{*\alpha'a'}_{A'}u^{\alpha'a'}(Y')+B^{*\alpha'a'}_{A'}d^{\alpha'a'}(Y')\right]}{\int \mathrm{e}^{-\bar{q}8Dq}}\\
&=-2\sum_{\alpha,a,\alpha',a'}\left[\Gamma^{\alpha a}_A\Gamma^{*\alpha'a'}_{A'}\frac{\int \mathrm{e}^{-\bar{q}8Dq}\bar{q}^{\alpha a}_1(Y)q^{\alpha'a'}_1(Y')}{\int \mathrm{e}^{-\bar{q}8Dq}}+\Gamma^{\alpha a}_AB^{*\alpha'a'}_{A'}\frac{\int \mathrm{e}^{-\bar{q}8Dq}\bar{q}^{\alpha a}_1(Y)q^{\alpha'a'}_2(Y')}{\int \mathrm{e}^{-\bar{q}8Dq}}\right.
\end{aligned}$$

$$+B_A^{\alpha a}\Gamma_{A'}^{*\alpha' a'}\frac{\int \mathrm{e}^{-\bar{q}8Dq}\bar{q}_2^{\alpha a}(Y)q_1^{\alpha' a'}(Y')}{\int \mathrm{e}^{-\bar{q}8Dq}}+B_A^{\alpha a}B_{A'}^{*\alpha' a'}\frac{\int \mathrm{e}^{-\bar{q}8Dq}\bar{q}_2^{\alpha a}(Y)q_2^{\alpha' a'}(Y')}{\int \mathrm{e}^{-\bar{q}8Dq}}\Bigg]$$

$$=\frac{1}{4}\sum_{\alpha,a,\alpha',a'}\Big[\Gamma_A^{\alpha a}\Gamma_{A'}^{*\alpha' a'}D^{-1}_{(Y'a'\alpha'1;Ya\alpha 1)}+\Gamma_A^{\alpha a}B_{A'}^{*\alpha' a'}D^{-1}_{(Y'a'\alpha'2;Ya\alpha 1)}$$

$$+B_A^{\alpha a}\Gamma_{A'}^{*\alpha' a'}D^{-1}_{(Y'a'\alpha'1;Ya\alpha 2)}+B_A^{\alpha a}B_{A'}^{*\alpha' a'}D^{-1}_{(Y'a'\alpha'2;Ya\alpha 2)}\Big]$$

Journal of Modern Physics, 2021, 12, 1822-1842
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

The Conformal Group Revisited

J.-F. Pommaret

CERMICS, Ecole des Ponts ParisTech, Paris, France
Email: jean-francois.pommaret@wanadoo.fr

How to cite this paper: Pommaret, J.-F. (2021) The Conformal Group Revisited. *Journal of Modern Physics*, **12**, 1822-1842.
https://doi.org/10.4236/jmp.2021.1213106

Received: October 18, 2021
Accepted: November 22, 2021
Published: November 25, 2021

Abstract

Since 100 years or so, it has been usually accepted that the *conformal group* could be defined in an arbitrary dimension n as the group of transformations preserving a non-degenerate flat metric up to a nonzero invertible point depending factor called "*conformal factor*". However, when $n \geq 3$, it is a finite dimensional Lie group of transformations with n *translations*, $n(n-1)/2$ *rotations*, 1 *dilatation* and n nonlinear transformations called *elations* by E. Cartan in 1922, that is a total of $(n+1)(n+2)/2$ transformations. Because of the Michelson-Morley experiment, the conformal group of space-time with 15 parameters is well known for the Minkowski metric and is the biggest group of invariance of the Minkowski constitutive law of electromagnetism (EM) in vacuum, even though the two sets of field and induction Maxwell equations are respectively invariant by any local diffeomorphism. As this last generic number is also well defined and becomes equal to 3 for $n=1$ or 6 for $n=2$, the purpose of this paper is to use modern mathematical tools such as the *Spencer operator* on systems of OD or PD equations, both with its restriction to their symbols leading to the *Spencer* δ-*cohomology*, in order to provide a *unique definition* that could be valid for any $n \geq 1$. The concept of an "*involutive system*" is crucial for such a new definition.

Keywords

Conformal Group, Lie Group, Lie Pseudogroup, Spencer Operator, Spencer Cohomology, Acyclicity, Involutive System, Maxwell Equations

1. Introduction

Using local notations, this paper is mainly concerned with the following two connected problems: Given a differential operator $\xi \xrightarrow{\mathcal{D}} \eta$, how can we find *compatibility conditions* (CC), that is how can we construct a sequence $\xi \xrightarrow{\mathcal{D}} \eta \xrightarrow{\mathcal{D}_1} \zeta$ such that $\mathcal{D}_1 \circ \mathcal{D} = 0$ and, among all such possible sequences, what are the "*best*" ones, at least among the generating ones and when could we say that the sequence obtained is "*exact*" in a purely formal way, that is using only computer algebra for testing such a property? The order of an operator will be indicated under its arrow.

The difficulty is that, physicists being more familiar with analysis, will say that a sequence is "*locally exact*" if one can find locally ξ such that $\mathcal{D}\xi = \eta$ whenever $\mathcal{D}_1\eta = 0$. However, they have in mind the property of the exterior derivative d and Maxwell equations in electromagnetism (EM), that is to say, using standard notations, the (local) possibility to introduce the EM potential A such that $dA = F$ whenever the EM field F is a closed 2-form with $dF = 0$.

The main purpose of this paper is to prove that the "things" may be much more delicate and that these problems are only rarely associated with exterior calculus. We use the notations that can be found at length in our many books ([1]-[6]) or papers ([7] [8] [9] [10] [11]).

Let $\mu = (\mu_1, \cdots, \mu_n)$ be a multi-index with *length* $|\mu| = \mu_1 + \cdots + \mu_n$, *class i* if $\mu_1 = \cdots = \mu_{i-1} = 0, \mu_i \neq 0$ and $\mu + 1_i = (\mu_1, \cdots, \mu_{i-1}, \mu_i + 1, \mu_{i+1}, \cdots, \mu_n)$. We set $y_q = \{ y_\mu^k \mid 1 \le k \le m, 0 \le |\mu| \le q \}$ with $y_\mu^k = y^k$ when $|\mu| = 0$. If E is a vector bundle over X with local coordinates (x^i, y^k) for $i = 1, \cdots, n$ and $k = 1, \cdots, m$, we denote by $J_q(E)$ the *q-jet bundle* of E with local coordinates simply denoted by (x, y_q) and *sections* $\xi_q : (x) \to (x, \xi^k(x), \xi_i^k(x), \xi_{ij}^k(x), \cdots)$ transforming like the section $j_q(\xi) : (x) \to (x, \xi^k(x), \partial_i \xi^k(x), \partial_{ij}\xi^k(x), \cdots)$ when ξ is an arbitrary section of E. Then both $\xi_q \in J_q(E)$ and $j_q(\xi) \in J_q(E)$ are over $\xi \in E$ and the *Spencer operator*, *which is defined on sections*, just allows to distinguish them by introducing a kind of "*difference*" through the operator $d : J_{q+1}(E) \to T^* \otimes J_q(E) : \xi_{q+1} \to j_1(\xi_q) - \xi_{q+1}$ with local components $(\partial_i \xi^k(x) - \xi_i^k(x), \partial_i \xi_j^k(x) - \xi_{ij}^k(x), \cdots)$ and more generally $(d\xi_{q+1})_{\mu,i}^k(x) = \partial_i \xi_\mu^k(x) - \xi_{\mu+1_i}^k(x)$. In a symbolic way, *when changes of coordinates are not involved*, it is sometimes useful to write down the components of d in the form $d_i = \partial_i - \delta_i$. The restriction of d to the kernel $S_{q+1}T^* \otimes E$ of the canonical projection $\pi_q^{q+1} : J_{q+1}(E) \to J_q(E)$ is *minus* the *Spencer map* $\delta = dx^i \wedge \delta_i : S_{q+1}T^* \otimes E \to T^* \otimes S_q T^* \otimes E$ and $\delta \circ \delta = 0$. The kernel of d is made by sections such that $\xi_{q+1} = j_1(\xi_q) = j_2(\xi_{q-1}) = \cdots = j_{q+1}(\xi)$. Finally, if $R_q \subset J_q(E)$ is a *system* of order q on E locally defined by linear equations $\Phi^\tau(x, y_q) \equiv a_k^{\tau\mu}(x) y_\mu^k = 0$, the *r-prolongation* $R_{q+r} = \rho_r(R_q) = J_r(R_q) \cap J_{q+r}(E) \subset J_r(J_q(E))$ is locally defined when $r = 1$ by the set of linear equations $\Phi^\tau(x, y_q) = 0$, $d_i \Phi^\tau(x, y_{q+1}) \equiv a_k^{\tau\mu}(x) y_{\mu+1_i}^k + \partial_i a_k^{\tau\mu}(x) y_\mu^k = 0$ and has *symbol* $g_{q+r} = R_{q+r} \cap S_{q+r}T^* \otimes E \subset J_{q+r}(E)$ if one looks at the *top order terms*. If $\xi_{q+1} \in R_{q+1}$ is over $\xi_q \in R_q$, differentiating the identity $a_k^{\tau\mu}(x) \xi_\mu^k(x) \equiv 0$ with respect to x^i and substracting the identity $a_k^{\tau\mu}(x) \xi_{\mu+1_i}^k(x) + \partial_i a_k^{\tau\mu}(x) \xi_\mu^k(x) \equiv 0$, we obtain the identity $a_k^{\tau\mu}(x) \left(\partial_i \xi_\mu^k(x) - \xi_{\mu+1_i}^k(x) \right) \equiv 0$ and thus the restriction $d : R_{q+1} \to T^* \otimes R_q$ ([1] [3] [4] [12]).

DEFINITION 1.1: g_q is said to be *s-acyclic* if the purely algebraic δ-cohomology $H_{q+r}^s(g_q)$ of $\cdots \xrightarrow{\delta} \wedge^s T^* \otimes g_{q+r} \xrightarrow{\delta} \cdots$ are such that

$H^1_{q+r}(g_q) = \cdots = H^s_{q+r}(g_q) = 0, \forall r \geq 0$ and *involutive* if it is *n*-acyclic. Also R_q is said to be *involutive* if it is *formally integrable* (FI), that is when the restriction $\pi^{q+r+1}_{q+r} : R_{q+r+1} \to R_{q+r}$ is an epimorphism $\forall r \geq 0$ or, equivalently, when all the equations of order $q+r$ are obtained by *r* prolongations only, $\forall r \geq 0$ and g_q is involutive. In that case, $R_{q+1} \subset J_1(R_q)$ is a canonical equivalent formally integrable first order involutive system on R_q with no zero order equations, called the *Spencer form*.

EXAMPLE 1.2: (*Classical Killing operator*)

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 $\gamma^k_{ij} = \frac{1}{2}\omega^{kr}\left(\partial_i\omega_{rj} + \partial_j\omega_{ir} - \partial_r\omega_{ij}\right)$ where (ω^{rs}) is the inverse matrix of (ω_{ij}) and the *formal Lie derivative* of gometric objects, we get the second order system $R_2 \subset J_2(T)$:

$$\begin{cases} \Omega_{ij} \equiv \left(L(\xi_1)\omega\right)_{ij} = \omega_{rj}(x)\xi^r_i + \omega_{ir}(x)\xi^r_j + \xi^r\partial_r\omega_{ij}(x) = 0 \\ \Gamma^k_{ij} \equiv \left(L(\xi_2)\gamma\right)^k_{ij} = \xi^k_{ij} + \gamma^k_{rj}(x)\xi^r_i + \gamma^k_{ir}(x)\xi^r_j - \gamma^r_{ij}(x)\xi^k_r + \xi^r\partial_r\gamma^k_{ij}(x) = 0 \end{cases}$$

with sections $\xi_2 : x \to \left(\xi^k(x), \xi^k_i(x), \xi^k_{ij}(x)\right)$ transforming like $j_2(\xi) : x \to \left(\xi^k(x), \partial_i\xi^k(x), \partial_{ij}\xi^k(x)\right)$. 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 ([4] [5]), 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 ([13]) which is only a particular example of the *Vessiot structure equations* discovered by E. Vessiot in 1904 ([14]), though in a quite different setting (See [4] and [15] for an explicit modern proof). Such a necessary condition for constructing an exact differential sequence could not have been used by any follower because the "*Spencer machinery*" has only been known after 1970 ([12]). Otherwise, if the metric does not satisfy this condition, CC may exist but have no link with the Riemann tensor ([10]). We may define the vector bundle F_1 in the short exact sequence made by the top row of the following commutative diagram:

$$
\begin{array}{ccccccccc}
 & & & 0 & & 0 & & & \\
 & & & \downarrow & & \downarrow & & & \\
 & 0 & \rightarrow & S_3T^*\otimes T & \rightarrow & S_2T^*\otimes F_0 & \rightarrow & F_1 & \rightarrow 0 \\
 & & & \downarrow & & \downarrow & & & \\
 & 0 & \rightarrow & T^*\otimes S_2T^*\otimes T & \rightarrow & T^*\otimes T^*\otimes F_0 & \rightarrow & 0 & \\
 & \downarrow & & \downarrow & & \downarrow & & & \\
0\rightarrow & \wedge^2T^*\otimes g_1 & \rightarrow & \underline{\wedge^2T^*\otimes T^*\otimes T} & \rightarrow & \wedge^2T^*\otimes F_0 & \rightarrow & 0 & \\
 & \downarrow & & \downarrow & & \downarrow & & & \\
0\rightarrow & \wedge^3T^*\otimes T & = & \wedge^3T^*\otimes T & \rightarrow & 0 & & & \\
 & \downarrow & & \downarrow & & & & & \\
 & 0 & & 0 & & & & &
\end{array}
$$

where the vertical δ-sequences are exact *but the first*, or, using a snake type diagonal chase, from the short exact sequence of vector bundles:

$$0\rightarrow F_1\rightarrow \wedge^2T^*\otimes g_1 \xrightarrow{\delta} \wedge^3T^*\otimes T\rightarrow 0$$

This result is first leading to the long exact sequence of vector bundles:

$$0\rightarrow R_3\rightarrow J_3(T)\rightarrow J_2(F_0)\rightarrow F_1\rightarrow 0$$

and to the *Riemann* operator $\mathcal{D}_1 : F_0 \xrightarrow{j_2} J_2(F_0)\rightarrow F_1$. As $g_2=0$, we also discover that F_1 is just the Spencer δ-cohomology $H^2(g_1)$ at $\wedge^2T^*\otimes g_1$ along the previous short exact sequence.

We get the *striking formulas* where the + signs are replaced by − signs:

$$
\begin{aligned}
dim(F_1) &= n^2(n+1)^2/4 - n^2(n+1)(n+2)/6 \\
&= n^2(n-1)^2/4 - n^2(n-1)(n-2)/6 \\
&= n^2(n^2-1)/12
\end{aligned}
$$

This result, first found as early as in 1978 ([9]), clearly exhibit *without indices* the two well known algebraic properties of the Riemann tensor as a section of the tensor bundle $\wedge^2T^*\otimes T^*\otimes T$.

It thus remains to exhibit the *Bianchi* operator exactly as we did for the *Riemann* operator, with the same historical comments already provided. However, now we know that R_1 is formally integrable (otherwise nothing could be achieved and we should start with a smaller system [1] [4] [6]), the construction of the linearized Janet-type differential sequence as a strictly exact differential sequence but *not* an involutive differential sequence because the system R_1 and thus the first order operator $\mathcal{D}$ are formally integrable though *not* involutive as g_1 is finite type with $g_2=0$ but not involutive. Doing one more prolongation only, we obtain the first order *Bianchi* operator $\mathcal{D}_2 : F_1 \xrightarrow{j_1} J_2(F_1)\rightarrow F_2$ as before, defining the vector bundle F_2 in the long exact sequence made by the top row of the following commutative diagram:

$$
\begin{array}{ccccccccccc}
 & & & 0 & & 0 & & 0 & & & \\
 & & & \downarrow & & \downarrow & & \downarrow & & & \\
 & 0 & \to & S_4T^*\otimes T & \to & S_3T^*\otimes F_0 & \to & T^*\otimes F_1 & \to & F_2 & \to 0\\
 & & & \downarrow & & \downarrow & & \| & & & \\
 & 0 & \to & T^*\otimes S_3T^*\otimes T & \to & T^*\otimes S_2T^*\otimes F_0 & \to & T^*\otimes F_1 & \to & 0 & \\
 & & & \downarrow & & \downarrow & & \downarrow & & & \\
 & 0 & \to & \wedge^2T^*\otimes S_2T^*\otimes T & \to & \wedge^2T^*\otimes T^*\otimes F_0 & \to & 0 & & & \\
 & \downarrow & & \downarrow & & \downarrow & & & & & \\
0\to & \wedge^3T^*\otimes g_1 & \to & \underline{\wedge^3T^*\otimes T^*\otimes T} & \to & \wedge^3T^*\otimes F_0 & \to & 0 & & & \\
 & \downarrow & & \downarrow & & \downarrow & & & & & \\
0\to & \wedge^4T^*\otimes T & = & \wedge^4T^*\otimes T & \to & 0 & & & & & \\
 & \downarrow & & \downarrow & & & & & & & \\
 & 0 & & 0 & & & & & & &
\end{array}
$$

where the vertical δ-sequences are exact *but the first*, or, using a snake type diagonal chase, from the short exact sequence:

$$0\to F_2\to\wedge^3T^*\otimes g_1\xrightarrow{\delta}\wedge^4T^*\otimes T\to 0$$

showing that $F_2=H^3(g_1)$ ([8] [9]). We have in particular for $n\geq 4$:

$$
\begin{aligned}
dim(F_2) &= n^2(n-1)^2(n-2)/12-n^2(n-1)(n-2)(n-3)/24\\
&= n^2(n+1)(n+2)(n+3)/24+n^3\left(n_1^2\right)/12\\
&\quad -n^2(n+1)(n+2)(n+3)/24\\
&= n^2\left(n^2-1\right)(n-2)/24
\end{aligned}
$$

and thus $dim(F_2)=(4\times 6)-(1\times 4)=(16\times 15\times 2)/24=20$ when $n=4$. This result also exhibits all the properties of the Bianchi identities as a section of the tensor bundle $\wedge^3T^*\otimes T^*\otimes T$. In arbitrary dimension, we finally obtain the differential sequence, which is *not* a Janet sequence:

$$0\to\Theta\to T\underset{1}{\overset{Killing}{\to}}F_0\underset{2}{\overset{Riemann}{\to}}F_1\underset{1}{\overset{Bianchi}{\to}}F_2$$

EXAMPLE 1.3: (*Conformal Killing operator*)

At first sight, it seems that similar methods could work in order to study the conformal Killing operator and, more generally, all conformal concepts will be described with a "hat", in order to provide the strictly exact differential sequence:

$$0\to\hat{\Theta}\to T\xrightarrow{\hat{\mathcal{D}}}\hat{F}_0\xrightarrow{\hat{\mathcal{D}}_1}\hat{F}_1\xrightarrow{\hat{\mathcal{D}}_2}\hat{F}_2$$

where $\hat{\mathcal{D}}_1$ is the *Weyl* operator with generating CC $\hat{\mathcal{D}}_2$. It is only in 2016 (see [9] and [15] for more details) that we have been able to recover all these operators and confirm with computer algebra that the orders of the operators involved highly depend on the dimension as follows:

- $n=3$: $3\underset{1}{\to}5\underset{3}{\to}5\underset{1}{\to}3\to 0$
- $n=4$: $4\underset{1}{\to}9\underset{2}{\to}10\underset{2}{\to}9\underset{1}{\to}4\to 0$
- $n\geq 5$: $5\underset{1}{\to}14\underset{2}{\to}35\underset{1}{\to}35\underset{2}{\to}14\underset{1}{\to}5\to 0$

These results are bringing the need to revisit entirely the mathematical foundations of conformal geometry, in particular when $n=3$ because the Weyl type operator is of third order and when $n=4$ because the Bianchi type operator is second order in this case contrary to the situation met when $n=5$. However, *surprisingly*, these results have *never* been acknowledged and the reader will not discover a single reference on such questions in the mathematical literature.

The reason is probably because these results are based on the following technical lemma that could not be even imagined without a deep knowledge *and* practice of the Spencer δ-cohomology (see [16] for details):

LEMMA 1.4: The symbol $\hat{g}_1$ defined by the linear equations:

$$\hat{\Omega}_{ij} \equiv \omega_{rj}(x)\xi_i^r + \omega_{ir}(x)\xi_j^r - \frac{1}{2}\omega_{ij}(x)\xi_r^r = 0$$

does not depend on any conformal factor, is finite type with $\hat{g}_3 = 0, \forall n \geq 3$ and is surprisingly such that $\hat{g}_2$ is 2-acyclic for $n \geq 4$ or even 3-acyclic when $n \geq 5$.

REMARK 1.5: In order to emphasize the reason for using Lie equations, we now provide the explicit form of the n infinitesimal relations with $1 \leq r,s,t \leq n$, whenever $n \geq 3$:

$$\boxed{\theta_s = -\frac{1}{2}x^2\delta_s^r\partial_r + \omega_{st}x^t x^r \partial_r \Rightarrow \partial_r\theta_s^r = n\omega_{st}x^t, [\theta_s,\theta_t] = 0}$$

where the underlying metric is used for the scalar product x^2 involved. It is easy to check that $\xi_2 \in S_2T^* \otimes T$ 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 the following formula where δ is the standard Kronecker symbol and $\xi_2 \in \hat{R}_2$:

$$\begin{aligned}\Gamma_{ij}^k &\equiv \left(L(\xi_2)\gamma\right)_{ij}^k = \xi_{ij}^k + \gamma_{rj}^k\xi_i^r + \gamma_{ir}^k\xi_j^r - \gamma_{ij}^r\xi_r^k + \xi^r\partial_r\gamma_{ij}^k \\ &= \delta_i^k A_j + \delta_j^k A_i - \omega_{ij}\omega^{kr}A_r\end{aligned}$$

We thus understand how important it is to use "*sections*" rather than "*solutions*".

Accordingly, a possible unification can be achieved through the "*fundamental diagram I*" relating together the *Spencer sequence* and the *Janet sequence* as follows in arbitrary dimension n for any *involutive* system $R_q \subseteq J_q(E)$ because these are the only existing canonical sequences ([1]):

$$\begin{array}{ccccccccccccccc}
 & & & & & & 0 & & 0 & & & & 0 & & 0 & & \\
 & & & & & & \downarrow & & \downarrow & & & & \downarrow & & \downarrow & & \\
 & & 0 & \to & \Theta & \xrightarrow{j_q} & C_0 & \xrightarrow{D_1} & C_1 & \xrightarrow{D_2} \cdots \xrightarrow{D_{n-1}} & C_{n-1} & \xrightarrow{D_n} & C_n & \to & 0 \\
 & & & & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & \to & E & \xrightarrow{j_q} & C_0(E) & \xrightarrow{D_1} & C_1(E) & \xrightarrow{D_2} \cdots \xrightarrow{D_{n-1}} & C_{n-1}(E) & \xrightarrow{D_n} & C_n(E) & \to & 0 \\
 & & & & \| & & \downarrow \Phi_0 & & \downarrow \Phi_1 & & \downarrow \Phi_{n-1} & & \downarrow \Phi_n & & \\
0 \to & \Theta & & \to & E & \xrightarrow{\mathcal{D}} & F_0 & \xrightarrow{\mathcal{D}_1} & F_1 & \xrightarrow{\mathcal{D}_2} \cdots \xrightarrow{\mathcal{D}_{n-1}} & F_{n-1} & \xrightarrow{\mathcal{D}_n} & F_n & \to & 0 \\
 & & & & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
 & & & & & & 0 & & 0 & & 0 & & 0 & &
\end{array}$$

where $C_0 = R_q \subset J_q(E) = C_0(E)$ and $dim(F_r) = dim(C_r(E)) - dim(C_r)$. Indeed, we have $dim(C_r) = dim(\wedge^r T^*) \times dim(R_q)$ for finite type involutive systems and we therefore notice that *the crucial point is to deal with involutive systems*. In the group framework, we have $E = T$ and, as we are dealing with finite type systems, it is thus sufficient to replace j_q and $R_q \subset J_q(E)$ by j_2 and $R_2 \subset J_2(T)$ with $g_2 = 0$ in the classical situation or by j_3 and $\hat{R}_3 \subset J_3(T)$ with $\hat{g}_3 = 0$ in the conformal situation, on the condition to be able to treat the specific cases $n = 1$ and $n = 2$.

Finally, as a different way to look at these questions, if *K* be a differential field containing $\mathbb{Q}$, we may introduce the ring $D = K[d] = K[d_1, \cdots, d_n]$ of differential operators with coefficients in *K* and consider a linear differential operator $\mathcal{D}$ with coefficients in *K*. If $\mathcal{D}_1$ generates the CC of $\mathcal{D}$, we have of course $\mathcal{D}_1 \circ \mathcal{D} = 0$. Taking the respective (formal) *adjoint* operators, we obtain therefore $ad(\mathcal{D}) \circ ad(\mathcal{D}_1) = 0$ but $ad(\mathcal{D})$ may not generate the CC of $ad(\mathcal{D}_1)$ and so on in *any* differential sequence where each operator generates the CC of the preceding one.

DEFINITION 1.6: If *M* is the differential module over *D* or simply *D*-module defined by $\mathcal{D}$, we set $ext_D^0(M) = hom_D(M, D)$. As for the other *extension modules*, they have been created in order to "*measure*" the previous gaps ([5]). In particular, we say that $ext_D^1(M) = 0$ if $ad(\mathcal{D})$ generates the CC of $ad(\mathcal{D}_1)$, that $ext_D^2(M) = 0$ if $ad(\mathcal{D}_1)$ generates the CC of $ad(\mathcal{D}_2)$ and so on. Moreover, if $\mathcal{D}$ is of finite type, then $ad(\mathcal{D})$ is surjective with $ext_D^0(M) = 0$. The simplest example is that of classical space geometry with $n = 3$ and $ad(grad) = -div$. Similar definitions are also valid for the Janet and Spencer sequences. Also, vanishing of the first extension module amounts to the existence of a local parametrization by potential-like functions ([7]).

According to a (difficult) theorem of (differential) homological algebra, *the extension modules only depend on M and not on the previous differential sequences used* ([17] [18]. They are used in agebraic geometry and have even been introduced in engineering sciences after 1990 (control theory) ([5] [6]). However, though the extension modules are the *only* intrinsic objects that can be associated with a differential module, they have surprisingly *never* been introduced in mathematical physics. The main problem is that a control system is controllable if and only if it is parametrizable by potentials while the systems involved can be parametrized in all classical physics (Cauchy or Maxwell equations are well known examples in [7]) apart from... Einstein equations ([8]). As for the tools involved, we let the reader compare ([2] [3]) to ([19] [20]).

After presenting two motivating examples in Section 2, such a procedure will be achieved in Section 3 in such a way that the Spencer sequences involved, being isomorphic to tensor products of the Poincaré sequence for the exterior derivative by finite dimensional Lie algebras, will have therefore vanishing zero, first and second extension modules when $n \geq 3$ ([4] [11]). For all results concerning differential modules, we refer the reader to the (difficult) references ([5]

[21] [22] [23]).

2. Two Motivating Examples

EXAMPLE 2.1

With $m=1, n=2, q=2, K=\mathbb{Q}$, let us consider the inhomogeneous second order operator:

$$Py \equiv d_{22}y = u, \quad Qy \equiv d_{12}y - y = v$$

We obtain at once through crossed derivatives:

$$y = d_{11}u - d_{12}v - v \Rightarrow \Theta = 0$$

and, by substituting, two fourth order CC for (u,v), namely:

$$\left\{\begin{array}{l} U \equiv d_{1122}u - d_{1222}v - d_{22}v - u = 0 \\ V \equiv d_{1112}u - d_{11}u - d_{1122}v = 0 \end{array}\right\} \Rightarrow W \equiv d_{12}V + V - d_{11}U = 0$$

However, the commutation relation $P \circ Q \equiv Q \circ P$ provides a single CC for (u,v), namely:

$$C \equiv d_{12}u - u - d_{22}v = 0$$

and we check at once $U = d_{12}C + C, V = d_{11}C$ while $C = d_{22}V - d_{12}U + U$, hat is: $(U=0, V=0) \Leftrightarrow (C=0)$.

Using corresponding notations, let us compare the two following differential sequences:

$$0 \rightarrow \Theta \rightarrow y \underset{2}{\overset{\mathcal{D}}{\rightarrow}} (u,v) \underset{4}{\overset{\mathcal{D}_1}{\rightarrow}} (U,V) \underset{2}{\overset{\mathcal{D}_2}{\rightarrow}} W \rightarrow 0 \quad (1)$$

$$0 \rightarrow \Theta \rightarrow y \underset{2}{\overset{\mathcal{D}}{\rightarrow}} (u,v) \underset{2}{\overset{\mathcal{D}_1'}{\rightarrow}} C \rightarrow 0 \quad (2)$$

Though the second order system considered is surely *not* FI because the 4 parametric jets of R_2 are (y, y_1, y_2, y_{11}) and the 4 (again !) parametric jets of R_3 are $(y, y_1, y_{11}, y_{111})$ but the 4 (again !) parametric jets of R_4 are $(y_1, y_{11}, y_{111}, y_{1111})$. More generally, we let the reader prove by induction that $dim(R_{2+r}) = 4, \forall r \geq 0$. The formal *r*-prolongation of (2), namely:

$$0 \rightarrow R_{r+4} \rightarrow J_{r+4}(y) \rightarrow J_{r+2}(u,v) \rightarrow J_r(C) \rightarrow 0$$

is exact because $4-(r+5)(r+6)/2+(r+3)(r+4)-(r+1)(r+2)/2=0$, even though the corresponding symbol sequence:

$$0 \rightarrow g_{r+4} \rightarrow S_{r+4}T^*(y) \rightarrow S_{r+2}T^*(u,v) \rightarrow S_rT^*(C) \rightarrow 0$$

is *not* exact because $(2(r+3)-(r+1))-((r+5)-1)=(r+5)-(r+4)=1 \neq 0$ because the system considered is not formally integrable.

On the contrary, the prolongations of (1) are *not* exact on the jet level. Indeed, the long sequence:

$$0 \rightarrow R_8 \rightarrow J_8(y) \rightarrow J_6(u,v) \rightarrow J_2(U,V) \rightarrow W \rightarrow 0$$

is *not* exact because we have $4-45+56-12+1=4 \neq 0$.

Now, considering the ring $D=\mathbb{Q}[d_1,d_2]$ of differential operators with coefficients in the trivial differential field $\mathbb{Q}$, we have the "exact" sequences of differential modules where $M=0$:

$$0 \to D \to D^2 \to D^2 \to D \xrightarrow{p} M \to 0 \qquad (1^*)$$

$$0 \to D \to D^2 \to D \xrightarrow{p} M \to 0 \qquad (2^*)$$

where p is the canonical residual projection. However, and this is a quite delicate point rarely known even by mathematicians, *a fortiori* by physicists, they are not "*strictly*" exact even if the Euler-Poincaré characteristics both vanish because $1-2+2-1=0$ and $12+1=0$ (see [15] for definitions and more details). Roughly speaking, it follows that the "*best*" differential sequences are obtained by using *only* formally integrable operators/systems in such a way that sequences on the jet level can be studied through their symbol sequences, the "canonical" ones by using *exclusively* involutive operators/systems in such a way that what happens with $\mathcal{D}$ also happens with $\mathcal{D}_1$ and so on. It follows that the sequences (2) or (2^*) are "*better*" than (1) or (1^*) because they provide *more information* on the generating CC.

However, the given system is *not* FI and it should be "*better*" to use another system providing *more information*. In particular, if we start wth a system $R_q \subset J_q(E)$ and set $R_{q+r}=\rho_r(R_q)=J_r(R_q)\cap J_{q+r}(E)$, it is known that (in general) one can find two integers $r,s\geq 0$ such that the system $R_{q+r}^{(s)}=\pi_{q+r}^{q+r+s}(R_{q+r})$ is formally integrable and even involutive with the same solutions ([1] [5] [6]). When all the operators are FI, the sequence is said to be *strictly* exact ([24]).

In the present situation, it should be "*better*" to replace R_2 by $R_2^{(4)}=0$ because $R_2^{(2)}$ is adding $y=0$ while $R_2^{(3)}$ is adding $y_1=0, y_2=0$ and $R_2^{(4)}$ is adding $y_{11}=0$. It follows that the Janet sequence for the injective trivially involutive operator j_2 is providing even *more information*, along with the fact that the Spencer bundles vanish in the "*fundamental diagram I*" ([1] [4] [5]).

We let the reader check that all the extension modules vanish because $M=0$ and to compare with the *totally different* involutive system defined by $y_{22}=0, y_{12}=0$ with $M\neq 0 \Rightarrow ext^0(M)=0$, $ext^1(M)\neq 0$, $ext^2(M)\neq 0$.

EXAMPLE 2.2

- *FIRST STEP* With $n=3, m=1, q=2$, let us consider the second order linear system $R_2\subset J_2(E)$ introduced by F.S. Macaulay in his 1916 book ([25]) (See also [6] for more details):

$$\Phi^3 \equiv y_{33}=0, \Phi^2 \equiv y_{23}-y_{11}=0, \Phi^1 \equiv y_{22}=0$$

Using muli-indices, we may introduce the operators $R=d_{33}, Q=d_{23}-d_{11}, P=d_{22}$. Taking into account the 3 commutation relations $[Q,R]=0,[R,P]=0,[P,Q]=0$ and the single Jacobi identity $[P,[Q,R]]+[Q,[R,P]]+[R,[P,Q]]=0, \forall(P,Q,R)$, we obtain at once the following locally and strictly exact sequence where the order of each operator is under its own arrow:

$$0 \to \Theta \to 1 \underset{2}{\overset{\mathcal{D}}{\to}} 3 \underset{2}{\overset{\mathcal{D}_1}{\to}} 3 \underset{2}{\overset{\mathcal{D}_2}{\to}} 1 \to 0$$

However, the first operator $\mathcal{D}$ involved cannot be involutive because it is finite type, that is $g_{q+r} = 0$ for a certain integer $r \geq 0$ as we must have an exact sequence $0 \to \wedge^{(n-1)} T^* \otimes g_{q+r-1} \to 0$ and so on. The first prolongation is obtained by adding the 9 PD equations:

$$y_{333} = 0, y_{233} = 0, y_{223} = 0, y_{222} = 0, y_{133} = 0,$$
$$y_{123} - y_{111} = 0, y_{122} = 0, y_{113} = 0, y_{112} = 0$$

and the second prolongation is obtained by adding the 15 PD equations $y_{ijkl} = 0$. We obtain therefore $dim(g_2) = 6 - 3 = 3$, $dim(g_3) = 1$, $g_4 = 0$. Nevertheless, the interesting fact is that g_3 is 2-acyclic without being 3-acycic and thus involutive. Indeed, we have the δ-sequences:

$$0 \to \wedge^2 T^* \otimes g_3 \overset{\delta}{\to} \wedge^3 T^* \otimes g_2 \to 0, \quad 0 \to \wedge^3 T^* \otimes g_3 \to 0$$

Using the letter v for the symbol coordinates, the mapping δ on the left is defined by:

$$v_{111,23} + v_{112,31} + v_{113,12} = v_{11,123},$$
$$v_{121,23} + v_{122,31} + v_{123,12} = v_{12,123},$$
$$v_{131,23} + v_{132,31} + v_{133,12} = v_{13,123}$$

that is to say $v_{111,23} = v_{11,23}$, $v_{111,12} = v_{12,123}$, $v_{111,31} = v_{13,123}$. The corresponding δ-map is thus injective *and* surjective, that is g_3 is 2-acyclic but cannot be also 3-acyclic because of the inequality, $dim\left(\wedge^3 T^* \otimes g_3\right) = dim(g_3) = 1 \neq 0$. The above sequence is thus very far from being a Janet sequence and we cannot compare it with the Spencer sequence.

- *SECOND STEP* In the example of Macaulay, we have at once $dim(R_2) = 7$ with the 7 parametric jets $(y, y_1, y_2, y_3, y_{11}, y_{12}, y_{13})$ and thus $dim(R_4) = dim(R_3) = 7 + 1 = 8 = 2^3$ with the only additional third order parametric jet (y_{111}). We notice that, when $n = 2$, the new system R_2 defined by $y_{22} = 0$, $y_{12} - y_{11} = 0$ is also finite type with $y_{ijr} = 0$ and thus $dim(R_3) = dim(R_2) = 4 = 2^2$ and we invite the reader to treat directly such an elementary example as an exercise and to compare (see [25] for this striking result on the powers of 2). Therefore, instead of starting with the previous second order operator $\mathcal{D}_1$ defined by R_2, we may now start afresh with the new third order operator $\mathcal{D}_1$ defined by R_3 which is not involutive again. We let the reader check as a tricky exercise or using computer algebra that one may obtain "*necessarily*" the following finite length differential sequence which is far from being a Janet sequence but for other reasons.

$$0 \to \Theta \to E \underset{3}{\overset{\mathcal{D}}{\to}} F_0 \underset{1}{\overset{\mathcal{D}_1}{\to}} F_1 \underset{2}{\overset{\mathcal{D}_2}{\to}} F_2 \underset{1}{\overset{\mathcal{D}_3}{\to}} F_3 \underset{1}{\overset{\mathcal{D}_4}{\to}} F_4 \underset{1}{\overset{\mathcal{D}_5}{\to}} F_5 \to 0$$

$$0 \to \Theta \to 1 \underset{3}{\overset{\mathcal{D}}{\to}} 12 \underset{1}{\overset{\mathcal{D}_1}{\to}} 21 \underset{2}{\overset{\mathcal{D}_2}{\to}} 46 \underset{1}{\overset{\mathcal{D}_3}{\to}} 72 \underset{1}{\overset{\mathcal{D}_4}{\to}} 48 \underset{1}{\overset{\mathcal{D}_5}{\to}} 12 \to 0$$

and we check that $1-12+21-46+72-48+12=0$. As g_3 is 2-acyclic, the third order operator $\mathcal{D}$ has a CC operator $\mathcal{D}_1$ of order 1 having a CC operator $\mathcal{D}_2$ of order 2 which is involutive, *totally by chance*, and we end with the Janet sequence for $\mathcal{D}_2$. Such a situation is the only one we have met during the last... 40 years !. (see [15], p 119-126 for more details).

- *THIRD STEP* We may finally start with the new operator $\mathcal{D}$ defined by the involutive system R_4 with symbol $g_4=0$. The following "*fundamental diagram I*" only depends on its left commutative square and $C_0=R_4$. Each horizontal sequence is formally exact and can be constructed step by step. The interest is that we have $C_r=\wedge^r T^*\otimes C_0$ because $g_4=0$. It is nevertheless, even today, not so well known that the three differential sequences appearing in this diagram can be constructed "*step by step*" or "*as a whole*" ([1] [4] [5] [6]). Accordingly, the reader not familiar with the formal theory of systems of PD equations may find difficult to deal with the following definitions of the Spencer bundles $C_r\subset C_r(E)$ and Janet bundles F_r for an involutive system $R_q\subset J_q(E)$ of order q over E:

$$\boxed{\begin{gathered} C_r=\wedge^r T^*\otimes R_q/\delta\left(\wedge^{r-1}T^*\otimes g_{q+1}\right)\\ C_r(E)=\wedge^r T^*\otimes J_q(E)/\delta\left(\wedge^{r-1}T^*\otimes S_{q+1}T^*\otimes E\right)\\ F_r=\wedge^r T^*\otimes J_q(E)/\left(\wedge^r T^*\otimes R_q+\delta\left(\wedge^{r-1}T^*\otimes S_{q+1}T^*\otimes E\right)\right)\end{gathered}}$$

For this reason, we prefer to use successive compatibility conditions, starting from the commutative square $D=\Phi\circ j_4$ on the left of the next diagram. The Janet tabular of the Macaulay system and its prolongations up to order 4 can be decomposed as follows ([26]):

$$\left\{\begin{array}{llll} 1 & \text{PDE} & \text{order } 4 & \text{class } 3\\ 4 & \text{PDE} & \text{order } 4 & \text{class } 2\\ 10 & \text{PDE} & \text{order } 4 & \text{class } 1\\ 9 & \text{PDE} & \text{order } 3 & \\ 3 & \text{PDE} & \text{order } 2 & \end{array}\right.\left|\begin{array}{ccc} 1 & 2 & 3\\ 1 & 2 & \bullet\\ 1 & \bullet & \bullet\\ \bullet & \bullet & \bullet\\ \bullet & \bullet & \bullet \end{array}\right|$$

The total number of different single "dots" provides the $4+20+27+9=60$ CC $\mathcal{D}_1$.

The total number of different couples of "dots" provides the $10+27+9=46$ CC $\mathcal{D}_2$.

The total number of different triples of "dots" provides the $9+3=12$ CC $\mathcal{D}_3$.

We obtain therefore the fiber dimensions of the successive Janet bundles in the Janet sequence.

The same procedure can be applied to the Spencer bundles in the Spencer sequence by introducing the new 8 parametric jet indeterminates:

$$z^1=y,z^2=y_1,z^3=y_2,z^4=y_3,z^5=y_{11},z^6=y_{12},z^7=y_{13},z^8=y_{111}$$

in the first order system defined by 24 PD equations (8 of class 3 + 8 of class 2 +

8 of class 1):

$$z_1^1 - z^2 = 0, z_2^1 - z^3 = 0, z_3^1 - z^4 = 0, \cdots, z_1^5 - z^8 = 0, \cdots,$$
$$z_3^6 - z^8 = 0, \cdots, z_3^7 = 0, \cdots, z_3^8 = 0$$

$$\begin{array}{ccccccccccccc}
 & & & & 0 & & 0 & & 0 & & 0 & & \\
 & & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
 & 0 & \rightarrow & \Theta & \xrightarrow{j_4} & C_0 & \xrightarrow{D_1} & C_1 & \xrightarrow{D_2} & C_2 & \xrightarrow{D_3} & C_3 & \rightarrow 0 \\
 & & & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
 & 0 & \rightarrow & E & \xrightarrow{j_4} & C_0(E) & \xrightarrow{D_1} & C_1(E) & \xrightarrow{D_2} & C_2(E) & \xrightarrow{D_3} & C_3(E) & \rightarrow 0 \\
 & & & \| & & \downarrow \Phi_0 & & \downarrow \Phi_1 & & \downarrow \Phi_2 & & \downarrow \Phi_3 & \\
0 \rightarrow & \Theta & \rightarrow & E & \xrightarrow{\mathcal{D}} & F_0 & \xrightarrow{\mathcal{D}_1} & F_1 & \xrightarrow{\mathcal{D}_2} & F_2 & \xrightarrow{\mathcal{D}_3} & F_3 & \rightarrow 0 \\
 & & & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
 & & & & & 0 & & 0 & & 0 & & 0 &
\end{array}$$

$$\begin{array}{ccccccccccccc}
 & & & & & 0 & & 0 & & 0 & & 0 & \\
 & & & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
 & 0 & \rightarrow & \Theta & \xrightarrow{j_4} & 8 & \xrightarrow{D_1} & 24 & \xrightarrow{D_2} & 24 & \xrightarrow{D_3} & 8 & \rightarrow 0 \\
 & & & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
 & 0 & \rightarrow & 1 & \xrightarrow{j_4} & 35 & \xrightarrow{D_1} & 84 & \xrightarrow{D_2} & 70 & \xrightarrow{D_3} & 20 & \rightarrow 0 \\
 & & & \| & & \downarrow \Phi_0 & & \downarrow \Phi_1 & & \downarrow \Phi_2 & & \downarrow \Phi_3 & \\
0 \rightarrow & \Theta & \rightarrow & 1 & \xrightarrow{\mathcal{D}} & 27 & \xrightarrow{\mathcal{D}_1} & 60 & \xrightarrow{\mathcal{D}_2} & 46 & \xrightarrow{\mathcal{D}_3} & 12 & \rightarrow 0 \\
 & & & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
 & & & & & 0 & & 0 & & 0 & & 0 &
\end{array}$$

The morphisms Φ_1, Φ_2, Φ_3 in the vertical short exact sequences are inductively induced from the morphism $\Phi_0 = \Phi$ in the first short exact vertical sequence on the left. The central horizontal sequence can be called "*hybrid sequence*" because it is at the same time a Spencer sequence for the first order system $J_5(E) \subset J_1(J_4(E))$ over $J_4(E)$ and a Janet sequence for the involutive injective operator $j_4 : E \rightarrow J_4(E)$. It can be constructed step by step, starting with the short exact sequence:

$$0 \rightarrow J_5(E) \rightarrow J_1(J_4(E)) \rightarrow C_1(E) \rightarrow 0$$

$$0 \rightarrow 56 \rightarrow 140 \rightarrow 84 \rightarrow 0$$

In actual practice, as the system $R_2 \subset J_2(E)$ is homogeneous, it is thus formally integrable and finite type because the system $R_4 = \rho_2(R_2) = ker(\Phi) \subset J_4(E)$ is trivially involutive with a symbol $g_4 = 0$. Accordingly, $\mathcal{D} = \Phi \circ j_4$ is an involutive operator of order 4 and we obtain a finite length Janet sequence which is formally exact both on the jet level and on the symbol level, that can only contain the successive first order operators $\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3$. For example, one can determine $\mathcal{D}_2 = \Psi_2 \circ j_1 : F_1 \rightarrow F_2$ just by counting the dimensions, either in the long exact jet sequence:

$$0 \to R_6 \to J_6(E) \to J_2(F_0) \to J_1(F_1) \xrightarrow{\Psi_2} F_2 \to 0$$
$$0 \to 8 \to 84 \to 270 \to 240 \to dim(F_2) \to 0$$

and obtain $dim(F_2) = -8 + 84 - 270 + 240 = 46$.

However, one can also use the fact that $dim(E) = 1$ and $g_4 = 0 \Rightarrow g_6 = 0$ while introducing the restriction $\sigma(\Psi_2)$ of Ψ_2 to $T^* \otimes F_1 \subset J_1(F_1)$ in the long exact symbol sequence:

$$0 \to S_6T^* \to S_2T^* \otimes F_0 \to T^* \otimes F_1 \xrightarrow{\sigma(\Psi_2)} F_2 \to 0$$
$$0 \to 28 \to 162 \to 180 \to dim(F_2) \to 0$$

in order to obtain again $dim(F_2) = 28 - 162 + 180 = 46$.

We wish good luck to anybody using Computer Algebra because one should have to deal with a matrix 540×600 in order to describe the prolongation morphism $J_3(F_0) \to J_2(F_1)$. Nevertheless, in order to give a hint, we recall the vanishing of the Euler-Poincaré characteristic as we can check successively:

$$8 - 24 + 24 - 8 = 0, -1 + 35 - 84 + 70 - 20 = 0, -1 + 27 - 60 + 46 - 12 = 0$$

In the case of finite type systems, the usefulness of the Spencer sequence is so evident, like on such an example, that it needs no comment.

We invite the reader to treat separately but similarly the system:

$$y_{33} - y_{11} = 0, y_{23} = 0, y_{22} - y_{11} = 0$$

and to compare the various extension modules.

3. Solution

According to the previous sections, it only remains to consider the two cases $n = 1$ and $n = 2$. For simplicity, we shall only consider the situation of the Euclidean metric and the corresponding linear systems. We let the reader treat by himself the nonlinear counterparts.

- *CASE* $n = 1$

With $\omega \neq 0$, we may consider a section $\xi_3 = (\xi(x), \xi_x(x), \xi_{xx}(x), \xi_{xxx})$ and introduce the classical Killing system $R_1 \subset J_1(T)$ by means of the *formal Lie derivative*:

$$\Omega \equiv L(\xi_1)\omega \equiv 2\omega\xi_x + \xi\partial_x\omega = 0$$

Similarly, with the Christoffel symbol $\gamma = \frac{1}{2\omega}\partial_x\omega$, we may consider:

$$\Gamma \equiv L(\xi_2)\gamma \equiv \xi_{xx} + \gamma\xi_x + \xi\partial_x\gamma = 0$$

The conformal Killing system can be defined with a conformal factor as:

$$\Omega \equiv L(\xi_1)\omega \equiv 2\omega\xi_x + \xi\partial_x\omega = 2A(x)\omega$$

and its first prolongation becomes:

$$\Gamma \equiv L(\xi_2)\gamma \equiv \xi_{xx} + \gamma\xi_x + \xi\partial_x\gamma = A_x(x)$$

The elimination of $A(x)$ or $A_x(x)$ does not provide any OD equation of order 1 or 2. Moreover, we let the reader check that $\xi_2 = j_2(\xi) \Rightarrow \partial_x A(x) - A_x(x) = 0$ as a way to understand the part plaid by the Spencer operator and the reason for introducing $2A(x)$. With more details, dividing the Killing system by 2ω, we get $\xi_x + \gamma\xi = A(x)$. Differentiating this OD equation, we get:

$$\partial_x \xi_x + \gamma \partial_x \xi + \partial_x \gamma \xi = \partial_x A(x)$$

and we just need to substract the OD equation $\Gamma = A_x(x)$ in order to get:

$$(\partial_x \xi_x - \xi_{xx}) + \gamma(\partial_x \xi - \xi_x) = \partial_x A(x) - A_x(x)$$

In order to escape from the previous situation while having a vanishing symbol $g_3 = 0$, we may consider the new *unusual prolongation*:

$$\xi_{xxx} + \gamma\xi_{xx} + 2(\partial_x \gamma)\xi_x + \xi \partial_{xx}\gamma = 0$$

and substract the second order OD equation $\Gamma = 0$ multiplied by γ while introducing the new geometric object $\nu = \partial_x \gamma - \frac{1}{2}\gamma^2$ in order to obtain the third order infinitesimal Lie equation:

$$L(\xi_3)\nu \equiv \xi_{xxx} + 2\nu\xi_x + \xi\partial_x \nu = 0$$

The nonlinear framework, not known today because the work of Vessiot is still not acknowledged, explains the successive inclusions $\gamma \in j_1(\omega), \nu \in j_1(\gamma)$. Indeed, if we consider the *translation group* $(y = x + a, a = cst)$ and the bigger *isometry group* $(y = x + a, y = -x + a, a = cst)$, the inclusion of groups of the real line:

$$\textit{translation group} \subset \textit{isometry group} \subset \textit{affine group} \subset \textit{projective group}$$

with the respective finite Lie equations in Lie form with the jet coordinates $(x, y, y_x, y_{xx}, y_{xxx})$:

$$\alpha(y)y_x = \alpha(x),\ \omega(y)(y_x)^2 = \omega(x),\ \frac{y_{xx}}{y_x} + \gamma(y)y_x = \gamma(x),$$

$$\frac{y_{xxx}}{y_x} - \frac{3}{2}\left(\frac{y_{xx}}{y_x}\right)^2 + \nu(y)(y_x)^2 = \nu(x)$$

where we recognize the Schwarzian third order differential invariant of the *projective group*.

Of course, we have $\alpha = 1 \Rightarrow \omega = 1 \Rightarrow \gamma = 0 \Rightarrow \nu = 0$ and the respective linearizations:

$$y_x = 1 \Rightarrow \xi_x = 0,\ y_{xx} = 0 \Rightarrow \xi_{xx} = 0,\ \frac{y_{xxx}}{y_x} - \frac{3}{2}\left(\frac{y_{xx}}{y_x}\right)^2 = 0 \Rightarrow \xi_{xxx} = 0$$

The Janet tabular of the conformal system order 3 can be decomposed as follows:

$$\{1 \quad \text{PDE} \quad \text{order } 3 \quad \text{class } 1 \ \boxed{1}$$

The total number of different single "dots" provides the 0 CC $\mathcal{D}_1$.

We obtain therefore the fiber dimensions of the successive Janet bundles in the Janet sequence.

The same procedure can be applied to the other canonical differential sequences.

When $n=1$, one has 3 parameters (1 translation + 1 dilatation + 1 elation) and we get the following "*fundamental diagram I*" only depending on the left commutative square:

$$\begin{array}{ccccccccccc}
 & & & & & & 0 & & 0 & & \\
 & & & & & & \downarrow & & \downarrow & & \\
 & & 0 & \rightarrow & \Theta & \xrightarrow{j_3} & 3 & \xrightarrow{D_1} & 3 & \rightarrow 0 & \textit{Spencer} \\
 & & & & & & \downarrow & & \| & & \\
 & & 0 & \rightarrow & 1 & \xrightarrow{j_3} & 4 & \xrightarrow{D_1} & 3 & \rightarrow 0 & \\
 & & & & \| & & \downarrow\Phi & & \downarrow & & \\
0 \rightarrow & & \Theta & \rightarrow & 1 & \xrightarrow{\mathcal{D}} & 1 & \rightarrow & 0 & & \textit{Janet} \\
 & & & & & & \downarrow & & & & \\
 & & & & & & 0 & & & &
\end{array}$$

In this diagram, the operator $j_3:\xi(x)\rightarrow\left(\xi(x)=\xi(x),\partial_x\xi(x)=\xi_x(x),\partial_{xx}\xi(x)=\xi_{xx}(x),\partial_{xxx}\xi(x)=\xi_{xxx}(x)\right)$ has compatibility conditions $D_1\xi_3=0$ induced by d and the space of solutions Θ of $D=\Phi\circ j_3:\xi(x)\rightarrow\partial_{xxx}\xi(x)$ is generated over the constants by the three infinitesimal generators:

$$\theta_1=\partial_x \quad (\text{translation}),\quad \theta_2=x\partial_x \quad (\text{dilatation}),\quad \theta_3=\frac{1}{2}x^2\partial_x \quad (\text{elation})$$

of the action and coincides with the projective group of the real line.

- *CASE* $n=2$

The classical approach is to consider the infinitesimal conformal Killing system for $n=2$ and eliminate the infinitesimal conformal factor $2A(x)$ as follows by introducing the *formal* and the *effective* Lie derivatives such that $L(j_1(\xi))=\mathcal{L}(\xi)$:

$$\Omega\equiv L(\xi_1)\omega=2A(x)\omega\Rightarrow\xi_1^1=A(x),\xi_2^1+\xi_1^2=0,\xi_2^2=A(x)$$
$$\Rightarrow\xi_2^2-\xi_1^1=0,\xi_2^1+\xi_1^2=0$$

that is to say the elimination of A is just producing locally the two well known Cauchy-Riemann equations allowing to define infinitesimal complex transformations of the plane, that is to say an infinite dimensional Lie pseudogroup which is by no way providing a finite dimensional Lie group. As such an operator has no *compatibility condition* (CC), we obtain by one prolongation $2\times2=4$ second order equations but another prolongation does not provide a zero symbol at order 3 and it is just such a delicate step that we have to overcome by adding $2\times4=8$ homogeneous third order PD equations. The *only possibility* is to consider the following system and to prove that it is defining a system of infinitesimal Lie equations leading to $2\times(1+2+3+4)-(2+4+8)=20-14=6$ infinitesimal generators.

$$\begin{cases}\xi_{ijr}^{k}=0\\ \xi_{22}^{2}-\xi_{12}^{1}=0,\xi_{22}^{1}+\xi_{12}^{2}=0,\xi_{12}^{2}-\xi_{11}^{1}=0,\xi_{12}^{1}+\xi_{11}^{2}=0\\ \xi_{2}^{2}-\xi_{1}^{1}=0,\xi_{2}^{1}+\xi_{1}^{2}=0\end{cases}$$

where the 4 second order PD equations can also be rewritten with $\Delta=d_{11}+d_{22}$ as:

$$\begin{aligned}&\Delta\xi^{2}\equiv\xi_{22}^{2}+\xi_{11}^{2}=0,\\ &\Delta\xi^{1}\equiv\xi_{22}^{1}+\xi_{11}^{1}=0,\\ &\xi_{12}^{2}-\xi_{11}^{1}=0,\\ &\xi_{12}^{1}+\xi_{11}^{2}=0\end{aligned}$$

The general solution of the 8 third order PD equations can be written with 12 arbitrary constant parameters as:

$$\xi^{1}=\frac{1}{2}a\left(x^{1}\right)^{2}+bx^{1}x^{2}+\frac{1}{2}c\left(x^{2}\right)^{2}+dx^{1}+ex^{2}+f$$

$$\xi^{2}=\frac{1}{2}\bar{a}\left(x^{1}\right)^{2}+\bar{b}x^{1}x^{2}+\frac{1}{2}\bar{c}\left(x^{2}\right)^{2}+\bar{d}x^{1}+\bar{e}x^{2}+g$$

Taking into account the first and second order PD equations, we must have the relations:

$$\bar{b}=a,\bar{c}=b,\bar{a}+b=0,\bar{b}+c=0,\bar{e}=d,\bar{d}+e=0$$

and the final number of parameters is indeed reduced to $2+1+1+2=6$ arbitrary parameters. Collecting the above results, we obtain the 6 infinitesimal generators:

$$a\to\frac{1}{2}\left(\left(x^{1}\right)^{2}-\left(x^{2}\right)^{2}\right)\partial_{1}+x^{1}x^{2}\partial_{2}$$

$$b\to x^{1}x^{2}\partial_{1}+\frac{1}{2}\left(\left(x^{2}\right)^{2}-\left(x^{1}\right)^{2}\right)\partial_{2}$$

$$-e\to x^{1}\partial_{2}-x^{2}\partial_{1},d\to x^{1}\partial_{1}+x^{2}\partial_{2}$$

$$f\to\partial_{1},g\to\partial_{2}$$

We find back the two infinitesimal generators of the elations, namely:

$$\begin{aligned}\theta_{1}&=-\frac{1}{2}\left(\left(x^{1}\right)^{2}+\left(x^{2}\right)^{2}\right)\partial_{1}+x^{1}\left(x^{1}\partial_{1}+x^{2}\partial_{2}\right)\\ &=\frac{1}{2}\left(\left(x^{1}\right)^{2}-\left(x^{2}\right)^{2}\right)\partial_{1}+x^{1}x^{2}\partial_{2}\end{aligned}$$

and θ_2 obtained by exchanging x^1 with x^2.

Contrary to the situation met when $n\geq 3$ where one starts with a groupoid of order 1 and obtains groupoids of order 2 or 3 after one or two prolongations, in the present situation, we have to check directly the commutation relations for the six infinitesimal generators already found, namely:

$$\left[\partial_1,\theta_1\right]=x^1\partial_1+x^2\partial_2,\left[\partial_2,\theta_1\right]$$
$$=x^1\partial_2-x^2\partial_1$$
$$\left[x^1\partial_2-x^2\partial_1,\theta_1\right]=-\theta_2,\left[x^1\partial_1+x^2\partial_2,\theta_1\right]$$
$$=\theta_1,\left[\theta_1,\theta_2\right]$$
$$=0$$

We have thus obtained in an unexpected way the desired 2 translations, 1 rotation, 1 dilatation and 2 elations of the conformal group when $n=2$.

At order one, we may consider the classical Killing system R_1 obtained by preserving ω, the Weyl system $\tilde{R}_1$ and the conformal system $\hat{R}_1$ with $R_1 \subset \tilde{R}_1 = \hat{R}_1 \subset J_1(T)$ and $dim\left(\tilde{R}_1/R_1\right)=1$. At order two, we have the strict inclusions $R_2 \subset \tilde{R}_2 \subset \hat{R}_2$ with $R_2=\rho_1(R_1)$ preserving $(\omega,\gamma)\simeq j_1(\omega)$, $\tilde{R}_2 \subset \rho_1\left(\tilde{R}_1\right)$ obtained by preserving $(\hat{\omega},\gamma)$ and $\hat{R}_2=\rho_1\left(\hat{R}_1\right)$ obtained by preserving $(\hat{\omega},\hat{\gamma})\simeq j_1(\hat{\omega})$. The main difference with the case $n\geq 3$ is that *now* $R_3=\rho_2(R_1)$ has a symbol $g_3=0$, $\tilde{R}_3=\rho_1\left(\tilde{R}_2\right)$ has also a symbol $\tilde{g}_3=0$ *but* that $\hat{R}_3 \subset \rho_1\left(\hat{R}_2\right)$ with strict inclusion in order to have *now* $\hat{g}_3=0$, even though $\rho_1(\hat{g}_2)\neq 0$. However, we are now able to deal with three *trivially involutive* systems having zero symbols and we have the strict inclusions $R_3 \subset \tilde{R}_3 \subset \hat{R}_3$ with respective dimensions $3<4<6$ according to the basic inequalities $n(n+1)/2<\left(n(n+1)/2\right)+1<(n+1)(n+2)/2$ valid in arbitrary dimension $n\geq 1$. The interest of this result is that we have for the Spencer bundles the strict inclusions $C_0 \subset \tilde{C}_0 \subset \hat{C}_0$ of the zero Spencer bundles, leading to the strict inclusions of the respective linear Spencer sequences because:

$$g_3=\tilde{g}_3=\hat{g}_3=0 \Rightarrow C_r=\wedge^r T^*\otimes C_0,\tilde{C}_r=\wedge^r T^*\otimes \tilde{C}_0,$$
$$\hat{C}_r=\wedge^r T^*\otimes \hat{C}_0 \Rightarrow C_r \subset \tilde{C}_r \subset \hat{C}_r$$

in agrement with many recent results ([21] [22] [23] [24]). As in Example 2.2, we let the reader introduce the 6 parametric jet indeterminates
$z^1=y^1,z^2=y^2,z^3=y_1^1,z^4=y_1^2,z^5=y_{11}^1,z^6=y_{11}^2$.

The Janet tabular of the conformal Killing system and its prolongations up to order 3 can be decomposed as follows:

$$\left\{\begin{array}{llll} 2 & \text{PDE} & \text{order } 3 & \text{class } 2 \\ 6 & \text{PDE} & \text{order } 3 & \text{class } 1 \\ 4 & \text{PDE} & \text{order } 2 & \\ 2 & \text{PDE} & \text{order } 1 & \end{array}\right. \quad \begin{array}{|cc|} 1 & 2 \\ 1 & \bullet \\ \bullet & \bullet \\ \bullet & \bullet \\ \hline \end{array}$$

The total number of different single "dots" provides the $6+8+4=18$ CC $\mathcal{D}_1$.

The total number of different couples of "dots" provides the $4+2=6$ CC $\mathcal{D}_2$.

We obtain therefore the fiber dimensions of the successive Janet bundles in the Janet sequence.

The same procedure can be applied to the other canonical differential sequences.

When $n=2$, one has 6 parameters (2 translations + 1 rotation + 1 dilatation

+ 2 elations) and we get the following "*fundamental diagram I*" only depending on the left commutative square:

$$
\begin{array}{ccccccccccccc}
 & & & & & 0 & & 0 & & 0 & & & \\
 & & & & & \downarrow & & \downarrow & & \downarrow & & & \\
 & 0 & \to & \Theta & \xrightarrow{j_3} & 6 & \xrightarrow{D_1} & 12 & \xrightarrow{D_2} & 6 & \to 0 & & \textit{Spencer} \\
 & & & & & \downarrow & & \downarrow & & \downarrow & & & \\
 & 0 & \to & 2 & \xrightarrow{j_3} & 20 & \xrightarrow{D_1} & 30 & \xrightarrow{D_2} & 12 & \to 0 & & \\
 & & & \| & & \downarrow \Phi_0 & & \downarrow \Phi_1 & & \downarrow \Phi_2 & & & \\
0 \to & \Theta & \to & 2 & \xrightarrow{\mathcal{D}} & 14 & \xrightarrow{\mathcal{D}_1} & 18 & \xrightarrow{\mathcal{D}_2} & 6 & \to 0 & & \textit{Janet} \\
 & & & & & \downarrow & & \downarrow & & \downarrow & & & \\
 & & & & & 0 & & 0 & & 0 & & &
\end{array}
$$

- *CASE* $n = 3$

The Janet tabular of the conformal Killing system and its prolongations up to order 3 can be decomposed as follows:

$$
\left\{
\begin{array}{llll}
3 & \text{PDE} & \text{order } 3 & \text{class } 3 \\
9 & \text{PDE} & \text{order } 3 & \text{class } 2 \\
18 & \text{PDE} & \text{order } 3 & \text{class } 1 \\
15 & \text{PDE} & \text{order } 2 & \\
5 & \text{PDE} & \text{order } 1 &
\end{array}
\right.
\quad
\begin{array}{|ccc|}
\hline
1 & 2 & 3 \\
1 & 2 & \bullet \\
1 & \bullet & \bullet \\
\bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet \\
\hline
\end{array}
$$

The total number of different single "dots" provides the $9+36+45+15=105$ CC $\mathcal{D}_1$.

The total number of different couples of "dots" provides the $18+45+15=78$ CC $\mathcal{D}_2$.

The total number of different triples of "dots" provides the $15+5=20$ CC $\mathcal{D}_3$.

We obtain therefore the fiber dimensions of the successive Janet bundles in the Janet sequence.

The same procedure can be applied to the other canonical differential sequences and we get the desired "*fundamental diagram I*" below:

$$
\begin{array}{cccccccccccccc}
 & & & & & 0 & & 0 & & 0 & & 0 & \\
 & & & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
 & 0 & \to & \Theta & \xrightarrow{j_3} & 10 & \xrightarrow{D_1} & 30 & \xrightarrow{D_2} & 30 & \xrightarrow{D_3} & 10 & \to 0 \\
 & & & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
 & 0 & \to & 3 & \xrightarrow{j_3} & 60 & \xrightarrow{D_1} & 135 & \xrightarrow{D_2} & 108 & \xrightarrow{D_3} & 30 & \to 0 \\
 & & & \| & & \downarrow \Phi_0 & & \downarrow \Phi_1 & & \downarrow \Phi_2 & & \downarrow \Phi_3 & \\
0 \to & \Theta & \to & 3 & \xrightarrow{\mathcal{D}} & 50 & \xrightarrow{\mathcal{D}_1} & 105 & \xrightarrow{\mathcal{D}_2} & 78 & \xrightarrow{\mathcal{D}_3} & 20 & \to 0 \\
 & & & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
 & & & & & 0 & & 0 & & 0 & & 0 &
\end{array}
$$

We have 10 parameters (3 translations, 3 rotations, 1 dilataion, 3 elations).

The computation of $dim(C_3(E)) = 30$ needs to determine the rank of a 1200×1350 matrix !

4. Conclusion

We have shown that the true important specific property of the conformal group, at least for applications to physics, is that, even if it is defined as a specific Lie pseudogroup of transformations, it is in fact a Lie group of transformations with a finite number $(n+1)(n+2)/2$ of parameters or infinitesimal generators when $n \geq 1$. Accordingly, in dimension $n = 1$, we have no OD equation of order 1 *and* 2, a result leading therefore to add 1 unexpected OD equation of order 3. Similarly, when $n = 2$, we obtain the Cauchy-Riemann PD equations defining an infinite dimensional Lie pseudogroup and we have therefore to add, again in a totally unexpected way, as many third order PD equations as the number of jet coordinates of strict order 3. When $n = 3$, the fact that the analogue of the Weyl operator for describing the CC of the conformal operator is of order 3 is rather un-pleasant but this is nothing compared to the fact that, when $n = 4$, the analogue of the Bianchi operator for describing the CC of the previous second order CC playing the part of the Weyl CC is of order 2 *again*. And we don't speak about the case $n = 5$ ([9] [15]). Though these results can be checked by means of computer algebra and are confirmed by the use of the fundamental diagram I, they do not seem to be known today. Accordingly, any *physical theory* (existence of gravitational waves or black holes...) which is not coherent with *differential homological algebra* (vanishing of the first and second extension modules for the Poincaré sequence in the previous examples...) must be revisited in the light of these new *mathematical tools*, even if it seems apparently well established ([8] [27] [28] [29] [30]).

Conflicts of Interest

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

References

[1] Pommaret, J.-F. (1978) Systems of Partial Differential Equations and Lie Pseudogroups. Gordon and Breach, New York. (Russian Translation: MIR, Moscow, 1983)

[2] Pommaret, J.-F. (1983) Differential Galois Theory. Gordon and Breach, New York.

[3] Pommaret, J.-F. (1988) Lie Pseudogroups and Mechanics. Gordon and Breach, New York.

[4] Pommaret, J.-F. (1994) Partial Differential Equations and Group Theory. Kluwer, Dordrecht. https://doi.org/10.1007/978-94-017-2539-2

[5] Pommaret, J.-F. (2001) Partial Differential Control Theory. Kluwer, Dordrecht. https://doi.org/10.1007/978-94-010-0854-9

[6] Pommaret, J.-F. (2001) Algebraic Analysis of Control Systems Defined by Partial Differential Equations. In: *Advanced Topics in Control Systems Theory*, Springer,

Berlin, Lecture Notes in Control and Information Sciences 311, Chapter 5, 155-223. https://doi.org/10.1007/11334774_5

[7] Pommaret, J.-F. (2010) *Acta Mechanica*, **215**, 43-55. https://doi.org/10.1007/s00707-010-0292-y

[8] Pommaret, J.-F. (2013) *Journal of Modern Physics*, **4**, 223-239. https://doi.org/10.4236/jmp.2013.48A022

[9] Pommaret, J.-F. (2O16) *Journal of Modern Physics*, **7**, 699-728. https://doi.org/10.4236/jmp.2016.77068

[10] Pommaret, J.-F. (2019) *Journal of Modern Physics*, **10**, 371-401. https://doi.org/10.4236/jmp.2019.103025

[11] Pommaret, J.-F. (2019) *Journal of Modern Physics*, **10**, 1454-1486. https://doi.org/10.4236/jmp.2019.1012097

[12] Spencer, D.C. (1965) *Bulletin of the American Mathematical Society*, **75**, 1-114.

[13] Eisenhart, L.P. (1926) Riemannian Geometry. Princeton University Press, Princeton. https://doi.org/10.1090/coll/008

[14] Vessiot, E. (1903) *Annales Scientifiques de l'École Normale Supérieure*, **20**, 411-451. https://doi.org/10.24033/asens.529

[15] Pommaret, J.-F. (2016) Deformation Theory of Algebraic and Geometric Structures. Lambert Academic Publisher (LAP), Saarbrucken. https://arxiv.org/abs/1207.1964 https://doi.org/10.1007/BFb0083506

[16] 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. https://doi.org/10.4236/jmp.2016.77068

[17] Northcott, D.G. (1966) An Introduction to Homological Algebra. Cambridge University Press, Cambridge.

[18] Rotman, J.J. (1979) An Introduction to Homological Algebra. Academic Press, Cambridge.

[19] Gasqui, J. and Goldschmidt, H. (1984) Déformations Infinitésimales des Structures Conformes Plates. Progress in Mathematics, Vol. 52, Birkhauser, Boston.

[20] Kumpera, A. and Spencer, D.C. (1972) Lie Equations. Ann. Math. Studies 73, Princeton University Press, Princeton. https://doi.org/10.1515/9781400881734

[21] Björk, J.-E. (1993) Analytic *D*-Modules and Applications. Mathematics and Its Applications, Vol. 247. Kluwer, Dordrecht. https://doi.org/10.1007/978-94-017-0717-6

[22] Kashiwara, M. (1995) Algebraic Study of Systems of Partial Differential Equations. Mémoires de la Société Mathématique de France, 63. (Transl. from Japanese of His 1970 Master's Thesis)

[23] Schneiders, J.-P. (1994) *Bulletin de la Société Royale des Sciences de Liège*, **63**, 223-295.

[24] Pommaret, J.-F. (2015) *Multidimensional Systems and Signal Processing*, **26**, 405-437. https://doi.org/10.1007/s11045-013-0265-0

[25] Macaulay, F.S. (1916) The Algebraic Theory of Modular Systems, Cambridge Tracts, Vol. 19. Cambridge University Press, London. Stechert-Hafner Service Agency, New-York, 1964. https://doi.org/10.3792/chmm/1263317740

[26] Janet, M. (1920) *Journal de Mathematiques*, **8**, 65-151.

[27] Pommaret, J.-F. (2017) *Journal of Modern Physics*, **8**, 2122-2158. https://doi.org/10.4236/jmp.2017.813130

[28] Pommaret, J.-F. (2018) New Mathematical Methods for Physics. Mathematical Physics Books, Nova Science Publishers, New York, 150 p.

[29] Pommaret, J.-F. (2020) Nonlinear Conformal Electromagnetism and Gravitation. https://arxiv.org/abs/2007.01710

[30] Pommaret, J.-F. (2021) *Journal of Modern Physics*, **12**, 453-482. https://doi.org/10.4236/jmp.2021.124032

Journal of Modern Physics, 2021, 12, 1843-1865
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Higher Order Periodic Base Pairs Opening in a Finite Stacking Enthalpy DNA Model

Nkeh Oma Nfor

Department of Physics, Higher Teacher Training College Bambili, The University of Bamenda, Bambili, Cameroon
Email: omnkeh@gmail.com, nfor.oma@uniba.cm

How to cite this paper: Nfor, N.O. (2021) Higher Order Periodic Base Pairs Opening in a Finite Stacking Enthalpy DNA Model. *Journal of Modern Physics*, **12**, 1843-1865.
https://doi.org/10.4236/jmp.2021.1213107

Received: October 26, 2021
Accepted: November 26, 2021
Published: November 29, 2021

Abstract

The dynamics of periodic base pairs opening in a finite stacking enthalpy DNA is investigated in this work. This is achieved by using the Joyeux-Buyukdagli DNA model, in which the polynomial approximations of the stacking interaction and Morse potential are expanded up to the fifth order nonlinear terms by using the Taylor series expansion technique. By incorporating the continuum limit approximation and the extended multiple scale asymptotic methods, higher order nonlinear Schrödinger amplitude equations are derived. In the limit of cubic nonlinearity, the periodic base pair configurations clearly depict the open state; with linear stability analysis exposing other periodic background modes that are vital in the DNA transcription, replication, and transmission of genetic codes. The higher order modes generally display a more robust and structurally stable wave profile, which epitomizes the base pair dynamics of the DNA molecule observed from experimental investigations. Prolonged time evolution of base pairs stretching greatly modifies the higher order modes of the DNA molecule, strongly suggesting that such modes may induce abnormalities like gene mutation which is responsible for numerous diseases.

Keywords

Base Pairs, Finite Stacking Enthalpy, Transcription, Genetic Codes

1. Introduction

The nonlinear dynamics of deoxyribonucleic acid (DNA) remains a very fascinating and active area of research in biophysics. This is because it provides the basis for understanding intrinsic processes like transcription, replication, and transmission of genetic codes [1]-[10]. In fact, the DNA is only found in the nucleus of living cells, and has the structure of a double-stranded macromolecule in the form of a double helix [1] [5]. A mastery of the interaction between nucleotides and water molecules is a key factor in understanding the double helix structure.

Concretely, bases are insoluble in water that is hydrophobic, while sugar and phosphate form bonds with water molecules that results to sugar-phosphate backbone. This backbone is generally aligned on the surface of a cylinder, while the bases are oriented toward its center. This configuration leads to a natural protection of the bases which carries most of the genetic codes, by the sugar-phosphate backbone [11]. It is important to note that ribonucleic acid (RNA) links the DNA with protein, hence leading to the effective control of the protein bio-synthesis by the DNA in the transcription process. Experimental results point to the fact that the transcription process is inextricably linked to variation of the DNA environmental temperature [12], during the process of denaturation or melting. The fluid medium that surrounds the DNA equally enhances molecular collisions that may trigger rotational, transverse, and longitudinal oscillations of nucleotides [13] [14].

Solitons are solutions of a widespread class of weakly disperse partial differential equations, and it generally originates from the balance between nonlinearity and dispersion. The soliton concept which emanates from the Fermi Pasta Ulam (FPU) paradox [15] [16], is increasingly being used to explain the complex dynamics of neural networks [17] [18] [19] [20] [21], optical fiber systems [22] [23] [24], and the local base pairs opening of DNA [25] [26]. Concretely, the local base pairs opening of DNA can be analytically captured as breather-like modes of small amplitude. These modes have fascinating properties owing to their small amplitude, like the induction of energy trapping as the breathers move along the DNA strand [27]. Some local dis-homogeneities can equally enhance the trapping mechanism [28], which is indicative of the fact that the properties of breathers could allow the formation of the transcription bubble after the interaction with the bound RNA-polymerase. From the seminal work of Englander *et al.*, the evolution of solitonic excitations in the DNA double chain play crucial roles in the transcription process [29]. Other simplified DNA models include the Y-model introduced by Yakushevich in 1989 [30], which has been improved upon and extensively studied [31] [32] [33]. According to the model, the DNA consists of two parallel chains of discs which are connected to each other with longitudinal and transverse springs. The rigidity of the longitudinal springs is higher than that of the transverse ones as they represent the covalent and hydrogen bonds, respectively. Another interesting model is the Plane-Base Rotator (PBR) model, initially proposed by Yomosa [34] [35] and improved by Homma and Takeno [36]. A degree of freedom characterizing base rotations in the plane perpendicular to the helical axis around the backbone structure is assumed in the PBR model, while the introduced Hamiltonian is based on the

Heisenberg's spin model for the ferromagnetic chain. The Peyrard-Bishop-Dauxois (PBD) model of DNA remains a very successfully model to analyze experiments on short DNA sequences [37], and able to mimic real denaturation curves identified by Raman spectroscopy [8]. The PBD model incorporates stacking interactions between neighboring base pairs to enhance the rigidity of secondary DNA structure. However, it should be noted that this stacking interaction inherent in the PBD model does not associate any characteristic energy in the important dynamics of the DNA system [8].

The quest to incorporate finite characteristic energies with phase transition triggered numerous research activities on DNA dynamics; tailored on capturing the appropriate phase transition predicted by statistical DNA models. This finally culminated with the brilliant works of Joyeux and Buyukdagli (BJ) model of DNA [2] [3] [4], which is based on site-specific stacking enthalpies. This model is very reliable because it reproduced exact experimental curves that ensured a sharp melting transition, as a result of the finiteness of the stacking interaction [2] [3] [4]. Carlos *et al.* exploited the JB model to investigate on the dynamics of discrete breathers which is governed by the extended discrete nonlinear Schrödinger equation [38]. Depending on the finite stacking parameters, compact bright solitary waves became more robust or quickly decomposed in the JB model of DNA [39]. On the other hand, Ying-Bo Yao *et al.* demonstrated that the JB system is capable of producing high-order envelope solitons; which can be viewed as high-order discrete breathers with zero group velocity at the center of the Brillouin zone [26].

Unlike the aforementioned studies which deal with spatially localized excitations, the present investigation seeks to explore periodic solutions that may characterize base pairs opening. Periodic wave train solutions gives a better understanding of myriad of bio-physical activities hitherto explained only by localized solutions. For example, Vargas *et al.* numerically exploited localized periodic solutions in the nerve model, to rigorously explain hyper-polarization, pulse trains, and refractory periods that were experimentally observed in the nerve of locust [18] [40]. Also, the energy released during the hydrolysis of adenosine triphosphate was shown to be transported via periodic soliton wave train in order to sustain important biological processes like enzyme catalysis and muscle contractions [41]. From a physiological standpoint, gene expression is very important in life because genetic codes are regularly transferred during protein synthesis [3] [5]. The effective stimulation of all the base pairs is essential for the DNA loop formation [38], regulation of gene expressions, and packaging of DNA into nucleosomes [5]. In fact, the RNA which is a key component in the transfer, transcription, and messaging in DNA, operates in few portions of the DNA sequence at the same time [5]. Such complex motion strongly suggests that it is more appropriate for the DNA base pairs opening to be considered as a spatial periodic activity, in order to holistically comprehend the DNA dynamics. We are therefore interested on how changes in the stacking parameters can generate new spatial periodic base pairs opening profiles, that may be responsible for some physiological abnormalities. Such open states are vulnerable to many external attacks that may cause reading or coding errors, which is responsible for numerous cancers induced by gene mutations [25] [42]. To the best of my knowledge, this is one of the rare studies that seek to model base pairs opening as a spatial periodic phenomenon.

In the present work, we demonstrate that the BJ model of DNA supports spatial periodic base pairs open configurations. Consequently, the organization of the paper is structured as follows. In section 2, we present the Hamiltonian of the BJ model which naturally leads us to the discrete equations of motion for the in-phase and out-of-phase motions, respectively. We expand terms of the stacking interaction and Morse potential in the out-of-phase equation of motion to the fifth order and implore the continuum limit approximation. Higher order amplitude equations are derived by using the extended multiple-scales asymptotic perturbation method. Analysis in section 3 is limited to terms up to the cubic order, in which periodic base pairs opening are captured to describe the open state configurations of the DNA. Stability analysis of these open states further reveal other localized background modes that are crucial in the transcription process. In section 4, we further explore higher order modes of the DNA open states configuration. This gives us a better opportunity to analytically and numerically investigate on the richer dynamics of DNA. Finally in section 5, we will summarize the important results obtained and articulate on some brighter perspectives.

2. Model and Equations of Motion

The dynamical DNA model proposed by Joyeux and Buyukdagli is more realistic than the PBD model. This is because the JB model is based on site-specific stacking enthalpies, with the finiteness of the stacking interaction sufficiently ensuring a sharp melting transition [2] [3] [4]. The DNA molecule in the JB model is considered as two elastic chains of nucleotides, that represent the double helix strand of the molecule. The nucleotides in the same strand experience nearest-neighbor interactions along the one dimensional chain configuration. The molecule is assumed homogeneous, with each strand linked to the other by hydrogen bonds which are modeled by the Morse potential. The longitudinal, rotational, and torsional motions of the DNA base pairs are all ignored, with focus only on the transverse motions. Concretely, this transverse displacements from the equilibrium position of the nucleotide pairs located in opposite strands are given by x_n and y_n. This naturally leads to the Hamiltonian of the BJ model as [2] [3]

$$H=\sum_n\left\{\frac{1}{2}m\left(\dot{x}_n^2+\dot{y}_n^2\right)+V_n\left(x_n,y_n\right)+W_n\left(x_n,y_n,x_{n-1},y_{n-1}\right)\right\}, \tag{1}$$

where

$$V_n(x_n, y_n) = D_0\left[e^{-a(x_n - y_n)} - 1\right]^2,$$

$$W_n(x_n, y_n, x_{n-1}, y_{n-1}) = \frac{\Delta H_n}{C}\left[2 - e^{-b(x_n - x_{n-1})^2} - e^{-b(y_n - y_n)^2}\right] + K_b\left[(x_n - x_{n-1})^2 + (y_n - y_{n-1})^2\right].$$

Each base pair position is represented by n with N being the total number of the base pairs of the chain, while m is the average mass of the nucleotides. The on-site potential $V_n(x_n, y_n)$ is due to the presence of hydrogen bonds which is described by the Morse potential of depth D_0 and width a. The Morse potential opposes the breaking of the hydrogen bonds because it is an increasing function of the distance between the two bases of a pair n. The first term in the potential $W_n(x_n, y_n, x_{n-1}, y_{n-1})$ describes the finite stacking interaction, while the second one models the stiffness of the sugar-phosphate backbone. Both terms are increasing functions of $|x_n - x_{n-1}|$ (as well as $|y_n - y_{n-1}|$), which implies that they oppose the de-stacking of the bases. The stacking potential which is approximated by a Gaussian hole of depth $\frac{\Delta H_n}{C}$, emanates from hydrophobic interactions with the solvent and electronic interactions between successive base pairs on the same strand. The backbone stiffness is taken as a harmonic potential of constant K_b, ensures that base pairs belonging to the same strand do not separate infinitely when approaching the melting temperature.

It is more convenient to introduce the coordinates u_n describing the movement of a center of mass of the nucleotide pair, and v_n, a stretching of the nucleotides belonging to the same pair defined as

$$u_n = \frac{x_n + y_n}{\sqrt{2}}, \quad \text{and} \quad v_n = \frac{x_n - y_n}{\sqrt{2}}. \tag{2}$$

The in-phase motion is actually governed by u_n, while v_n represent the out-of-phase motion. From the Hamiltonian (1), it is possible to obtain two nonlinear discrete differential equations describing the transverse in-phase and out-of-phase dynamics of the DNA molecular chain respectively given as [25]

$$\ddot{u}_n = \frac{2K_b}{m}\left[(u_{n+1} - u_n) + (u_{n-1} - u_n)\right] + \frac{2b\Delta H_n}{mC}\left[(u_{n+1} - u_n)e^{-b(u_{n+1} - u_n)^2} + (u_{n-1} - u_n)e^{-b(u_{n-1} - u_n)^2}\right], \tag{3a}$$

$$\ddot{v}_n = \frac{2K_b}{m}\left[(v_{n+1} - v_n) + (v_{n-1} - v_n)\right] + \frac{2b\Delta H_n}{mC}\left[(v_{n+1} - v_n)e^{-b(v_{n+1} - v_n)^2} - (v_{n-1} - v_n)e^{-b(v_{n-1} - v_n)^2}\right] + \frac{2\sqrt{2}aD_0}{m}e^{-a\sqrt{2}v_n}\left[e^{-a\sqrt{2}v_n} - 1\right]. \tag{3b}$$

From a comparative analysis between Eqns (3a) and (3b), it is clear that the effects of nonlinearity are more pronounced in the out-of-phase motion Equation (3b). This is because Equation (3b) which mimics base pair stretching, incorporates hydrogen bonds interactions which is a vital component in the DNA dynamics. On the other hand the in-phase equation of motion (3a), is more associated with strong covalent bonds with little or no effects on the holistic DNA dynamics [25]. Hence forth, we will neglect Equation (3a) and deal only with the out-of-phase motion Equation (3b), and adopt the following experimental parameters [2] [4]: $m = 300$ amu, $D_0 = 0.04\ \text{eV}$, $a = 4.45\ \text{Å}^{-1}$, $\Delta H_n = 0.44\ \text{eV}$, $b = 0.10\ \text{Å}^{-2}$, $K_b = 10^{-5}\ \text{eV}\cdot\text{Å}^{-2}$, and $C = 2.00$. It is further assumed that the base pair oscillations are quite large enough to induce inharmonicity, but still inadequate to destroy the hydrogen bonds because the plateau of the Morse potential is not attained. Based on this assumption, the base nucleotides should oscillate around the bottom of the Morse potential. Consequently one can expand the exponential functions up to the fifth order in the Taylor series approximation, and rewrite Equation (3b) as

$$\begin{aligned}\ddot{v}_n = {} & \frac{2}{m}\left(K_b + \frac{b\Delta H_n}{C}\right)\left(v_{n+1} - 2v_n + v_{n-1}\right) \\ & - \frac{2b^2\Delta H_n}{mC}\left[\left(v_{n+1}-v_n\right)^3 - \left(v_n - v_{n-1}\right)^3\right] \\ & + \frac{b^3\Delta H_n}{mC}\left[\left(v_{n+1}-v_n\right)^5 - \left(v_n - v_{n-1}\right)^5\right] \\ & - \frac{2a^2D_0}{m}\left[v_n - \frac{3a}{2}v_n^2 + \frac{7a^2}{6}v_n^3 - \frac{15a^3}{24}v_n^4 + \frac{31a^4}{120}v_n^5\right].\end{aligned} \tag{4}$$

The discrete coupled nonlinear Equation (4) is non integrable, but an approximation can be implored which preserves the nonlinearity of the system and reduce Equation (4) to an integrable form of a partial differential equation in order to obtain analytic solutions. Let us assume that the base pairs stretching $v_n(t)$ changes only slightly from one site to the next such that $v_n(t) = v(z = nr, t)$, where z is a dimensionless variable that measures the position along the DNA strand. r is a measure of the equilibrium distance between two successive neighboring nucleotides in the same strand, with numerical value of 3.4Å in a real DNA molecule [43]. Hence by considering a slow spatial variation of $v(z,t)$, and exploiting a Taylor expansion around $z = nr$, leads to

$$v_{n\pm1}(t) \approx v \pm r\frac{\partial v}{\partial z} + \frac{r^2}{2}\frac{\partial^2 v}{\partial z^2} + \cdots. \tag{5}$$

The continuum limit approximation (5), transforms Equation (4) to

$$\begin{aligned}\frac{\partial^2 v}{\partial t^2} = {} & \left[\frac{2r^2}{m}\left(K_b + \frac{b\Delta H_n}{C}\right) + \frac{6r^4b^2\Delta H_n}{mC}\left(\frac{\partial v}{\partial z}\right)^2 + \cdots\right]\frac{\partial^2 v}{\partial z^2} \\ & - \frac{2a^2D_0}{m}\left[v - \frac{3a}{2}v^2 + \frac{7a^2}{6}v^3 - \frac{5a^3}{8}v^4 + \frac{31a^4}{120}v^5\right],\end{aligned} \tag{6}$$

with all the key features involved in the DNA dynamics maintained, and terms of order $\bigcirc\left(r^5\right)$ or higher are neglected. It is more convenient to transform Equation (6) into a an appealing form by considering the dimensionless variables [38] [39]

$$V = av, t = \sqrt{\frac{m}{a^2 D_0}}\tau, K_0 = \left(\frac{2r^2}{a^2 D_0}\right)\left(K_b + \frac{b\Delta H_n}{C}\right), K_1 = \frac{6r^4 b^2 \Delta H_n}{a^4 C D_0},$$
$$\omega^2 = 2, \alpha = 3/2, \beta = -7/6, \gamma = 5/8, \sigma = -31/120. \tag{7}$$

Equation (6) is now written in the dimensionless form

$$\frac{\partial^2 V}{\partial \tau^2} = \left[K_0 + K_1\left(\frac{\partial V}{\partial z}\right)^2 + \cdots\right]\frac{\partial^2 V}{\partial z^2} - \omega^2\left[V - \alpha V^2 - \beta V^3 - \gamma V^4 - \sigma V^5\right], \tag{8}$$

in which in the new dimensionless time unit, we have that $\tau = 1.00$, corresponds to $t = 0.198\ \text{ps}$.

To find the solution of Equation (8), we must first obtain a useful and manageable equation by assuming a more appropriate ansatz; which we postulate that the asymptotic series is more generalized while preserving the essential features of the DNA system. This can be done using reductive perturbative analysis, which is more robust in that it can work for a wide range of problems. In fact, we explore the extended multiple-scales asymptotic approach to reduce Equation (8) to higher order nonlinear Schrödinger amplitude equations. The main idea behind the asymptotic approach is to introduce fast and slow time and spatial variables into Equation (8), by exploiting the perturbation parameter $\varepsilon \ll 1$. The hierarchies of new independent variables replacing z and τ are

$$\begin{cases} t_0 = \tau, t_2 = \varepsilon^2 \tau, t_4 = \varepsilon^4 \tau, \cdots \\ z_1 = \varepsilon z, z_3 = \varepsilon^3 z, \cdots \end{cases}$$

so that the τ - and z-derivatives are replaced by

$$\frac{\partial}{\partial \tau} = \frac{\partial}{\partial t_0} + \varepsilon^2 \frac{\partial}{\partial t_2} + \varepsilon^4 \frac{\partial}{\partial t_4} + \cdots,$$
$$\frac{\partial}{\partial z} = \varepsilon \frac{\partial}{\partial z_1} + \varepsilon^3 \frac{\partial}{\partial z_3} + \cdots. \tag{9}$$

According to the extended reductive perturbation method, we consider the following ansatz for the solution of $V(z,\tau)$ as [44]

$$\begin{aligned} V(z,\tau) = {} & \varepsilon\left(\psi e^{i\omega t_0} + \psi^* e^{-i\omega t_0}\right) + \varepsilon^2\left(G_2 + H_2 e^{2i\omega t_0} + H_2^* e^{-2i\omega t_0}\right) \\ & + \varepsilon^3\left(F_3 e^{i\omega t_0} + F_3^* e^{-i\omega t_0} + J_3 e^{3i\omega t_0} + J_3^* e^{-3i\omega t_0}\right) \\ & + \varepsilon^4\left(G_4 + H_4 e^{2i\omega t_0} + H_4^* e^{-2i\omega t_0} + K_4 e^{4i\omega t_0} + K_4^* e^{-4i\omega t_0}\right), \end{aligned} \tag{10}$$

where the amplitudes $\psi, G_2, H_2, F_3, J_3, G_4, H_4, K_4$ and their complex conjugates $\psi^*, H_2^*, F_3^*, J_3^*, H_4^*, K_4^*$ are functions of (z_1, t_2, z_3, t_4), while ω stands for the angular frequency. Upon substitution of the ansatz (10) into the DNA Equation (8) yields a series of inhomogeneous equations at different orders of $\left[\varepsilon, e^{i\omega t_0}\right]$.

Grouping terms to orders $\left[\varepsilon^2, e^{0i\omega t_0}\right]$, $\left[\varepsilon^2, e^{2i\omega t_0}\right]$, and $\left[\varepsilon^3, e^{i\omega t_0}\right]$, respectively yields the equations

$$0 = -\omega^2 G_2 + \omega^2 \alpha |\psi|^2, \tag{11a}$$

$$-4\omega^2 H_2 = -\omega^2 H_2 + \omega^2 \alpha \psi^2, \tag{11b}$$

$$2i\omega\frac{\partial\psi}{\partial t_2} = K_0\frac{\partial^2\psi}{\partial z_1^2} + 2\alpha\omega^2\left[\psi G_2 + \psi^* H_2 + \psi G_2^*\right] + 3\beta\omega^2|\psi|^2\psi. \tag{11c}$$

Simplification of Equation (11a) gives $G_2 = \alpha|\psi|^2$, while that of Equation (11b) result in $H_2 = -\alpha\psi^2/3$, and Equation (11c) is eventually reduced to

$$i\frac{\partial\psi}{\partial t_2} - P\frac{\partial^2\psi}{\partial z_1^2} + Q|\psi|^2\psi = 0, \tag{12}$$

where

$$\begin{cases} P = \dfrac{r^2}{a^2\omega D_0}\left[K_b + \dfrac{b\Delta H_n}{C}\right], \\ Q = -\dfrac{\omega}{6}\left[10\alpha^2 + 9\beta\right]. \end{cases}$$

The dispersion coefficient *P*, and nonlinear coefficient *Q*, are inextricably linked to the intrinsic constants $r, a, \omega, D_0, K_b, b, \Delta H_n, C, \alpha, \beta$, which are vital in the DNA dynamics; because it determines the type of wave profile in the system. The NLS Equation (12) is a prominent equation used to model a plethora of weakly nonlinear quasi-harmonic wave packets, like the propagation of electromagnetic waves in optical fibers [22] [23] [24], and DNA base pairs opening [25] [26].

This work is quite unique because it considers higher order modes of the DNA dynamics by incorporating the terms F_3, J_3, G_4, H_4, K_4, which highly depend on the type of solution emanating from Equation (12). Consequently, terms to orders $\left[\varepsilon^3, e^{3i\omega t_0}\right]$, $\left[\varepsilon^4, e^{0i\omega t_0}\right]$, and $\left[\varepsilon^4, e^{2i\omega t_0}\right]$, respectively gives

$$-9\omega^2 J_3 = -\omega^2 J_3 + 2\alpha\omega^2\psi H_2 + \beta\omega^2\psi^3, \tag{13}$$

$$\begin{aligned} 0 = {} & K_0\left(\frac{\partial^2 G_2}{\partial z_1^2} + \frac{\partial^2 G_2^*}{\partial z_1^2}\right) - \omega^2\left(G_4 + G_4^*\right) + 6\gamma\omega^2|\psi|^4 \\ & + \alpha\omega^2\left[2\psi F_3^* + 2\psi^* F_3 + \left(G_2 + G_2^*\right)^2 + 2|H_2|^2\right] \\ & + 3\beta\omega^2\left[\psi^2 H_2^* + \psi^{*2} H_2 + 2|\psi|^2\left(G_2 + G_2^*\right)\right], \end{aligned} \tag{14}$$

$$\begin{aligned} -4\omega^2 H_4 + 4i\omega\frac{\partial H_2}{\partial t_2} = {} & K_0\frac{\partial^2 H_2}{\partial z_1^2} - \omega^2 H_4 + 4\gamma\omega^2|\psi|^2\psi^2 \\ & + 3\beta\omega^2\left[\psi^2\left(G_2 + G_2^*\right) + 2|\psi|^2 H_2\right] \\ & + \alpha\omega^2\left[2\psi F_3 + 2\psi^* J_3 + 2G_2 H_2 + 2G_2^* H_2\right]. \end{aligned} \tag{15}$$

Finally, grouping terms to orders $\left[\varepsilon^4, e^{4i\omega t_0}\right]$ and $\left[\varepsilon^5, e^{i\omega t_0}\right]$, respectively generates the equations

$$-16\omega^2 K_4 = -\omega^2 K_4 + \alpha\omega^2 H_2^2 + 2\alpha\omega^2\psi J_3 + \gamma\omega^2\psi^4 + 3\beta\omega^2\psi^2 H_2, \tag{16}$$

and

$$
\begin{aligned}
&\frac{\partial^2\psi}{\partial t_2^2}+2i\omega\frac{\partial\psi}{\partial t_4}+2i\omega\frac{\partial F_3}{\partial t_2}\\
&=K_0\frac{\partial^2 F_3}{\partial z_1^2}+2K_0\frac{\partial^2\psi}{\partial z_1\partial z_3}+10\sigma\omega^2|\psi|^4\psi\\
&\quad+4\gamma\omega^2\left[\psi^3H_2^*+3|\psi|^2\psi\left(G_2+G_2^*\right)+3|\psi|^2\psi^*H_2\right]+3\beta\omega^2\left[\psi\left(G_2+G_2^*\right)^2\right.\\
&\quad\left.+2\psi^*H_2\left(G_2+G_2^*\right)+2|H_2|^2\psi+\psi^{2*}J_3+\psi^2F_3^*+2|\psi|^2F_3\right]\\
&\quad+2\alpha\omega^2\left[\psi\left(G_4+G_4^*\right)+\psi^*H_4+F_3^*H_2+F_3\left(G_2+G_2^*\right)+H_2^*J_3\right].
\end{aligned}
\tag{17}
$$

Equations (12) to (17) will be used to comprehensively study periodic base pair opening in a DNA double strand, in order to give satisfactory explanations to the transcription and replication processes from a biophysical perspective. However the analysis in section III will be limited just to Equation (12), in which we focus only on the periodic DNA dynamics in the cubic limit [2]-[8] [14]. The higher order modes that deals with Equations (13) to (17), is quite innovative and will be considered separately in section 4.

3. Dynamics of Base Pairs Opening in the Cubic Limit

3.1. Periodic Solution of the Nonlinear Schrödinger Amplitude Equation

We now consider the profile of the solution dictated by the derived amplitude Equation (12). Based on experimental parameters [2] [4], it is very clear that the dispersion coefficient $P>0$, and nonlinear coefficient $Q<0$. Consequently the system of Equation (13) can only support bright solitons as a result of the process of modulational instability [14] [15] [17] [19] [20] [21], because $PQ<0$. Plane waves gradually evolve into nonlinear periodic modes, that leads to energy activation in a DNA double strand chain by the process of modulational instability (MI). The MI process equally leads to the spontaneous emission of breather-like modes, and generally thrives in a DNA chain as a result of the dynamic interplay between nonlinearity (emanating from the hydrogen bonds) and dispersion (induced by stacking interaction and sugar-phosphate backbone stiffness).

In order to look for stationary solutions to Equation (12), we assume an ansatz of the form

$$\psi\left(z_1,t_2\right)=a\left(z_1\right)e^{ibt_2}, \tag{18}$$

where b is the modulation frequency of the envelope and $a\left(z_1\right)$ is a real constant that represent the amplitude of the field envelope. Upon substituting Equation (18) into Equation (12) gives

$$-ba-P\frac{\partial^2 a}{\partial z_1^2}+Qa^3=0, \tag{19}$$

which can be conveniently transformed to a first-order integral equation

$$\left(\frac{\mathrm{d}a}{\mathrm{d}z_1}\right)^2 = -\frac{b}{P}a^2 + \frac{Q}{2P}a^4 + C. \tag{20}$$

It is important to note that the integration constant C is a key parameter that determines the nature of solution of the amplitude $a(z_1)$. For a localized profile solution in which $a(z_1)$ vanishes as $z_1 \to \pm\infty$, C naturally turns to zero. Consequently the solution to Equation (20) is given by

$$a(z_1) = \sqrt{\frac{2b}{Q}}\operatorname{sech}\left[\sqrt{\frac{-b}{P}}z_1\right]. \tag{21}$$

It is more appropriate to set $b = \left(2u_e u_c - u_e^2\right)/4P$, where u_e and u_c are real numbers which respectively represent envelope and carrier wave velocities measured in units of Å per dimensionless time τ [25]. Hence, solution (18) can now be re-written in the form [25]

$$\psi(z_1,t_2) = \sqrt{\frac{2u_e u_c - u_e^2}{2PQ}}\operatorname{sech}\left[\sqrt{\frac{u_e^2 - 2u_e u_c}{4P^2}}z_1\right]\mathrm{e}^{i\left[\frac{u_e^2 - 2u_e u_c}{4P}\right]t_2}. \tag{22}$$

When the integration constant C is nonzero, the single pulse solution (22) becomes very unstable and difficult to be sustained in the system. Consequently, Equation (12) now admits periodic solution of the form [18] [24]

$$\psi(z_1,t_2) = a_0 dn\left[Lz_1,k\right]\mathrm{e}^{i\zeta t_2}, \tag{23}$$

where dn is a Jacobi elliptic function of modulus $k(0 \le k \le 1)$,

$$a_0 = u_e\sqrt{\frac{1-2\eta}{-2\left(2-k^2\right)PQ}}, L = a_0\sqrt{\frac{-Q}{2P}}, \zeta = -PL^2, \tag{24}$$

and $\eta = u_c/u_e$, with $0 \le \eta < 0.5$. The solution of Equation (8) in the cubic limit (*i.e.* $\gamma = \sigma = F_3 = J_3 = G_4 = H_4 = K_4 = 0$), now gives

$$V(z,\tau) = 2a_{0\varepsilon}dn\left[L_\varepsilon z,k\right]\cos(\Omega\tau) + 2a_{0\varepsilon}^2 dn^2\left[L_\varepsilon z,k\right]\left\{\frac{\alpha}{2} - \frac{\alpha}{3}\cos(2\Omega\tau)\right\} + \mathrm{O}\left(\varepsilon^3\right), \tag{25}$$

where

$$a_{0\varepsilon} = \varepsilon u_e\sqrt{\frac{1-2\eta}{-2\left(2-k^2\right)PQ}}, L_e = \varepsilon a_{0\varepsilon}\sqrt{\frac{-Q}{2P}}, \zeta_\varepsilon = -PL_\varepsilon^2, \Omega = \omega + \varepsilon^2\zeta_\varepsilon, t_0 = \tau. \tag{26}$$

The constants $a_{0\varepsilon}, 1/L_e$, and Ω, are respectively the amplitude, width, and angular frequency of the soliton solution that mimics the DNA open state configuration in the cubic limit. They are all dependent on the perturbation parameter ε, elliptic modulus k, and the experimental parameters of the DNA system as depicted in the contour plot in **Figure 1**. The amplitude $a_{0\varepsilon}$ increases with increase in ε, and more appreciable for $\varepsilon > 0.7$ and $k > 0.8$. The width of the soliton $1/L_e$, is measured in Å and more feasible for $\varepsilon > 0.7$, irrespective of the values of k. Lastly, an increase in ε generally diminishes the angular frequency Ω, of the soliton. Based on the contour plot in **Figure 1**, we will

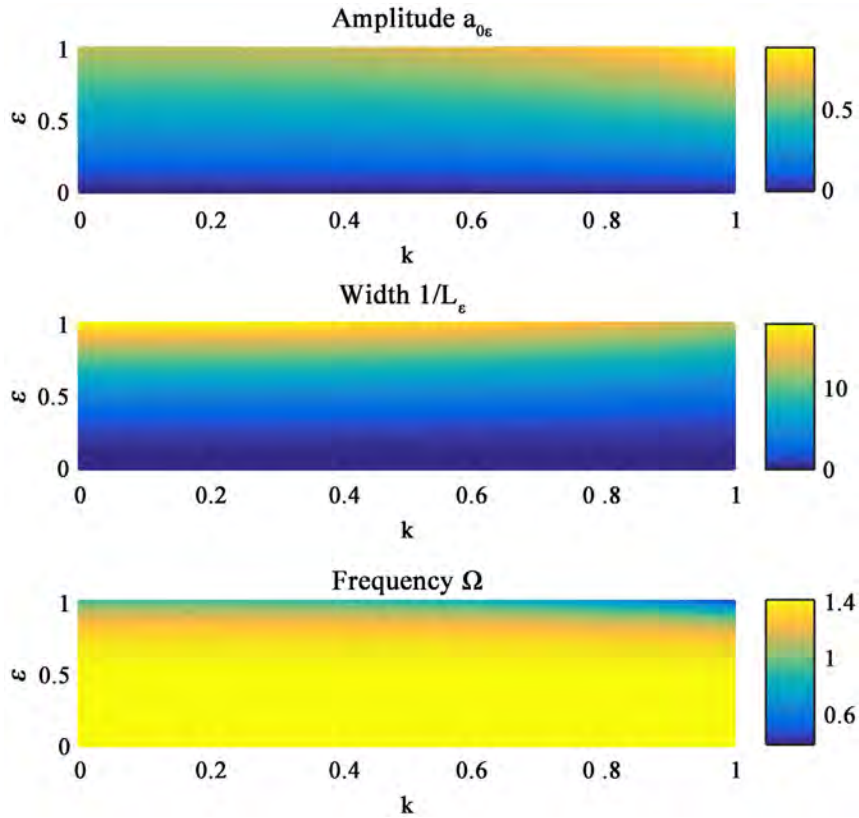


Figure 1. (color online) Variations of $a_{0\varepsilon}, 1/L_e$, and Ω, with elliptic modulus modulus k and perturbation parameter ε. The experimental values of the constants are $m = 300$ amu, $D_0 = 0.04$ eV, $a = 4.45$ Å^{-1}, $\Delta H_n = 0.44$ eV, $b = 0.10$ Å^{-2}, $K_b = 10^{-5}$ eV·Å^{-2}, $r = 3.4$ Å, $\omega = \sqrt{2}$, $C = 2.00$, $u_e = 1.00$ Å, $u_c = 0.00$ Å, $\alpha = 3/2$, and $\beta = -7/6$.

henceforth adopt $\varepsilon = 1.00$, $k > 0.8$, without loss of generality. The choice of the values of ε and *k*, coupled with the experimental parameters given in **Figure 1**, is best suited for us to analytically describe the base pair stretching that characterize the open state.

Figure 2 depicts the initial stages of the base pairs stretching in the cubic limit, which represent the breathing modes in the DNA molecular chain. We observe spatial periodic modes of the DNA base pairs stretching in **Figure 2(a)**, for $k = 0.88$, and **Figure 2(b)**, for $k = 0.98$. However for $k = 1.00$, a more localized open state mode is observed in **Figure 2(c)**; similar to the higher order discrete breather mode [26], and spatial compaction profile [39]. The spatiotemporal profile in **Figure 2** generally portrays a very stable structural wave features, probably because it is still evolving at the early stages (*i.e.* τ varies from 0.00 to 60.0).

Concretely at the initial stage for $k = 0.88$, and $k = 0.98$, as in **Figure 3(a)** and **Figure 3(b)** respectively, the breathing modes of the DNA is purely periodic. These modes degenerate to a strong secant hyperbolic excitation in **Figure 3(c)** for $k = 1.00$. These excitations mainly mimics the open state configuration, and a precursor for the transcription, replication, and transmission of genetic codes

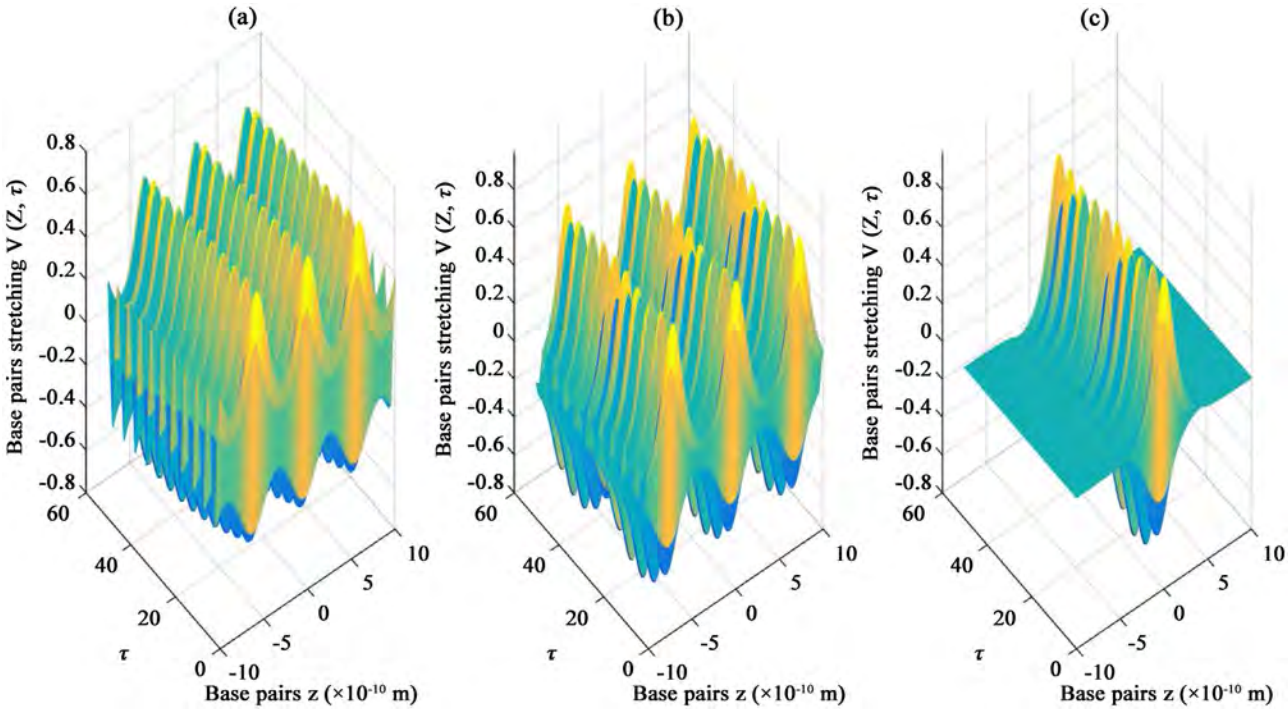


Figure 2. (color online) Three dimensional plot at the initial stages of base pairs stretching, according to solution (25). Parameters are: $m = 300$ amu, $D_0 = 0.04$ eV, $a = 4.45$ Å^{-1}, $\Delta H_n = 0.44$ eV, $b = 0.10$ Å^{-2}, $K_b = 10^{-5}$ eV·Å^{-2}, $r = 3.4$ Å, $\omega = \sqrt{2}$, $C = 2.00$, $u_e = 1.00$ Å, $u_c = 0.40$ Å, $\alpha = 3/2$, $\beta = -7/6$, and $\varepsilon = 1.00$. This is for elliptic modulus: (a) $k = 0.88$, (b) $k = 0.98$, (d) $k = 1.00$.

in the DNA double strand chain [1]-[8]. After a period of time $t = 100$ ps ($\tau = 505.05$), the modes in **Figures 3(a)-(c)** gradually changes to give the profile shown in **Figures 3(d)-(f)**; where the nucleotide stretching becomes more pronounced especially as in **Figure 3(f)** for $k = 1.00$. The base pairs stretching which is governed by the analytical solution (25), equally evolve to give a structurally robust breather modes in **Figures 3(g)-(i)** after $t = 200$ ps ($\tau = 1010.10$).

We further plot the long-time evolution of base pair stretching, at different times as shown in **Figure 4**. The periodic and localized modes in **Figure 4**, are also known as the DNA fluctuational opening; best described as precursor states for the local denaturation observed during DNA transcription. It equally captures the thermal denaturation process, based on the finite stacking enthalpy DNA model. The structural variations of the nucleotide base pairs stretching in **Figure 3** and **Figure 4**, simply points to the instability of the DNA open states. Consequently, we will carry out a linear stability analysis in the proceeding subsection in order to test the robustness of these spatial periodic DNA modes.

3.2. Stability Analysis

In the preceding subsection, we obtained localized periodic wave trains that mimic base pairs stretching in the finite stacking enthalpy DNA molecular chain. To discuss the stability of these periodic breathing modes, one must superimpose a small perturbation on this solution and analyze the evolution of the

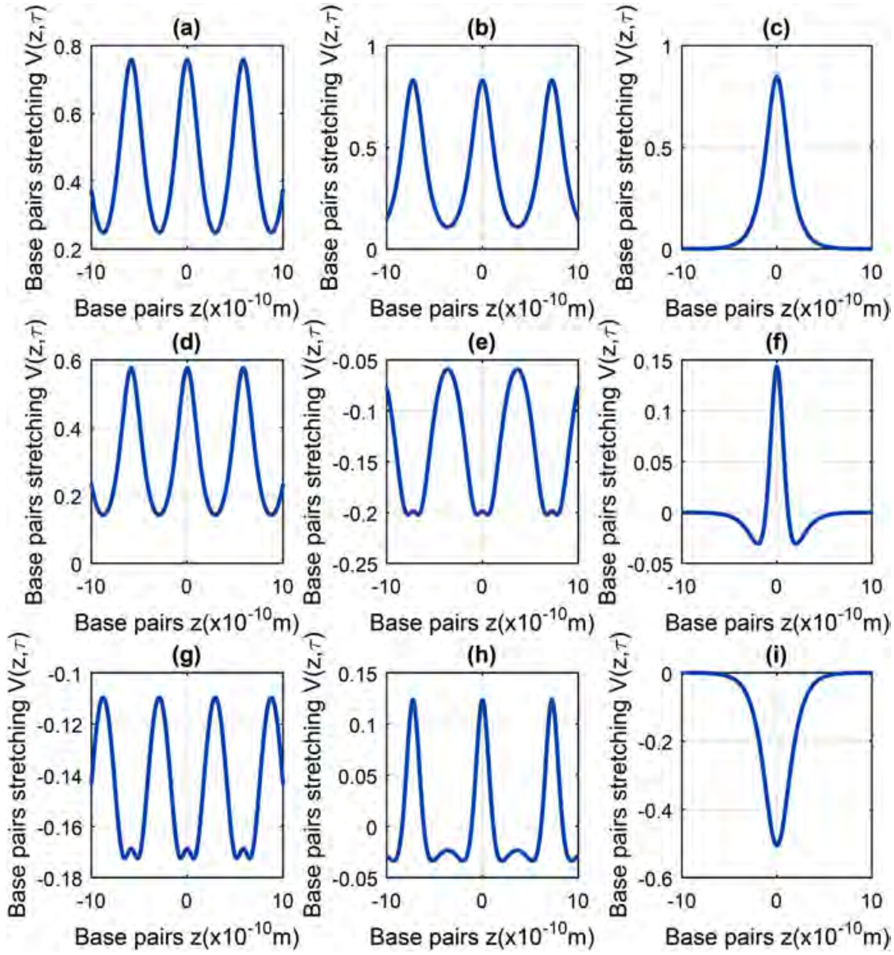


Figure 3. Base pairs stretching according to solution (25), for experimental values: $m = 300$ amu, $D_0 = 0.04$ eV, $a = 4.45$ Å^{-1}, $\Delta H_n = 0.44$ eV, $b = 0.10$ Å^{-2}, $K_b = 10^{-5}$ eV $\cdot$ Å^{-2}, $r = 3.4$ Å, $\omega = \sqrt{2}$, $C = 2.00$, $u_e = 1.00$ Å, $u_c = 0.40$ Å, $\alpha = 3/2$, $\beta = -7/6$, and $\varepsilon = 1.00$. This is for dimensionless time (τ) and elliptic modulus (k), as follows: (a) $\tau = 0.00$, $k = 0.88$, (b) $\tau = 0.00$, $k = 0.98$, (c) $\tau = 0.00$, $k = 1.00$, (d) $\tau = 505.05$, $k = 0.88$, (e) $\tau = 505.05$, $k = 0.98$, (f) $\tau = 505.05$, $k = 1.00$, (g) $\tau = 1010.10$, $k = 0.88$, (h) $\tau = 1010.10$, $k = 0.98$, (i) $\tau = 1010.10$, $k = 1.00$.

perturbation. Note that stability analysis is an important issue related to the study of nonlinear dynamical systems because it provides an effective way of testing the robustness of the soliton trains against small perturbation in the amplitude. Stability analysis is applied in a diverse manner, based on the complexity of the physical system under review and the type of solution involved.

In order to investigate the linear stability of the spatial periodic soliton mode, we consider small perturbations $a_p(z_1)$ to the amplitude of the DNA excitation mode denoted by $a_0(z_1)$, so that

$$\psi(z_1, t_2) = \left[a_0(z_1) + \varepsilon a_p(z_1)\right] e^{ibt_2}, \tag{27}$$

where $\varepsilon \ll 1$. After the nonlinear interactions, the resultant internal modes of vibration carrying the genetic codes is obtained by substituting Equation (27) into Equation (12) and considering terms to the various orders of ε:

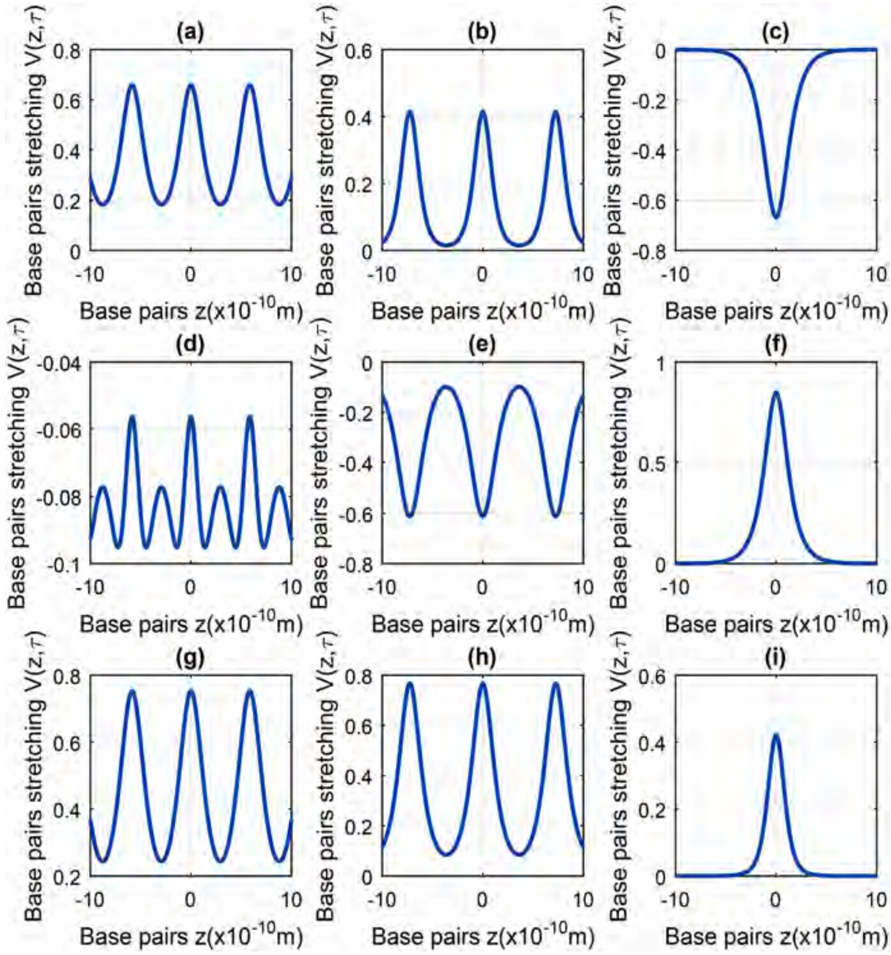


Figure 4. Parameters are the same as in **Figure 3**, but for: (a) $\tau = 2525.25$, $k = 0.88$, (b) $\tau = 2525.25$, $k = 0.98$, (c) $\tau = 2525.25$, $k = 1.00$, (d) $\tau = 2700.00$, $k = 0.88$, (e) $\tau = 2700.00$, $k = 0.98$, (f) $\tau = 2700.00$, $k = 1.00$, (g) $\tau = 3030.30$, $k = 0.88$, (h) $\tau = 3030.30$, $k = 0.98$, (i) $\tau = 3030.30$, $k = 1.00$.

Order ε^0,

$$P\frac{\mathrm{d}^2 a_0}{\mathrm{d}z_1^2} - Qa_0^3 + ba_0 = 0. \tag{28}$$

Order ε^1,

$$P\frac{\mathrm{d}^2 a_p}{\mathrm{d}z_1^2} - 3Qa_0^2 a_p + ba_p = 0. \tag{29}$$

We have already obtained periodic solution of Equation (28) as

$$a_0(z_1) = \sqrt{\frac{2b}{Q(2-k^2)}}\, dn\left[\sqrt{\frac{-b}{P(2-k^2)}}z_1, k\right], \tag{30}$$

which can be substituted into Equation (29) and simplified to obtain [18] [24]

$$\frac{\mathrm{d}^2 a_p}{\mathrm{d}\xi^2} + \left[E(k) - 6k^2 sn^2(\xi, k)\right]a_p = 0, \tag{31}$$

where $\lambda = \sqrt{\frac{-b}{P\left(2-k^2\right)}}$, $\xi = \lambda z_1$, and $E\left(k\right) = \frac{6P\lambda^2 + b}{P\lambda^2}$. Equation (31) is known as the Lamé equation of the second order with five distinct localized solutions [18] [24]

$$a_{p1}\left(\xi\right) = A_1\left(k\right)cn\left(\xi,k\right)dn\left(\xi,k\right), b_1 = \left(5-k^2\right)\left|P\right|\lambda^2, \tag{32a}$$

$$a_{p2}\left(\xi\right) = A_2\left(k\right)cn\left(\xi,k\right)sn\left(\xi,k\right), b_2 = \left(2-k^2\right)\left|P\right|\lambda^2, \tag{32b}$$

$$a_{p3}\left(\xi\right) = A_3\left(k\right)sn\left(\xi,k\right)dn\left(\xi,k\right), b_3 = \left(5-4k^2\right)\left|P\right|\lambda^2, \tag{32c}$$

$$\begin{aligned} a_{p4,p5}\left(\xi\right) &= A_{4,5}\left(k\right)\left[sn^2\left(\xi,k\right) - \frac{1+k^2}{3k^2} \mp \frac{\sqrt{1-k^2\left(1-k^2\right)}}{3k^2}\right], \\ b_{4,5} &= \left[2-k^2 \pm \frac{\sqrt{1-k^2\left(1-k^2\right)}}{2}\right]\left|P\right|\lambda^2. \end{aligned} \tag{32d}$$

The amplitudes $A_i\left(k\right)$ for $i = 1,2,3,4,5$ can be obtained by using appropriate orthogonality and normalization relations [45], and the profile of these perturbations are given in **Figure 5**. The spatial period and length of the perturbed solitons $a_{pi}\left(\xi\right)$, is controlled by the elliptic modulus *k*. In fact for $k = 1.00$, the separation between solitons become infinite and periodic train degenerates to a localized mode. From the profiles of these single solitons on the right column in **Figure 5**, it is clear that the distinct localized soliton modes are both symmetric and asymmetric with respect to the origin.

The five bound states given in **Figure 5** shows the various genetic code structures, that clearly signifies the complex nature of the transcription and replication processes. Gene transfer under appropriate physiological conditions is indeed possible only in one of these modes. Concretely from the standpoint of biophysics, these localized modes describe internal oscillations in the DNA structure which is capable of exposing all hidden modes that may be responsible for gene mutations or other physiological disorders [25] [42]. These bound states can also be associated with radiation-carrying excitations in the background of the DNA open state, by virtue of their nonzero energies [18].

4. Higher Order Stretching of Nucleotide Base Pairs

In order to give a comprehensive account of the base pairs stretching that mimics the DNA open state, we now incorporate all the terms in the ansatz (10). However we set $F_3 = 0$, without of loss of generality, since we are dealing with periodic modes. The term F_3 generally introduces secularity which makes the solution cumbersome, and blurs the basic physics that characterize periodic open state dynamics of finite stacking enthalpy DNA. It is important to note that in the investigation of stationary breather modes of generalized nonlinear Klein-Gordon lattices, $F_3 \neq 0$ [44].

Simplification of Equations (13) to (19), naturally lead us to the following

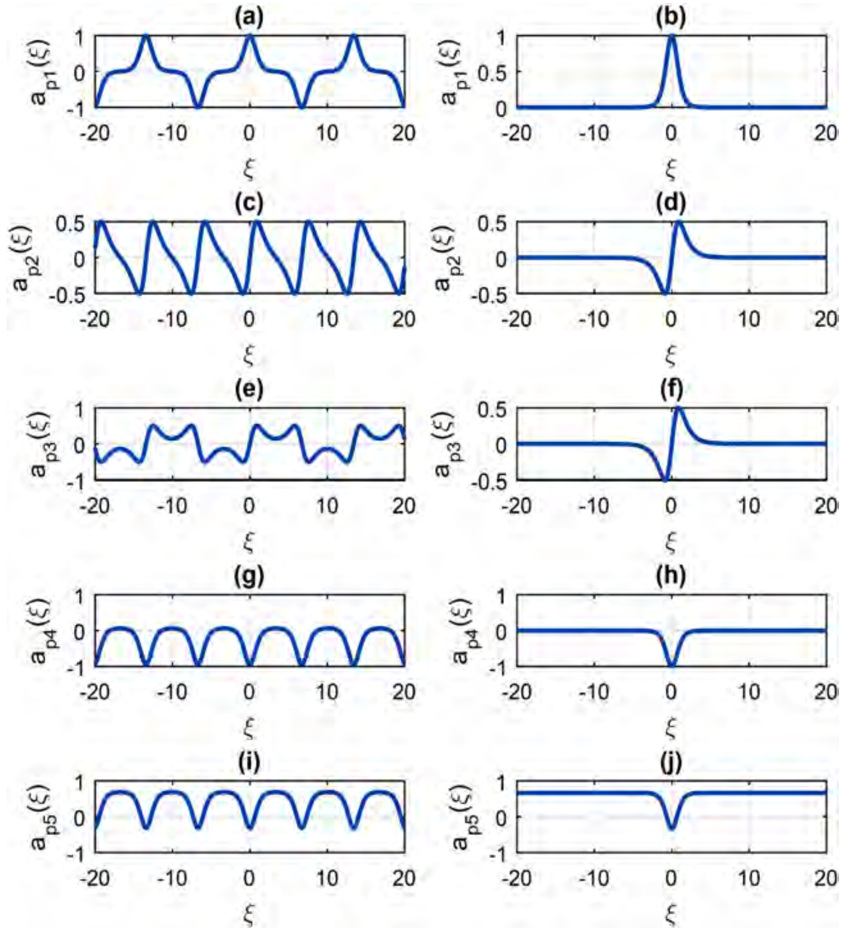


Figure 5. Profiles of bound state solutions (32), of the second order Lamé Equation (31) with $A_i(k)=1.0$, for $i=1,2,3,4,5$. The left column is for $k=0.98$, while the right is for $k=1.00$.

important quantities:

$$J_3=\frac{1}{24}\left[2\alpha^2-3\beta\right]\psi^3, \tag{33a}$$

$$G_4=\frac{K_0\alpha}{\omega^2}\left[\frac{\partial^2|\psi|^2}{\partial z_1^2}\right]+\left[3\gamma+5\alpha\beta+\frac{19}{9}\alpha^3\right]|\psi|^4, \tag{33b}$$

$$H_4=\frac{2K_0\alpha}{9\omega^2}\left[\frac{\partial\psi^2}{\partial z_1}-\psi\frac{\partial^2\psi}{\partial z_1^2}\right]-\left[\frac{4}{3}\gamma+\frac{31}{12}\alpha\beta+\frac{59}{54}\alpha^3\right]|\psi|^2\psi^2, \tag{33c}$$

$$K_4=\left[\frac{1}{12}\alpha\beta-\frac{1}{15}\gamma-\frac{1}{54}\alpha^3\right]\psi^4. \tag{33d}$$

Consequently, Equation (17) can now be re-written as

$$\begin{aligned}2i\omega\frac{\partial\psi}{\partial t_4}+\frac{\partial^2\psi}{\partial t_2^2}&=2K_0\frac{\partial^2\psi}{\partial z_1\partial z_3}+4K_0\alpha^2\psi\frac{\partial^2|\psi|^2}{\partial z_1^2}+\frac{2K_0\alpha}{\omega^2}\left[\psi^*\frac{\partial\psi^2}{\partial z_1}-|\psi|^2\frac{\partial^2\psi}{\partial z_1^2}\right]\\&+\omega^2\left[\frac{84}{3}\alpha\gamma+\frac{635}{36}\alpha^2\beta-\frac{1005}{162}\alpha^4-\frac{3}{8}\beta^2+10\sigma\right]\psi|\psi|^4.\end{aligned} \tag{34}$$

Based on the values obtained for J_3, G_4, H_4, K_4, we can conveniently modify solution (23) to now read [44]

$$\psi(z_1,t_2,z_3,t_4) = a_0\, dn\left[Lz_1 + L_3 z_3, k\right] \mathrm{e}^{i(\zeta t_2 + \zeta_4 t_4)}, \tag{35}$$

where L_3 and ζ_4 are real constants. To determine the values of L_3 and ζ_4, we substitute the modified solution (35) into Equation (34) and group coefficients of terms of the orders $dn^0[\,], dn^2[\,]$ and so on. After some rigorous calculations and simplifications, this leads us to

$$L_3 = La_0^2\alpha\left[4\alpha\left(3-2k^2\right) + \frac{1}{\omega}\left(k^2-1\right)\right], \tag{36a}$$

$$\zeta_4 = \frac{K_0\alpha L^2 a_0^2\left(k^2-1\right)^2\left[1-4\alpha\omega^2\left(4-k^2\right)\right]}{\omega^3} - \frac{\zeta}{2\omega}. \tag{36b}$$

The most appropriate form of the analytic solution of the higher order stretching of nucleotide base pairs is now obtained by substituting the solution (35), into the ansatz (10) to have

$$\begin{aligned} V(z,\tau) &= 2a'_{0\varepsilon} dn\left[L'_\varepsilon z,k\right]\cos\left(\Omega'\tau\right) + a'^2_{0\varepsilon} dn^2\left[L'_\varepsilon z,k\right]\left\{\frac{3}{2} - \cos\left(2\Omega'\tau\right)\right\} \\ &+ \frac{2}{3}a'^3_{0\varepsilon} dn^3\left[L'_\varepsilon z,k\right]\cos\left(3\Omega'\tau\right) - a'^4_{0\varepsilon} dn^4\left[L'_\varepsilon z,k\right]\left\{\frac{1}{2} + \frac{23}{50}\cos\left(4\Omega'\tau\right)\right\} \\ &+ \frac{33}{50}\varepsilon a'^3_{0\varepsilon} dn^3\left[L'_\varepsilon z,k\right]\cos\left(\left[3\Omega' - \sqrt{2}\right]\tau\right) + 3\varepsilon^2 K_0 L^2 a'^2_{0\varepsilon} dn\left[L'_\varepsilon z,k\right]\left\{dn^3\left[L'_\varepsilon z,k\right]\right. \\ &\left. + \left(k^2-1\right)dn\left(L'_\varepsilon z,k\right) - \frac{2}{L}k^2 sn\left[L'_\varepsilon z,k\right]cn\left[L'_\varepsilon z,k\right]\right\}\cos\left(2\Omega'\tau\right) \\ &+ \frac{3}{2}\varepsilon^2 K_0 L^2 a'^2_{0\varepsilon} dn\left[L'_\varepsilon z,k\right]\left\{k^2 - 1 + 2\left(1-k^2\right)dn^2\left[L'_\varepsilon z,k\right] - 2dn^4\left[L'_\varepsilon z,k\right]\right\}. \end{aligned} \tag{37}$$

The constants in solution (37) are

$$a'_{0\varepsilon} = \varepsilon u_e\sqrt{\frac{1-2\eta}{-2\left(2-k^2\right)PQ}},\quad L'_\varepsilon = \varepsilon a'_{0\varepsilon}\sqrt{\frac{-Q}{2P}}\left\{1+\varepsilon^2 a'^2_{0\varepsilon}\left(25.9-16.9k^2\right)\right\},$$

$$\Omega' = \sqrt{2} + 0.325\varepsilon^4 Q a'^2_{0\varepsilon} + 0.53\varepsilon^4 K_0 L^2 a_0^2\left(k^2-1\right)^2\left(12k^2-47\right), \tag{38}$$

and the values of L and a_0 are given in Equation (24).

The profiles of the higher order modes are given in **Figure 6**. It clearly depicts the stretching of DNA double strand. Such open states lead to a better representation of the base pairs stretching that generally precedes the transcription and replication processes. Furthermore, it equally depicts a more accurate energy activator for RNA-polymerase transport during the periodic opening of DNA double strand chain, thereby exposing more bases out of the stack. As shown in (**Figure 6(a)**, **Figure 6(d)**, **Figure 6(g)**) for $k = 0.88$, and (**Figure 6(b)**, **Figure 6(e)**, **Figure 6(h)**) for $k = 0.98$, we observe that more base pairs are experiencing a very structurally stable open state configurations. However for $k = 1.00$ as in (**Figure 6(c)**, **Figure 6(f)**, **Figure 6(i)**), only few base pairs open up during the transcription process as experimentally confirmed with DNA double helix

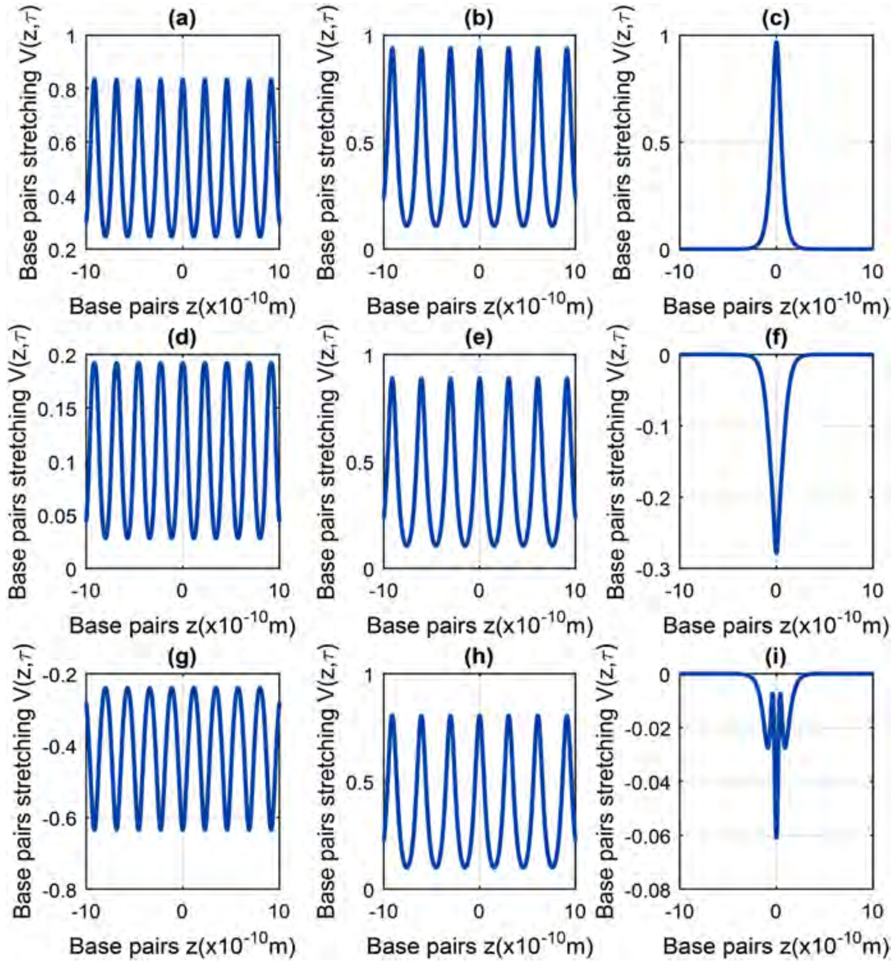


Figure 6. Higher order base pairs stretching according to solution (37), for experimental values: $m = 300$ amu, $D_0 = 0.04$ eV, $a = 4.45$ Å^{-1}, $\Delta H_n = 0.44$ eV, $b = 0.10$ Å^{-2}, $K_b = 10^{-5}$ eV·Å^{-2}, $r = 3.4$ Å, $\omega = \sqrt{2}$, $C = 2.00$, $u_e = 1.00$ Å, $u_c = 0.40$ Å, $\alpha = 3/2$, $\beta = -7/6$, $\gamma = 5/8$, $\sigma = -31/120$, and $\varepsilon = 1.00$. This is for: (a) $\tau = 0.00$, $k = 0.88$, (b) $\tau = 0.00$, $k = 0.98$, (c) $\tau = 0.00$, $k = 1.00$, (d) $\tau = 505.05$, $k = 0.88$, (e) $\tau = 505.05$, $k = 0.98$, (f) $\tau = 505.05$, $k = 1.00$, (g) $\tau = 1010.10$, $k = 0.88$, (h) $\tau = 1010.10$, $k = 0.98$, (i) $\tau = 1010.10$, $k = 1.00$.

[46] [47]. This mainly confirms that the soliton solution (37), gives a more accurate analytic representation within theoretical limits [48]; of stable periodic open states under appropriate physiological conditions.

The long time evolution of the gradual unzipping of the DNA molecule is captured in Figure 7, which is characterized by minimal distortion of the periodic modes. It is important to note that the stretching of the base pairs in Figure 6(i) and Figure 7(c), Figure 7(f) for $k = 1.00$, induces minor distortions in the open state by the slight splitting of a single pulse. This may have long term effects during the transcription and replication processes, as the open state becomes more susceptible to external attack. Such attacks may alter some parameters of the DNA system, hence distorting the reading of genetic codes and induces gene mutations which is responsible for numerous diseases [25] [42] [46]

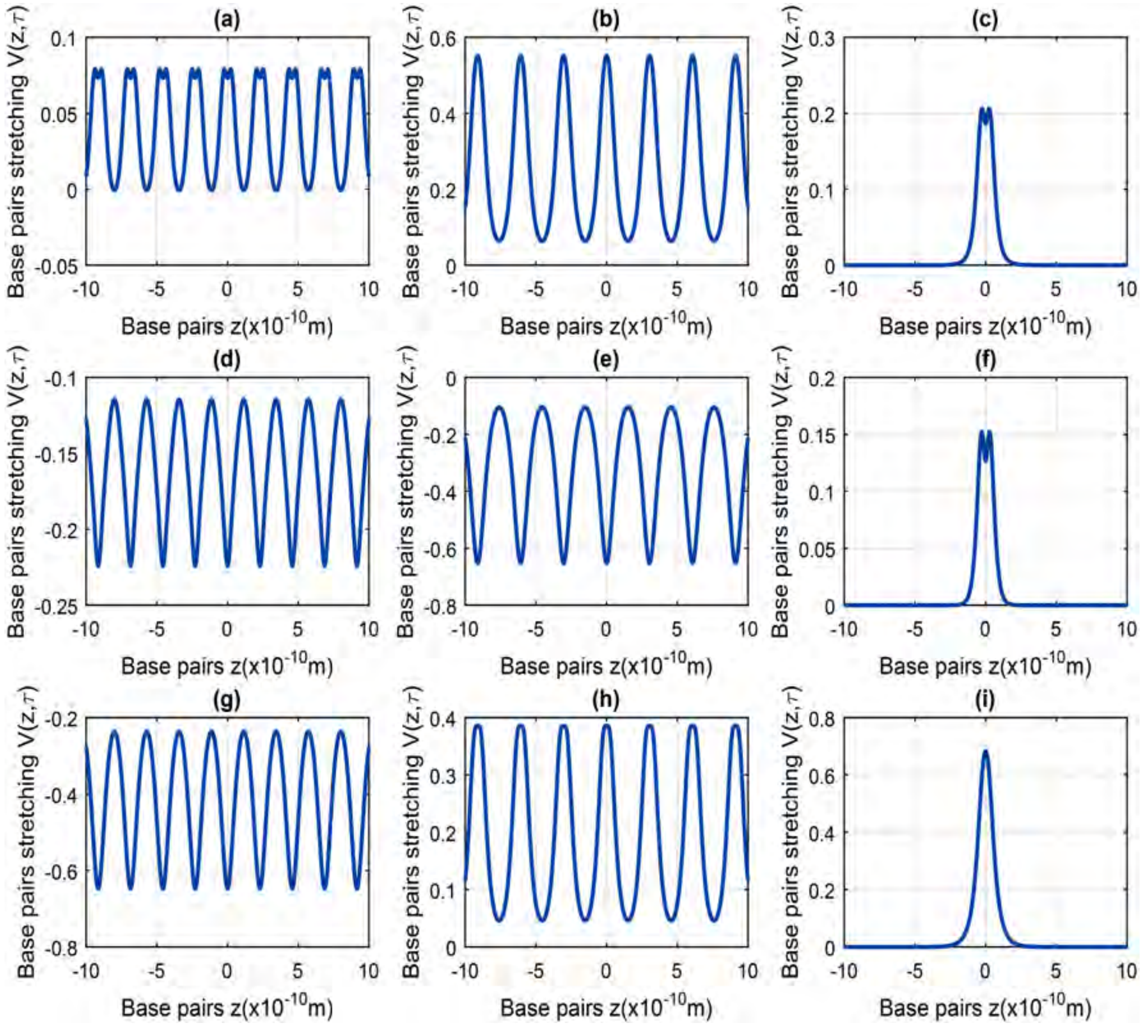


Figure 7. Parameters are the same as in **Figure 6**, but for: (a) $\tau = 2525.25$, $k = 0.88$, (b) $\tau = 2525.25$, $k = 0.98$, (c) $\tau = 2525.25$, $k = 1.00$, (d) $\tau = 2700.00$, $k = 0.88$, (e) $\tau = 2700.00$, $k = 0.98$, (f) $\tau = 2700.00$, $k = 1.00$, (g) $\tau = 3030.30$, $k = 0.88$, (h) $\tau = 3030.30$, $k = 0.98$, (i) $\tau = 3030.30$, $k = 1.00$.

[47]. A careful observation shows that the structural stability and number of solitons in the base pairs stretching of **Figure 6** and **Figure 7**, supersedes that of their cubic limit counterparts in **Figure 3** and **Figure 4** respectively. Since all the experimental values used are the same, the comparison strongly suggest that the higher order modes solution (37), reflects a more realistic open state configuration during the transcription and replication processes.

5. Discussion and Conclusion

The nucleotide is an elementary unit which consists of sugar, phosphate, and base. The segment of a DNA which is responsible for biosynthesis of a single polypeptide chain is called a gene, which averagely contains about 900 to 1500 nucleotide base pairs. During the transcription and replication processes, these genes are effectively transferred to single stranded shorter RNA. In fact, bases freely interact with enzymes during the local opening of DNA chain and therefore play a key regulatory role in the transcription process. This study seeks to expose the various forms of periodic base pairs stretching of DNA, in order to elucidate on intrinsic processes like transmission, transcription, and replication of genetic codes. Concretely, the DNA remains the basic organ of a cell which stores all vital information that ensures the effective growth and reproduction of all living organisms. Some distorted profile of base pairs stretching identified in this study may make the system so vulnerable to external attack, and hinders the smooth flow of genetic codes.

We effectively derived higher order nonlinear Schrödinger amplitude equations from the BJ model of DNA, by using the extended multiple scale asymptotic methods. Periodic solutions of these amplitude equations were used to mimic the open state configurations of a DNA strand under appropriate physiological conditions. The BJ model can also provide valuable information relating to the thermodynamic properties of DNA, by showing that the finite enthalpy may be responsible for the DNA denaturation. The stability analysis shows the existence of other background modes that may provide a possible physical mechanism for the effect of finite enthalpy stacking on DNA dynamics to be investigated. The stacking interaction of the BJ model equally provides both linear and nonlinear coupling parameters in the system which can independently control the dynamics and stability of the periodic solutions. We believe that this work opens up new vision on the concept of nonlinear periodic waves in DNA, and can also be exported in the study of many other physical systems.

In perspective, it will be very interesting to investigate on the impact of the helicoidal interactions on the periodic stretching of base pairs. The helicoidal term can best be appreciated because of the twisted nature of a real DNA molecule. Secondly, the interactions between the oscillating nucleotides and the aqueous environment naturally induce frictional forces, that must not be neglected in order to holistically describe the DNA dynamics. Lastly, quantum theory needs to be fully incorporated in order to comprehensively describe the DNA dynamics at the level much smaller than the nucleotides [49].

Acknowledgements

N. Oma Nfor appreciates the enriching discussions with S. Bruno Yamgoué, on the dynamics of DNA molecule.

Conflicts of Interest

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

References

[1] Englander, S.W., Kallenbach, N.R., Heeger, A.J., Krumhansl, J.A. and Litwin, S. (1980) *Proceedings of the National Academy of Sciences of the United States of America*, **77**, 7222. https://doi.org/10.1073/pnas.77.12.7222

[2] Joyeux, M. and Buyukdagli, S. (2005) *Physical Review E*, **72**, Article ID: 051902. https://doi.org/10.1103/PhysRevE.72.051902

[3] Joyeux, M. and Buyukdagli, S. (2008) *Physical Review E*, **77**, Article ID: 031903. https://doi.org/10.1103/PhysRevE.77.031903

[4] Joyeux, M. and Florescu, A.-M. (2009) *Journal of Physics: Condensed Matter*, **21**, Article ID: 34101. https://doi.org/10.1088/0953-8984/21/3/034101

[5] Yakusevich, L. (2004) Nonlinear Physics of DNA. Wiley Series in Nonlinear Sciences. John Wiley and Sons, Weinheim.

[6] Dauxois, T., Peyrard, M. and Bishop, A.R. (1993) *Physical Review E*, **47**, 684-695. https://doi.org/10.1103/PhysRevE.47.684

[7] Dauxois, T. and Peyrard, M. (1995) *Physical Review E*, **51**, 4027-4040. https://doi.org/10.1103/PhysRevE.51.4027

[8] Peyrard, M. (2004) *Nonlinearity*, **17**, R1-R40. https://doi.org/10.1088/0951-7715/17/2/R01

[9] El Kinani, R., Kaidi, H. and Barka, N. (2020) *Open Journal of Biophysics*, **10**, 129-149. https://doi.org/10.4236/ojbiphy.2020.103011

[10] Biegeleisen, K. (2016) *Open Access Library Journal*, **3**, 1-24. https://doi.org/10.4236/oalib.1103114

[11] Watson, J.D. (1976) Molecular Biology of the Gene. 3rd Edition, W. A. Benjamin, Menlo Park.

[12] Gotoh, O. (1983) *Advances in Biophysics*, **16**, 3. https://doi.org/10.1016/0065-227X(83)90007-2

[13] Yakushevich, L.V., Savin, A.V. and Manevitch, L.I. (2002) *Physical Review E*, **66**, Article ID: 016614. https://doi.org/10.1103/PhysRevE.66.016614

[14] Okaly, J.B., Ndzana, F.-II, Woulaché, R.L. and Kofané, T.C. (2019) *The European Physical Journal Plus*, **134**, 598. https://doi.org/10.1140/epjp/i2019-12992-3

[15] Nfor, N.O., Yamgoué, S.B. and Moukam Kakmeni, F.M. (2021) *Chinese Physics B*. https://doi.org/10.1088/1674-1056/abbbf3

[16] Fermi, E., Pasta, J. and Ulam, S. (1965) Los Alamos Report LA-1940 (1955), Published Later. In: Segr, E., Ed., *Collected Papers of Enrico Fermi*, University of Chicago Press, Chicago.

[17] Nfor, N.O., Ghomsi, P.G. and Moukam Kakmeni, F.M. (2018) *Physical Review E*, **97**, Article ID: 022214. https://doi.org/10.1103/PhysRevE.97.022214

[18] Achu, G.F., Mkam, S.E., Moukam Kakmeni, F.M. and Tchawoua, C. (2018) *Physical Review E*, **98**, Article ID: 022216. https://doi.org/10.1103/PhysRevE.98.022216

[19] Nfor, N.O. and Mokoli, M.T. (2016) *Journal of Modern Physics*, **7**, 1166-1180. https://doi.org/10.4236/jmp.2016.710106

[20] Achu, G.F., Moukam Kakmeni, F.M. and Dikandé, A.M. (2018) *Physical Review E*, **97**, Article ID: 012211. https://doi.org/10.1103/PhysRevE.97.012211

[21] Moukam Kakmeni, F.M., Inack, E.M. and Yamakou, E.M. (2014) *Physical Review E*, **89**, Article ID: 052919. https://doi.org/10.1103/PhysRevE.89.052919

[22] Agrawal, G.P. (2001) Nonlinear Fiber Optics. Third Edition, Academic Press, Cambridge.

[23] Kivshar, Y.S. and Agrawal, G.P. (2003) Optical Solitons: From Fibers to Photonic Crystals. Academic Press, San Diego. https://doi.org/10.1016/B978-012410590-4/50012-7

[24] Fandio Jubgang, D.J., Dikandé, A.M. and Sunda-Meya, A. (2015) *Physical Review A*, **92**, Article ID: 053850. https://doi.org/10.1103/PhysRevA.92.053850

[25] Okaly, J.B., Mvogo, A., Tabi, C.B., Ekobena Fouda, H.P. and Kofané, T.C. (2020) *Physical Review E*, **102**, Article ID: 062402. https://doi.org/10.1103/PhysRevE.102.062402

[26] Ying-Bo, Y., Xiao-Yun, W. and Bing, T. (2016) *Journal of Biological Physics*, **42**, 213-222. https://doi.org/10.1007/s10867-015-9402-3

[27] Bang, O. and Peyrard, M. (1996) *Physical Review E*, **53**, 4143. https://doi.org/10.1103/PhysRevE.53.4143

[28] Ting, J.L. and Peyrard, M. (1996) *Physical Review E*, **53**, 1011.

[29] Englander, S.W., Kallenbach, N.R., Heeger, A.J., Krumhansl, J.A. and Litwin, S. (1980) *Proceedings of the National Academy of Sciences of the United States of America*, **77**, 7222. https://doi.org/10.1073/pnas.77.12.7222

[30] Yakushevich, L.V. (1989) *Physics Letters A*, **136**, 413-417. https://doi.org/10.1016/0375-9601(89)90425-8

[31] Yakushevich, L.V. (1994) *Physica D*, **79**, 77-86. https://doi.org/10.1016/0167-2789(94)90038-8

[32] Gaeta, S. (2006) *Physical Review E*, **74**, Article ID: 021921. https://doi.org/10.1103/PhysRevE.74.021921

[33] Gaeta G. (2007) *Journal of Nonlinear Mathematical Physics*, **14**, 57-81. https://doi.org/10.2991/jnmp.2007.14.1.6

[34] Yomosa, S. (1983) *Physical Review A*, **27**, 2120-2125. https://doi.org/10.1103/PhysRevA.27.2120

[35] Yomosa, S. (1984) *Physical Review A*, **30**, 474-480. https://doi.org/10.1103/PhysRevA.30.474

[36] Homma, S. and Takeno, S. (1984) *Progress of Theoretical Physics*, **72**, 679-693. https://doi.org/10.1143/PTP.72.679

[37] Campa, A. and Giansanti, A. (1998) *Physical Review E*, **58**, 3585. https://doi.org/10.1103/PhysRevE.58.3585

[38] Gninzanlong, C.L., Ndjomatchoua, F.T. and Tchawoua, C. (2018) *Chaos*, **28**, Article ID: 043105. https://doi.org/10.1063/1.5009147

[39] Ndjoko, P.B., Bilbault, J.M., Binczak, S. and Kofané, T.C. (2012) *Physical Review E*, **85**, Article ID: 011916. https://doi.org/10.1103/PhysRevE.85.011916

[40] Vargas, E.V., Ludu, A., Hustert, R., Gumrich, P., Jackson, A.D. and Heimburg, T. (2011) *Biophysical Chemistry*, **153**, 159. https://doi.org/10.1016/j.bpc.2010.11.001

[41] Nde Aboringong, E.N. and Dikandé, A.M. (2018) *The European Physical Journal E*, **41**, 35. https://doi.org/10.1140/epje/i2018-11640-9

[42] Daniel, M. and Vasumathi, V. (2008) *Physics Letters A*, **372**, 5144. https://doi.org/10.1016/j.physleta.2008.05.063

[43] Okaly, J.B., Ndzana, F.-II, Woulaché, R.L., Tabi, C.B. and Kofané, T.C. (2019) *Chaos*, **29**, Article ID: 093103. https://doi.org/10.1063/1.5098341

[44] Jonathan, A.D.W. (1998) *Journal of Physics A*: *Mathematical and General*, **31**, 3301-3323. https://doi.org/10.1088/0305-4470/31/14/016

[45] Dikandé, A.M. (1999) *Physica Scripta*, **60**, 291. https://doi.org/10.1238/Physica.Regular.060a00291

[46] Plopper, G. (2012) Principles of Cell Biology. Jones and Bartlett Learning, Burling-

ton.

[47] Balajee, A.S. and Phil, M. (2006) DNA Repair and Human Disease. Springer Science, New York.

[48] Fakhretdinov, M.I. and Zakiŕyanov, F.K. (2012) *Russian Physics Journal*, **54**, 1304. https://doi.org/10.1007/s11182-012-9747-z

[49] Chevizovich, D., Michieletto, D., Mvogo, A., Zakiryanov, F. and Zdravković, S. (2020) *Royal Society Open Science*, **7**, Article ID: 200774. https://doi.org/10.1098/rsos.200774

Journal of Modern Physics, 2021, 12, 1866-1921
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Relativistic-Covariant Energy-Momentum Tensor for Homogeneous Anisotropic Dispersive Media

Alfred Wünsche

Institut für Physik, Humboldt-Universität, Berlin, Germany
Email: alfred.wuensche@physik.hu-berlin.de

How to cite this paper: Wünsche, A. (2021) Relativistic-Covariant Energy-Momentum Tensor for Homogeneous Anisotropic Dispersive Media. *Journal of Modern Physics*, **12**, 1866-1921.
https://doi.org/10.4236/jmp.2021.1213108

Received: October 7, 2021
Accepted: November 27, 2021
Published: November 30, 2021

Abstract

A new relativistically covariant approach is discussed for the derivation of local conservation theorems for homogeneous anisotropic and, in particular, dispersive media. We start from a three-dimensional operator equation for the electric field and obtain mainly by coordinate-invariant methods the results basically expressed by the slowly varying amplitudes of the electric field. Apart from local energy and momentum conservation formulated by the energy-momentum conservation, we find a local conservation theorem for the action which is more general and which is the only one which remains also true for inhomogeneous media.

Keywords

Spatial and Frequency Dispersion, Permittivity Tensor, Group Velocity, Cold Plasma, Coordinate-Invariant Methods, Lorentz Group, Lorentz Invariants

Some 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, $\boldsymbol{a} \cdot \boldsymbol{b}$ dyadic products;

Latin letters as indices, three-dimensional, e.g., $i = 1,2,3$, ε_{ijk} is Levi-Civita pseudo-tensor;

Greek letters as indices, four-dimensional, e.g., $\mu = 1,2,3,4$;

$\alpha \equiv 4\pi$ (for *CGS* or Gauss system of units, α is often in the denominator of formulae).

1. Introduction

The four-dimensional energy-momentum tensor was introduced 1913 by Einstein with notation $\Theta_{\mu\nu}$ in [1] (republished in [2]) for the relativistically covariant generalization of the local conservation theorems of energy and momentum in differential form[1]. Three-dimensional parts of this tensor (Maxwell stress tensor) were already known before and used earlier. These local conservation theorems of energy and momentum are as it is well known a consequence of the homogeneity or translation invariance of (three-dimensional Euclidean) space and of time.

Beginning from 1915 Einstein published his General relativity theory (see, e.g., [2]) in which the symmetric energy-momentum tensor $T_{\mu\nu}$ is a kind of source term for curvature of space-time expressed by the symmetric Ricci tensor $R_{\mu\nu}$ which results from Riemann curvature tensor which is a four-valent tensor by contraction over two of its four indices [3] [4]. Due, in particular, to Pauli [5] [6] who wrote in his very young year of 21 an encyclopedic article about the new General relativity theory which requires a symmetric energy-momentum tensor a long discussion of the right energy-momentum tensor, the symmetric Abraham tensor or the non-symmetric Minkowski tensor, began where Pauli brought arguments in favor of the Abraham tensor. However, the Abraham tensor was only derived for isotropic media without taking into account dispersion and a more general symmetric tensor for all subgroups of the three-dimensional orthogonal group cannot exist. Pauli obviously recognized his incorrectness and in a later republication of his article shortly before his death he corrected it in the remark (see Section 13 here). Now the problem of the right energy-momentum tensor for media seems to be decided in favor of the Minkowski tensor.

Our main older sources for considerations to the energy-momentum tensor and its parts were, in particular, Landau and Lifshits [3] [7], Agranovich and Ginzburg [8] [9] [10] and Silin and Rukhadze [11] where, in particular, the dispersion was taken into account and apart from already cited encyclopedic article of Pauli, the work of Tolman [12], Sommerfeld [13], von Laue [14], Fock [15], Skobeltsyn [16] and Ugarov [17] [18], the (astonishingly modern) work of Tamm [19] (1st Ed. 1929) and the monographs of Møller [20] and of Jackson [21] for the older development of electrodynamics and Relativity theory. The number of interesting text-books with treatment of these topics grew rapidly in the following time among them, e.g., [22] [23] [24] [25] [26] which organically take into account dispersion. Long before but less known is the coordinate-invariant approach which was developed, first, mainly in the work of F.I. Fyodorov [27] [28] [29] (see also [30] [31]). Many articles of later time did not use these methods for reflection and refraction problems where, in particular, they are advantageous and derived with other (mostly coordinate) methods new re-

[1]Contravariant and covariant components of tensors are not yet distinguished in this article by upper and lower indices and $\Theta_{\mu\nu}$ is meant as contravariant tensor. The corresponding covariant energy-momentum tensor with notation $T_{\mu\nu}$ is introduced there by the relation $T_{\mu\nu} = g_{\mu\alpha} g_{\nu\beta} \Theta_{\alpha\beta}$ with $g_{\mu\nu}$ the covariant metric tensor ($\gamma_{\mu\nu}$ corresponding contravariant tensor). Both these tensors are symmetric ones. All this was prepared for the transition to curvilinear coordinates which is discussed in Mathematical Part II written together with M. Grossmann.

sults or rederived older results. In [32] [33] were given new arguments in favor of the Abraham tensor.

Apart from Abraham or Minkowski tensor there exist yet other serious problems with the energy-momentum tensor. As it is well known, the energy-momentum tensor is not uniquely determined by the requirement to satisfy a differential conservation theorem, e.g., [3] [20]. Practically, one has to do in every case of a calculated energy-momentum tensor with a whole class of equivalent energy-momentum tensors. One main use of this non-uniqueness is the possibility to remove the highly oscillatory terms in the approximation to waves with slowly varying amplitudes (beams) if we first insert the whole electromagnetic field with a positive and corresponding negative peak in the frequency distribution. We discuss this in Section 14 but mention already here that the energy-momentum tensor for media (only vacuum excluded) is basically non-symmetric and that by using the non-uniqueness it cannot be reduced to a symmetric one. Another difficulty for dispersive media is that their energy-momentum tensor cannot be derived starting from a Lagrange function as consequence of translation invariance of the medium in space and time (Noether theorem) as it is standard for electrodynamics of the vacuum. For inhomogeneous media the energy-momentum tensor does not exist at all in a local conservation law.

In present article we consider first the equations of macroscopic electrodynamics as averaged from microscopic electrodynamics (Section 2) then the constitutive equations in the concept of media with spatial and frequency dispersion and the symmetry of the permittivity tensor for neglect of dissipation (Section 3) and its symmetry under presence of discrete symmetries (space inversion, time inversion and their product, Section 4). Then we derive a three-dimensional operator equation for the electric field (Section 5) which is relativistically covariant but on the first glance it may seem to be paradoxical that this is possible. From this operator equation we derive a local conservation theorem for action (Section 6) and for energy and momentum (Section 7) and discuss the obtained energy-momentum tensor (Section 8). A peculiarity of our approach is that we obtain basically all results expressed by the electric field alone and after limiting transition to plane monochromatic waves we make the transition to more usual representation by the electric and magnetic field (Section 9). Then we consider the role which the group velocity plays (Section 10). After this we discuss the neglect of dispersion and the calculation of the group velocity in this case (Section 11). Next, we consider the special case of a cold plasma (Section 12). In the discussion of controversial opinions to local conservation of angular momentum we show that a complete symmetry of the four-dimensional energy-momentum tensor is not necessary but only symmetry of the stress tensor (Section 13). The non-uniqueness of the energy-momentum tensor is considered under new aspects (Section 14). Connected with the general non-symmetry of the energy-momentum tensor arise some difficulties for the General relativity theory (Section 15) and, finally, we mention some possibilities for generalizations of the discussed material (Section 16). In two Appendices we have separated the derivation of the relativistic covariance of our approach and the transformation of the energy-momentum tensor under special Lorentz transformations. A short paper to some of these problems can be found in [34] (see "Remark" and "Acknowledgements" at the end of present article).

2. Maxwell Equations of Macroscopic Electrodynamics in Two Concepts

The basis of our derivations is the following Maxwell equations of macroscopic electrodynamics

$$\begin{aligned}&\left[\nabla,\boldsymbol{E}(\boldsymbol{r},t)\right]+\frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{B}(\boldsymbol{r},t)=\boldsymbol{0},\quad \nabla\boldsymbol{B}(\boldsymbol{r},t)=0\\&\left[\nabla,\boldsymbol{B}(\boldsymbol{r},t)\right]-\frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{E}(\boldsymbol{r},t)=\frac{4\pi}{c}\frac{\partial}{\partial t}\boldsymbol{P}(\boldsymbol{r},t),\quad \nabla\boldsymbol{E}(\boldsymbol{r},t)=-4\pi\nabla\boldsymbol{P}(\boldsymbol{r},t),\end{aligned}\tag{2.1}$$

where $\boldsymbol{E}$ and $\boldsymbol{B}$ are the averaged electric field $\boldsymbol{E}^{\text{micro}}$ and magnetic field $\boldsymbol{B}^{\text{micro}}$ of microscopic electrodynamics in the electron theory of Lorentz [35][2] (republ. in [36]; see also [37] [38]) in the sense of the transition from microscopic to macroscopic electrodynamics ([7], de Groot and Suttorp [39], (II. section 3))

$$\boldsymbol{E}(\boldsymbol{r},t)\equiv\overline{\boldsymbol{E}^{\text{micro}}}(\boldsymbol{r},t),\quad \boldsymbol{B}(\boldsymbol{r},t)\equiv\overline{\boldsymbol{B}^{\text{micro}}}(\boldsymbol{r},t).\tag{2.2}$$

This transition can include different averaging processes, for example, spatial, temporal and statistical ones (denoted by overlining of the corresponding quantity). The averaged microscopic current density $\boldsymbol{j}^{\text{micro}}$ and charge density ρ^{micro} are expressed by only one macroscopic quantity $\boldsymbol{P}$, called polarization, in the rank of following definition

$$\overline{\boldsymbol{j}^{\text{micro}}}(\boldsymbol{r},t)\equiv\frac{\partial}{\partial t}\boldsymbol{P}(\boldsymbol{r},t),\quad \overline{\rho^{\text{micro}}}(\boldsymbol{r},t)\equiv-\nabla\boldsymbol{P}(\boldsymbol{r},t),\tag{2.3}$$

where the necessary validity of the continuity equation for microscopic current and charge densities

$$\nabla\overline{\boldsymbol{j}^{\text{micro}}}(\boldsymbol{r},t)+\frac{\partial}{\partial t}\overline{\rho^{\text{micro}}}(\boldsymbol{r},t)=0,\tag{4}$$

is taken into account. Such an identification is possible in almost all cases with exception of some static cases (e.g., electrostatics, stationary currents, magnetostatics) which have to be considered in this concept as limiting cases. As usual, we define the "electric induction"[3] $\boldsymbol{D}(\boldsymbol{r},t)$ by

$$\boldsymbol{D}(\boldsymbol{r},t)\equiv\boldsymbol{E}(\boldsymbol{r},t)+4\pi\boldsymbol{P}(\boldsymbol{r},t).\tag{2.5}$$

From vectorial equations in (2.1) follows then by forming the divergence

[2]Lorentz denotes microscopic fields with small letters corresponding to the Capital letters commonly used in macroscopic electrodynamics (but $\boldsymbol{d}$ instead of $\boldsymbol{e}$ for microscopic electric field). Most authors use $\boldsymbol{H}^{\text{micro}}$ instead of $\boldsymbol{B}^{\text{micro}}$ but since there is no difference between them in microscopic theory this is only of some didactic importance.

[3]We follow here in terminology Landau and Lifshits [7] (Russian editions) to distinguish the "electric induction" from such notions as, e.g., "dielectric displacement" which are mostly used in a more special sense.

$$\frac{\partial}{\partial t}\nabla \boldsymbol{B}(\boldsymbol{r},t)=\boldsymbol{0},\quad \frac{\partial}{\partial t}\nabla \boldsymbol{D}(\boldsymbol{r},t)=0, \tag{2.6}$$

that are the scalar equations in (2.1) differentiated with respect to the time. Therefore with exclusion of the static limiting case, the scalar equations from (2.1) in

$$\begin{aligned}&\left[\nabla,\boldsymbol{E}(\boldsymbol{r},t)\right]+\frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{B}(\boldsymbol{r},t)=0,\quad \nabla\boldsymbol{B}(\boldsymbol{r},t)=0,\\&\left[\nabla,\boldsymbol{B}(\boldsymbol{r},t)\right]-\frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{D}(\boldsymbol{r},t)=0,\quad \nabla\boldsymbol{D}(\boldsymbol{r},t)=0,\end{aligned} \tag{2.7}$$

are redundant and the only field equations to take into account are the vectorial equations.

One can often find in literature forms of the equations of macroscopic electrodynamics where the averaged current and charge density are not fully included only into one quantity $\boldsymbol{P}(\boldsymbol{r},t)$ called polarization and defined by (2.3) but into some different quantities, for example, electric polarization in a more special sense $\boldsymbol{P}'(\boldsymbol{r},t)$ and magnetization $\boldsymbol{M}(\boldsymbol{r},t)$ according to, e.g., de Groot and Suttorp [39] and Bloembergen [40] (chap. 3, Eqs. (3.3), (3.5)) (the "Netherland school" together with H.A. Lorentz) but also many other authors

$$\begin{aligned}&\overline{\boldsymbol{j}^{\text{micro}}}(\boldsymbol{r},t)\equiv\frac{\partial}{\partial t}\boldsymbol{P}'(\boldsymbol{r},t)+c\left[\nabla,\boldsymbol{M}(\boldsymbol{r},t)\right]+\cdots,\\&\overline{\rho^{\text{micro}}}((\boldsymbol{r},t)\equiv-\nabla\boldsymbol{P}'(\boldsymbol{r},t)+\cdots,\end{aligned} \tag{2.8}$$

and in nonlinear optics sometimes additionally into electric quadrupole density and higher electric and magnetic multipole densities (indicated by additional points). They can be joined then in different ways with the electric field $\boldsymbol{E}(\boldsymbol{r},t)$ and magnetic field $\boldsymbol{B}(\boldsymbol{r},t)$ to new quantities, for example, to a more special dielectric displacement $\boldsymbol{D}'(\boldsymbol{r},t)$ and to a new field $\boldsymbol{H}(\boldsymbol{r},t)$ as follows

$$\boldsymbol{D}'(\boldsymbol{r},t)\equiv\boldsymbol{E}(\boldsymbol{r},t)+4\pi\boldsymbol{P}'(\boldsymbol{r},t),\quad \boldsymbol{H}(\boldsymbol{r},t)\equiv\boldsymbol{B}(\boldsymbol{r},t)-4\pi\boldsymbol{M}(\boldsymbol{r},t), \tag{2.9}$$

which obey then the following field equations instead of (2.7)

$$\begin{aligned}&\left[\nabla,\boldsymbol{E}(\boldsymbol{r},t)\right]+\frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{B}(\boldsymbol{r},t)=\boldsymbol{0},\quad \nabla\boldsymbol{B}(\boldsymbol{r},t)=0,\\&\left[\nabla,\boldsymbol{H}(\boldsymbol{r},t)\right]-\frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{D}'(\boldsymbol{r},t)=\boldsymbol{0},\quad \nabla\boldsymbol{D}'(\boldsymbol{r},t)=0.\end{aligned} \tag{2.10}$$

The field $\boldsymbol{H}(\boldsymbol{r},t)$ is mostly called magnetic field but it is not the averaged microscopic magnetic field $\overline{\boldsymbol{B}^{\text{micro}}}(\boldsymbol{r},t)$ and therefore not the "genuine" magnetic field [7] (chap IV, section 29, after Eq. (29.8)) (in our treatment with equations (2.7) we have $\boldsymbol{H}\equiv\boldsymbol{B}$ and $\boldsymbol{M}$ is included into $\boldsymbol{P}$). However, the greater symmetry of (2.10) in comparison to (2.7) is deceptive and to fix the separation (2.8) is then difficult and not fully unique without additional conventions, in particular, for high frequencies under presence of dispersion in the constitutive equations. To see this we recommend to read also the very instructive section 79 in Landau and Lifshits [7]. All such more special schemes can be transformed to the general scheme which we prefer and which is characterized by Equations

(2.3), (2.5) and (2.7) and by constitutive relations considered in next Section and only some static cases have to be dealt with then as limiting cases.

3. Linear Constitutive Equations for Homogeneous Anisotropic Dispersive Media

The Maxwell equations of macroscopic electrodynamics (2.1) or (2.7) form a closed system of equations only together with constitutive equations which depend on the kind of the considered medium. We discuss this now.

The general linear constitutive relation between the electric induction $\boldsymbol{D}(\boldsymbol{r},t)$ and the electric field $\boldsymbol{E}(\boldsymbol{r},t)$ for spatially and temporally homogeneous media is (summation convention over equal indices; $\boldsymbol{\rho} \equiv \boldsymbol{r}-\boldsymbol{r}'$, $\tau = t-t'$; $\exp\left(-\boldsymbol{\rho}\nabla - \tau\frac{\partial}{\partial t}\right)$ is displacement operator of arguments of a function of $(\boldsymbol{r},t)$ to $(\boldsymbol{r}-\boldsymbol{\rho},t-\tau)$)

$$\begin{aligned} D_i(\boldsymbol{r},t) &= \int \mathrm{d}^3\rho \wedge \mathrm{d}\tau\, \hat{\varepsilon}_{ij}(\boldsymbol{\rho},\tau) E_j(\boldsymbol{r}-\boldsymbol{\rho},t-\tau) \\ &= \int \mathrm{d}^3\rho \wedge \mathrm{d}\tau\, \hat{\varepsilon}_{ij}(\boldsymbol{\rho},\tau) \exp\left(-\boldsymbol{\rho}\nabla - \tau\frac{\partial}{\partial t}\right) E_j(\boldsymbol{r},t), \end{aligned} \tag{3.1}$$

where the real-valued tensor function $\hat{\varepsilon}_{ij} = \hat{\varepsilon}_{ij}(\boldsymbol{\rho},\tau)$ characterizes the material properties and where the integration is written as going over the whole space-time and restrictions of this integration (e.g., to prehistory, causality) are thought to be included by vanishing of this tensor function in certain regions. These restrictions, for example, to the prehistory of the field evolution lead to properties of analyticity and thus to relations between real and imaginary part of the Fourier transform of $\hat{\varepsilon}_{ij}(\boldsymbol{\rho},\tau)$ which are called Kramers-Kronig relations which we do not discuss here (e.g., [7]). Relation (3.1) means that the most general linear constitutive relations are also nonlocal in space that describes the spatial dispersion[4]. The homogeneity of the medium is expressed by the property that $\hat{\varepsilon}_{ij}(\boldsymbol{\rho},\tau)$ does not explicitly depend on the considered space-time point $(\boldsymbol{r},t)$ but only on the differences $(\boldsymbol{\rho},\tau)$ to this point.

Using the Fourier transform $\varepsilon_{ij}(\boldsymbol{k},\omega)$ of $\hat{\varepsilon}_{ij}(\boldsymbol{\rho},\tau)$ according to [9] [10]

$$\begin{aligned} \varepsilon_{ij}(\boldsymbol{k},\omega) &= \int \mathrm{d}^3\rho \wedge \mathrm{d}\tau\, \hat{\varepsilon}_{ij}(\boldsymbol{\rho},\tau) \mathrm{e}^{-\mathrm{i}(\boldsymbol{k\rho}-\omega\tau)}, \\ \hat{\varepsilon}_{ij}(\boldsymbol{\rho},\tau) &= \frac{1}{(2\pi)^4}\int \mathrm{d}^3k \wedge \mathrm{d}\omega\, \varepsilon_{ij}(\boldsymbol{k},\omega) \mathrm{e}^{\mathrm{i}(\boldsymbol{k\rho}-\omega\tau)}, \end{aligned} \tag{3.2}$$

the constitutive Equation (3.1) can be represented by

$$D_i(\boldsymbol{r},t) = \varepsilon_{ij}\left(-\mathrm{i}\nabla, \mathrm{i}\frac{\partial}{\partial t}\right) E_j(\boldsymbol{r},t), \tag{3.3}$$

where $\varepsilon_{ij}(\boldsymbol{k},\omega)$ denotes the complex permittivity tensor. After Fourier transformation of the electric field $\boldsymbol{E}(\boldsymbol{r},t)$ (analogously $\boldsymbol{D}(\boldsymbol{r},t)$ and $\boldsymbol{B}(\boldsymbol{r},t)$) ac-

[4]According to Silin and Rukhadse [11] (p.14) the notion "spatial dispersion" was introduced by Gertsenshteyn. Clearly, the name "spatial dispersion" is not analogous to "frequency dispersion" which then has to be better named "temporal dispersion" or vice versa the "spatial dispersion" then "wave-vector dispersion".

cording to

$$\begin{aligned} \boldsymbol{E}(\boldsymbol{k},\omega) &= \int \mathrm{d}^3 r \wedge \mathrm{d}t\, \boldsymbol{E}(\boldsymbol{r},t)\, \mathrm{e}^{-\mathrm{i}(\boldsymbol{kr}-\omega t)}, \\ \boldsymbol{E}(\boldsymbol{r},t) &= \frac{1}{(2\pi)^4} \int \mathrm{d}^3 k \wedge \mathrm{d}\omega\, \boldsymbol{E}(\boldsymbol{k},\omega)\, \mathrm{e}^{\mathrm{i}(\boldsymbol{kr}-\omega t)}, \end{aligned} \tag{3.4}$$

the constitutive relation (3.3) takes on the well-known form (e.g., [7] [8] [9] [10] [11] and others)

$$D_i(\boldsymbol{k},\omega) = \varepsilon_{ij}(\boldsymbol{k},\omega) E_j(\boldsymbol{k},\omega). \tag{3.5}$$

The dispersion of the medium is here expressed by the dependence of the permittivity tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)$ on wave vector $\boldsymbol{k}$ and frequency ω and the anisotropy by its tensor character.

The electric field $\boldsymbol{E}(\boldsymbol{r},t)$ and the electric induction $\boldsymbol{D}(\boldsymbol{r},t)$ are real quantities. From this follows for the Fourier transform of the electric field $\boldsymbol{E}(\boldsymbol{r},t)$ (analogously $\boldsymbol{D}(\boldsymbol{r},t)$ and $\boldsymbol{B}(\boldsymbol{r},t)$)

$$\boldsymbol{E}(\boldsymbol{r},t) = \left(\boldsymbol{E}(\boldsymbol{r},t)\right)^*, \quad \Leftrightarrow \quad \boldsymbol{E}(\boldsymbol{k},\omega) = \left(\boldsymbol{E}\left(-\boldsymbol{k}^*,-\omega^*\right)\right)^*. \tag{3.6}$$

As a consequence, the permittivity tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)$ possesses the general symmetry property

$$\varepsilon_{ij}(\boldsymbol{k},\omega) = \left(\varepsilon_{ij}\left(-\boldsymbol{k}^*,-\omega^*\right)\right)^*. \tag{3.7}$$

Local or differential conservation laws of energy and momentum can only be derived under the condition that the medium is lossless which means that it does not have any dissipation or accumulation or transmission of energy and momentum to other frequencies and wave vectors. As the later derivations show, the condition for this is the following symmetry

$$\hat{\varepsilon}_{ij}(\boldsymbol{\rho},\tau) = \hat{\varepsilon}_{ji}(-\boldsymbol{\rho},-\tau), \tag{3.8}$$

which after Fourier transformation according to (3.2) and in connection with (3.7) takes on the following form

$$\varepsilon_{ij}(\boldsymbol{k},\omega) = \varepsilon_{ji}(-\boldsymbol{k},-\omega) = \left(\varepsilon_{ji}\left(\boldsymbol{k}^*,\omega^*\right)\right)^*, \tag{3.9}$$

and which for dispersive media can only be satisfied approximately for certain regions of wave vector and frequency. Such kind of conditions are closely related to Onsager conditions for quasi-stationary processes [7] (section 21, Ed. 1982) but instead of a rigorous derivation from basic principles we prefer again that one can conclude this from the necessary conditions for the most general possibility of derivation of differential conservation laws of energy and momentum. These conditions should not be confused with the influence of point group symmetries on the medium properties which additionally may be present or may not. The symmetry conditions which are related to different inversion symmetries are discussed in next Section.

To treat spatial dispersion it is mostly appropriate to make a Taylor-series expansion of $\varepsilon_{ij}(\boldsymbol{k},\omega)$ in powers of the full wave-vector $\boldsymbol{k}$ (*i.e.*, at $\boldsymbol{k}_0 = \boldsymbol{0}$)

with notations in [9] (p. 155) and for the inverse tensor $\varepsilon_{ij}^{-1}(\boldsymbol{k},\omega)$ in the new chapter XII in [7] (Ed. 1982) (Landau and Lifshits prefer there mainly to work with this inverse tensor)

$$\begin{aligned}\varepsilon_{ij}(\boldsymbol{k},\omega) &= \varepsilon_{ij}(\omega)+\mathrm{i}\gamma_{ijk}(\omega)k_k+\alpha_{ijkl}(\omega)k_k k_l+\cdots,\\ \varepsilon_{ij}^{-1}(\boldsymbol{k},\omega) &= \varepsilon_{ij}^{-1}(\omega)+\mathrm{i}\delta_{ijk}(\omega)k_k+\beta_{ijkl}(\omega)k_k k_l+\cdots,\end{aligned} \tag{3.10}$$

that means for the expansion (3.10)

$$\varepsilon_{ij}(\omega)=\varepsilon_{ji}(-\omega),\quad \gamma_{ijk}(\omega)=-\gamma_{jik}(-\omega),\quad \alpha_{ijkl}(\omega)=\alpha_{jikl}(-\omega),\cdots. \tag{3.11}$$

The approximate transition from one to the inverse tensor in (3.10) is easily to make.

We mention here shortly that in the most common treatment of macroscopic electrodynamics with two constitutive equations $D_i(\boldsymbol{k},\omega)=\varepsilon_{ij}(\omega)E_j(\boldsymbol{k},\omega)$ and $B_i(\boldsymbol{k},\omega)=\mu_{ji}(\omega)H_j(\boldsymbol{k},\omega)$ the more general tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)$ corresponds to the special one (ε_{ijk} is Levi-Civita pseudo-tensor)

$$\varepsilon_{ij}(\boldsymbol{k},\omega)=\varepsilon_{ij}(\omega)+\frac{c^2}{\omega^2}\varepsilon_{ikm}\varepsilon_{jln}\left(\delta_{mn}-\mu_{mn}^{-1}(\omega)\right)k_k k_l. \tag{3.12}$$

This shows that a possible magnetization appears here as effect of spatial dispersion of second order in wave-vector $\boldsymbol{k}$ that is important for the energy-momentum tensor and also for the boundary conditions at such medium. Furthermore, we see that the tensor $\alpha_{ijkl}(\omega)$ in (3.10) is more general and usually contains more non-vanishing terms than this special tensor proportional to $k_k k_l$ in (3.12)[5]. Only for magnetostatics this concept not used in present article is less appropriate.

4. Additional Restrictions of Tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)$ for Discrete Symmetries of Spatial and Time Inversion

We now consider the most simple discrete symmetries of order 2.

1) Spatial inversion (presence of symmetry center)

The presence of spatial inversion that means invariance of the medium with respect to the transformation $\boldsymbol{r}\to-\boldsymbol{r}$ of the coordinates where due to the homogeneity the chosen coordinate origin is arbitrary and due to the property of $\boldsymbol{E}$ and $\boldsymbol{D}$ to be genuine vectors changing their sign under this transformation (in contrast, $\boldsymbol{B}$ is a pseudo-vector) leads to

$$\varepsilon_{ij}(\boldsymbol{k},\omega)=\varepsilon_{ij}(-\boldsymbol{k},\omega)=\left(\varepsilon_{ij}\left(\boldsymbol{k}^*,-\omega^*\right)\right)^*, \tag{4.1}$$

where in the second step (3.7) was used in addition. The first part of this condition means that for media with spatial inversion the components of the permittivity tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)$ are mutually independent from each other and are even functions of the wave vector $\boldsymbol{k}$, whereas it does not mean a restriction for its dependence on the frequency ω. In composition with the condition (3.9) for absence of dissipation we have

[5]For a deeper understanding we recommend here again the very instructive section 79 in [7] (Ed. 1982).

$$\varepsilon_{ij}(\boldsymbol{k},\omega)=\varepsilon_{ij}(-\boldsymbol{k},\omega)=\varepsilon_{ji}(\boldsymbol{k},-\omega)=\varepsilon_{ji}(-\boldsymbol{k},\omega), \tag{4.2}$$

which relates different components of the permittivity tensor and is not true for regions of $\boldsymbol{k}$ and ω where dissipation is not negligible. If we deal with spatial dispersion by expansion of $\varepsilon_{ij}(\boldsymbol{k},\omega)$ in powers of the wave vector $\boldsymbol{k}$ then we can use mainly the first part $\varepsilon_{ij}(\boldsymbol{k},\omega)=\varepsilon_{ij}(-\boldsymbol{k},\omega)$ of these symmetry conditions which are true also in case of dissipation.

2) Time inversion (nonmagnetic symmetry classes)

The presence of time inversion that means of invariance of the medium with respect to the transformation $t\to -t$ taking into account that $\boldsymbol{E}$ and $\boldsymbol{D}$ do not change their sign under this transformation (in contrast, $\boldsymbol{B}$ changes it) leads to

$$\varepsilon_{ij}(\boldsymbol{k},\omega)=\varepsilon_{ij}(\boldsymbol{k},-\omega)=\left(\varepsilon_{ij}\left(-\boldsymbol{k}^{*},\omega^{*}\right)\right)^{*}, \tag{4.3}$$

where in addition (3.7) is used in last equality. The first part of this condition means that for media with spatial inversion the components of the permittivity tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)$ are mutually independent from each other even functions of the frequency ω, whereas it does not mean a restriction for its dependence on the wave vector $\boldsymbol{k}$. In composition with the condition (3.9) for absence of dissipation we have here

$$\varepsilon_{ij}(\boldsymbol{k},\omega)=\varepsilon_{ij}(\boldsymbol{k},-\omega)=\varepsilon_{ji}(-\boldsymbol{k},\omega)=\varepsilon_{ji}(-\boldsymbol{k},-\omega). \tag{4.4}$$

In expansions of the permittivity tensor $\varepsilon_{ij}(k,\omega)$ in powers of $\boldsymbol{k}$ one may use here mainly the part $\varepsilon_{ij}(\boldsymbol{k},\omega)=\varepsilon_{ji}(-\boldsymbol{k},\omega)$ for simplifications which, however, are true only under neglect of dissipation.

3) Product of spatial inversion with time inversion (nongyrotropic media)

The presence of the product of spatial inversion with time inversion (including, evidently, the case of presence of both symmetry elements separately and therefore also of their product) leads to the symmetry

$$\varepsilon_{ij}(\boldsymbol{k},\omega)=\varepsilon_{ij}(-\boldsymbol{k},-\omega)=\left(\varepsilon_{ij}\left(\boldsymbol{k}^{*},\omega^{*}\right)\right)^{*}, \tag{4.5}$$

where again the condition (3.7) is used in last equality. In composition with the condition for absent dissipation (3.9) we find

$$\varepsilon_{ij}(\boldsymbol{k},\omega)=\varepsilon_{ij}(-\boldsymbol{k},-\omega)=\varepsilon_{ji}(-\boldsymbol{k},-\omega)=\varepsilon_{ji}(\boldsymbol{k},\omega). \tag{4.6}$$

The most interesting part of this relation $\varepsilon_{ij}(\boldsymbol{k},\omega)=\varepsilon_{ji}(\boldsymbol{k},\omega)$ describes complete symmetry of the permittivity tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)$ and this symmetry, by definition, is called non-gyrotropy of a medium and is connected with the symmetry element of product of spatial inversion with time inversion of the medium but not necessarily with both symmetries separately. Clearly, a non-gyrotropic medium possesses this symmetry only in regions of wave vector and frequency where dissipation which is not included in the symmetric part can be neglected. On the other side, gyrotropic media are such media which do not possess this symmetry of the permittivity tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)$.

Usually, if nothing is said in crystal optics it is meant that the medium possesses time inversion as symmetry element in addition to the spatial symmetry elements of one of the considered 32 crystal classes which are then called non-magnetic crystal classes and for which (4.3) is true. From the 32 crystal classes 11 do not possess a symmetry center and 11 possess it and therefore the spatial inversion as symmetry element and thus are non-gyrotropic and (4.1) and (4.5) are true for them in addition. There are 90 magnetic crystal classes from which 32 are trivial ones and correspond to the usual crystal classes but without time inversion as symmetry element.

Furthermore, there exist 58 magnetic crystal classes which contain time inversion not directly as symmetry element but in the form of the product of time inversion with the elements of a coset to an invariant subgroup of one of the 11 groups with only rotations [7] [41]. The 90 magnetic classes form the basis for the symmetry classification of ferromagnetics and anti-ferromagnetics. This concerns natural absence of time inversion as symmetry element but this absence can be generated also artificially under the influence of the medium by an external magnetic field from a primarily non-magnetic class. In the same way, among the 122 crystal classes (magnetic and non-magnetic ones) there are 32 classes with symmetry center for which (4.1) is true and 90 without symmetry center. This is contained in a compact form in **Figure 1** copied from our paper [41].

5. Elimination of Magnetic Field and Three-Dimensional Operator Equation with Relativistic Covariance for the Electric Field

By differentiation of the second vectorial equation in (2.7) with respect to time and using the first vectorial equation, the magnetic field can be eliminated and using the constitutive Equation (3.3) we obtain the following equation for the electric field

$$\begin{aligned} 0 &= \left(\nabla_i \nabla_j - \nabla^2 \delta_{ij}\right) E_j(\boldsymbol{r},t) + \frac{1}{c^2}\frac{\partial^2}{\partial t^2} D_i(\boldsymbol{r},t) \\ &= \left\{\nabla_i \nabla_j - \nabla^2 \delta_{ij} + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\varepsilon_{ij}\left(-\mathrm{i}\nabla, \mathrm{i}\frac{\partial}{\partial t}\right)\right\} E_j(\boldsymbol{r},t). \end{aligned} \tag{5.1}$$

This equation for the electric field contains the full information about the electromagnetic field in the medium with exception of some static cases which have to be considered as limiting cases.

Since Equation (5.1) carries the full information about the electromagnetic field the conservation theorems may be derived from it that possesses considerable advantages, in particular, taking into account the dispersion as we will demonstrate this in the following. For such derivations, roughly speaking, we have to multiply this equation from the left with other electric fields $E_i'(\boldsymbol{r},t)$. However, in this way we do not get expressions which are relativistic-covariant (they multiply by factors under Lorentz transformations). This shortage can be removed if we divide this equation by $\frac{1}{c^2}\frac{\partial^2}{\partial t^2}$ as will be shown in Appendix A that makes

Die 122 verallgemeinerten Kristallklassen

		Punktgruppen erster Art	Direkte Produkte von G mit $\bar{1}, \underline{1}, \underline{\bar{1}}$			Direktes Produkt mit $\bar{1}\underline{1}$	Untergruppen von G vom Index 2	Punktgruppen G' (isomorph zu G)			Direkte Produkte von G' mit $\bar{1}, \underline{1}, \underline{\bar{1}}$			Punktgruppen G'' (isomorph zu G)				
	Elementezahl	n	$2n$	$2n$	$2n$	$4n$	Hilfsspalte	n	n	n	$2n$	$2n$	$2n$	n	Hilfsspalten			
Kristallsystem	n	G^*	$G \times \bar{1}^*$	$G \times \underline{1}$	$G \times \underline{\bar{1}}$	$G \times \bar{1} \times \underline{1}$	H	$H+\bar{1}(G-H)^*$	$H+\underline{1}(G-H)$	$H+\underline{\bar{1}}(G-H)$	$\{H+\underline{1}(G-H)\}\times\bar{1} = \{H+\underline{\bar{1}}(G-H)\}\times\bar{1}$	$\{H+\bar{1}(G-H)\}\times\underline{1} = \{H+\underline{\bar{1}}(G-H)\}\times\underline{1}$	$\{H+\bar{1}(G-H)\}\times\underline{\bar{1}} = \{H+\underline{1}(G-H)\}\times\underline{\bar{1}}$	G''	$G=H_1+H_2+H_3+H_4$ $G''=H_1+\bar{1}H_2+\underline{1}H_3+\underline{\bar{1}}H_4$			
triklin	1	$1 \equiv C_1$	$\bar{1} \equiv C_i$	$\underline{1}$	$\underline{\bar{1}}$	$\bar{1}\underline{1}$												
monoklin	2	$2 \equiv C_2$	$2/m \equiv C_{2h}$	$2\underline{1}$	$2/\underline{m}$	$2/m\underline{1}$	$1 \equiv C_1$	$m \equiv C_s$	$\underline{2}$	$\underline{m}$	$\underline{2/m}$	$m\underline{1}$	$\underline{2}/m$		H_1	H_1+H_2	H_1+H_3	H_1+H_4
rhombisch	4	$222 \equiv D_2$	$mmm \equiv D_{2h}$	$222\underline{1}$	$\underline{m}\underline{m}\underline{m}$	$mmm\underline{1}$	$2 \equiv C_2$	$2mm \equiv C_{2v}$	$2\underline{2}\underline{2}$	$2\underline{m}\underline{m}$	$m\underline{m}\underline{m}$	$2mm\underline{1}$	$\underline{m}mm$	$\underline{2}\underline{m}m$	1	2	2	2
trigonal	3	$3 \equiv C_3$	$\bar{3} \equiv S_6$	$3\underline{1}$	$\underline{\bar{3}}$	$\bar{3}\underline{1}$												
	6	$32 \equiv D_3$	$\bar{3}m \equiv D_{3d}$	$32\underline{1}$	$\underline{\bar{3}}\underline{m}$	$\bar{3}m\underline{1}$	$3 \equiv C_3$	$3m \equiv C_{3v}$	$3\underline{2}$	$3\underline{m}$	$\bar{3}\underline{m}$	$3m\underline{1}$	$\underline{\bar{3}}m$					
tetragonal	4	$4 \equiv C_4$	$4/m \equiv C_{4h}$	$4\underline{1}$	$4/\underline{m}$	$4/m\underline{1}$	$2 \equiv C_2$	$\bar{4} \equiv S_4$	$\underline{4}$	$\underline{\bar{4}}$	$\underline{4}/m$	$\bar{4}\underline{1}$	$\underline{4}/\underline{m}$					
	8	$422 \equiv D_4$	$4/mmm \equiv D_{4h}$	$422\underline{1}$	$4/\underline{m}\underline{m}\underline{m}$	$4/mmm\underline{1}$	$4 \equiv C_4$	$4mm \equiv C_{4v}$	$4\underline{2}\underline{2}$	$4\underline{m}\underline{m}$	$4/m\underline{m}\underline{m}$	$4mm\underline{1}$	$4/\underline{m}mm$	$\bar{4}\underline{2}\underline{m}$	2	4	222	222
														$\underline{4}m\underline{m}$		222	4	222
							$222 \equiv D_2$	$\bar{4}2m \equiv D_{2d}$	$\underline{4}2\underline{2}$	$\underline{\bar{4}}2\underline{m}$	$\underline{4}/mm\underline{m}$	$\bar{4}2m\underline{1}$	$\underline{4}/\underline{m}\underline{m}m$	$\underline{\bar{4}}\underline{2}m$		222	222	4
hexagonal	6	$6 \equiv C_6$	$6/m \equiv C_{6h}$	$6\underline{1}$	$6/\underline{m}$	$6/m\underline{1}$	$3 \equiv C_3$	$\bar{6} \equiv C_{3h}$	$\underline{6}$	$\underline{\bar{6}}$	$\underline{6}/\underline{m}$	$\bar{6}\underline{1}$	$\underline{6}/m$					
	12	$622 \equiv D_6$	$6/mmm \equiv D_{6h}$	$622\underline{1}$	$6/\underline{m}\underline{m}\underline{m}$	$6/mmm\underline{1}$	$6 \equiv C_6$	$6mm \equiv C_{6v}$	$6\underline{2}\underline{2}$	$6\underline{m}\underline{m}$	$6/m\underline{m}\underline{m}$	$6mm\underline{1}$	$6/\underline{m}mm$	$\bar{6}\underline{m}\underline{2}$	3	6	32	32
														$\underline{6}m\underline{m}$		32	6	32
							$32 \equiv D_3$	$\bar{6}m2 \equiv D_{3h}$	$\underline{6}2\underline{2}$	$\underline{\bar{6}}\underline{m}2$	$\underline{6}/\underline{m}mm$	$\bar{6}m2\underline{1}$	$\underline{6}/mm\underline{m}$	$\underline{\bar{6}}m\underline{2}$		32	32	6
kubisch	12	$23 \equiv T$	$m3 \equiv T_h$	$23\underline{1}$	$\underline{m}3$	$m3\underline{1}$												
	24	$432 \equiv O$	$m3m \equiv O_h$	$432\underline{1}$	$\underline{m}3\underline{m}$	$m3m\underline{1}$	$23 \equiv T$	$\bar{4}3m \equiv T_d$	$\underline{4}3\underline{2}$	$\underline{\bar{4}}3\underline{m}$	$m3\underline{m}$	$\bar{4}3m\underline{1}$	$\underline{m}3m$					
ohne Symmetriezentrum		11					×	10							= 21*			= 90
				11	11		×		10	10		10	10	7	= 69			
mit Symmetriezentrum			11				×								= 11*			= 32
						11	×				10				= 21			
magnetische Klassen		11	11				×	10							= 32*			= 90
					11		×		10	10	10		10	7	= 58			
nichtmagnetische Klassen				11		11	×					10			= 32			= 32
gyrotrope Klassen		11	11				×	10							= 32*			= 90
				11			×		10	10	10	10		7	= 58			
nichtgyrotrope Klassen					11	11	×						10		= 32			= 32

* gewöhnliche Punktgruppen ohne Zeitinversion

Identitäten: $C_2 \equiv D_1$, $C_i \equiv S_2$, $C_s \equiv C_{1h} = C_{1V} \equiv S_1$, $D_2 \equiv V$, $C_{2h} \equiv D_{1d} \equiv S_{2V}$, $C_{2V} \equiv D_{1h}$, $D_{2h} \equiv V_h$, $D_{2d} \equiv V_d \equiv S_{4V}$, $C_{3h} \equiv S_3$, $S_6 \equiv C_{3i}$, $D_{3d} \equiv S_{6V}$.

Figure 1. The 122 magnetic and non-magnetic crystal classes (from [41]). Hilfsspalte = auxiliary column.

difficulties only in the limiting transition to vanishing frequencies. Thus starting from Maxwell equations and constitutive equations as intermediate step (5.1) we arrive at the following vectorial equation for the electric field

$$L_{ij}\left(-\mathrm{i}\nabla,\mathrm{i}\frac{\partial}{\partial t}\right)E_j\left(\boldsymbol{r},t\right)=0, \tag{5.2}$$

where the tensor operator $L_{ij}\left(-\mathrm{i}\nabla,\mathrm{i}\frac{\partial}{\partial t}\right)$ is defined in the following way

$$L_{ij}\left(-\mathrm{i}\nabla,\mathrm{i}\frac{\partial}{\partial t}\right)\equiv\frac{c^2\left(\nabla_i\nabla_j-\boldsymbol{\nabla}^2\delta_{ij}\right)}{\frac{\partial^2}{\partial t^2}}+\varepsilon_{ij}\left(-\mathrm{i}\nabla,\mathrm{i}\frac{\partial}{\partial t}\right). \tag{5.3}$$

The vectorial equation for the electric field (5.2) together with definition (5.3) forms a closed system of equations of macroscopic electrodynamics of homogeneous media and, moreover, is relativistic-covariant (contrary to (5.1)) and are appropriate for the derivation of the energy-momentum tensor. After Fourier transformation of the electric field according to (3.4) we obtain from (5.2) the equation for the Fourier components $\boldsymbol{E}\left(\boldsymbol{k},\omega\right)$ of the electric field and then the magnetic field $\boldsymbol{B}\left(\boldsymbol{k},\omega\right)$

$$L_{ij}\left(\boldsymbol{k},\omega\right)E_j\left(\boldsymbol{k},\omega\right)=0,\quad B_i\left(\boldsymbol{k},\omega\right)=\frac{c}{\omega}\varepsilon_{ijk}k_jE_k\left(\boldsymbol{k},\omega\right), \tag{5.4}$$

with the tensor operator $L_{ij}\left(\boldsymbol{k},\omega\right)$ in this equation defined by

$$\begin{aligned}L_{ij}\left(\boldsymbol{k},\omega\right)&\equiv\frac{c^2\left(k_ik_j-\boldsymbol{k}^2\delta_{ij}\right)}{\omega^2}+\varepsilon_{ij}\left(\boldsymbol{k},\omega\right)\\&=\frac{c^2}{\omega^2}\left\{k_ik_j-\left(\boldsymbol{k}^2-\frac{\omega^2}{c^2}\right)\delta_{ij}\right\}+4\pi\chi_{ij}\left(\boldsymbol{k},\omega\right),\end{aligned} \tag{5.5}$$

where $\chi_{ij}\left(\boldsymbol{k},\omega\right)\equiv\frac{\varepsilon_{ij}\left(\boldsymbol{k},\omega\right)-\delta_{ij}}{4\pi}$ is the general susceptibility tensor. Since the tensor $\chi_{ij}\left(\boldsymbol{k},\omega\right)$ may be a complicated function of the wave-vector $\boldsymbol{k}$ and, in particular, of the frequency ω it is hardly possible to write down a Lagrange function for the system and the usual formalism of derivation of the energy-momentum tensor from such function is almost impossible.

Equations (5.4) with operator (5.5) as transformed Equation (5.2) possess also a relativistically covariant form in three-dimensional orthogonal coordinates for arbitrary inertial systems $\mathcal{I}'$

$$L'_{ij}\left(\boldsymbol{k}',\omega'\right)E'_j\left(\boldsymbol{k}',\omega'\right)=0, \tag{5.6}$$

with

$$L'_{ij}\left(\boldsymbol{k}',\omega'\right)\equiv\frac{c^2\left(k'_ik'_j-\boldsymbol{k}'^2\delta_{ij}\right)}{\omega'^2}+\varepsilon'_{ij}\left(\boldsymbol{k}',\omega'\right), \tag{5.7}$$

where $\varepsilon'_{ij}\left(\boldsymbol{k}',\omega'\right)$ is related to $\varepsilon_{ij}\left(\boldsymbol{k},\omega\right)$ by a transformation which we derive in detail in **Appendix A**. For an inertial system $\mathcal{I}'$ moving with velocity $\boldsymbol{V}$ in the inertial system $\mathcal{I}$ according to the special Lorentz transformation (A.14) it possesses the form (A.20) [42]

$$\underbrace{\varepsilon'_{ij}(\boldsymbol{k}',\omega')-\delta_{ij}}_{=4\pi\chi'_{ij}(\boldsymbol{k}',\omega')} = \left\{\frac{V_iV_k}{\boldsymbol{V}^2}+\gamma\left(\delta_{ik}-\frac{V_iV_k}{\boldsymbol{V}^2}+\frac{\boldsymbol{k}'\boldsymbol{V}\delta_{ik}-V_ik'_k}{\omega'}\right)\right\}$$
$$\cdot\left\{\frac{V_jV_l}{\boldsymbol{V}^2}+\gamma\left(\delta_{jl}-\frac{V_jV_l}{\boldsymbol{V}^2}+\frac{\boldsymbol{k}'\boldsymbol{V}\delta_{jl}-V_jk'_l}{\omega'}\right)\right\}\underbrace{\left(\varepsilon_{kl}(\boldsymbol{k},\omega)-\delta_{kl}\right)}_{=4\pi\chi_{ij}(\boldsymbol{k},\omega)}, \tag{5.8}$$

with the following relations between $(\boldsymbol{k}',\omega')$ and $(\boldsymbol{k},\omega)$ and their inversion by $\boldsymbol{V}\to-\boldsymbol{V}$ (see (A.15)

$$\boldsymbol{k}'=\boldsymbol{k}+\left((\gamma-1)\frac{\boldsymbol{k}\boldsymbol{V}}{\boldsymbol{V}^2}-\gamma\frac{\omega}{c^2}\right)\boldsymbol{V},\quad \omega'=\gamma(\omega-\boldsymbol{k}\boldsymbol{V}),$$
$$\boldsymbol{k}=\boldsymbol{k}'+\left((\gamma-1)\frac{\boldsymbol{k}'\boldsymbol{V}}{\boldsymbol{V}^2}+\gamma\frac{\omega'}{c^2}\right)\boldsymbol{V},\quad \omega=\gamma(\omega'+\boldsymbol{k}'\boldsymbol{V}), \tag{5.9}$$

and with relativistic invariant

$$\boldsymbol{k}'^2-\frac{\omega'^2}{c^2}=\boldsymbol{k}^2-\frac{\omega^2}{c^2}. \tag{5.10}$$

In $\varepsilon_{kl}(\boldsymbol{k},\omega)$ on the right-hand side of (5.8) one has yet to express the arguments $(\boldsymbol{k},\omega)$ by the arguments $(\boldsymbol{k}',\omega')$ according to the transformation formulae (5.9) to have the same variables on both sides. This means that the transformed permittivity tensor $\varepsilon'_{ij}(\boldsymbol{k}',\omega')$ becomes dependent on the wave vector $\boldsymbol{k}'$ even in case that the primary permittivity tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)$ in the resting system of the medium does not depend on the wave vector $\boldsymbol{k}$. However, this dependence on the wave vector $\boldsymbol{k}$ in the system moving with velocity $\boldsymbol{V}$ which formally means spatial dispersion of the medium is of some other kind than the natural dependence of the permittivity of a medium on wave vector $\boldsymbol{k}$ in resting system and it is not reasonable to expand it in a Taylor series in $\boldsymbol{k}$. For the formal derivation of the energy-momentum tensor these differences are not of importance.

The transformation formulae (5.8) for the permittivity tensor from one to another inertial system and for (5.9) simplify essentially in non-relativistic approximation $\frac{|\boldsymbol{V}|}{c}\ll 1$ if we neglect quadratic and higher terms in $\frac{\boldsymbol{V}}{c}$ in comparison to linear terms in $\frac{\boldsymbol{V}}{c}$ (e.g., $\gamma\to 1$) that we do not write down. From transformations (5.8) together with (5.9) we see that if we change at the same time the signs of $\boldsymbol{k}$ and ω this also changes at the same time the signs of $\boldsymbol{k}'$ and ω' according to

$$(\boldsymbol{k},\omega)\to(-\boldsymbol{k},-\omega),\quad\Leftrightarrow\quad(\boldsymbol{k}',\omega')\to(-\boldsymbol{k}',-\omega'). \tag{5.11}$$

As expected this means that the condition (3.9) for the dissipation-free case transforms into a corresponding condition for the dissipation-free case in an arbitrary inertial system $\mathcal{I}'$ moving with velocity $\boldsymbol{V}$ in inertial system $\mathcal{I}$

$$\varepsilon_{ij}(\boldsymbol{k},\omega)=\varepsilon_{ji}(-\boldsymbol{k},-\omega),\quad\Leftrightarrow\quad\varepsilon'_{ij}(\boldsymbol{k}',\omega')=\varepsilon'_{ji}(-\boldsymbol{k}',-\omega'), \tag{5.12}$$

and is therefore invariant with respect to Lorentz transformations as one could have to expect for such a physical property.

It is seen that the condition (3.9) for absent losses can be continued to the following condition for $L_{ij}(\boldsymbol{k},\omega)$

$$L_{ij}(\boldsymbol{k},\omega)=L_{ji}(-\boldsymbol{k},-\omega)=\left(L_{ji}\left(\boldsymbol{k}^{*},\omega^{*}\right)\right)^{*}. \tag{5.13}$$

Therefore, if (5.13) is satisfied we can write down in addition to (5.2) the following equation for the electric field

$$L_{ij}\left(\mathrm{i}\nabla,-\mathrm{i}\frac{\partial}{\partial t}\right)E_{i}(\boldsymbol{r},t)=0. \tag{5.14}$$

The two Equations (5.2) and (5.14) form the basis of our derivations of local conservation laws and it possesses a great advantage that we have only one field function for the electric field in these equations in comparison to the electric and magnetic field in the common derivations.

We now consider quasiplane and quasimonochromatic waves in the form

$$\boldsymbol{E}(\boldsymbol{r},t)=\boldsymbol{E}_{0}(\boldsymbol{r},t)\mathrm{e}^{\mathrm{i}(\boldsymbol{k}_{0}\boldsymbol{r}-\omega_{0}t)}+\boldsymbol{E}_{0}^{*}(\boldsymbol{r},t)\mathrm{e}^{-\mathrm{i}(\boldsymbol{k}_{0}\boldsymbol{r}-\omega_{0}t)},\quad\left(\boldsymbol{k}_{0}=\boldsymbol{k}_{0}^{*},\,\omega_{0}=\omega_{0}^{*}\right), \tag{5.15}$$

where $\boldsymbol{E}_{0}(\boldsymbol{r},t)$ is a slowly varying complex vectorial amplitude and $\boldsymbol{k}_{0}$ a mean wave vector and ω_{0} a mean frequency. We suppose that $\boldsymbol{k}_{0}$ and ω_{0} are real and exclude in such way, but only for simplicity, evanescent waves with complex values of these quantities which may exist even in lossless media (for example, waves under total reflection in the lossless optically thinner medium or surface waves). The inclusion of such waves would complicate the following considerations but does not destroy the existence of local conservation theorems.

The approximations which we make in the following are that due to slowness of changing of the amplitudes $\boldsymbol{E}_{0}(\boldsymbol{r},t)$ in such way that we may take into account in expansions only a small number of spatial and temporal derivatives of these amplitudes. This means that the wave vectors and frequencies in the Fourier decomposition of the quasiplane and quasimonochromatic wave are concentrated around $\boldsymbol{k}_{0}$ and ω_{0} (and, clearly, around $-\boldsymbol{k}_{0}$ and $-\omega_{0}$) and the two complex conjugated parts in (5.15) are well separated. The supposition and at once approximation in the following is that we can deal with both parts as independent solutions of the wave equation for the electric field. This is apparently equivalent to some averaging procedure over terms with rapidly varying frequencies and wave vectors which then vanish from the equations such as made in [7] and is justified for quasiplane and quasimonochromatic waves.

If we insert the first part from the right-hand side of (5.15) as independent solution into Equation (5.2) we obtain the following equation for the slowly varying complex amplitude $\boldsymbol{E}_{0}(\boldsymbol{r},t)$

$$\begin{aligned}0&=L_{ij}\left(\boldsymbol{k}_{0}-\mathrm{i}\nabla,\omega_{0}+\mathrm{i}\frac{\partial}{\partial t}\right)E_{0,j}(\boldsymbol{r},t)\\&=\left\{\left(L_{ij}\right)_{0}-\mathrm{i}\left(\frac{\partial L_{ij}}{\partial k_{l}}\right)_{0}\nabla_{l}+\mathrm{i}\left(\frac{\partial L_{ij}}{\partial\omega}\right)_{0}\frac{\partial}{\partial t}\right.\end{aligned}$$

$$-\frac{1}{2}\left(\left(\frac{\partial^2 L_{ij}}{\partial k_l \partial k_m}\right)_0 \nabla_l \nabla_m - 2\left(\frac{\partial^2 L_{ij}}{\partial k_l \partial \omega}\right)_0 \nabla_l \frac{\partial}{\partial t} + \left(\frac{\partial^2 L_{ij}}{\partial \omega^2}\right)_0 \frac{\partial^2}{\partial t^2}\right) + \cdots\Bigg\} E_{0,j}(\boldsymbol{r},t), \quad (16)$$

where index "0" means that the corresponding derivatives have to be taken for $\boldsymbol{k} = \boldsymbol{k}_0$ and $\omega = \omega_0$ (e.g., $(L_{ij})_0 \equiv L_{ij}(\boldsymbol{k}_0, \omega_0)$). In the following, we take the derivatives of the slowly varying amplitudes up to the second order but before this we introduce a shorter relativistic-covariant notation of the equations for $\boldsymbol{E}_0(\boldsymbol{r},t)$ and $\boldsymbol{E}_0^*(\boldsymbol{r},t)$.

Our derivation of the energy-momentum tensor is similar to the derivation of approximate equations for beam solutions with the only difference that in last case the determinant of L_{ij} has to be taken as starting point for the expansion to get the equation for the main component of $\boldsymbol{E}_0(\boldsymbol{r},t)$.

6. Local Action Conservation in Relativistic Covariant Form

The derivation of local (or differential) laws of action conservation and of other local conservation theorems becomes much more concise if we introduce for abbreviation the following four-dimensional notations of special theory of relativity[6]

$$\begin{aligned} &r \equiv (\boldsymbol{r}, \mathrm{i}ct), \quad k \equiv \left(\boldsymbol{k}, \frac{\mathrm{i}\omega}{c}\right), \quad r^2 = \boldsymbol{r}^2 - c^2t^2, \quad k^2 = \boldsymbol{k}^2 - \frac{\omega^2}{c^2}, \quad \nabla \equiv \left(\nabla, \frac{1}{\mathrm{i}c}\frac{\partial}{\partial t}\right), \\ &-\mathrm{i}\nabla \mathrm{e}^{\mathrm{i}k_0 r} = -\mathrm{i}\left(\nabla, \frac{1}{\mathrm{i}c}\frac{\partial}{\partial t}\right)\mathrm{e}^{\mathrm{i}(\boldsymbol{k}_0\boldsymbol{r} - \omega_0 t)} = \mathrm{e}^{\mathrm{i}(\boldsymbol{k}_0\boldsymbol{r} - \omega_0 t)}\left(\boldsymbol{k}_0 - \mathrm{i}\nabla, \frac{\mathrm{i}}{c}\left(\omega_0 + \mathrm{i}\frac{\partial}{\partial t}\right)\right). \end{aligned} \quad (6.1)$$

An advantage of the four-dimensional formalism is that we obtain the results in relativistic covariant form.

We may write the equations for the electric field (5.2) and (5.14) in the concise form[7]

$$L_{ij}(-\mathrm{i}\nabla) E_j(r) = 0, \quad L_{ij}(\mathrm{i}\nabla) E_i(r) = 0. \quad (6.2)$$

with $L_{ij}(k)$ in relativistic-covariant form (see Appendix A, (25))

[6]Modern development mainly for preparing the transition to General relativity theory favors to use only representations by real components for space-time. This makes it necessary to distinguish between contravariant and covariant components of vectors and tensors but this becomes very inconvenient for our purposes. According to Pauli [5] (Part III, p. 71), the historically older notation $(x^4 =) x_4 = \mathrm{i}ct$ was first used by Poincaré in 1906 in Journal "R.C. Circ. mat. Palermo **21**, 129" for vectors $x = (\boldsymbol{x} \equiv (x,y,z), t)$ of space-time later called Minkowski space (after Minkowski's Lecture "Raum und Zeit" in Köln in 1908, published in 1909 [43] (republ. [44, 36]) in the year of his premature death). Apart from basic Maxwell equations, Einstein up to 1912 preferred to write down all four-dimensional vectorial relations separately in the 4 components and mentions $x_4 = \mathrm{i}ct$ only shortly in 1910 with reference to Minkowski [43] that can be traced from the collection of Einstein's scientific papers [1] (p. 138, for corresponding article). Minkowski on his part finds it in [43] to be a "very pregnant manner" to cloth $s = \sqrt{-1}ct$ in the "mystic formula" $3 \times 10^5\,\mathrm{km} = \sqrt{-1}\,\mathrm{secs}$.

[7]Deviating from (6.1) we now used in the notations of the argument of the electric field E_j and of the operator function L_{ij} the four-dimensional notations in the form $r = (\boldsymbol{r}, t)$ and $k = (\boldsymbol{k}, \omega)$ which we do not want to change here into that of (6.1). However, by comparison with second line in (1) it seems that this does not cause problems.

$$L_{ij}(k) \equiv L_{ij}(\boldsymbol{k},\omega) = \frac{c^2}{\omega^2}\left(k_i k_j - \boldsymbol{k}^2\delta_{ij}\right) + \varepsilon_{ij}(\boldsymbol{k},\omega), \tag{6.3}$$

and k substituted by $-\mathrm{i}\nabla$. The quasiplane and quasimonochromatic waves (5.15) take on the shorter form

$$E_j(r) = E_{0,j}(r)\mathrm{e}^{\mathrm{i}k_0 r} + E^*_{0,j}(r)\mathrm{e}^{-\mathrm{i}k_0 r}. \tag{6.4}$$

In the same approximation as in (5.16) we obtain from the first of Equation (6.2)

$$\begin{aligned} 0 &= L_{ij}(k_0 - \mathrm{i}\nabla)E_{0,j}(r) \\ &= \left\{\left(L_{ij}\right)_0 - \mathrm{i}\left(\frac{\partial L_{ij}}{\partial k_\lambda}\right)_0 \nabla_\lambda - \frac{1}{2!}\left(\frac{\partial^2 L_{ij}}{\partial k_\lambda \partial k_\mu}\right)_0 \nabla_\lambda \nabla_\mu + \cdots\right\} E_{0,j}(r), \end{aligned} \tag{6.5}$$

and from the second equation

$$\begin{aligned} 0 &= L_{ij}(k_0 + \mathrm{i}\nabla)E^*_{0,i}(r) \\ &= \left\{\left(L_{ij}\right)_0 + \mathrm{i}\left(\frac{\partial L_{ij}}{\partial k_\lambda}\right)_0 \nabla_\lambda - \frac{1}{2!}\left(\frac{\partial^2 L_{ij}}{\partial k_\lambda \partial k_\mu}\right)_0 \nabla_\lambda \nabla_\mu + \cdots\right\} E^*_{0,i}(r). \end{aligned} \tag{6.6}$$

We first derive a conservation theorem which is even more fundamental than the theorem for energy-momentum conservation since it may be extended to inhomogeneous media.

If we multiply (6.5) by $E^*_{0,i}(r)$ and (6.6) by $E_{0,j}(r)$ and form the difference of the obtained equations then it can be written as the 4-divergence of a 4-vector function in the following way

$$\begin{aligned} 0 &= -\mathrm{i}\left\{E^*_{0,i}(r)L_{ij}(k_0 - \mathrm{i}\nabla)E_{0,j}(r) - E_{0,j}(r)L_{ij}(k_0 + \mathrm{i}\nabla)E^*_{0,i}(r)\right\} \\ &= \nabla_\lambda \left\{-\left(\frac{\partial L_{ij}}{\partial k_\lambda}\right)_0 E^*_{0,i}(r)E_{0,j}(r)\right. \\ &\quad \left. + \frac{\mathrm{i}}{2}\left(\frac{\partial^2 L_{ij}}{\partial k_\lambda \partial k_\mu}\right)_0 \left(E^*_{0,i}(r)\nabla_\mu E_{0,j}(r) - E_{0,j}(r)\nabla_\mu E^*_{0,i}(r)\right) + \cdots\right\}. \end{aligned} \tag{6.7}$$

The terms are explicitly written down up to first-order derivatives of the slowly varying amplitudes but the higher-order terms on the right-hand side can also be represented as 4-divergence of a 4-vector $T_\lambda(r)$ and (6.7) possess the form of a vanishing 4-divergence and can be included in the local conservation law

$$\nabla_\lambda T_\lambda(r) = 0, \tag{6.8}$$

with $T_\lambda(r)$ defined by (remind: $\alpha \equiv 4\pi$ in CGS or Gauss system)

$$\begin{aligned} \alpha T_\lambda(r) &= -\left(\frac{\partial L_{ij}}{\partial k_\lambda}\right)_0 E^*_{0,i}(r)E_{0,j}(r) \\ &\quad + \frac{\mathrm{i}}{2}\left(\frac{\partial^2 L_{ij}}{\partial k_\lambda \partial k_\mu}\right)_0 \left(E^*_{0,i}(r)\nabla_\mu E_{0,j}(r) - E_{0,j}(r)\nabla_\mu E^*_{0,i}(r)\right) + \cdots. \end{aligned} \tag{6.9}$$

In three-dimensional separation according to the definition

$$T_\lambda(r)=\left(T_l(\boldsymbol{r},t),\mathrm{i}cs(\boldsymbol{r},t)\right), \tag{6.10}$$

the local conservation theorem (6.8) takes on the form

$$\nabla_l T_l(\boldsymbol{r},t)+\frac{\partial}{\partial t}s(\boldsymbol{r},t)=0. \tag{6.11}$$

From their dimensions, $s(\boldsymbol{r},t)$ can be identified with the action density and $T_l(\boldsymbol{r},t)$ with the vector field of action flow density.

From (6.10) and (6.9) we find in three-dimensional representation up to explicitly given first-order derivatives of the slowly varying amplitudes $\boldsymbol{E}_0(\boldsymbol{r},t)$ of the electric field which last take into account the diffraction of beams

$$\begin{aligned}\alpha T_l(\boldsymbol{r},t)=&-\left(\frac{\partial L_{ij}}{\partial k_l}\right)_0 E_{0,i}^*(\boldsymbol{r},t)E_{0,j}(\boldsymbol{r},t)\\&+\frac{\mathrm{i}}{2}\left(\frac{\partial^2 L_{ij}}{\partial k_l\partial k_m}\right)_0\left(E_{0,i}^*(\boldsymbol{r},t)\nabla_m E_{0,j}(\boldsymbol{r},t)-E_{0,j}(\boldsymbol{r},t)\nabla_m E_{0,i}^*(\boldsymbol{r},t)\right)\\&-\frac{\mathrm{i}}{2}\left(\frac{\partial^2 L_{ij}}{\partial k_l\partial\omega}\right)_0\left(E_{0,i}^*(\boldsymbol{r},t)\frac{\partial}{\partial t}E_{0,j}(\boldsymbol{r},t)-E_{0,j}(\boldsymbol{r},t)\frac{\partial}{\partial t}E_{0,i}^*(\boldsymbol{r},t)\right)+\cdots,\end{aligned}$$

$$\begin{aligned}\alpha s(\boldsymbol{r},t)=&\left(\frac{\partial L_{ij}}{\partial\omega}\right)_0 E_{0,i}^*(\boldsymbol{r},t)E_{0,j}(\boldsymbol{r},t)\\&-\frac{\mathrm{i}}{2}\left(\frac{\partial^2 L_{ij}}{\partial\omega\partial k_m}\right)_0\left(E_{0,i}^*(\boldsymbol{r},t)\nabla_m E_{0,j}(\boldsymbol{r},t)-E_{0,j}(\boldsymbol{r},t)\nabla_m E_{0,i}^*(\boldsymbol{r},t)\right)\\&+\frac{\mathrm{i}}{2}\left(\frac{\partial^2 L_{ij}}{\partial\omega^2}\right)_0\left(E_{0,i}^*(\boldsymbol{r},t)\frac{\partial}{\partial t}E_{0,j}(\boldsymbol{r},t)-E_{0,j}(\boldsymbol{r},t)\frac{\partial}{\partial t}E_{0,i}^*(\boldsymbol{r},t)\right)+\cdots.\end{aligned} \tag{6.12}$$

Before discussing these expressions we derive the local form of energy-momentum conservation.

7. Local Energy and Momentum Conservation in Relativistic Covariant Form

In analogy to (6.7) we consider the following combination which can be represented as the 4-divergence of a second-rank 4-tensor

$$\begin{aligned}0=&-\mathrm{i}\left\{E_{0,i}^*(r)\left(k_{0,\kappa}-\mathrm{i}\nabla_\kappa\right)L_{ij}\left(k_0-\mathrm{i}\nabla\right)E_{0,j}(r)\right.\\&\left.-E_{0,j}(r)\left(k_{0,\kappa}+\mathrm{i}\nabla_\kappa\right)L_{ij}\left(k_0+\mathrm{i}\nabla\right)E_{0,i}^*(r)\right\}\\=&\nabla_\lambda\left\{-\left(\frac{\partial\left(k_\kappa L_{ij}\right)}{\partial k_\lambda}\right)_0 E_{0,i}^*(r)E_{0,j}(r)\right.\\&\left.+\frac{\mathrm{i}}{2}\left(\frac{\partial^2\left(k_\kappa L_{ij}\right)}{\partial k_\lambda\partial k_\mu}\right)_0\left(E_{0,i}^*(r)\nabla_\mu E_{0,j}(r)-E_{0,j}(r)\nabla_\mu E_{0,i}^*(r)\right)+\cdots\right\}.\end{aligned} \tag{7.1}$$

Thus we obtained a local conservation theorem of the form

$$\nabla_\lambda T_{\kappa\lambda}(r)=0, \tag{7.2}$$

with a tensor function $T_{\kappa\lambda}(r)$ which explicitly written down up to terms with first-order derivatives of the slowly varying amplitudes is given by

$$\begin{aligned}\alpha T_{\kappa\lambda}(r) &= -\left(\frac{\partial\left(k_\kappa L_{ij}\right)}{\partial k_\lambda}\right)_0 E_{0,i}^*(r) E_{0,j}(r) \\ &\quad + \frac{\mathrm{i}}{2}\left(\frac{\partial^2\left(k_\kappa L_{ij}\right)}{\partial k_\lambda \partial k_\mu}\right)_0 \left(E_{0,i}^*(r)\nabla_\mu E_{0,j}(r) - E_{0,j}(r)\nabla_\mu E_{0,i}^*(r)\right) + \cdots.\end{aligned} \tag{7.3}$$

The four-dimensional covariance of $T_{\kappa\lambda}(r)$ with respect to index λ is the same as in the action 4-vector $T_\lambda(r)$ and the covariance with respect to index κ is evident from construction (7.1) with 4-wave vector k_κ. That this is connected with homogeneity (or translation invariance) in space and time is easily seen since in case of absence of this symmetry it is impossible to have globally constant wave vectors and frequencies as used in the derivation. Thus we have the justification to call $T_{\kappa\lambda}(r)$ the energy-momentum tensor of homogeneous anisotropic dispersive media in the approximation of quasiplane and quasimonochromatic waves. In general, the tensor $T_{\kappa\lambda}(r)$ is non-symmetric

$$T_{\kappa\lambda}(r) \neq T_{\lambda\kappa}(r), \tag{7.4}$$

and is, in general, not equivalent to a symmetric one that means it is intrinsically non-symmetric.

We now transform the energy-momentum tensor $T_{\kappa\lambda}(r)$ to another form which is interesting for the physical interpretation. For this purpose we use the identities

$$\begin{aligned}\left(\frac{\partial\left(k_\kappa L_{ij}\right)}{\partial k_\lambda}\right)_0 &= k_{0,\kappa}\left(\frac{\partial L_{ij}}{\partial k_\lambda}\right)_0 + \delta_{\kappa\lambda}\left(L_{ij}\right)_0, \\ \left(\frac{\partial^2\left(k_\kappa L_{ij}\right)}{\partial k_\lambda \partial k_\mu}\right)_0 &= k_{0,\kappa}\left(\frac{\partial^2 L_{ij}}{\partial k_\lambda \partial k_\mu}\right)_0 + \delta_{\kappa\mu}\left(\frac{\partial L_{ij}}{\partial k_\lambda}\right)_0 + \delta_{\kappa\lambda}\left(\frac{\partial L_{ij}}{\partial k_\mu}\right)_0.\end{aligned} \tag{7.5}$$

Inserting this into (7.3) and using the representation (6.9) of $T_\lambda(r)$ and the Equations (6.5) and (6.6) for the slowly varying amplitudes we obtain up to first-order derivatives of these amplitudes

$$\begin{aligned}\alpha T_{\kappa\lambda}(r) &= \alpha k_{0,\kappa} T_\lambda(r) - \delta_{\kappa\lambda}\left(L_{ij}\right)_0 E_{0,i}^*(r) E_{0,j}(r) \\ &\quad + \frac{\mathrm{i}}{2}\left(\frac{\partial L_{ij}}{\partial k_\lambda}\right)_0 \left(E_{0,i}^*(r)\nabla_\kappa E_{0,j}(r) - E_{0,j}(r)\nabla_\kappa E_{0,i}^*(r)\right) + \cdots.\end{aligned} \tag{7.6}$$

In the limiting transition from the slowly varying amplitudes to constant amplitudes the terms with derivatives of these amplitudes vanish and $\left(L_{ij}\right)_0 E_{0,j} = 0$ and we obtain the factorization

$$T_{\kappa\lambda}(r) = k_{0,\kappa} T_\lambda(r), \tag{7.7}$$

of the energy-momentum tensor. This is in full analogy to a homogeneous particle flow as discussed in e.g. [3] [12] (see also [34] and below) where, however, macroscopic electrodynamics provides a greater variety of possible dependencies of the momentum of one particle on the group velocity than classical mechanics. Taking seriously this analogy to a homogeneous particle flow this leads in a natural way to a quantization of the electrodynamic flow and to its interpretation as a flow of quasiparticles. The energy-momentum tensor $T_{\kappa\lambda}(r)$ in higher approximations according to (7.6) does not fully factorize into the product $k_{0,\kappa}T_{\lambda}(r)$ and the remaining terms are important at such space-time points where the 4-gradient of the slowly varying amplitudes of the electric field components is important. This may be interpreted as the tendency that energy and momentum flow at these points are forced to choose deviating directions in comparison to the homogeneous particle flow and expresses some interaction of the particles within the flow or some (direction-dependent) pressure or stress. This is in rough agreement with the diffraction of beams, for example, of Gaussian beams which cannot remain to be focused over the whole length of the beam.

8. Three-Dimensional Representation of Energy-Momentum Tensor

We now make the transition to the three-dimensional separation of the terms in the local laws of momentum and of energy conservation. The 4-dimensional energy-momentum tensor can be separated into three-dimensional parts in the following way defining (in common sense) the introduced new quantities on the right-hand side

$$T_{\kappa\lambda}(r)=\begin{pmatrix} T_{kl}(r) & T_{k4}(r) \\ T_{4l}(r) & T_{44}(r) \end{pmatrix} \equiv \begin{pmatrix} T_{kl}(\boldsymbol{r},t) & \mathrm{i}cg_k(\boldsymbol{r},t) \\ \frac{\mathrm{i}}{c}S_l(\boldsymbol{r},t) & -w(\boldsymbol{r},t) \end{pmatrix}. \tag{8.1}$$

Then from (7.2) we find the following differential law of momentum conservation

$$\nabla_l T_{kl}(\boldsymbol{r},t)+\frac{1}{\mathrm{i}c}\frac{\partial}{\partial t}T_{k4}(\boldsymbol{r},t)=0, \quad \Leftrightarrow \quad \nabla_l T_{kl}(\boldsymbol{r},t)+\frac{\partial}{\partial t}g_k(\boldsymbol{r},t)=0, \tag{8.2}$$

where $T_{kl}(\boldsymbol{r},t)$ is the (Maxwell) stress tensor and $g_k(\boldsymbol{r},t)$ the momentum density[8]. Furthermore, the following differential law of energy conservation holds

$$\nabla_l T_{4l}(\boldsymbol{r},t)+\frac{1}{\mathrm{i}c}\frac{\partial}{\partial t}T_{44}(\boldsymbol{r},t)=0, \quad \Leftrightarrow \quad \nabla_l S_l(\boldsymbol{r},t)+\frac{\partial}{\partial t}w(\boldsymbol{r},t)=0, \tag{8.3}$$

where $S_l(\boldsymbol{r},t)$ is the energy flow density (Poynting vector) and $w(\boldsymbol{r},t)$ the energy density.

According to (7.3) and (7.6) taking into account (8.1) the stress tensor possesses the form

[8]The three-dimensional stress tensor T_{kl} is sometimes defined with opposite sign. Our sign of T_{kl} agrees with that in the same notation $T_{kl}=T^{kl}$ in Landau and Lifshits [3] (Ed. 1962) and with $T_{kl}\equiv-\sigma_{kl}$ in later editions (e.g., [45] from 1988). Apparently, the notation σ_{kl} agrees also with respect to sign to the same notation in [13] and in [10].

$$\begin{aligned}\alpha T_{kl}(\boldsymbol{r},t) = &-\left(\frac{\partial(k_k L_{ij})}{\partial k_l}\right)_0 E^*_{0,i}(\boldsymbol{r},t)E_{0,j}(\boldsymbol{r},t)\\ &+\frac{\mathrm{i}}{2}\left(\frac{\partial^2(k_k L_{ij})}{\partial k_l\partial k_m}\right)_0\left(E^*_{0,i}(\boldsymbol{r},t)\nabla_m E_{0,j}(\boldsymbol{r},t)-E_{0,j}(\boldsymbol{r},t)\nabla_m E^*_{0,i}(\boldsymbol{r},t)\right)\\ &-\frac{\mathrm{i}}{2}\left(\frac{\partial^2(k_k L_{ij})}{\partial k_l\partial \omega}\right)_0\left(E^*_{0,i}(\boldsymbol{r},t)\frac{\partial}{\partial t}E_{0,j}(\boldsymbol{r},t)-E_{0,j}(\boldsymbol{r},t)\frac{\partial}{\partial t}E^*_{0,i}(\boldsymbol{r},t)\right)+\cdots,\end{aligned} \tag{8.4}$$

and the momentum density is

$$\begin{aligned}\alpha g_k(\boldsymbol{r},t) = &\left(\frac{\partial(k_k L_{ij})}{\partial \omega}\right)_0 E^*_{0,i}(\boldsymbol{r},t)E_{0,j}(\boldsymbol{r},t)\\ &-\frac{\mathrm{i}}{2}\left(\frac{\partial^2(k_k L_{ij})}{\partial \omega\partial k_m}\right)_0\left(E^*_{0,i}(\boldsymbol{r},t)\nabla_m E_{0,j}(\boldsymbol{r},t)-E_{0,j}(\boldsymbol{r},t)\nabla_m E^*_{0,i}(\boldsymbol{r},t)\right)\\ &+\frac{\mathrm{i}}{2}\left(\frac{\partial^2(k_k L_{ij})}{\partial \omega^2}\right)_0\left(E^*_{0,i}(\boldsymbol{r},t)\frac{\partial}{\partial t}E_{0,j}(\boldsymbol{r},t)-E_{0,j}(\boldsymbol{r},t)\frac{\partial}{\partial t}E^*_{0,i}(\boldsymbol{r},t)\right)+\cdots.\end{aligned} \tag{8.5}$$

For the energy flow density we find from (7.3) and (7.6) taking into account (8.1)

$$\begin{aligned}\alpha S_l(\boldsymbol{r},t) = &-\left(\frac{\partial(\omega L_{ij})}{\partial k_l}\right)_0 E^*_{0,i}(\boldsymbol{r},t)E_{0,j}(\boldsymbol{r},t)\\ &+\frac{\mathrm{i}}{2}\left(\frac{\partial^2(\omega L_{ij})}{\partial k_l\partial k_m}\right)_0\left(E^*_{0,i}(\boldsymbol{r},t)\nabla_m E_{0,j}(\boldsymbol{r},t)-E_{0,j}(\boldsymbol{r},t)\nabla_m E^*_{0,i}(\boldsymbol{r},t)\right)\\ &-\frac{\mathrm{i}}{2}\left(\frac{\partial^2(\omega L_{ij})}{\partial k_l\partial \omega}\right)_0\left(E^*_{0,i}(\boldsymbol{r},t)\frac{\partial}{\partial t}E_{0,j}(\boldsymbol{r},t)-E_{0,j}(\boldsymbol{r},t)\frac{\partial}{\partial t}E^*_{0,i}(\boldsymbol{r},t)\right)+\cdots,\end{aligned} \tag{8.6}$$

and for the energy density

$$\begin{aligned}\alpha w(\boldsymbol{r},t) = &\left(\frac{\partial(\omega L_{ij})}{\partial \omega}\right)_0 E^*_{0,i}(\boldsymbol{r},t)E_{0,j}(\boldsymbol{r},t)\\ &-\frac{\mathrm{i}}{2}\left(\frac{\partial^2(\omega L_{ij})}{\partial \omega\partial k_m}\right)_0\left(E^*_{0,i}(\boldsymbol{r},t)\nabla_m E_{0,j}(\boldsymbol{r},t)-E_{0,j}(\boldsymbol{r},t)\nabla_m E^*_{0,i}(\boldsymbol{r},t)\right)\\ &+\frac{\mathrm{i}}{2}\left(\frac{\partial^2(\omega L_{ij})}{\partial \omega^2}\right)_0\left(E^*_{0,i}(\boldsymbol{r},t)\frac{\partial}{\partial t}E_{0,j}(\boldsymbol{r},t)-E_{0,j}(\boldsymbol{r},t)\frac{\partial}{\partial t}E^*_{0,i}(\boldsymbol{r},t)\right)+\cdots.\end{aligned} \tag{8.7}$$

The terms with spatial and temporal derivatives of the slowly varying amplitudes describe in addition to the stable form of propagation of a wave group its diffraction.

Integral forms of the conservation of momentum and energy in time follow from integration of the conservation theorems within a volume V with surface S and normal unit-vector $\boldsymbol{N}$ directed to the inside of the surface S by (Gauss theorem)

$$\begin{aligned}\frac{\partial}{\partial t}P_k(t) &\equiv \frac{\partial}{\partial t}\int_V \mathrm{d}V(\boldsymbol{r})\,g_k(\boldsymbol{r},t) = -\int_V \mathrm{d}V(\boldsymbol{r})\nabla_l T_{kl}(\boldsymbol{r},t) = +\oint_S \mathrm{d}S(\boldsymbol{r})T_{kl}(\boldsymbol{r},t)N_l(\boldsymbol{r}),\\ \frac{\partial}{\partial t}E(t) &\equiv \frac{\partial}{\partial t}\int_V \mathrm{d}V(\boldsymbol{r})\,w(\boldsymbol{r},t) = -\int_V \mathrm{d}V(\boldsymbol{r})\nabla_l S_l(\boldsymbol{r},t) = +\oint_S \mathrm{d}S(\boldsymbol{r})S_l(\boldsymbol{r},t)N_l(\boldsymbol{r}),\end{aligned} \tag{8.8}$$

with $P_\kappa \equiv \left(P_k, P_4 = \mathrm{i}\dfrac{E}{c}\right)$ the four-vector of momentum[9].

9. Limiting Transition to Plane Monochromatic Waves in Anisotropic Dispersive Media

In the limiting transition from quasiplane and quasimonochromatic waves to plane monochromatic waves the slowly varying amplitudes become constant amplitudes $\boldsymbol{E}_0(\boldsymbol{r},t)\to\boldsymbol{E}_0$ and the energy-momentum tensor becomes independent on the space-time points $(\boldsymbol{r},t)$ that means $T_{\kappa\lambda}(\boldsymbol{r},t)\to T_{\kappa\lambda}$. The propagation of the wave as a wave packet with the group velocity in this limiting transition becomes the more invisible the nearer it comes to a plane monochromatic wave.

From (7.3) together with (7.5) or from (7.6) follows in this limiting transition in relativistic covariant form

$$\alpha T_{\kappa\lambda} = -E_{0,i}^{*}\left(\frac{\partial\left(k_\kappa L_{ij}\right)}{\partial k_\lambda}\right)_0 E_{0,j} = -k_{0,\kappa}E_{0,i}^{*}\left(\frac{\partial L_{ij}}{\partial k_\lambda}\right)_0 E_{0,j} = k_{0,\kappa}\alpha T_\lambda. \tag{9.1}$$

In the three-dimensional separation expressed by the formulae (8.4), (8.5), (8.6) and (8.7) this limiting transition results in the relations

$$\alpha T_{\kappa\lambda} = \begin{pmatrix} -E_{0,i}^{*}\left(\dfrac{\partial\left(k_k L_{ij}\right)}{\partial k_l}\right)_0 E_{0,j} & \mathrm{i}cE_{0,i}^{*}\left(\dfrac{\partial\left(k_k L_{ij}\right)}{\partial\omega}\right)_0 E_{0,j} \\ -\dfrac{\mathrm{i}}{c}E_{0,i}^{*}\left(\dfrac{\partial\left(\omega L_{ij}\right)}{\partial k_l}\right)_0 E_{0,j} & -E_{0,i}^{*}\left(\dfrac{\partial\left(\omega L_{ij}\right)}{\partial\omega}\right)_0 E_{0,j}\end{pmatrix} = \alpha\begin{pmatrix} k_{0,k}T_l & \mathrm{i}ck_{0,k}s \\ \dfrac{\mathrm{i}}{c}\omega_0 T_l & -\omega_0 s\end{pmatrix}, \tag{9.2}$$

where the action flow density T_l and the action density s according to (6.12) become

$$\alpha T_l = -E_{0,i}^{*}\left(\frac{\partial L_{ij}}{\partial k_l}\right)_0 E_{0,j},\quad \alpha s = E_{0,i}^{*}\left(\frac{\partial L_{ij}}{\partial\omega}\right)_0 E_{0,j}. \tag{9.3}$$

[9]The sign of $N_l(\boldsymbol{r})$ at the surface element $\mathrm{d}S$ can be verified from the derivative of the characteristic function $\theta_V(\boldsymbol{r}) \equiv \begin{cases}1, & \boldsymbol{r}\in V\\ 0, & \boldsymbol{r}\notin V\end{cases}$ of the volume V which is $\nabla_l\theta_V(\boldsymbol{r}) = N_l(\boldsymbol{r})\delta_S(\boldsymbol{r})$ with normal unit vector $N_l(\boldsymbol{r})$ directed to the inside of $\mathrm{d}S(\boldsymbol{r})$ as generalization of the step function and of the delta function as its derivative according to

$$\begin{aligned}\int_V \mathrm{d}V(\boldsymbol{r})\left(\nabla_l A_l(\boldsymbol{r})\right) &= \int \mathrm{d}V(\boldsymbol{r})\theta_V(\boldsymbol{r})\left(\nabla_l A_l(\boldsymbol{r})\right)\\ &= \underbrace{\int \mathrm{d}V(\boldsymbol{r})\nabla_l\left(\theta_V(\boldsymbol{r})A_l(\boldsymbol{r})\right)}_{=0} - \int \mathrm{d}V(\boldsymbol{r})A_l(\boldsymbol{r})\left(\nabla_l\theta_V(\boldsymbol{r})\right)\\ &= -\oint_S \mathrm{d}S(\boldsymbol{r})A_l(\boldsymbol{r})N_l(\boldsymbol{r}),\end{aligned}$$

where integrals without given boundaries go over the whole space.

We will show in the following that these expressions are not in contradiction to known expressions for the energy-momentum tensor (mostly more special or otherwise formulated ones).

If we use the explicit form of $L_{ij}(\boldsymbol{k},\omega)$ given in (5.5) we obtain from (9.2)

$$\begin{aligned}
\alpha T_{kl} &= \frac{c^2}{\omega_0^2} E_{0,i}^* \left\{ \left(\boldsymbol{k}_0^2 \delta_{ij} - k_{0,i} k_{0,j} \right) \delta_{kl} + k_{0,k} \left(2k_{0,l} \delta_{ij} - k_{0,i} \delta_{jl} - k_{0,j} \delta_{il} \right) \right\} E_{0,j} \\
&\quad - E_{0,i}^* \left(\frac{\partial \left(k_k \varepsilon_{ij} \right)}{\partial k_l} \right)_0 E_{0,j}, \\
\alpha g_k &= 2\frac{c^2}{\omega_0^3} k_{0,k} E_{0,i}^* \left(\boldsymbol{k}_0^2 \delta_{ij} - k_{0,i} k_{0,j} \right) E_{0,j} + k_{0,k} E_{0,i}^* \left(\frac{\partial \varepsilon_{ij}}{\partial \omega} \right)_0 E_{0,j}, \\
\alpha S_l &= \frac{c^2}{\omega_0} E_{0,i}^* \left(2k_{0,l} \delta_{ij} - k_{0,i} \delta_{jl} - k_{0,j} \delta_{il} \right) E_{0,j} - \omega_0 E_{0,i}^* \left(\frac{\partial \varepsilon_{ij}}{\partial k_l} \right)_0 E_{0,j}, \\
\alpha w &= \frac{c^2}{\omega_0^2} E_{0,i}^* \left(\boldsymbol{k}_0^2 \delta_{ij} - k_{0,i} k_{0,j} \right) E_{0,j} + E_{0,i}^* \left(\frac{\partial \left(\omega \varepsilon_{ij} \right)}{\partial \omega} \right)_0 E_{0,j},
\end{aligned} \tag{9.4}$$

and from (9.3) for action flow density T_l and action density s

$$\begin{aligned}
\alpha T_l &= \frac{c^2}{\omega_0^2} E_{0,i}^* \left(2k_{0,l} \delta_{ij} - k_{0,i} \delta_{jl} - k_{0,j} \delta_{il} \right) E_{0,j} - E_{0,i}^* \left(\frac{\partial \varepsilon_{ij}}{\partial k_l} \right)_0 E_{0,j}, \\
\alpha s &= 2\frac{c^2}{\omega_0^3} E_{0,i}^* \left(\boldsymbol{k}_0^2 \delta_{ij} - k_{0,i} k_{0,j} \right) E_{0,j} + E_{0,i}^* \left(\frac{\partial \varepsilon_{ij}}{\partial \omega} \right)_0 E_{0,j}.
\end{aligned} \tag{9.5}$$

As already discussed, in the transition to the factorized form in (9.1) and (9.2) we used the equations for the slowly varying amplitudes which after transition to plane monochromatic waves become the algebraic equations $\left(L_{ij}\right)_0 E_{0,j} = 0$ and $E_{0,i}^* \left(L_{ij}\right)_0 = 0$. Instead of these equations we can also directly use the equations of macroscopic electrodynamics (2.7) which for plane monochromatic waves with real wave vector $\boldsymbol{k}_0$ and real frequency ω_0 take on the form[10]

$$\begin{aligned}
\boldsymbol{B}_0 &= \frac{c}{\omega_0} \left[\boldsymbol{k}_0, \boldsymbol{E}_0 \right], \quad \boldsymbol{D}_0 = -\frac{c}{\omega_0} \left[\boldsymbol{k}_0, \boldsymbol{B}_0 \right], \quad \Rightarrow \quad \boldsymbol{k}_0 \boldsymbol{B}_0 = \boldsymbol{k}_0 \boldsymbol{D}_0 = 0, \\
\boldsymbol{B}_0^* &= \frac{c}{\omega_0} \left[\boldsymbol{k}_0, \boldsymbol{E}_0^* \right], \quad \boldsymbol{D}_0^* = -\frac{c}{\omega_0} \left[\boldsymbol{k}_0, \boldsymbol{B}_0^* \right], \quad \Rightarrow \quad \boldsymbol{k}_0 \boldsymbol{B}_0^* = \boldsymbol{k}_0 \boldsymbol{D}_0^* = 0.
\end{aligned} \tag{9.6}$$

Using these equations, we can transform (9.4) exactly to the following "mixed" forms of representation with the amplitudes of the electric and magnetic field $\boldsymbol{E}_0$ and $\boldsymbol{B}_0$ and the electric induction $\boldsymbol{D}_0$ which dominate in their kind in literature (compare also [7] [9] [10]).

$$\begin{aligned}
\alpha T_{kl} &= \left(\boldsymbol{B}_0^* \boldsymbol{B}_0 + \frac{1}{2} \left(\boldsymbol{E}_0^* \boldsymbol{D}_0 + \boldsymbol{D}_0^* \boldsymbol{E}_0 \right) \right) \delta_{kl} - \left(B_{0,k}^* B_{0,l} + B_{0,k} B_{0,l}^* + E_{0,k}^* D_{0,l} + E_{0,k} D_{0,l}^* \right) \\
&\quad - k_{0,k} \boldsymbol{E}_0^* \left(\frac{\partial \boldsymbol{\varepsilon}}{\partial k_l} \right)_0 \boldsymbol{E}_0 = \alpha k_{0,k} T_l,
\end{aligned}$$

[10]We emphasize again that this restriction to real wave vectors and frequencies is not a principal restriction for lossless media but simplifies our derivations considerably since it does not introduce additional difficulties with inhomogeneous (evanescent) waves in lossless media which necessarily are to be discussed without this restriction.

$$
\begin{aligned}
\alpha g_k &= \frac{1}{c}\left(\left[\boldsymbol{D}_0^*, \boldsymbol{B}_0\right]_k + \left[\boldsymbol{D}_0, \boldsymbol{B}_0^*\right]_k\right) + k_{0,k}\boldsymbol{E}_0^*\left(\frac{\partial \boldsymbol{\varepsilon}}{\partial \omega}\right)_0 \boldsymbol{E}_0 = \alpha k_{0,k} s, \\
\alpha S_l &= c\left(\left[\boldsymbol{E}_0^*, \boldsymbol{B}_0\right]_l + \left[\boldsymbol{E}_0, \boldsymbol{B}_0^*\right]_l\right) - \omega_0 \boldsymbol{E}_0^*\left(\frac{\partial \boldsymbol{\varepsilon}}{\partial k_l}\right)_0 \boldsymbol{E}_0 = \alpha\omega_0 T_l, \\
\alpha w &= \boldsymbol{B}_0^*\boldsymbol{B}_0 + \frac{1}{2}\left(\boldsymbol{E}_0^*\boldsymbol{D}_0 + \boldsymbol{D}_0^*\boldsymbol{E}_0\right) + \omega_0\boldsymbol{E}_0^*\left(\frac{\partial \boldsymbol{\varepsilon}}{\partial \omega}\right)_0 \boldsymbol{E}_0 = \alpha\omega_0 s.
\end{aligned} \tag{9.7}
$$

For the action flow density T_l and the action density *s*, we find

$$
\begin{aligned}
\alpha T_l &= \frac{c}{\omega_0}\left(\left[\boldsymbol{E}_0^*, \boldsymbol{B}_0\right]_l + \left[\boldsymbol{E}_0, \boldsymbol{B}_0^*\right]_l\right) - E_0^*\left(\frac{\partial \boldsymbol{\varepsilon}}{\partial k_l}\right)_0 \boldsymbol{E}_0, \\
\alpha s &= \frac{1}{\omega_0}\left(\boldsymbol{B}_0^*\boldsymbol{B}_0 + \frac{1}{2}\left(\boldsymbol{E}_0^*\boldsymbol{D}_0 + \boldsymbol{D}_0^*\boldsymbol{E}_0\right)\right) + \boldsymbol{E}_0^*\left(\frac{\partial \boldsymbol{\varepsilon}}{\partial \omega}\right)_0 \boldsymbol{E}_0 \\
&= \frac{c}{\omega_0^2}\boldsymbol{k}_0\left(\left[\boldsymbol{E}_0^*, \boldsymbol{B}_0\right] + \left[\boldsymbol{E}_0, \boldsymbol{B}_0^*\right]\right) + \boldsymbol{E}_0^*\left(\frac{\partial \boldsymbol{\varepsilon}}{\partial \omega}\right)_0 \boldsymbol{E}_0.
\end{aligned} \tag{9.8}
$$

The appearance of ω_0 in the denominators for action flow density T_l and action density *s* shows that they are formed in nonlocal way by the fields that in the space-time picture is impossible to express by quadratic local field combinations only and which, perhaps, is a reason that they did not find much attention (exception: similar considerations to adiabatic invariance).

We see that all parts of the energy-momentum tensor in (9.7) contain a part with origin from the dispersion of the medium. The momentum density g_k which possesses the direction of the mean wave vector $k_{0,k}$ and the energy density *w* are modified by terms $\left(\frac{\partial \varepsilon_{ij}}{\partial \omega}\right)_0$ with derivatives of the permittivity tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)$ with respect to frequency ω (this goes back to Brillouin in 1921; see also [7] [9] but many other, in principle, excellent monographs on electrodynamics and optics do not take this into account). The stress tensor T_{kl} and the energy flow density S_l are modified by terms $\left(\frac{\partial \varepsilon_{ij}}{\partial k_l}\right)_0$ with derivatives of the permittivity tensor with respect to the wave vector $\boldsymbol{k}$ which are non-vanishing only in case of presence of spatial dispersion (see, e.g., [9] [10] [11]). The terms from frequency dispersion may become very important in the neighborhood of eigenfrequencies of the medium (e.g., such as used for laser transitions) and do not represent in this case only a small correction to the terms without dispersion. The terms with spatial dispersion are non-vanishing, for example, for media with natural optical activity or for hot gases and plasmas. If a medium possesses only frequency dispersion $\varepsilon_{ij}(\boldsymbol{k},\omega) \equiv \varepsilon_{ij}(\omega)$ in inertial system $\mathcal{I}$ then by transition to an inertial system $\mathcal{I}'$ where this medium is moving the new permittivity tensor $\varepsilon'_{ij}(\boldsymbol{k}',\omega')$ in $\mathcal{I}'$ depends apart from transformed frequency ω' also on transformed wave vector $\boldsymbol{k}'$ and appears there as medium with "unnatural" spatial dispersion (see **Appendix A**).

The trace of the energy-momentum tensor which is a relativistic invariant is non-vanishing taking into account the dispersion. From the limiting case of plane monochromatic waves in (9.7) we find

$$\alpha\left(\langle \mathrm{T}\rangle - w\right) = -k_{0,k}E_{0,i}^{*}\left(\frac{\partial \varepsilon_{ij}}{\partial k_k}\right)_0 E_{0,j} - \omega_0 E_{0,i}^{*}\left(\frac{\partial \varepsilon_{ij}}{\partial \omega}\right)_0 E_{0,j}. \tag{9.9}$$

If we neglect dispersion the trace of the energy-momentum tensor becomes vanishing as it is seen from this expression. Due to factorization of the stress tensor $T_{kl} = k_{0,k}T_l$ and of the whole energy-momentum tensor $T_{\kappa\lambda}$ they possess only one non-vanishing eigenvalue and the quantity $[T] \equiv \frac{1}{2}\left(\langle \mathrm{T}\rangle^2 - \langle \mathrm{T}^2\rangle\right)$ involved in the second invariant $[T] \equiv [\mathrm{T}] - \langle \mathrm{T}\rangle w + \boldsymbol{Sg}$ of the four-dimensional tensor T (see (B.8) in **Appendix B**) is vanishing. Therefore, using the form (9.7) in considered approximation one can check the vanishing of the second invariant of the energy-momentum tensor

$$[T] = \underbrace{[\mathrm{T}]}_{=0} - \left(\langle \mathrm{T}\rangle w - \boldsymbol{Sg}\right) = 0, \tag{9.10}$$

which is a Lorentz-invariant and thus this relation is true in arbitrary inertial systems. Due to factorization (9.1) of the energy-momentum tensor in considered approximation we find

$$\boldsymbol{Sg} - \langle \mathrm{T}\rangle w = 0, \tag{9.11}$$

remaining true after Lorentz transformation.

The energy-momentum tensor (9.7) is intrinsically non-symmetric expressed by relation (7.4) also under neglect of dispersion. In general, for anisotropic media the momentum density g_k and the energy flow density S_l possess different directions and there is no way to remove this but also the stress tensor T_{kl} is non-symmetric for anisotropic media. From the two old proposals for this tensor which are the Minkowski tensor and the Abraham tensor (see, e.g., [5] [10]) the tensor (9.2) is nearer to the Minkowski tensor and makes the transition to it in case of neglected dispersion. However, this problem of the correct tensor did not genuinely exist in our derivations since under the condition (3.9) that the medium is lossless the local form of the conservation laws could be formulated as exact vanishing of a 4-divergence of an energy-momentum tensor.

We can subdivide the energy-momentum tensor $T_{\kappa\lambda}$ in (9.7) in additive way into a pure electromagnetic field tensor $T_{\kappa\lambda}^{(\mathrm{F})}$ which contains only the electric field $\boldsymbol{E}$ and magnetic field $\boldsymbol{B}$ and a field-matter interaction tensor $T_{\kappa\lambda}^{(\mathrm{I})}$ which contains in addition the polarization $\boldsymbol{P} = \frac{1}{4\pi}\left(\boldsymbol{D} - \boldsymbol{E}\right)$ and derivatives of the permittivity tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)$. The field part which is quadratic in the electromagnetic field is then a symmetric tensor $T_{\kappa\lambda}^{(\mathrm{F})} = T_{\lambda\kappa}^{(\mathrm{F})}$ and equal in form to the tensor for vacuum. The interaction part $T_{\kappa\lambda}^{(\mathrm{I})}$ which is bilinear in field and polarization or contains derivatives of the permittivity tensor is non-symmetric. Their explicit forms may be taken from (9.7). It should be emphasized that such a subdivision remains to be formally since each of the two parts does not separately obey a local conservation law.

In the transition to vacuum $\varepsilon_{ij}(\boldsymbol{k},\omega)\to\delta_{ij}$ from (9.7) we obtain a symmetric energy-momentum tensor and it possesses at the same time the most simple form in comparison to equivalent ones since it does not contain parts with the rapidly varying phase factors $\mathrm{e}^{\pm \mathrm{i}2(\boldsymbol{k}_0\boldsymbol{r}-\omega_0 t)}$ and admits a direct physical interpretation by considerations about the group velocity.

10. Group Velocity in Energy-Momentum Tensor for Anisotropic Dispersive Media and Its Calculation

A wave packet in a homogeneous medium propagates in first approximation with shape stability and without diffraction with the group velocity and therefore energy and momentum of this wave packet should propagate also with the group velocity. The introduction of the group velocity into the energy-momentum tensor in the limiting case of plane monochromatic wave reveals a simple basic structure of this tensor (see also, [7] [9] [10]).

Plane monochromatic waves with real wave vector and real frequency satisfy Equation (5.4) and together with the condition (5.13) for absent losses this can be written in operator form (*i.e.* without vectorial indices $L_{ij}(\boldsymbol{k},\omega)\to \mathrm{L}(\boldsymbol{k},\omega)$) as the following eigenvalue equations for right-hand and left-hand eigenvectors of $\mathrm{L}(\boldsymbol{k},\omega)$ to eigenvalue zero

$$\mathrm{L}(\boldsymbol{k},\omega)\boldsymbol{E}(\boldsymbol{k},\omega)=\boldsymbol{0},\quad \boldsymbol{E}^*(\boldsymbol{k},\omega)\mathrm{L}(\boldsymbol{k},\omega)=\boldsymbol{0}, \tag{10.1}$$

with the operator $\mathrm{L}(\boldsymbol{k},\omega)$ defined by (5.5) ($\boldsymbol{a}\cdot\boldsymbol{b}$ is dyadic product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$)

$$\mathrm{L}(\boldsymbol{k},\omega)\equiv\frac{c^2}{\omega^2}\left(\boldsymbol{k}\cdot\boldsymbol{k}-\boldsymbol{k}^2\mathrm{I}\right)+\boldsymbol{\varepsilon}(\boldsymbol{k},\omega). \tag{10.2}$$

The necessary condition for solutions of the operator Equations (10.1) is the vanishing of the determinant $\left|\mathrm{L}(\boldsymbol{k},\omega)\right|$ of the operator $\mathrm{L}(\boldsymbol{k},\omega)$

$$\left|\mathrm{L}(\boldsymbol{k},\omega)\right|=0, \tag{10.3}$$

which in coordinate-invariant notation is explicitly given by [27] [28] [31][11]

$$\begin{aligned}\left|\mathrm{L}(\boldsymbol{k},\omega)\right|&=\frac{c^4}{\omega^4}\boldsymbol{k}^2(\boldsymbol{k\varepsilon k})-\frac{c^2}{\omega^2}\left(\langle\boldsymbol{\varepsilon}\rangle\boldsymbol{k\varepsilon k}-\boldsymbol{k\varepsilon}^2\boldsymbol{k}\right)+|\boldsymbol{\varepsilon}|\\&=\frac{c^4}{\omega^4}\boldsymbol{k}^2(\boldsymbol{k\varepsilon k})-\frac{c^2}{\omega^2}\left(\langle\bar{\boldsymbol{\varepsilon}}\rangle\boldsymbol{k}^2-\boldsymbol{k}\bar{\boldsymbol{\varepsilon}}\boldsymbol{k}\right)+|\boldsymbol{\varepsilon}|,\quad \boldsymbol{\varepsilon}\equiv\boldsymbol{\varepsilon}(\boldsymbol{k},\omega),\end{aligned} \tag{10.4}$$

where $\langle\boldsymbol{\varepsilon}\rangle$ denotes the trace and $|\boldsymbol{\varepsilon}|$ the determinant of the permittivity tensor and where $\boldsymbol{k\varepsilon k}\equiv k_i\varepsilon_{ij}k_j$ and $\boldsymbol{k\varepsilon}^2\boldsymbol{k}\equiv k_i\varepsilon_{ij}\varepsilon_{jk}k_k$ in notation with three-dimensional vector indices. The vanishing of the determinant (10.4) de-

[11]Fyodorov [27], Eq. (17.21) expresses the second sum term in round brackets by the inverse permittivity tensor $\boldsymbol{\varepsilon}^{-1}$ or, more precisely, by the related complementary tensor $\bar{\boldsymbol{\varepsilon}}\equiv|\boldsymbol{\varepsilon}|\boldsymbol{\varepsilon}^{-1}=\boldsymbol{\varepsilon}^2-\langle\boldsymbol{\varepsilon}\rangle\boldsymbol{\varepsilon}+[\boldsymbol{\varepsilon}]\mathrm{I}$ to $\boldsymbol{\varepsilon}$ with the identity $\langle\bar{\boldsymbol{\varepsilon}}\rangle=[\boldsymbol{\varepsilon}]\equiv\frac{1}{2}\left(\langle\boldsymbol{\varepsilon}\rangle^2-\langle\boldsymbol{\varepsilon}^2\rangle\right)$ as here additionally given. Other (historically older) formulations of vanishing of this determinant in the form of the Fresnel equation in coordinates of the principal axes of the permittivity tensor are known (e.g., Szivessy [46] (Eqs. (47) and (48)) or Born and Wolf [47] (chap. XV.2.2, Eq. (21)) which correspond to (3) and (4) in coordinates of the principal axes and the same is true for Eq. (97.10) in [7]).

scribes a three-dimensional (hyper-)surface in the four-dimensional space of variables $(\boldsymbol{k},\omega)$ which is called dispersion surface.

The dispersion Equation (10.3) can be resolved with respect to one of the 4 components of $(\boldsymbol{k},\omega)$, for example, in the form of the frequency ω as function of the wave vector $\boldsymbol{k}$ that means

$$\omega=\omega(\boldsymbol{k}). \tag{10.5}$$

The group velocity $\boldsymbol{v}$ is then defined by

$$\boldsymbol{v}\equiv\frac{\partial\omega}{\partial\boldsymbol{k}}(\boldsymbol{k})=\boldsymbol{v}(\boldsymbol{k}). \tag{10.6}$$

It is a "regular" velocity also in the relativistic theorem of addition of velocities. Inserting $\omega=\omega(\boldsymbol{k})$ in the dispersion Equation (10.3) we get a scalar identity as a function of the wave vector $\boldsymbol{k}$ from which after differentiation with respect to the wave vector follows a vector identity

$$\left|\mathrm{L}(\boldsymbol{k},\omega(\boldsymbol{k}))\right|=0,\quad\Rightarrow\quad\left(\frac{\partial|\mathrm{L}|}{\partial\boldsymbol{k}}\right)_{(\boldsymbol{k},\omega(\boldsymbol{k}))}+\left(\frac{\partial|\mathrm{L}|}{\partial\omega}\right)_{(\boldsymbol{k},\omega(\boldsymbol{k}))}\frac{\partial\omega}{\partial\boldsymbol{k}}(\boldsymbol{k})=\boldsymbol{0}, \tag{10.7}$$

with arguments of involved functions of $(\boldsymbol{k},\omega)$ taken at $(\boldsymbol{k},\omega(\boldsymbol{k}))$ and we obtain for the group velocity (vectorial indices of $\boldsymbol{v}$ and of $\frac{\partial}{\partial\boldsymbol{k}}$ correspond to each other on left- and right-hand sides)

$$\boldsymbol{v}(k)=-\frac{\left(\frac{\partial|\mathrm{L}|}{\partial\boldsymbol{k}}\right)_{(\boldsymbol{k},\omega(\boldsymbol{k}))}}{\left(\frac{\partial|\mathrm{L}|}{\partial\omega}\right)_{(\boldsymbol{k},\omega(\boldsymbol{k}))}}=-\frac{\left\langle\overline{\mathrm{L}}\frac{\partial\mathrm{L}}{\partial\boldsymbol{k}}\right\rangle_{(\boldsymbol{k},\omega(\boldsymbol{k}))}}{\left\langle\overline{\mathrm{L}}\frac{\partial\mathrm{L}}{\partial\omega}\right\rangle_{(\boldsymbol{k},\omega(\boldsymbol{k}))}}=-\frac{\left(\boldsymbol{E}^*\frac{\partial\mathrm{L}}{\partial\boldsymbol{k}}\boldsymbol{E}\right)_{(\boldsymbol{k},\omega(\boldsymbol{k}))}}{\left(\boldsymbol{E}^*\frac{\partial\mathrm{L}}{\partial\omega}\boldsymbol{E}\right)_{(\boldsymbol{k},\omega(\boldsymbol{k}))}}. \tag{10.8}$$

We applied here the relation $\frac{\partial|\mathrm{A}|}{\partial\lambda}=\left\langle\overline{\mathrm{A}}\frac{\partial\mathrm{A}}{\partial\lambda}\right\rangle$ for the differentiation of the determinant $|\mathrm{A}|$ of an arbitrary operator A with respect to a variable λ where $\overline{\mathrm{A}}$ denotes the complementary operator to the operator A which satisfies the relations $\mathrm{A}\overline{\mathrm{A}}=\overline{\mathrm{A}}\mathrm{A}=|\mathrm{A}|\mathrm{I}$. The complementary operator $\overline{\mathrm{A}}$ to A is determined in components by $\overline{A}_{li}=\frac{1}{2!}\varepsilon_{ijk}\varepsilon_{lmn}A_{jm}A_{kn}$ or due to Hamilton-Cayley identity $\mathrm{A}^3-\langle\mathrm{A}\rangle\mathrm{A}^2+[\mathrm{A}]\mathrm{A}-|\mathrm{A}|\mathrm{I}=0$ in operator form by $\overline{\mathrm{A}}\equiv\mathrm{A}^2-\langle\mathrm{A}\rangle\mathrm{A}+[\mathrm{A}]\mathrm{I}$ where $[\mathrm{A}]\equiv\frac{1}{2}\left(\langle\mathrm{A}\rangle^2-\langle\mathrm{A}^2\rangle\right)=\langle\overline{\mathrm{A}}\rangle$ denotes the second invariant of A and I is the identity operator.

The explicit form of the complementary operator $\overline{\mathrm{L}}(\boldsymbol{k},\omega)$ to $\mathrm{L}(\boldsymbol{k},\omega)$ which by its vanishing determines the optic axes is (with $(\boldsymbol{\varepsilon k})_i\equiv\varepsilon_{ik}k_k$ and $(\boldsymbol{k\varepsilon})_j\equiv k_l\varepsilon_{lj}$ in coordinates)

$$\overline{\mathrm{L}}(\boldsymbol{k},\omega)=\frac{c^4}{\omega^4}(\boldsymbol{k}^2)\boldsymbol{k}\cdot\boldsymbol{k}-\frac{c^2}{\omega^2}\left(\langle\boldsymbol{\varepsilon}\rangle\boldsymbol{k}\cdot\boldsymbol{k}-\boldsymbol{\varepsilon k}\cdot\boldsymbol{k}-\boldsymbol{k}\cdot\boldsymbol{k\varepsilon}+(\boldsymbol{k\varepsilon k})\mathrm{I}\right)+\overline{\boldsymbol{\varepsilon}}, \tag{10.9}$$

with trace equal to ($\langle\overline{\mathrm{A}}\rangle=[\mathrm{A}]$ is a three-dimensional operators identity for arbitrary A)

$$\left\langle \overline{\mathrm{L}}(\boldsymbol{k},\omega)\right\rangle = \left[\mathrm{L}(\boldsymbol{k},\omega)\right] = \frac{c^4}{\omega^4}\left(\boldsymbol{k}^2\right)^2 - \frac{c^2}{\omega^2}\left(\langle\boldsymbol{\varepsilon}\rangle\boldsymbol{k}^2 + \boldsymbol{k}\boldsymbol{\varepsilon}\boldsymbol{k}\right) + [\boldsymbol{\varepsilon}]. \tag{10.10}$$

From the Hamilton-Cayley identity in the form $\mathrm{L}\overline{\mathrm{L}} = \overline{\mathrm{L}}\mathrm{L} = |\mathrm{L}|\mathrm{I}$ it becomes clear that on the dispersion surface $|\mathrm{L}| = 0$ each non-vanishing vector $\overline{\mathrm{L}}\boldsymbol{a} \neq \boldsymbol{0}$ with appropriately chosen vector $\boldsymbol{a}$ is right-hand eigenvector of L to eigenvalue 0 and thus a solution $\boldsymbol{E}$ of Equation (10.1) and each non-vanishing vector $\boldsymbol{b}\overline{\mathrm{L}} \neq \boldsymbol{0}$ is left-hand eigenvector of L to eigenvalue 0 and thus a solution $\boldsymbol{E}^*$ of second Equation (10.1). Therefore, the normalized operator $\overline{\mathrm{L}}$ in case of $|\mathrm{L}| = 0$ as follows by applying the Hamilton-Cayley identity is the normalized dyadic product $\boldsymbol{e}\cdot\boldsymbol{e}^* \propto \boldsymbol{E}\cdot\boldsymbol{E}^*$ and

$$\left(|\mathrm{L}| = 0\right), \quad \Pi \equiv \frac{\overline{\mathrm{L}}}{\left\langle \overline{\mathrm{L}}\right\rangle} = \frac{\overline{\mathrm{L}}}{[\mathrm{L}]} = \boldsymbol{e}\cdot\boldsymbol{e}^*, \quad \Pi^2 = \Pi, \quad \langle\Pi\rangle = \boldsymbol{e}^*\boldsymbol{e} = 1, \tag{10.11}$$

is projection operator for the determination of non-degenerate solutions of Equations (10.1). This explains the last part of the formulae (10.8) for the group velocity (taking into account the general identity $\left\langle(\boldsymbol{a}\cdot\boldsymbol{b})\mathrm{A}\right\rangle = \boldsymbol{b}\mathrm{A}\boldsymbol{a}$). In case of $\overline{\mathrm{L}} = 0$, ($\Rightarrow [\mathrm{L}] = 0$), the first and second part of relations (10.8) become indeterminate due to vanishing numerator and denominator but the last part with representation by the electric field amplitudes remains true. This singular case is the case of optic axes or binormals which we do not further discuss here since it leads us far from our proper aim.

With the first or second part of formulae (10.8), one can find explicit formulae for the group velocity which express it as a function of wave vector $\boldsymbol{k}$, frequency ω and medium properties involved in the permittivity tensor $\boldsymbol{\varepsilon} \equiv \boldsymbol{\varepsilon}(\boldsymbol{k},\omega)$. Although not difficult to obtain, however, they are long taking into account the dispersion and, therefore, we will not write them down (we give them in next Section under neglect of dispersion). Instead of this we will use last part of (10.8) which reveals interesting relations to the action 4-vector and to the energy-momentum tensor. According to (10.8), the group velocity $\boldsymbol{v}_0$ at the considered point $\left(\boldsymbol{k} = \boldsymbol{k}_0, \omega = \omega_0 \equiv \omega(\boldsymbol{k}_0)\right)$ of the dispersion surface (10.5) is determined by

$$v_{0,l} = -\frac{\boldsymbol{E}_0^*\left(\dfrac{\partial\mathrm{L}}{\partial k_l}\right)_0 \boldsymbol{E}_0}{\boldsymbol{E}_0^*\left(\dfrac{\partial\mathrm{L}}{\partial\omega}\right)_0 \boldsymbol{E}_0} = \frac{T_l}{s}, \tag{10.12}$$

where we used the representation (6.12) for action flow density $\boldsymbol{T}$ and action density s in the limiting case of plane monochromatic waves. The energy-momentum tensor for this limiting case of plane monochromatic waves can now be represented in the form (see also next Section)

$$\begin{aligned} T_{kl} &= s k_{0,k} v_{0,l}, \quad g_k = s k_{0,k}, \\ S_l &= s\omega_0 v_{0,l}, \quad w = s\omega_0, \end{aligned} \quad \Leftrightarrow \quad T_{\kappa\lambda} = s\begin{pmatrix} k_{0,k}v_{0,l} & \mathrm{i}ck_{0,k} \\ \dfrac{\mathrm{i}}{c}\omega_0 v_{0,l} & -\omega_0 \end{pmatrix}. \tag{10.13}$$

From this it can be easily seen for this limiting case

$$\mathrm{T} = \boldsymbol{g} \cdot \boldsymbol{v}_0, \quad \boldsymbol{S} = w\boldsymbol{v}_0, \quad \Rightarrow \quad \frac{T_{kl}}{g_k} = \frac{S_l}{w} = v_{0,l}, \tag{10.14}$$

which means that the three-dimensional stress tensor T which in considered case is a dyadic product is proportional to the momentum density $\boldsymbol{g}$ and the energy flow density $\boldsymbol{S}$ to the energy density w with the group velocity $\boldsymbol{v}_0$ as the proportionality factor in analogy to (10.12).

As it is well known [3], the velocity $\boldsymbol{v}_0$ is not spatial part of a relativistic covariant 4-vector but with following modification by the factor

$$\gamma_0 \equiv \frac{1}{\sqrt{1 - \frac{\boldsymbol{v}_0^2}{c^2}}}, \tag{10.15}$$

one obtains the relativistic covariant 4-vector of velocity $u_{0,\lambda}$

$$u_{0,\lambda} \equiv \left(u_{0,l}, u_{0,4}\right) = \gamma_0 \left(v_{0,l}, \mathrm{i}c\right), \quad \Rightarrow \quad u_0^2 \equiv u_{0,\lambda} u_{0,\lambda} = -c^2. \tag{10.16}$$

Using it the energy-momentum tensor (10.13) may be represented in the following relativistic covariant form

$$T_{\kappa\lambda} = s_0 k_{0,\kappa} u_{0,\lambda} = s_0 \begin{pmatrix} k_{0,k} u_{0,l} & \mathrm{i}\gamma_0 c k_{0,k} \\ \frac{\mathrm{i}}{c}\omega_0 u_{0,l} & -\gamma_0 \omega_0 \end{pmatrix}, \quad s_0 = s\sqrt{1 - \frac{\boldsymbol{v}_0^2}{c^2}} \equiv \frac{s}{\gamma_0}, \tag{10.17}$$

where s_0 is the action density in the inertial system where the wave packet is resting. This is in analogy to a homogeneous particle flow in classical hydrodynamics without interaction of the particles (or without inner pressure) for which the energy-momentum tensor possesses the form

$$T_{\kappa\lambda} = n_0 p_{0,\kappa} u_{0,\lambda}, \quad p_{0,\kappa} = m_0 u_{0,\kappa}, \quad n_0 = n\sqrt{1 - \frac{\boldsymbol{v}_0^2}{c^2}} \equiv \frac{n}{\gamma_0}, \tag{10.18}$$

where $p_{0,\kappa} = m_0 u_{0,\kappa}$ is the momentum of one particle, m_0 its rest mass and n_0 the particle density in the inertial system where the particles rest (e.g., [12]). The analogy of (10.17) to (10.18) for a homogeneous particle flow suggests (with knowledge of quantum theory) to interpret the first as homogeneous flow of quasiparticles and to introduce an abbreviation $\hbar$ according to

$$\hbar \equiv \frac{s_0}{n_0} = \frac{s}{n}, \tag{10.19}$$

as action of one particle independently of the considered inertial system (*i.e.*, as a Lorentz invariant and, moreover, even as adiabatic Lorentz invariant as may be shown) and we may write

$$p_{0,\kappa} = \frac{s_0}{n_0} k_{0,\kappa} = \hbar k_{0,\kappa}, \tag{10.20}$$

as 4-vector of the momentum of quasiparticles in agreement with quantum theory. Then we have analogous expressions for the energy-momentum tensors of a homogeneous flow without pressure in classical hydrodynamics on one side and in macroscopic electrodynamics on the other side where the last provides a richer variety of possible functional dependencies $\boldsymbol{v} \equiv \boldsymbol{v}(\boldsymbol{k})$ (or $\boldsymbol{v} \equiv \boldsymbol{v}(\boldsymbol{p})$) or of their inversion $\boldsymbol{k} \equiv \boldsymbol{k}(\boldsymbol{v})$ (or $\boldsymbol{p} \equiv \boldsymbol{p}(\boldsymbol{v})$) than (relativistic) classical mechanics which considers only

$$\underbrace{p}_{=\hbar k} \equiv \underbrace{\left(\boldsymbol{p}, \mathrm{i}\frac{E}{c}\right)}_{=\hbar\left(\boldsymbol{k}, \mathrm{i}\frac{\omega}{c}\right)} = m_0 \left(\frac{\boldsymbol{v}}{\sqrt{1-\frac{\boldsymbol{v}^2}{c^2}}}, \mathrm{i}\frac{c}{\sqrt{1-\frac{\boldsymbol{v}^2}{c^2}}} \right) \equiv m_0 u, \tag{10.21}$$

with m_0 the rest mass of one particle. Usually, the relation between 4-vectors p and u in macroscopic electrodynamics is a 4-tensorial one with tensor components depending on components of $\boldsymbol{v}$ separately where this cannot be expressed by only relativistic scalars such as m_0. A certain exception is formed by transverse waves in a cold isotropic plasma (Section 12).

It was mentioned but not explicitly shown that the local action conservation (6.8) or (6.11) is a more general conservation law than the local energy-momentum conservation (7.2) or (8.2) together with (8.3) and holds also for inhomogeneous media (in general, spatially and temporarily inhomogeneous). If we suppose that the action conservation is true for an inhomogeneous medium that means $\nabla_\lambda T_\lambda(r) = 0$ it is informative to see how the energy-momentum conservation is lost for such a medium in case of propagation of almost plane monochromatic waves as here considered. We may assume that in a weakly inhomogeneous medium as main effect the 4-wave vector k_0 becomes dependent on the considered space-time point $r \equiv (\boldsymbol{r}, t)$ within the medium that means $k_0 = k_0(r)$. Then we find for the 4-divergence of the energy-momentum tensor under the supposition that local action conservation $\nabla_\lambda T_\lambda(r) = 0$ holds

$$\begin{aligned}
\nabla_\lambda T_{\kappa\lambda}(r) &= \nabla_\lambda \left(k_{0,\kappa}(r) T_\lambda(r)\right) \\
&= \frac{\partial k_{0,\kappa}}{\partial r_\lambda}(r) \underbrace{T_\lambda(r)}_{=s_0(r)u_{0,\lambda}(r)} + k_{0,\kappa}(r) \underbrace{\nabla_\lambda T_\lambda(r)}_{=0} \\
&= \frac{\partial k_{0,\kappa}}{\partial r_\lambda}(r) T_\lambda(r).
\end{aligned} \tag{10.22}$$

The right-hand side is non-vanishing that corresponds to local non-conservation of energy-momentum and the 4-divergence of the energy-momentum tensor (if we overtake its formula from the homogeneous medium) becomes a linear combination of the components of the action vector $T_\lambda(r)$.

11. Neglect of Spatial and Frequency Dispersion and Group Velocity

In the considerations of Section 10 about the group velocity and the representation of the energy-momentum tensor in the limiting case of plane monochromatic waves in analogy to that for a homogeneous particle flow we did not use the explicit form of the determinant $\left|\mathrm{L}(\boldsymbol{k}, \omega)\right|$ of the wave-equation operator

$\mathrm{L}(\boldsymbol{k},\omega)$. In the following, we will make some explicit calculations of the group velocity under neglect of the dispersion (spatial and frequency one).

Neglecting spatial and temporal dispersion of the medium means that we consider

$$\boldsymbol{\varepsilon}(\boldsymbol{k},\omega) \equiv \boldsymbol{\varepsilon}_0 \tag{11.1}$$

as a constant permittivity tensor $\boldsymbol{\varepsilon}_0$ in the inertial system of the resting medium or, at least, as a good approximation for a neighborhood of the considered mean wave vector $\boldsymbol{k}_0$ and mean frequency ω_0. Using the first of the relations for the group velocity in (10.8) and a transformation of the denominator by means of the dispersion equation $|\mathrm{L}(\boldsymbol{k},\omega)| = 0$, we find

$$\boldsymbol{v} = \omega \frac{2(\boldsymbol{k}\boldsymbol{\varepsilon}_0\boldsymbol{k})\boldsymbol{k} + \boldsymbol{k}^2(\boldsymbol{\varepsilon}_0\boldsymbol{k} + \boldsymbol{k}\boldsymbol{\varepsilon}_0) - \frac{\omega^2}{c^2}\left(\langle\boldsymbol{\varepsilon}_0\rangle(\boldsymbol{\varepsilon}_0\boldsymbol{k} + \boldsymbol{k}\boldsymbol{\varepsilon}_0) - \left(\boldsymbol{\varepsilon}_0^2\boldsymbol{k} + \boldsymbol{k}\boldsymbol{\varepsilon}_0^2\right)\right)}{4\boldsymbol{k}^2(\boldsymbol{k}\boldsymbol{\varepsilon}_0\boldsymbol{k}) - 2\frac{\omega^2}{c^2}\left(\langle\boldsymbol{\varepsilon}_0\rangle\boldsymbol{k}\boldsymbol{\varepsilon}_0\boldsymbol{k} - \boldsymbol{k}\boldsymbol{\varepsilon}_0^2\boldsymbol{k}\right)}, \tag{11.2}$$

where we emphasize that the permittivity tensor $\boldsymbol{\varepsilon}_0$ herein is, in general, not a symmetric tensor that includes gyrotropy of the medium (see also (4.5)). Furthermore, in general, the directions of $\boldsymbol{k}$ and $\boldsymbol{v}$ in anisotropic media are different. From (11.2) follows immediately for the scalar product of wave vector with group velocity

$$\boldsymbol{k}\boldsymbol{v} = \omega, \tag{11.3}$$

that proves to be equivalent to vanishing of the trace of the energy-momentum tensor under neglect of dispersion (see (9.9) in connection with (9.7) and (10.12)). According to (15) this also means that the frequency ω' in the inertial system $\mathcal{I}' = \mathcal{I}_0$ which moves with the group velocity $\boldsymbol{v}$ in the inertial system $\mathcal{I}$ of the resting medium (*i.e.* $\boldsymbol{V} = \boldsymbol{v}$) vanishes and due to (A.17) that the wave vector is transformed in the following way $\left(\gamma^2 \equiv \left(1 - \frac{\boldsymbol{v}^2}{c^2}\right)^{-1}\right)$

$$\omega_0' = 0, \quad [\boldsymbol{k}_0', \boldsymbol{v}] = [\boldsymbol{k}, \boldsymbol{v}], \quad \boldsymbol{k}_0'\boldsymbol{v} = \gamma\left(\boldsymbol{k}\boldsymbol{v} - \omega\frac{\boldsymbol{v}^2}{c^2}\right) = \gamma\omega\left(1 - \frac{\boldsymbol{v}^2}{c^2}\right) = \frac{\omega}{\gamma}. \tag{11.4}$$

However, already the presence of frequency dispersion (and, moreover, of spatial dispersion) destroys these relations since we have then additional terms in the denominator of the right-hand side in (11.2) which contain the derivatives $\frac{\partial\boldsymbol{\varepsilon}}{\partial\omega}$ of the permittivity tensor $\boldsymbol{\varepsilon} \equiv \boldsymbol{\varepsilon}(\omega)$.

In case of neglected dispersion, the operator $\mathrm{L}(\boldsymbol{k},\omega)$ becomes a function of only a vector $\boldsymbol{n}$ which is called refraction vector (and, clearly, of medium properties contained in $\boldsymbol{\varepsilon}_0$) according to

$$\mathrm{L}(\boldsymbol{k},\omega) \to \mathrm{L}(\boldsymbol{n}) \equiv \boldsymbol{n}\cdot\boldsymbol{n} - \boldsymbol{n}^2\mathrm{I} + \boldsymbol{\varepsilon}_0, \quad \boldsymbol{n} \equiv \frac{c}{\omega}\boldsymbol{k}. \tag{11.5}$$

The dispersion equation $|\mathrm{L}(\boldsymbol{n})| = 0$ leads to the following fourth-order equation in the components of the refraction vector $\boldsymbol{n}$ (e.g., [9], Eq. (2.22))

$$\left|\mathrm{L}(\boldsymbol{n})\right| = \left(\boldsymbol{n}^2 - \langle \boldsymbol{\varepsilon}_0 \rangle\right)\boldsymbol{n}\boldsymbol{\varepsilon}_0\boldsymbol{n} + \boldsymbol{n}\boldsymbol{\varepsilon}_0^2\boldsymbol{n} + \left|\boldsymbol{\varepsilon}_0\right| = 0. \tag{11.6}$$

The dispersion surface describes now a two-dimensional surface in the three-dimensional space of refraction vectors $\boldsymbol{n}$ which is better accessible for visualization than the three-dimensional (hyper-) surface in four-dimensional $(\boldsymbol{k},\omega)$ -space described by Equation (10.3).

Under neglect of the dispersion, the group velocity $\boldsymbol{v}$ depends only on the quotient $\frac{\boldsymbol{k}}{\omega}$ as one can explicitly see from (11.2) and, therefore, on the refraction vector $\frac{c}{\omega}\boldsymbol{k} \equiv \boldsymbol{n}$. It is favorable to normalize the group velocity by the light velocity c and to introduce together with the refraction vector $\boldsymbol{n}$ a ray vector $\boldsymbol{s}$ by (our notations agree with that of Landau and Lifshits, Vol. VIII [7])[12]

$$\boldsymbol{n} \equiv \frac{c\boldsymbol{k}}{\omega}, \quad \boldsymbol{s} \equiv \frac{\boldsymbol{v}}{c}, \quad \Leftrightarrow \quad \boldsymbol{k} \equiv \frac{\omega}{c}\boldsymbol{n}, \quad \boldsymbol{v} \equiv c\boldsymbol{s}, \tag{11.7}$$

and from (11.3) follows

$$\boldsymbol{n}\boldsymbol{s} = 1. \tag{11.8}$$

If we substitute now the operator $\mathrm{L}(\boldsymbol{k},\omega)$ by the operator $\mathrm{L}(\boldsymbol{n})$ according to $\mathrm{L}(\boldsymbol{k},\omega) \to \mathrm{L}(\boldsymbol{n})$, then we have to substitute derivatives of these operators and of their functions according to $\frac{\partial}{\partial k_l} \to \frac{\partial n_k}{\partial k_l}\frac{\partial}{\partial n_k} = \frac{c}{\omega}\frac{\partial}{\partial n_l}$, $\frac{\partial}{\partial \omega} \to \frac{\partial n_k}{\partial \omega}\frac{\partial}{\partial n_k} = -\frac{n_k}{\omega}\frac{\partial}{\partial n_k}$. Applied to the ray vector $\boldsymbol{s} \equiv \frac{\boldsymbol{v}}{c}$ this means that formula (10.8) can now be substituted by

$$\boldsymbol{s} = \frac{\frac{\partial|\mathrm{L}|}{\partial \boldsymbol{n}}}{n_k \frac{\partial|\mathrm{L}|}{\partial n_k}} = \frac{\left\langle \overline{\mathrm{L}}\frac{\partial \mathrm{L}}{\partial \boldsymbol{n}} \right\rangle}{n_k \left\langle \overline{\mathrm{L}}\frac{\partial \mathrm{L}}{\partial n_k} \right\rangle} = \frac{\boldsymbol{E}_0^* \frac{\partial \mathrm{L}}{\partial \boldsymbol{n}} \boldsymbol{E}_0}{n_k \boldsymbol{E}_0^* \frac{\partial \mathrm{L}}{\partial n_k} \boldsymbol{E}_0}. \tag{11.9}$$

where vectorial indices of $\boldsymbol{s}$ and of $\frac{\partial}{\partial \boldsymbol{n}}$ in numerator correspond to each other (or $s_l = \frac{\boldsymbol{E}_0^* \frac{\partial \mathrm{L}}{\partial n_l} \boldsymbol{E}_0}{n_k \boldsymbol{E}_0^* \frac{\partial \mathrm{L}}{\partial n_k} \boldsymbol{E}_0}$ and $\boldsymbol{E}_0^* \frac{\partial \mathrm{L}}{\partial n_l} \boldsymbol{E}_0 \equiv E_{0,i}^* \frac{\partial L_{ij}}{\partial n_l} E_{0,j}$). The first part of this equation shows the well-known property that in case of neglected dispersion the ray vector is proportional to the gradient $\frac{\partial|\mathrm{L}|}{\partial \boldsymbol{n}}$ of the dispersion surface $\left|\mathrm{L}(\boldsymbol{n})\right| = 0$ at the considered point and the denominator determines its norma-

[12]Many authors, however, denote with $\boldsymbol{n}$ a unit vector in direction of $\boldsymbol{k}$, for example, Fyodorov [37] [38], the initiator of coordinate-invariant methods, who denotes refraction vectors by $\boldsymbol{m}$ and ray vectors by $\boldsymbol{p}$ and, furthermore, denotes with $\boldsymbol{s}$ a unit vector in direction of the ray vector. Born and Wolf [47] denote with $\boldsymbol{s}$ a unit vector in direction of the wave vector $\boldsymbol{k}$ and with $\boldsymbol{t}$ a unit vector in direction of the ray vector. In our notation we have conveniently $\boldsymbol{n}^2 = n^2$ where $n \equiv |\boldsymbol{n}|$ is the index of refraction of the wave.

lization according to $\boldsymbol{ns}=1$. With the explicit form of the determinant $|\mathrm{L}(\boldsymbol{n})|$ given in (11.6) from which we easily calculate its derivatives with respect to the refraction vector $\boldsymbol{n}$ we find for the ray vector

$$\boldsymbol{s}=\frac{2(\boldsymbol{n}\boldsymbol{\varepsilon}_0\boldsymbol{n})\boldsymbol{n}+\left(\boldsymbol{n}^2-\langle\boldsymbol{\varepsilon}_0\rangle\right)(\boldsymbol{n}\boldsymbol{\varepsilon}_0+\boldsymbol{\varepsilon}_0\boldsymbol{n})+\boldsymbol{\varepsilon}_0^2\boldsymbol{n}+\boldsymbol{n}\boldsymbol{\varepsilon}_0^2}{2\left(\left(2\boldsymbol{n}^2-\langle\boldsymbol{\varepsilon}_0\rangle\right)\boldsymbol{n}\boldsymbol{\varepsilon}_0\boldsymbol{n}+\boldsymbol{n}\boldsymbol{\varepsilon}_0^2\boldsymbol{n}\right)}\equiv\boldsymbol{s}(\boldsymbol{n}), \tag{11.10}$$

confirmed by the calculation from (11.2) using the definition of the ray vector. However, the main purpose of these calculations was to establish the algebraic structure of the ray vectors $\boldsymbol{s}$ that means its connection to the invariants of $\mathrm{L}\equiv\mathrm{L}(\boldsymbol{n})$ and to its complementary operator $\overline{\mathrm{L}}$. The inversion of the vectorial function $\boldsymbol{s}=\boldsymbol{s}(\boldsymbol{n})$ to a vectorial function $\boldsymbol{n}=\boldsymbol{n}(\boldsymbol{s})$ by means of duality relations between ray and refraction quantities or otherwise is up to now only solved under the restriction $\boldsymbol{\varepsilon}_0=\boldsymbol{\varepsilon}_0^{\mathrm{T}}$ that means to nongyrotropy of the medium.

Using the refraction vector $\boldsymbol{n}$ and the ray vector $\boldsymbol{s}$, the 4-wave vector k and the 4-vector of velocity u (see (10.16)) can be represented as follows

$$\begin{aligned} &k\equiv\left(\boldsymbol{k},\mathrm{i}\frac{\omega}{c}\right)=\frac{\omega}{c}(\boldsymbol{n},\mathrm{i}), \quad\Rightarrow\quad k^2=\boldsymbol{k}^2-\frac{\omega^2}{c^2}=\frac{\omega^2}{c^2}\left(\boldsymbol{n}^2-1\right),\\ &u\equiv\left(\frac{\boldsymbol{v}}{\sqrt{1-\frac{\boldsymbol{v}^2}{c^2}}},\mathrm{i}\frac{c}{\sqrt{1-\frac{\boldsymbol{v}^2}{c^2}}}\right)=c\left(\frac{\boldsymbol{s}}{\sqrt{1-\boldsymbol{s}^2}},\mathrm{i}\frac{1}{\sqrt{1-\boldsymbol{s}^2}}\right), \quad\Rightarrow\quad u^2=-c^2. \end{aligned} \tag{11.11}$$

The scalar product of 4-wave vector and 4-velocity becomes vanishing in case of neglected dispersion

$$ku=\frac{\boldsymbol{kv}-\omega}{\sqrt{1-\frac{\boldsymbol{v}^2}{c^2}}}=\omega\frac{\boldsymbol{ns}-1}{\sqrt{1-\boldsymbol{s}^2}}=0. \tag{11.12}$$

The scalar products of 4-vectors k^2,u^2 and ku are Lorentz invariants and values for them which are calculated in one inertial system such as here in the inertial system of the resting medium remain the same in arbitrary other inertial systems.

If we apply this to the mean wave vector $\boldsymbol{k}_0$ and mean frequency ω_0 in the expressions for the energy-momentum tensor in first approximation in (9.7), we find for neglected dispersion

$$T_{\kappa\kappa}=T_{kk}-w=0, \quad\Leftrightarrow\quad \langle\mathrm{T}\rangle\equiv T_{kk}=w, \tag{11.13}$$

which means that the trace of the energy-momentum tensor $T_{\kappa\lambda}$ is vanishing in such approximation. This is well known for the general energy-momentum tensor in vacuum electrodynamics [3]. Using this together with (9.11) we find for neglected dispersion

$$\langle\mathrm{T}\rangle=w, \quad \boldsymbol{Sg}=\langle\mathrm{T}\rangle w \quad\Rightarrow\quad \boldsymbol{Sg}=\langle\mathrm{T}\rangle^2=w^2\geq 0. \tag{11.14}$$

This means, in particular, that the scalar product of energy flow density $\boldsymbol{S}$ with momentum density $\boldsymbol{g}$ is equal to the square of the energy density w.

Furthermore, under neglect of dispersion and in approximation of plane monochromatic waves, we find from (9.8)

$$k_{0,\lambda}T_{\lambda}=k_{0,l}T_{l}-\omega_{0}s=0,\quad\Rightarrow\quad k_{0,l}T_{kl}=\omega_{0}g_{k},\quad k_{0,l}S_{l}=\omega_{0}w,\tag{11.15}$$

that, however, is no more true taking into account the dispersion of the medium. First relation in (11.15) states that the scalar product $k_{0,\lambda}T_{\lambda}$ of mean 4-wave vector $k_{0,\lambda}$ with action 4-vector T_{λ} vanishes meaning that they are mutually orthogonal in Minkowski space. Last relation in (11.14) suggests that the scalar product of energy flow density $\boldsymbol{S}$ with momentum density $\boldsymbol{g}$ should be greater than zero or otherwise these two vectors form an angle greater than $\frac{\pi}{2}$.

As a special case we consider now transversal waves in a resting isotropic medium under neglect of dispersion with the constant permittivity tensor $\varepsilon_{0,ij}=\varepsilon_{0}\delta_{ij}$ for which we find the following relations between wave vector $\boldsymbol{k}$ and group velocity $\boldsymbol{v}$

$$\boldsymbol{\varepsilon}=\varepsilon_{0}I,\quad\Rightarrow\quad\boldsymbol{k}^{2}=\frac{\omega^{2}}{c^{2}}\varepsilon_{0},\quad\boldsymbol{v}=\frac{c^{2}}{\omega\varepsilon_{0}}\boldsymbol{k}=\frac{c}{\sqrt{\varepsilon_{0}}}\frac{\boldsymbol{k}}{|\boldsymbol{k}|},\quad\boldsymbol{k}\boldsymbol{v}-\omega=0,\tag{11.16}$$

and between ray vector $\boldsymbol{s}$ and refraction vector $\boldsymbol{n}$

$$\boldsymbol{s}\equiv\frac{\boldsymbol{v}}{c}=\frac{c\boldsymbol{k}}{\omega\varepsilon_{0}}\equiv\frac{\boldsymbol{n}}{\varepsilon_{0}},\quad\Leftrightarrow\quad k\equiv\left(\boldsymbol{k},i\frac{\omega}{c}\right)=\frac{\omega\varepsilon_{0}}{c^{2}}\left(\boldsymbol{v},i\frac{c}{\varepsilon_{0}}\right).\tag{11.17}$$

For the energy-momentum tensor then follows according to (10.13)

$$\begin{aligned}T(r)&\equiv\begin{pmatrix}\mathrm{T}&\mathrm{i}c\boldsymbol{g}\\\frac{\mathrm{i}}{c}\boldsymbol{S}&-w\end{pmatrix}=s(r)\begin{pmatrix}\boldsymbol{k}\cdot\boldsymbol{v}&\mathrm{i}c\boldsymbol{k}\\\frac{\mathrm{i}}{c}\omega\boldsymbol{v}&-\omega\end{pmatrix}\\&=s(r)\omega\begin{pmatrix}\varepsilon_{0}\boldsymbol{s}\cdot\boldsymbol{s},&\mathrm{i}\varepsilon_{0}\boldsymbol{s}\\\mathrm{i}\boldsymbol{s}&-1\end{pmatrix}=s(r)\frac{\omega}{\varepsilon_{0}}\begin{pmatrix}\boldsymbol{n}\cdot\boldsymbol{n}&\mathrm{i}\varepsilon_{0}\boldsymbol{n}\\\mathrm{i}\boldsymbol{n}&-\varepsilon_{0}\end{pmatrix},\end{aligned}\tag{11.18}$$

where $s(r)$ is the action density in the system of the resting medium (in contrast to ray vector $\boldsymbol{s}$). This tensor is non-symmetric but its spatial part $\mathrm{T}=s(r)\boldsymbol{k}\cdot\boldsymbol{v}$ is symmetric since the vectors $\boldsymbol{k}$ and $\boldsymbol{v}$ possess the same direction (for the necessity of this symmetry for isotropic media see Section 14). The trace $\langle T\rangle$ of the four-dimensional tensor T is vanishing

$$\langle T(r)\rangle=s(r)(\boldsymbol{k}\boldsymbol{v}-\omega)=0,\tag{11.19}$$

that is only true under neglect of the dispersion. In the next Section we discuss the energy-momentum tensor for a special isotropic medium but without neglect of the dispersion.

12. Energy-Momentum Tensor of a Cold Isotropic Plasma and Transverse Photons with Scalar Rest Mass

As one of the simplest models including frequency dispersion in explicit form we now consider a cold isotropic plasma. Its energy-momentum tensor shows the peculiarity that it is symmetric. Without an external magnetic field it possesses the following permittivity tensor, e.g., [7] [11]

$$\varepsilon_{ij}(\boldsymbol{k},\omega)=\varepsilon(\omega)\delta_{ij},\quad \varepsilon(\omega)=1-\frac{\omega_p^2}{\omega^2},\quad \omega_p\equiv\sqrt{\frac{4\pi n_e e^2}{m_e}},\tag{12.1}$$

where ω_p denotes the plasma frequency which, e.g., for an electron plasma is expressed expressed by the plasma parameters n_e, e, m_e (electron density, electron charge and electron mass).

The dispersion equation $\varepsilon(\omega)=0$ for longitudinal waves with its resolution $\omega=\omega_p$ does not include the wave vector $\boldsymbol{k}$ and therefore the group velocity $\boldsymbol{v}$ vanishes. This means that longitudinal localized excitations cannot propagate or decay in the approximation of absent spatial dispersion and absent losses. The energy-momentum tensor for these excitations can be obtained from (10.13) by setting $\boldsymbol{v}_0=\boldsymbol{0}$ for the group velocity and using relation (10.19) between action and particle (excitation) density.

The dispersion equation for transverse waves $c^2\boldsymbol{k}^2=\omega^2\varepsilon(\omega)$ with real wave vector $\boldsymbol{k}=\boldsymbol{k}^*$ is

$$c^2\boldsymbol{k}^2-\omega^2+\omega_p^2=0,\quad\Rightarrow\quad \frac{\omega}{\omega_p}=\frac{\sqrt{\omega_p^2+c^2\boldsymbol{k}^2}}{\omega_p}\geq 1.\tag{12.2}$$

From (12.2) follows for the group velocity $\boldsymbol{v}$ of transverse waves or of their quasiparticles

$$\boldsymbol{v}\equiv\frac{\partial\omega}{\partial\boldsymbol{k}}=\frac{c^2\boldsymbol{k}}{\omega}=c\sqrt{\frac{c^2\boldsymbol{k}^2}{\omega_p^2+c^2\boldsymbol{k}^2}}\frac{\boldsymbol{k}}{|\boldsymbol{k}|}\equiv\boldsymbol{v}(\boldsymbol{k}),\quad\Rightarrow\quad\frac{|\boldsymbol{v}|}{c}=\sqrt{\frac{c^2\boldsymbol{k}^2}{\omega_p^2+c^2\boldsymbol{k}^2}}\leq 1,\tag{12.3}$$

with $|\boldsymbol{k}|\equiv\sqrt{\boldsymbol{k}^2}$ and $|\boldsymbol{v}|\equiv\sqrt{\boldsymbol{v}^2}$. Due to isotropy, the group velocity $\boldsymbol{v}$ possesses the direction of the wave vector $\boldsymbol{k}$ and $\boldsymbol{v}\equiv\boldsymbol{v}(\boldsymbol{k})$ can be converted to the vectorial function $\boldsymbol{k}\equiv\boldsymbol{k}(\boldsymbol{v})$ according to

$$\boldsymbol{k}=|\boldsymbol{k}|\frac{\boldsymbol{v}}{|\boldsymbol{v}|}=\frac{\omega_p}{c^2}\frac{\boldsymbol{v}}{\sqrt{1-\frac{\boldsymbol{v}^2}{c^2}}}\equiv\boldsymbol{k}(\boldsymbol{v}),\quad\Rightarrow\quad \boldsymbol{k}\boldsymbol{v}=\omega\frac{\boldsymbol{v}^2}{c^2},\quad\frac{c^2\boldsymbol{k}^2}{\omega_p^2}=\frac{\frac{\boldsymbol{v}^2}{c^2}}{1-\frac{\boldsymbol{v}^2}{c^2}}.\tag{12.4}$$

We introduce an abbreviation γ in analogy to the procedure for special Lorentz transformations by

$$\gamma\equiv\frac{1}{\sqrt{1-\frac{\boldsymbol{v}^2}{c^2}}}=\frac{\sqrt{\omega_p^2+c^2\boldsymbol{k}^2}}{\omega_p}\geq 1,\quad\Rightarrow\quad\frac{\boldsymbol{v}^2}{c^2}=\frac{\gamma^2-1}{\gamma^2},\quad \boldsymbol{k}\boldsymbol{v}-\omega=-\frac{\omega_p}{\gamma}.\tag{12.5}$$

which is equivalent to $\omega=\gamma\omega_p$ with ω_p the minimal possible frequency for velocity $\boldsymbol{v}=\boldsymbol{0}$.

The second-order derivatives of $\omega\equiv\omega(\boldsymbol{k})$ for transverse waves in a cold plasma consist of a transverse part proportional to $\delta_{kl}-\frac{k_k k_l}{|\boldsymbol{k}|^2}$ and a longitudinal part proportional to $\frac{k_k k_l}{|\boldsymbol{k}|^2}$ given by

$$\frac{\partial^2 \omega}{\partial k_k \partial k_l} = \frac{c^2}{\omega}\left(\delta_{kl} - \frac{k_k k_l}{|\boldsymbol{k}|^2} + \frac{\omega_p^2}{\omega^2}\frac{k_k k_l}{|\boldsymbol{k}|^2}\right). \tag{12.6}$$

They are responsible for diffraction perpendicular and parallel to the direction of propagation and we call $\frac{\partial^2 \omega}{\partial k_k \partial k_l}$ the diffraction tensor.

For the 4-wave vector k we find from (12.4) the following proportionality to the 4-velocity u

$$k \equiv \left(\boldsymbol{k}, \mathrm{i}\frac{\omega}{c}\right) = \frac{\omega_p}{c^2}(\gamma \boldsymbol{v}, \mathrm{i}\gamma c) = \frac{\omega_p}{c^2}(\boldsymbol{u}, \mathrm{i}c) \equiv \frac{m_p}{\hbar}u, \tag{12.7}$$

where we defined a "rest" mass m_p which is independent on the velocity $\boldsymbol{v}$ of the moving particles (plasmons) and depends beside physical constants only on the plasma frequency ω_p by

$$m_p c^2 \equiv \hbar \omega_p. \tag{12.8}$$

From (12.7) follows for the squared modulus k^2 of the 4-wave vector k which is a Lorentz-invariant

$$k^2 = \boldsymbol{k}^2 - \frac{\omega^2}{c^2} = -\frac{\omega_p^2}{c^2} = -\frac{m_p^2 c^2}{\hbar^2}, \tag{12.9}$$

independently of the frequency ω in the resting plasma. Therefore, it can also serve as a model for only one kind of transverse photons with the same rest mass independently on the frequency and it seems to be clear that it is the only case of a medium which provide this. Furthermore, we find

$$ku = \gamma(\boldsymbol{k}\boldsymbol{v} - \omega) = -\omega_p = -\frac{m_p c^2}{\hbar}, \tag{12.10}$$

in contrast to its vanishing (11.12) in case of neglected dispersion.

For the energy-momentum tensor in three-dimensional representation and written in coordinate-invariant way, *i.e.*, without indices (see **Appendix B**) and furthermore if we omit the indices "0" at $\boldsymbol{k}$ and ω_0 we find from (12.3) the relation $\boldsymbol{k} = \frac{\omega}{c^2}\boldsymbol{v} = \frac{\omega_p}{c^2}\boldsymbol{u}$ and using the definition of m_p by (12.8)

$$T \equiv \begin{pmatrix} \mathrm{T} & \mathrm{i}c\boldsymbol{g} \\ \frac{\mathrm{i}}{c}\boldsymbol{S} & -w \end{pmatrix} = s_0 \begin{pmatrix} \boldsymbol{k}\cdot\boldsymbol{v} & \mathrm{i}c\boldsymbol{k} \\ \frac{\mathrm{i}}{c}\omega\boldsymbol{v} & -\omega \end{pmatrix} = n_0 m_p \begin{pmatrix} \boldsymbol{u}\cdot\boldsymbol{u} & \mathrm{i}c\boldsymbol{u} \\ \mathrm{i}c\boldsymbol{u} & -c^2 \end{pmatrix}, \tag{12.11}$$

where $s_0 = n_0\hbar$ is the action density and n_0 the particle density with mass m_p defined in (12.8) in the inertial system $\mathcal{I}_0$ of the resting plasma. The four-dimensional energy-momentum tensor (12.11) is completely symmetric and can be represented in the factorized form

$$T_{\kappa\lambda} = n_0 m_p u_\kappa u_\lambda = \frac{n_0}{m_p}\hbar^2 k_\kappa k_\lambda = T_{\lambda\kappa}, \tag{12.12}$$

Its trace is is represented by (see **Appendix B**)

$$\langle T \rangle = \langle \mathrm{T} \rangle - w = s_0 \left(\boldsymbol{kv} - \omega \right) = -s_0 \frac{\omega_p}{\gamma} = -n_0 \frac{m_p c^2}{\gamma} \leq 0. \tag{12.13}$$

with the trace of the three-dimensional part of the stress tensor T

$$\langle \mathrm{T} \rangle = s_0 \boldsymbol{kv} = \left(\sqrt{n_0 m_p} \gamma \boldsymbol{v} \right)^2 \geq 0, \quad w = s_0 \omega \geq 0. \tag{12.14}$$

Since the four-dimension tensor (or operator) T as well as its spatial part T factorize they possess only one non-vanishing eigenvalue and we have

$$[T] = 0, \quad |T| = 0, \quad \|T\| = 0, \quad [\mathrm{T}] = 0, \quad |\mathrm{T}| = 0, \tag{12.15}$$

From (B.8) follows then

$$\boldsymbol{Sg} = \langle \mathrm{T} \rangle w \geq 0, \quad \boldsymbol{S}\mathrm{T}\boldsymbol{g} = \langle \mathrm{T} \rangle^2 w \geq 0, \quad \boldsymbol{S}\mathrm{T}^2 \boldsymbol{g} = \langle \mathrm{T} \rangle^3 w \geq 0. \tag{12.16}$$

The derivations were made for real wave vector $\boldsymbol{k}$ and since $\boldsymbol{S}$ and $\boldsymbol{g}$ are proportional to $\boldsymbol{k}$ in this case the positivity of $\boldsymbol{Sg}$ is also understandable from this side.

In the inertial system $\mathcal{I}'$ where the excitation rests that means which moves with group velocity $\boldsymbol{v}$ in $\mathcal{I}_0$ and thus where $\boldsymbol{v} \to \boldsymbol{v}' = \boldsymbol{0}$ we have the energy momentum tensor

$$T' = \begin{pmatrix} 0 & 0 \\ 0 & -w' \end{pmatrix} = s_0' \begin{pmatrix} 0 & 0 \\ 0 & -\omega' \end{pmatrix} = n_0' \begin{pmatrix} 0 & 0 \\ 0 & -m_p c^2 \end{pmatrix},$$

$$w' = n_0 m_p c^2 = n_0 \hbar \omega_p = s_0 \omega_p, \quad n_0' = \gamma n_0, \quad \langle T' \rangle = -w'. \tag{12.17}$$

In the transition $\mathcal{I} \to \mathcal{I}' = \mathcal{I}_0$ to this system, we have to transform (see transformation formulae in Appendix A with $\boldsymbol{V} = \boldsymbol{v}$): $\boldsymbol{k} \to \boldsymbol{k}' = \boldsymbol{0}$, $\omega \to \omega' = \omega_p$, $\boldsymbol{E}_0 \to \boldsymbol{E}_0 = \sqrt{1 - \frac{\boldsymbol{v}^2}{c^2}} \boldsymbol{E}_0$. Therefore, the phase factor $\exp\left\{ \mathrm{i}\left(\boldsymbol{k}_0 \boldsymbol{r} - \omega_0 t \right) \right\}$ transforms according to

$$\exp\left\{ \mathrm{i}\left(\boldsymbol{kr} - \omega t \right) \right\} \to \exp\left\{ \mathrm{i}\left(\boldsymbol{k}'\boldsymbol{r}' - \omega' t' \right) \right\} = \exp\left(-\mathrm{i}\omega_p t' \right), \tag{12.18}$$

and the excitation appears in $\mathcal{I}'$ as a pure oscillation of the electric field in time with plasma frequency ω_p which due to vanishing wave vector cannot be classified as transverse or longitudinal one but is its unification. The specialized formula (A.20) for the susceptibility in the system $\mathcal{I}_0$ which moves with group velocity $\boldsymbol{v}_0$ of a certain excitation in $\mathcal{I}$ is relatively complicated. The transformation to this system makes only one considered wave to a resting excitation, whereas all other ones are not resting. Therefore, in the system $\mathcal{I}' = \mathcal{I}_0$ only the excitation with $k_0' = 0, \omega_0' = \omega_p$ is simple, whereas all others remain complicated and propagate with some group velocity.

The action of a cold plasma onto a flow of light particles propagating in vacuum is that it makes them to a flow of transverse quasiparticles and equips the last with a rest mass $m_p = \frac{\hbar \omega_p}{c^2}$. Therefore, it may serve as a concrete model for the transition from massless particles to particles with rest mass. A well-known model for equipping massless particles with a rest mass is the Higgs mechanism by symmetry breaking in gauge field theories and this model is similar. The result for the energy-momentum tensor in the limiting case of plane monochromatic waves (12.12) is formulated in Lorentz-covariant form and contains only Lorentz invariants (n_0, s_0 are relativistic scalars and m_p an invariant). It agrees with the energy-momentum tensor for a homogeneous flow of relativistic point-like particles in classical mechanics (e.g., [3] [5] [12] [15]). In the transition to vacuum $n_e \to 0 \Rightarrow \omega_p \to 0, m_p \to 0$, the rest mass m_p goes to zero. In this case we do not have an inertial system where the excitation is resting and the 4-velocity $u_{0,\lambda}$ is diverging due to $\boldsymbol{v}^2 \to c^2$ and the action density s_0 in the transition to a resting system goes to zero, *i.e.* $s_0 \to Ss_0' = 0$ (see (10.17) and (10.18)). Formula (10.17) for the energy-momentum tensor is then no more applicable but it can be substituted by (10.13) together with the relation $c^2\boldsymbol{k}_0 = \omega_0\boldsymbol{v}_0$ from (12.3) that leads to the energy-momentum tensor for quasiplane and quasimonochromatic waves in vacuum. The appearance of a rest mass for the elementary transverse excitations in a plasma is a collective effect of the interaction of the charged particles and it vanishes in the transition to vacuum (Lorentz invariance). The cold plasma may serve as orientation for a relativistic covariant electromagnetic theory which provides transverse photons with a certain rest

If we combine formula (12.8) for the rest mass of the quasiparticles with the expression for the plasma frequency of a cold electron plasma in (12.1) then we find

$$m_p = \frac{\hbar\omega_p}{c^2} = \frac{\hbar}{c^2}\sqrt{\frac{4\pi n_e e^2}{m_e}} = \frac{\hbar}{c}\sqrt{4\pi n_e r_e}, \quad \left(r_e \equiv \frac{e^2}{m_e c^2}\right), \tag{12.19}$$

where $r_e \approx 2.82\times10^{-13}\,\mathrm{cm}$ denotes the classical electron radius. Thus the mass m_p is proportional to the square root of the electron density n_e but is not in a simple relation to the electron mass m_e without taking into account the electron charge *e*. It is proportional to the reciprocal square root of m_e if we fix the charge *e*. For plasma frequencies ω_p in the visible region of about $\nu_p \equiv \frac{\omega_p}{2\pi} \approx 5\times10^{14}\ \mathrm{Hz}\ \left(\hat{=}\,\mathrm{sec}^{-1}\right)$ (for alkali metals they are a little higher and are much higher for most other metals) we find according to (12.8) a mass of about $m_p \approx 4\times10^{-33}\,\frac{\mathrm{erg}\cdot\mathrm{sec}^2}{\mathrm{cm}^2} = 4\times10^{-33}\,\mathrm{g}$ that is by a factor ≈ 2×10^5 smaller than the rest mass of an electron which is of about $m_e \approx 10^{-27}\,\mathrm{g}$ corresponding to a rest energy of about ≈ 0.5 MeV. Since the appearance of a rest mass is a collective effect (quasi-particles), we cannot separate different parts of energy and momentum from the pure field and from the moving particles (electrons) on the background of the heavier ions considered as resting and making the medium (plasma) macroscopically neutral.

We mention that an energy momentum tensor of the form (12.17) with only one nonvanishing component T_{44} in the energy density part for a point-like resting particle is the starting point for establishing the direct connection of Newton's gravitation law with Einstein's equations of general relativity (see, e.g.,

[3], section 99, Eq. (99.1)).

A warm plasma with spatial inversion as symmetry element possesses a transversal and a longitudinal part of the permittivity tensor proportional to $\delta_{ij} - \frac{k_i k_j}{\boldsymbol{k}^2}$ and to $\frac{k_i k_j}{\boldsymbol{k}^2}$ with parameters $\varepsilon_t\left(|\boldsymbol{k}|,\omega\right)$ and $\varepsilon_l\left(|\boldsymbol{k}|,\omega\right)$ and in case of gyrotropy a further term.

13. Angular Momentum Conservation in Resting Isotropic Media

The most prominent supporter of the Abraham tensor in old time was Pauli [5] in his younger years. He considers explicitly only isotropic non-dispersive media and it is not clear how these results may be generalized to anisotropic media since no proposal for this exists. In his encyclopedic article published in the age of 21, Pauli [5] sees in electron-theoretical considerations of Abraham a weighty argument in favor of the symmetric Abraham tensor but he discusses also advantages and disadvantages of the Minkowski tensor. Since Pauli's article is about Special and General relativity theory and since the last requires a symmetric energy-momentum tensor as source term in Einstein's equations it is understandable that Pauli looked mainly for arguments in favor of the symmetric Abraham tensor. However, in his late years, apparently, he changed his opinion and favored the non-symmetric Minkowski tensor as the correct one. This can be seen from the supplementary notes made by Pauli in 1956, two years before his death, to the re-edition of his encyclopedic article [5]. In Note 11 with reference to von Laue [14] Pauli praised emphatically the Minkowski tensor as the right one and it seems that he wants to correct his earlier opinion[13]. For anisotropic media which were never explicitly considered by Pauli in this regard it is clear that the energy-momentum tensor cannot be symmetric since the momentum density is in direction of the mean wave vector and the energy-flow density (Poynting vector) in direction of the ray vector which, in general, are not parallel to each other as it is well known from experimental and theoretical crystal optics.

In recent time the Abraham tensor was declared in papers of Leonhardt and coworkers [32] [33] as the correct one. It is easily to conjecture that the same as the young Pauli they want to have a symmetric energy-momentum tensor because the General Relativity theory requires such but they should ask themselves how it can be generalized as such symmetric tensor to general anisotropic media. As mentioned most authors favor the Minkowski tensor as the correct one also for its relativistic covariance but many of them do not consider anisotropic me-

[13]The full text of Note 11 in [5] is:"*M. v. Laue [see his Relativitätstheorie, Vol. 1 (6th edn., 1955) § 19] has shown that only the unsymmetric energy-momentum tensor of Minkowski is correct for a phenomenological description of moving bodies (just as it is in crystals at rest). His argument also emphasizes the validity of the addition theorem of velocities for the ray-velocity (see Eq. (312) of the text), which is in agreement only with this unsymmetric tensor.". The Editors of the Russ. Transl. V.L. Ginzburg and V.P. Frolov make further remarks to this problem with four additional citations, in particular, [16]* [17].

dia where the problem becomes more clear though more difficult. The reason for different views to this tensor is different separations of ponderomotive forces in the conservation theorems which in this case do not possess the exact form of local conservation laws. This is discussed in detail by Ginzburg [10]. The discussion of the relations between the Abraham and the Minkowski tensor is usually restricted to isotropic media and we do not know an explicit more general form of the symmetric Abraham tensor for anisotropic dispersive media. For anisotropic media it is evident that the momentum density should possess the direction of the mean wave vector and the energy flow density should be in direction of the group velocity of quasiplane and quasimonochromatic waves which, in general, are different for anisotropic media or, in other case, essential parts of crystal optics would be wrong. This is provided if the momentum density is proportional to $[\boldsymbol{D},\boldsymbol{B}]$ and the energy flow density proportional to $[\boldsymbol{E},\boldsymbol{B}]$ as in the Minkowski tensor. Taking into account the dispersion both expression have to be modified as was shown ([9] and Section 9 of present article). In the case of taking into account the dispersion the constitutive relations bring into play additional derivatives of the electric (and in certain cases of the magnetic) field which have to be taken into account in the derivation of conservation laws. Our strategy is to formulate the differential conservation laws without taking into account absorption (dissipation or absorption or even amplification, open system) as exact vanishing of 4-divergences. With dissipation this is impossible. The condition for the permittivity tensor to describe a dissipation-less medium was discussed in Section 3.

Even in the special case of an isotropic medium and under neglect of dispersion (spatial and temporal ones) that means in case of the constitutive relations $\boldsymbol{D}_0 = \varepsilon_0 \boldsymbol{E}_0, \boldsymbol{D}_0^* = \varepsilon_0 \boldsymbol{E}_0^*$ with constant scalar ε_0 the energy-momentum tensor $T_{\kappa\lambda}$ remains, in general, non-symmetric, in particular (see next Section)

$$T_{k4} = \mathrm{i}cg_k = \mathrm{i}\varepsilon_0 \frac{S_k}{c} = \varepsilon_0 T_{4k} \neq T_{4k}, \tag{13.1}$$

that is nonsymmetric for $\varepsilon_0 \neq 1$. However, the stress tensor T_{kl} is symmetric in this case in the inertial system where the isotropic medium is resting (in moving systems it is no more isotropic)

$$T_{kl} = T_{lk}. \tag{13.2}$$

Moreover, this symmetry remains to be true also in the general case of taking into account the dispersion as can be seen from (10.13) and (10.8) since for isotropic media the group velocity $\boldsymbol{v}$ and the wave vector $\boldsymbol{k}$ possess in general case the same direction (see below). This partial symmetry of the three-dimensional part T_{kl} (and only this) of the full four-dimensional energy-momentum tensor $T_{\kappa\lambda}$ is necessary for the existence of a local law of angular-momentum conservation in isotropic media due to invariance with respect to the three-dimensional rotation group in this inertial system where the medium rests.

Multiplying the differential momentum conservation (8.2) by $\varepsilon_{ijk} r_j$ (ε_{ijk} is three-dimensional Levi-Civita pseudo-tensor) we may transform this according to

$$0=\varepsilon_{ijk}r_j\underbrace{\left(\nabla_l T_{kl}+\frac{\partial}{\partial t}g_k\right)}_{=0}=\nabla_l\left(\varepsilon_{ijk}r_jT_{kl}\right)+\frac{\partial}{\partial t}\left(\varepsilon_{ijk}r_jg_k\right)-\underbrace{\varepsilon_{ijk}T_{kj}}_{=0}, \tag{13.3}$$

and obtain

$$\nabla_l S_{il}+\frac{\partial}{\partial t}m_i=0,\quad S_{il}\equiv\varepsilon_{ijk}r_jT_{kl},\quad m_i\equiv\varepsilon_{ijk}r_jg_k=\left[\boldsymbol{r},\boldsymbol{g}\right]_i, \tag{13.4}$$

where we used the symmetry (13.2) and where this last symmetry is evidently required for the vanishing of $\varepsilon_{ijk}T_{kj}$. Herein, $m_i\equiv m_i\left(\boldsymbol{r},t\right)=\left[\boldsymbol{r},\boldsymbol{g}\left(\boldsymbol{r},t\right)\right]$ means the angular momentum density and $S_{il}\equiv S_{il}\left(\boldsymbol{r},t\right)=\varepsilon_{ijk}r_jg_k\left(\boldsymbol{r},t\right)$ a non-symmetric tensor which is the analogue to the stress tensor for momentum conservation (angular-momentum flow density). Since the momentum density takes into account the polarization of the electromagnetic field the angular momentum density comprises both the orbital and the spin angular momentum.

The vanishing of the right-hand side of (13.4) possesses the form of a differential or local conservation theorem. We also see that a pure translation $\boldsymbol{r}\rightarrow\boldsymbol{r}'=\boldsymbol{r}+\boldsymbol{a}$ with a constant vector $\boldsymbol{a}$ (displacement of coordinate origin) does not disturb the local conservation theorem (13.4) although this changes m_i and S_{il}. Despite the symmetry (13.2) the whole tensor $T_{\kappa\lambda}$ is non-symmetric ($T_{\kappa\lambda}\neq T_{\lambda\kappa}$) due to (13.1) and under Lorentz transformations the spatial part T_{kl} becomes also nonsymmetric ($T'_{kl}\neq T'_{lk}$) as consequence that isotropy of a medium resting in the inertial system $\mathcal{I}$ is lost in inertial systems $\mathcal{I}'$ of the moving medium (see Appendix B, in particular, (B.3)). Due to $\varepsilon_{ijk}T'_{kj}\neq 0$ a local law of angular-momentum conservation cannot be formulated then in analogy to (13.4) for such inertial systems. One can convert this conclusion. If the four-dimensional energy-momentum tensor $T_{\kappa\lambda}$ would be symmetric in the inertial system of the resting isotropic medium then it remains to be symmetric also in arbitrary other inertial systems and one would be able to prove a local conservation law of angular momentum for an arbitrary inertial system that is evidently wrong (with exception of vacuum). This excludes the symmetric Abraham tensor as candidate for the energy-momentum tensor from the beginning in contradiction to confusing remarks in [5]. We did not find explicit expressions in literature for a symmetric Abraham tensor in the general anisotropic case without or with dispersion where already the same directions of momentum density and energy flow density would be in striking contradiction to known experimental facts.

With exception of the vacuum the angular momentum conservation in isotropic media at rest cannot be extended to a more general four-dimensional conservation theorem such as it was possible for energy and momentum conservation. The reason is that invariance with respect to Lorentz transformations requires a permittivity tensor $\varepsilon_{ij}\left(\boldsymbol{k},\omega\right)=\delta_{ij}$ of the medium which determines the electrodynamic vacuum. We mention that our derivation (13.3) of the necessary symmetry $T_{kl}=T_{lk}$ for local angular momentum conservation for isotropic media corresponds to the derivation of the necessary symmetry of the four-dimensional tensor $T_{\kappa\lambda} = T_{\lambda\kappa}$ for the local conversation of the four-angular momentum given in [3] (section 32, Eq. 32.10 in Russ. Ed. from 1988).

14. Non-Uniqueness of the Energy-Momentum Tensor and Its Role for Finding the Simplest One

As is known the energy-momentum tensor in the local conservation law of energy and momentum (and, similarly, the action 4-vector in local action conservation) is not unique [3]. However, we will suggest that this is not a very strong problem, in particular, not in the approximation of quasiplane and quasimonochromatic waves. We do not strive in this Section for high generality of our considerations and try to illuminate the problems of non-uniqueness only by some remarks.

The general form of non-uniqueness of the energy momentum tensor $T_{\kappa\lambda}(r) \sim T'_{\kappa\lambda}(r)$ in the local conservation theorem is described by an arbitrary third-rank four-tensor function $\psi_{\kappa\lambda\mu}(r)$ which is antisymmetric in the last two indices in the following form [3]

$$T'_{\kappa\lambda}(r) = T_{\kappa\lambda}(r) + \nabla_\mu \psi_{\kappa\lambda\mu}(r), \quad \psi_{\kappa\lambda\mu}(r) = -\psi_{\kappa\mu\lambda}(r), \tag{14.1}$$

from which immediately follows

$$\nabla_\lambda T'_{\kappa\lambda}(r) = \nabla_\lambda T_{\kappa\lambda}(r) = 0. \tag{14.2}$$

In three-dimensional separation this means for the stress tensor $T_{kl}(\boldsymbol{r},t)$ and the momentum density $g_k(\boldsymbol{r},t)$ and for the energy flow density $S_l(\boldsymbol{r},t)$ and the energy density $w(\boldsymbol{r},t)$

$$\begin{aligned} T'_{kl}(\boldsymbol{r},t) &= T_{kl}(\boldsymbol{r},t) + \varepsilon_{lmn}\nabla_m \psi_{kn}(\boldsymbol{r},t) + \frac{\partial}{\partial t}\chi_{kl}(\boldsymbol{r},t), \\ g'_k(\boldsymbol{r},t) &= g_k(\boldsymbol{r},t) - \nabla_l \chi_{kl}(\boldsymbol{r},t), \\ S'_l(\boldsymbol{r},t) &= S_l(\boldsymbol{r},t) + \varepsilon_{lmn}\nabla_m \psi_n(\boldsymbol{r},t) + \frac{\partial}{\partial t}\chi_l(\boldsymbol{r},t), \\ w'(\boldsymbol{r},t) &= w(\boldsymbol{r},t) - \nabla_l \chi_l(\boldsymbol{r},t), \end{aligned} \tag{14.3}$$

with the following separation of $\psi_{\kappa\lambda\mu}(r)$ into arbitrary three-dimensional tensor or pseudo-tensor functions, respectively

$$\begin{aligned} \psi_{kn}(\boldsymbol{r},t) &\equiv \frac{1}{2}\varepsilon_{lmn}\psi_{klm}(\boldsymbol{r},t), \quad \chi_{kl}(\boldsymbol{r},t) \equiv -\frac{\mathrm{i}}{c}\psi_{kl4}(\boldsymbol{r},t) = \frac{\mathrm{i}}{c}\psi_{k4l}(\boldsymbol{r},t), \\ \psi_n(\boldsymbol{r},t) &\equiv -\mathrm{i}\frac{c}{2}\varepsilon_{lmn}\psi_{4lm}(\boldsymbol{r},t), \quad \chi_l(\boldsymbol{r},t) \equiv -\psi_{4l4}(\boldsymbol{r},t) = \psi_{44l}(\boldsymbol{r},t). \end{aligned} \tag{14.4}$$

For neglected dispersion ($\boldsymbol{\varepsilon}_0$ constant permittivity tensor in $D_i = \varepsilon_{0,ij}E_j$) and without losses ($\varepsilon_{0,ij} = \varepsilon^*_{0,ji}$) one can derive the following well-known general expressions for the parts of the energy-momentum tensor here denoted by $T'_{\kappa\lambda}$ in contrast to our $T_{\kappa\lambda}$ ($\alpha = 4\pi$ in Gauss system)

$$\alpha T'_{kl} = \frac{1}{2}(\boldsymbol{BB} + \boldsymbol{DE})\delta_{kl} - (B_k B_l + E_k D_l),$$

$$\alpha g'_k = \frac{1}{c}\varepsilon_{kmn}D_m B_n = \frac{1}{c}[\boldsymbol{D},\boldsymbol{B}]_k,$$

$$
\begin{aligned}
&\alpha S_l' = c\varepsilon_{lmn}E_m B_n = c\left[\boldsymbol{E},\boldsymbol{B}\right]_l,\\
&\alpha w' = \frac{1}{2}\left(\boldsymbol{BB}+\boldsymbol{DE}\right) = \frac{1}{2}\left(\boldsymbol{B}^2+\boldsymbol{E}\boldsymbol{\varepsilon}_0\boldsymbol{E}\right),
\end{aligned}
\tag{14.5}
$$

where $\boldsymbol{E}$ and $\boldsymbol{B}$ mean here the full electric and magnetic field and $\boldsymbol{D}$ the full electric induction which depend on $(\boldsymbol{r},t)$. If we insert into these expressions plane monochromatic waves of the form $\boldsymbol{E} = \boldsymbol{E}_0 \mathrm{e}^{\mathrm{i}(\boldsymbol{k}_0\boldsymbol{r}-\omega_0 t)} + \boldsymbol{E}_0^* \mathrm{e}^{-\mathrm{i}(\boldsymbol{k}_0\boldsymbol{r}-\omega_0 t)}$ with real $\boldsymbol{k}_0$ and ω_0 and constant vector amplitudes $\boldsymbol{E}_0$ and analogously for $\boldsymbol{B}$ and $\boldsymbol{D}$ we get besides constant terms in the energy-momentum tensor also terms with the rapidly varying phase factors $\mathrm{e}^{\pm\mathrm{i}2(\boldsymbol{k}_0\boldsymbol{r}-\omega_0 t)}$. It is not possible to generalize in some simple way the form (14.5) of the energy-momentum tensor containing the full fields to the case of taking into account the dispersion. In particular, the corresponding expressions cannot be local in the fields that means cannot be taken only at the same space-time points $(\boldsymbol{r},t)$ (see Section 3). Therefore, we have to restrict us in the following discussion of the non-uniqueness concerning the terms with rapidly varying phase factors $\mathrm{e}^{\pm\mathrm{i}2(\boldsymbol{k}_0\boldsymbol{r}-\omega_0 t)}$ to the neglect of dispersion.

The suppression of terms with rapidly varying phase factors was made in [34] for local energy conservation (and analogously possible for momentum conservation) in noncovariant form. Here we will show that it can be made also in relativistic covariant form. Under neglect of dispersion and in the limiting case of constant amplitudes $\boldsymbol{E}_0$ we can add to our $T_{\kappa\lambda}$ terms with rapidly varying phase factors possessing the form of a four-divergence of a function $\psi_{\kappa\lambda\mu}(r)$ with antisymmetry in last two indices as shown in (14.1) in the following way (recall $k_0^2 \equiv k_{0,\mu}k_{0,\mu}$, $k_0 r \equiv k_{0,\nu}r_\nu$)

$$
\begin{aligned}
\alpha T_{\kappa\lambda}' &= \alpha T_{\kappa\lambda} + \nabla_\mu\left\{\frac{\mathrm{i}}{4k_0^2}k_{0,\kappa}\left(\left(\frac{\partial L_{ij}}{\partial k_\lambda}\right)_0 k_{0,\mu} - k_{0,\lambda}\left(\frac{\partial L_{ij}}{\partial k_\mu}\right)_0\right)E_{0,i}E_{0,j}\mathrm{e}^{\mathrm{i}2k_0 r} + c.c.\right\}\\
&= \alpha T_{\kappa\lambda} - \frac{1}{2k_0^2}k_{0,\kappa}\left(\left(\frac{\partial L_{ij}}{\partial k_\lambda}\right)_0 k_0^2 - k_{0,\lambda}\left(\frac{\partial L_{ij}}{\partial k_\mu}\right)_0 k_{0,\mu}\right)E_{0,i}E_{0,j}\mathrm{e}^{\mathrm{i}2k_0 r} + c.c.\\
&= \alpha T_{\kappa\lambda} - \frac{1}{2}k_{0,\kappa}\left(\frac{\partial L_{ij}}{\partial k_\lambda}\right)_0 E_{0,i}E_{0,j}\mathrm{e}^{\mathrm{i}2k_0 r} + c.c.,
\end{aligned}
\tag{14.6}
$$

with $T_{\kappa\lambda}$ denoting the energy-momentum tensor in (9.1) without terms with rapidly varying phase factors and represented by the slowly varying amplitudes

$$
\alpha T_{\kappa\lambda} = -k_{0,\kappa}\left(\frac{\partial L_{ij}}{\partial k_\lambda}\right)_0 E_{0,i}^* E_{0,j},
\tag{14.7}
$$

and, in addition, neglecting derivatives $\left(\frac{\partial \varepsilon_{ij}}{\partial k_\lambda}\right)_0$ in it corresponding to neglect of dispersion. We used in (14.6) the vanishing of the following expression

$$
\begin{aligned}
\left(\frac{\partial L_{ij}}{\partial k_\mu}\right)_0 k_{0,\mu} &= \left(\frac{\partial L_{ij}}{\partial k_m}\right)_0 k_{0,m} - \mathrm{i}c\left(\frac{\partial L_{ij}}{\partial \omega}\right)_0\left(\mathrm{i}\frac{\omega_0}{c}\right)\\
&= \frac{c^2}{\omega_0^2}\left(k_{0,i}\delta_{jm} + \delta_{im}k_{0,j} - 2k_{0,m}\delta_{ij}\right)k_{0,m} - 2\frac{c^2}{\omega_0^3}\left(k_{0,i}k_{0,j} - \boldsymbol{k}^2\delta_{ij}\right)\omega_0\\
&= 0,
\end{aligned}
\tag{14.8}
$$

which is true for the special form of L_{ij} given in (5.5) only under neglect of dispersion and is related to (11.12) or also to (11.3). Thus we have represented terms with rapidly varying phase factors in the energy-momentum tensor (14.6) by a four-divergence $\nabla_\mu \psi_{\kappa\lambda\mu}(r)$ with explicitly given $\psi_{\kappa\lambda\mu}(r) = -\psi_{\kappa\mu\lambda}(r)$ and comparison of (14.6) with the noncovariant form in [30] shows that the choice of $\psi_{\kappa\lambda\mu}(r)$ itself for removing one and the same terms is to certain extent also not fully unique. Evidently, substitutions $\psi_{\kappa\lambda\mu}(r) \to \psi'_{\kappa\lambda\mu}(r) = \psi_{\kappa\lambda\mu}(r) + \nabla_\nu \psi_{\kappa\lambda\mu\nu}(r)$ in (14.1) with arbitrary $\psi_{\kappa\lambda\mu\nu}(r)$ which is fully antisymmetric in last three indices (λ, μ, ν) provide equivalent possibilities.

Taking into account the dispersion, the expressions (14.5) for the parts of the energy-momentum tensor are no more true. However, one can find from Maxwell equations vanishing quadratic expressions in the field which do not possess the form of local conservation of energy and momentum but are related to it and by averaging these expressions over space and time, we get local conservation laws of energy and momentum which take into account the dispersion in some approximation of quasiplane and quasimonochromatic waves that is demonstrated in [7] for frequency dispersion in the energy density w and energy flow density $\boldsymbol{S}$ where there is obtained the expression for w in (9.7) ($\boldsymbol{S}$ is not altered in comparison to the usual one due to neglected spatial dispersion). In the derivations of local conservation theorems one cannot work with constant amplitudes and the amplitudes $\boldsymbol{E}_0, \boldsymbol{B}_0$ and $\boldsymbol{D}_0$ are, at least, slowly varying amplitudes. If we take expressions of the kind in (14.6) for removing the terms with rapidly varying phase factors proportional to $\mathrm{e}^{\pm i2(\boldsymbol{k}_0\boldsymbol{r} - \omega_0 t)}$ then one has also to differentiate the slowly varying amplitudes and there remain some new terms with these phase factors which, however, are small compared with the main terms with these phase factors before. In a second step and successively in higher steps one can try to remove also these smaller terms. However, there is no possibility to remove any parts in the energy-momentum tensor (9.7) which do not contain such rapidly varying phase factors without creating new terms with rapidly varying phase factors or terms which grow in space and time in unreasonable way (e.g., linearly). Clearly, we can derive higher-order approximations of the energy-momentum tensor than in (9.7) (see sections 5-7) and can try to express them not only by the group velocity $v_l \equiv \frac{\partial \omega}{\partial k_l}$ as done but also by the higher derivatives $\frac{\partial^2 \omega}{\partial k_l \partial k_m}, \frac{\partial^3 \omega}{\partial k_l \partial k_m \partial k_n}$ and so on (this is not yet done) but they also do not contain such rapidly varying phase factors. Therefore, the nonuniqueness cannot be used to make energy-momentum tensors without rapidly varying phase factors to symmetric ones and nonsymmetric energy-momentum tensors of such kind remain intrinsically nonsymmetric. This suggests also that expressions of the kind (9.7) or (9.4) and of their generalization possess some distinguished position with the possibility of direct physical interpretation among all other equivalent energy-momentum tensors in local conservation theorems. Our derivations provided these expressions directly without the necessity of suppression of terms by the discussed non-uniqueness.

There is another case where the non-uniqueness of the energy-momentum tensor seems to be of great importance. These are evanescent or inhomogeneous waves with complex wave vector $\boldsymbol{k}_0 = k_0' + \mathrm{i}k_0''$ and (or) complex frequency $\omega_0 = \omega_0' + \mathrm{i}\omega_0''$ leading besides the periodic phase factors $\mathrm{e}^{\pm \mathrm{i}2(\boldsymbol{k}'\boldsymbol{r} - \omega' t)}$ to exponential factors $\mathrm{e}^{-2(\boldsymbol{k}_0''\boldsymbol{r} - \omega_0'' t)}$ in the energy-momentum tensor. Such waves are present, for example, under total reflection in the optically thinner medium and in surface waves to both sides of a boundary plane. The exact generalization of the local energy-momentum conservation to such cases is possible. Using the additional factors $\mathrm{e}^{-2(\boldsymbol{k}_0''\boldsymbol{r} - \omega_0'' t)}$ in such waves provides further possibilities to remove terms in the energy-momentum tensor which are difficult to interpret and to get equivalent tensors but this makes the problems of non-uniqueness more complex.

Summarizing, it seems to us that the non-uniqueness of the energy-momentum tensor can mainly be used to remove or to change terms with periodically or exponentially rapidly changing phase factors, whereas the others are hardly to touch. This problem of non-uniqueness has little to do with the discussion of the correctness of the Minkowski or the Abraham tensor which in absence of dispersion was decided in favor of the Minkowski tensor.

15. Difficulties for General Relativity Theory Connected with General Asymmetry of Energy-Momentum Tensor in Media

The energy-momentum tensor $T_{\kappa\lambda}$ forms the source term in Einstein's gravitation equations which determine the metric tensor $g_{\kappa\lambda}$ and thus also the curvature of a Riemannian space-time as a generalization of Minkowski's space-time. Ricci tensor and thus Einstein tensor in these equations are symmetric ones and, consequently, the energy-momentum tensor has also to be symmetric. Since macroscopic electrodynamics is an averaged microscopic electrodynamics it can be assumed that its energy-momentum tensor provides the source term for a correspondingly averaged gravitation field in the medium and requires boundary conditions in case of transition to vacuum with a sharp boundary. The connection of the classical energy-momentum tensor in Minkowski space as source of a curvature in Einstein's equations is patchwork since it starts from a pseudo-Euclidian space but it cannot be fully wrong concerning its symmetry. We will shortly discuss some difficulties which result from this for General relativity theory.

If it would be possible to extend the Ricci tensor in Einstein's equations to a possible non-symmetric one that up to now did not be achieved then, nevertheless, there remain some serious problems. The energy-momentum tensor in the local conservation laws is not uniquely defined (see Section 14) and there arises the problem which of these tensors provides the right source term in Einstein's equations of General relativity theory. Moreover, in spatially and (or) temporally inhomogeneous media such a tensor in local conservation laws does not exist at all. In these cases only the local conservation theorem of action remains with a four-dimensional vector of action and action-flow density. One may expect that then this four-vector must be involved in some way as source in generalized Einstein equation but this to our knowledge was also not found up to now.

The unification of the basic laws of physics is a steady desire of physicists. After Einstein's General relativity theory in 1916 it was the problem of its unification with the experimentally well established Maxwell theory of electromagnetism. First successful trials in this direction with extension of the dimensionality of the space-time to 5 dimensions were the Kaluza-Klein theories from about 1920 on. With the foundation of the rigorous quantum theory in about 1925 it became the problem of unification of quantum theory with electromagnetism and gravity from which only the first part found a satisfactory solution in quantum electrodynamics and from this more or less only the microscopic quantum electrodynamics of charged particles is well elaborated. In the sixties and seventies the standard model of elementary particles and fields was established which unified the electromagnetic and the weak with the strong interactions in satisfactory way but with the new problem of symmetry breaking and of the experimental proof of the theoretical Higgs particles. Thus the problem as it represents to us at this time became already the unification of the standard model with a quantum theory of gravitation where great but up to now not fulfilled hopes were set in the development of string theories. The non-symmetry of the energy-momentum tensor for electromagnetic excitations in anisotropic dispersive and, in general, not even homogeneous media in classical electrodynamics as represented here adds a further serious problem because already Einstein's gravitation theory in existing form is not consistent with the electrodynamics of continuous media since first requires a symmetric energy-momentum tensor. This non-symmetry of the energy-momentum tensor is intrinsic and cannot be removed by considering the current and charge distributions of media on the background of the vacuum. Since in the principal correctness of the existing classical electrodynamics of continuous media cannot be doubt the least which is required is some extension or generalization of the General relativity theory if not a more basically new theory. As it seems to us this problem has to be solved before a successful unification with the other fundamental forces of nature can be accomplished.

The General relativity theory has great success for explanation of astronomical observations and for cosmology. It is a beautiful theory which is considered as experimentally verified. It must not be incorrect due to some of the shown classical difficulties and we hope that they can be overcome in the course of time by generalization or somehow in other way and that it remains true as an approximation.

16. Possible Additions and Generalizations

The preceding theory of the energy-momentum tensor can be extended in different directions. In particular, the following possibilities seem to be interesting (partially already elaborated):

1) Statistical model for permittivity tensor for gases and (warm) plasmas with spatial dispersion.

2) Non-statistical permittivity tensor for solids with spatial dispersion (but averaged over the space).

3) Inclusion of inhomogeneous (evanescent) waves in lossless media that means solutions of the wave equations where both the mean wave vector $\boldsymbol{k}_0$ and, possibly, the mean frequency ω_0 are complex quantities (is more a technical than a principal problem).

4) For homogeneous media with losses one cannot derive exact differential conservation theorems but one can derive (not in fully unique way) equations of the kind

$$\nabla_\lambda T_{\kappa\lambda}(r) = f_\kappa(r), \tag{16.1}$$

or in three-dimensional separation with $f_\kappa(r) = \left(f_k(\boldsymbol{r},t), \frac{\mathrm{i}}{c} q(\boldsymbol{r},t)\right)$

$$\nabla_l T_{kl}(\boldsymbol{r},t) + \frac{\partial}{\partial t} g_k(\boldsymbol{r},t) = f_k(\boldsymbol{r},t), \quad \nabla_l S_l(\boldsymbol{r},t) + \frac{\partial}{\partial t} w(\boldsymbol{r},t) = q(\boldsymbol{r},t), \tag{16.2}$$

where $\boldsymbol{f}(\boldsymbol{r},t)$ can be interpreted as a force density and $q(\boldsymbol{r},t)$ as a density of loss or gain of electromagnetic energy.

5) Derivation of a local conservation theorem for inhomogeneous media with $\hat{\varepsilon}_{ij} \equiv \hat{\varepsilon}_{ij}(\boldsymbol{r},\boldsymbol{r}',t,t')$. In this case local conservation of energy and momentum is not possible but the action conservation is possible (adiabatic invariants).

6) Specialization of permittivity tensor, for example, to polaritons with $\varepsilon_{ij}(\boldsymbol{k},\omega) = \frac{\omega^2 - \omega_l^2}{\omega^2 - \omega_t^2}\delta_{ij}$.

7) Taking into account higher-order derivatives of the slowly varying electric field amplitudes.

8) Quantum-mechanical generalization.

17. Conclusion

In present article, we developed a relativistically covariant approach to the local conservation theorems for homogeneous anisotropic media with dispersion of general permittivity tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)$ and to the calculation of the four-dimensional energy-momentum tensor[14]. The limiting case to plane monochromatic waves is discussed and the results, in particular, are demonstrated for the special case of cold plasma. In the usual approach for such problems, the starting point is a Lagrange function, but this approach is hardly applicable with inclusion of the dispersion. Our calculations are made in a coordinate-invariant operator approach

[14] I used my unpublished paper in German from about 1979 (see Remark) which I translated in preparation to a Conference in 2004 into English (see [34]) and to which I added since this time new material not fully included here and I am convinced that the content contains elements worth to be published also now.

with an operator equation only for the electric field and therefore with results which are basically expressed by the electric field.

Remark

A shorter paper of this theme (in particular, without any statements to difficulties for General relativity theory due to asymmetry of the energy-momentum tensor and to application of a plasma) written in German with nearly all basic formulae as now was made in about 1979 but was rejected from Editor of Annalen der Physik in GDR Professor Gustav Richter with wrong arguments. His main wrong argument was that in my formulae for the limiting case to plane monochromatic waves stands the derivative of the permittivity with respect to the frequency that he declared as wrong "for physical reason". However, the correctness of these formulae was already known from cited monographs and papers, in particular, of Landau and Lifshits and of Agranovich and Ginzburg. When I wanted to explain G. Richter who was also Member of our Institute at that time (in age of a few years below 65) in personal talk why mentioned formulae are correct for beams in limiting case he became very angry. Similar things happened shortly before when I wanted to publish my paper about generalized boundary conditions which, finally, was published after intervention by a prominent physicist of GDR from Editorial Board of "Annalen" and recently I published a continuation of this topic. When I tried to send the mentioned paper about boundary conditions to a Western journal I never got an answer. One could not check whether or not it was really sent since mail to Western countries went before this in an open couvert to the Chief and through Security or was a response withheld. I found now the hand-written comments of G. Richter and will pose them into my Home-page.

Acknowledgements

Shortly, after my active time in the Institute at the end of 1999, already as pensioner I went to an Optics Conference in Szeged (Hungary) and gave a lecture about the "Energy-momentum tensor for... dispersive media...". The difficulty here was that in the Proceedings after the conference only 6 pages in a given small Latex-frame should be published with which I have had great difficulties but the 9 pages which I sent were accepted. Thank you very much M.G. Benedict from Szeged! A further forum in the new decennary I found also in the Group of Peřinová and Peřina from Olomouc where I could give a seminar lecture about mentioned topics. Thank you also very much!

Conflicts of Interest

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

References

[1] Einstein, A. (1913) *Zeitschrift für Angewandte Mathematik und Physik*, **62**,

225-261.

[2] Tamm, I.E., Smorodinsky, J.A. and Kuznetsov, B.G. (1965) Albert Einstein, Sobranyje nauchnykh truduv, I (Collection of Scientific Works, Part I). Nauka, Moskva.

[3] Landau, L.D. and Lifshitz, E.M. (1962) The Classical Theory of Fields (4th Revised Edition). Nauka, Moskva.

[4] Rashevski, P.K. (1964) Riemann Geometry and Tensor Analysis. 2nd Edition, Nauka, Moskva. (In Russian)

[5] Pauli, W. (1921) Relativitatstheorie. Enzyklopadie der mathematischen Wissenschaften, Band V, Heft IV, Art. 19, B.G. Teubner, Leipzig.

[6] Lorentz, H.A., Weyl, H. and Minkowski, H. (1922) Das Relativitatsprinzip, B.G. Teubner, Leipzig. https://doi.org/10.1007/978-3-663-16170-7

[7] Landau, L.D. and Lifshitz, E.M. (1982) Electrodynamics of Continuous Media. 2nd Edition, Nauka, Moskva.

[8] Agranovich, V.M. and Ginzburg, V.L. (1962) *Nauk*, **76**-**77**, 643, 663. https://doi.org/10.3367/UFNr.0076.196204b.0643

[9] Agranovich, V.M. and Ginzburg, V.L. (1979) Spatial Dispersion in Crystal Optics and the Theory of Excitons. Second Edition, Nauka, Moskva.

[10] Ginzburg, V.L. (1981) Theoretical Physics and Astrophysics. Nauka, Moskva.

[11] Silin, V.P. and Rukhadze, A.A. (1961) Elektromagnitnye svoistva plasma i plasmapodobnykh sred (Electromagnetic Properties of Plasma and Plasma-Like Media). Gosatomisdat, Moskva.

[12] Tolman, R.Ch. (1969) Relativity Thermodynamics and Cosmology. Clarendon Press, Oxford.

[13] Sommerfeld, A. (1961) Elektrodynamik (3. Auflage). Akad. Verlagsgesellschaft Geest and Portig, Leipzig.

[14] von Laue, M. (1955) Die Relativitatstheorie, vol. 1, Die spezielle Relativitatstheorie. 6th Edition, Vieweg, Braunschweig.

[15] Fock, V. (1961) Theory of Space, Time and Gravitation. 2nd Edition, Fizmatgiz, Moskva.

[16] Skobeltsyn, D.V. (1973) *Uspekhi Fizicheskikh Nauk*, **110**, 253-292. https://doi.org/10.3367/UFNr.0110.197306d.0253

[17] Ginsburg, V.L. and Ugarov, V.A. (1976) *Uspekhi Fizicheskikh Nauk*, **118**, 175-188. https://doi.org/10.3367/UFNr.0118.197601e.0175

[18] Ugarov, V.A. (1977) Spetsialnaya teoriya otnositel'nosti (Special Relativity Theory). 2nd Edition, Nauka, Moskva.

[19] Tamm, I.Ye. (1976) Osnovy teorii elektrichestva (Foundations of the Theory of Elctricity). 9th Edition, Nauka, Moskva.

[20] Møller, C. (1972) The Theory of Relativity. Clarendon Press, Oxford.

[21] Jackson, J.D. (1962) Classical Electrodynamics. John Wiley, New York. https://doi.org/10.1063/1.3057859

[22] Felsen, L.B. and Marcuvitz, N. (1973) Radiation and Scattering of Waves. Prentice-Hall, Englewood Cleffs.

[23] Yariv, A. and Yeh, P. (1984) Optical Waves in Crystals. Wiley-Interscience, New York.

[24] Melrose, D.B. and McPhedran, R.C. (1991) Electromagnetic Processes in Dispersive Media. Cambridge University Press, Cambridge.

https://doi.org/10.1017/CBO9780511600036

[25] Jimenez, J.L. and Campos, I. (1995) The Balance Equations of Energy and Momentum in Classical Electrodynamics. In: Barrett, T.W. and Grimes, D.M., Eds., *Advanced Electromagnetism*, World Scientific, Singapore, 464-495. https://doi.org/10.1142/9789812831323_0016

[26] Schwinger, J., DeRaad Jr., L.L., Milton, K.A. and Tsai, W.-Y. (1998) Classical Electrodynamics. Westview Press (Perseus Books Group, U.S.A.), Boulder.

[27] Fyodorov, F.I. (1958) Optika Anisotropnykh Sred, Optics of Anisotropic Media. Izdatelstvo Akademya nauk BSSR, Minsk. (In Russian)

[28] Fyodorov, F.I. (1976) Teorya girotropii (Theroy of gyrotropy). Nauka i tekhnika, Minsk.

[29] Fyodorov, F.I. (1979) Gruppa lorentsa (Lorentz Group). Nauka, Moskva.

[30] Wünsche, A. (1970) *Annals of Physics* (*Leipzig*), **25**, 201-214. https://doi.org/10.1002/andp.19704800207

[31] Wünsche, A. (1978) *Annals of Physics* (*Leipzig*), **35**, 303-320. https://doi.org/10.1002/andp.19784900407

[32] Leonhardt, U. (2000) *Physical Review A*, **62**, Article ID: 012111. https://doi.org/10.1103/PhysRevA.62.012111

[33] Leonhardt, U. and Piwnicki, P. (2001) *Journal of Modern Optics*, **48**, 977-988. https://doi.org/10.1080/09500340108230969

[34] Wünsche, A. (2005) *Acta Physica Hungarica B*, **23**, 151-162. https://doi.org/10.1556/APH.23.2005.3-4.5

[35] Lorentz, H.A. (1904) *van Wetenschappen te Amsterdam*, **6**, 809.

[36] Tyapkin, A.A. (1973) The Principle of Relativity (with a Preface of D.I. Blokhintsev). Atomisdat, Moscow.[15]

[37] Lorentz, H.A. (1904) Weiterbildung der Maxwellschen Theorie. Elektronentheorie. In: *Enzyklopadie der mathematischen Wissenschaften*, Bd. 2, Teubner, Leipzig, Art. 14, 245.

[38] Lorentz, H.A. (1915) Theory of Electrons. 2nd Edition, Teubner, Leipzig.

[39] de Groot, S.R. and Suttorp, L.G. (1972) Foundations of Electrodynamics. North Holland Publ. Comp., Amsterdam.

[40] Bloembergen, N. (1965) Nonlinear Optics. W.A. Benjamin, Inc., New York.

[41] Wünsche, A. (1971) *Monatsberichte der Deutschen Akademie der Wissenschaften zu Berlin*, **13**, 479-492.

[42] Hebenstreit, H. and Suchy, K. (1979) *Zeitschrift für Naturforschung A*, **34**, 1147-1157. https://doi.org/10.1515/zna-1979-1001

[43] Minkowski, H. (1909) *Physikalische Zeitschrift*, **10**, 104-111.

[44] Lorentz, H.A., Einstein, A., Minkowski, H. and Weyl, H. (1952) The Principle of Relativity. Dover, Mineola.

[45] Landau, L.D. and Lifshitz, E.M. (1988) The Classical Theory of Fields. 7nd Edition,

[15]In comparison to [6], this is a collection of original articles only to Special relativity theory (SRT) and contains in addition articles by H. Poincaré, H.A. Lorentz, M. Planck and some others than in [6] which, partially, are more difficult to find in the original and in a second part it contains articles to the history of the principle of relativity. The Editor of this collection is controversially judged because he tries to lower the main merits of Einstein in the development of SRT and to suggest that other scientists (e.g., H. Poincaré) made this earlier as Einstein that, however, is not convincing (see also [48]).

Nauka, Moskva.

[46] Szivessy, G. (1928) Kristalloptik. In: Geiger, H. and Scheel, K., Eds., *Handbuch der Physik*, *Band XX*, Springer, Berlin, 635-904. https://doi.org/10.1007/978-3-642-90780-7_11

[47] Born, M. and Wolf, E. (1999) Principles of Optics. Seventh Edition, Cambridge University Press, Cambridge.

[48] Ginzburg, V.L. (1974) Kak i kto sozdal teoriyu otnocitel'nosti? (How and Who Created Relativity Theory?) In: *Einshteynovski sbornik (Einsteinian Anthology)*, Nauka, Moscow, 351-384.

Appendix A

Derivations to relativistic-covariant treatment of macroscopic electrodynamics of anisotropic dispersive media

Macroscopic electrodynamics is relativistically invariant only for the vacuum that means for permittivity tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)=\varepsilon(\omega)\delta_{ij}$ with $\varepsilon(\omega)=1$. However, for general $\varepsilon_{ij}(\boldsymbol{k},\omega)$ it can be formulated in relativistic covariant form according to Minkowski (1908) [7] (Eqs. (76.9)-(76.11)) and many others. This form was less appropriate for our derivations of the energy-momentum tensor. Our derivation rests more on the invariance of $\boldsymbol{B}^2-\boldsymbol{E}^2$ after Fourier transformation from space-time to wavevector-frequency representation, e.g., [3] [13] [15].

The starting point is Equation (5.4) for the electric field of plane monochromatic waves with the definition (5.5) of the wave-equation operator that means

$$L_{ij}(\boldsymbol{k},\omega)E_j(\boldsymbol{k},\omega)=0,\quad L_{ij}(\boldsymbol{k},\omega)\equiv\frac{c^2}{\omega^2}\left(k_ik_j-\boldsymbol{k}^2\delta_{ij}\right)+\varepsilon_{ij}(\boldsymbol{k},\omega). \tag{A.1}$$

As discussed, it contains the full information of the macroscopic electromagnetic field together with linear constitutive equations for homogeneous anisotropic dispersive media expressed by the permittivity tensor $\varepsilon_{ij}(\boldsymbol{k},\omega)\equiv\delta_{ij}+4\pi\chi_{ij}(\boldsymbol{k},\omega)$ with $\chi_{ij}(\boldsymbol{k},\omega)$ the general susceptibility tensor.

We consider an arbitrary special Lorentz transformation Λ which transforms a space-time vector $r=\begin{pmatrix}\boldsymbol{r}\\ r_4=\mathrm{i}ct\end{pmatrix}$ in inertial system[16] $\mathcal{I}$ into a new space-time vector $r'=\begin{pmatrix}\boldsymbol{r}'\\ r_4'=\mathrm{i}ct\end{pmatrix}$ in inertial system $\mathcal{I}'$ according to

$$\begin{aligned}&r'_\mu=\Lambda_{\mu\nu}r_\nu,\quad r_\nu=\left(\Lambda^{-1}\right)_{\nu\mu}r'_\mu=r'_\mu\Lambda_{\mu\nu},\\ &\nabla_\mu=\Lambda_{\mu\nu}\nabla'_\nu,\quad \nabla'_\nu=\left(\Lambda^{-1}\right)_{\nu\mu}\nabla_\mu=\nabla_\mu\Lambda_{\mu\nu},\quad \left(\Lambda_{\mu\lambda}\Lambda_{\nu\lambda}=\delta_{\mu\nu}\right).\end{aligned} \tag{A.2}$$

where $\Lambda_{\mu\nu}(\boldsymbol{V})$ with $\boldsymbol{V}$ as the relative velocity of $\mathcal{I}'$ in $\mathcal{I}$ possesses the well-known form

$$\Lambda_{\mu\nu}(\boldsymbol{V})=\begin{pmatrix}\delta_{mn}+(\gamma-1)\dfrac{V_mV_n}{\boldsymbol{V}^2} & i\gamma\dfrac{V_m}{c}\\ -i\gamma\dfrac{V_n}{c} & \gamma\end{pmatrix},\quad \gamma\equiv\frac{1}{\sqrt{1-\dfrac{\boldsymbol{V}^2}{c^2}}}\geq 1, \tag{A.3}$$

with $\Lambda^{-1}_{\mu\nu}(\boldsymbol{V})=\Lambda_{\mu\nu}(-\boldsymbol{V})$. For the 4-wave vector $k\equiv\left(\boldsymbol{k},\mathrm{i}\dfrac{\omega}{c}\right)$ the analogous transformation formula

$$k'_\mu=\Lambda_{\mu\nu}k_\nu,\quad k_\nu=\left(\Lambda^{-1}\right)_{\nu\mu}k'_\mu=k'_\mu\Lambda_{\mu\nu}, \tag{A.4}$$

holds. The antisymmetric electromagnetic field tensor $F_{\mu\nu}(r)\equiv F_{\mu\nu}=-F_{\nu\mu}$ in space-time representation and separated in three-dimensional form together with the transformation relations is (ε_{lmn} three-dimensional Levi-Civita pseudo-tensor)

[16] Einstein denotes "inertial systems" with letter *K*, likely, from the German "Koordinatensystem".

$$F_{\mu\nu} = \begin{pmatrix} \varepsilon_{mnl} B_l & -\mathrm{i}E_m \\ \mathrm{i}E_n & 0 \end{pmatrix}, \quad F'_{\kappa\lambda} = \Lambda_{\kappa\mu} \Lambda_{\lambda\nu} F_{\mu\nu}, \tag{A.5}$$

and the basic equations of macroscopic electrodynamics in arbitrary inertial systems can be written ($\varepsilon_{\kappa\lambda\mu\nu}$ four-dimensional Levi-Civita pseudo-tensor)

$$\varepsilon_{\kappa\lambda\mu\nu} \nabla_\lambda F_{\mu\nu} = 0, \quad \nabla_\nu F_{\mu\nu} = 4\pi j_\mu, \tag{A.6}$$

where $j_\mu = j_\mu(r)$ is the 4-vector of current density in space-time representation. We now make the transition to the Fourier transforms of the field functions (definitions see (3.4)).

From the transformation formula for the 4-vector of current density $j(k) = (\boldsymbol{j}(\boldsymbol{k},\omega), \mathrm{i}c\rho(\boldsymbol{k},\omega))$ using the definition (2.3) which leads to $j(k) = (-\mathrm{i}\omega \boldsymbol{P}(\boldsymbol{k},\omega), c\boldsymbol{k}\boldsymbol{P}(\boldsymbol{k},\omega))$, we obtain the following transformation formula for the polarization ($P_k(k) \equiv P_k(\boldsymbol{k},\omega), P'_i(k') \equiv P'_i(\boldsymbol{k}',\omega')$)

$$P'_i(k') = \frac{\omega}{\omega'} \left(\Lambda_{ik} + \mathrm{i}\frac{c}{\omega} \Lambda_{i4} k_k \right) P_k(k) = -\mathrm{i}\frac{c}{\omega'} k'_\mu \left(\Lambda_{ik} \Lambda_{\mu 4} - \Lambda_{i4} \Lambda_{\mu k} \right) P_k\left(\Lambda^{-1} k'\right). \tag{A.7}$$

After transformation of the known formula for the tensor of the electromagnetic field into a corresponding formula for the Fourier components and then after the elimination of the magnetic field by means of the first vectorial equation in (2.7) we arrive after some intermediate calculations to the following transformation formula for the electric field ($E_l(k) \equiv E_l(\boldsymbol{k},\omega)$, $E'_j(k') \equiv E'_{\boldsymbol{j}}(\boldsymbol{k}',\omega')$)

$$E_l(k) = E_l\left(\Lambda^{-1} k'\right) = -\mathrm{i}\frac{c}{\omega'} k'_\nu \left(\Lambda_{\nu 4} \Lambda_{jl} - \Lambda_{\nu l} \Lambda_{j4} \right) E'_j(k'). \tag{A.8}$$

An analogous formula which connects alone the components of the magnetic field before and after the Lorentz transformation does not exist without using the permittivity tensor but one may calculate the magnetic field from the electric field in each system via the Fourier-transformed Maxwell equation $B_j(k) = \frac{c}{\omega} \varepsilon_{jkl} k_k E_l(k)$ and one can widely work with the electric field alone.

From the definition of the susceptibility tensor before and after the Lorentz transformation

$$P_k(k) = \chi_{kl}(k) E_l(k), \quad P'_i(k') = \chi'_{ik}(k') E'_j(k'), \tag{A.9}$$

using (A.7) and (A.8) we obtain the following transformation formula

$$\chi'_{ij}(k') = -\frac{c^2}{\omega'^2} k'_\mu k'_\nu \left(\Lambda_{\mu 4} \Lambda_{ik} - \Lambda_{\mu k} \Lambda_{i4} \right) \left(\Lambda_{\nu 4} \Lambda_{jl} - \Lambda_{\nu l} \Lambda_{j4} \right) \chi_{kl}\left(\Lambda^{-1} k'\right). \tag{A.10}$$

Using the definition of the permittivity tensor before and after the Lorentz transformation, we obtain from (A.10) the corresponding transformation formula for the permittivity tensor

$$\varepsilon_{kl}(k) \equiv \delta_{kl} + 4\pi \chi_{kl}(k), \quad \varepsilon'_{ij}(k') \equiv \delta_{ij} + 4\pi \chi'_{ij}(k'). \tag{A.11}$$

Thus in the inertial system $\mathcal{I}'$ the medium appears as a homogeneous anisotropic and dispersive one with the permittivity tensor $\varepsilon'_{ij}(k') \equiv \varepsilon'_{ij}(\boldsymbol{k}',\omega')$.

For a three-dimensional orthogonal transformation

$$\Lambda_{\mu\nu} = \begin{pmatrix} \Lambda_{mn} & \Lambda_{m4} \\ \Lambda_{4n} & \Lambda_{44} \end{pmatrix} = \begin{pmatrix} R_{mn} & 0 \\ 0 & 1 \end{pmatrix}, \quad \left(R_{ml} R_{nl} = \delta_{mn}\right), \tag{A.12}$$

we find the usual transformation formula of a second-rank tensor function

$$\chi'_{ij}\left(\boldsymbol{k}', \omega'\right) = R_{ik} R_{jl} \chi_{kl}\left(\mathrm{R}^{-1}\boldsymbol{k}', \omega'\right), \tag{A.13}$$

where $\mathrm{R} = \left(\mathrm{R}^{\mathrm{T}}\right)^{-1}$ denotes the three-dimensional rotation operator.

The transformation of space-time vectors after separation of a part parallel and perpendicular of $\boldsymbol{r}$ to the velocity $\boldsymbol{V}$ can be written

$$\boldsymbol{r}' = \boldsymbol{r} + V\left(\left(\gamma - 1\right)\frac{\boldsymbol{Vr}}{\boldsymbol{V}^2} - \gamma t\right), \quad t' = \gamma\left(t - \frac{\boldsymbol{Vr}}{c^2}\right), \tag{A.14}$$

and, correspondingly, of wave vector and frequency according to (see (A.4) together with (A.3))

$$\boldsymbol{k}' = \boldsymbol{k} + \left(\left(\gamma - 1\right)\frac{\boldsymbol{kV}}{\boldsymbol{V}^2} - \gamma\frac{\omega}{c^2}\right)\boldsymbol{V}, \quad \omega' = \gamma\left(\omega - \boldsymbol{kV}\right), \tag{A.15}$$

with relativistic invariants

$$\boldsymbol{r}'^2 - c^2 t'^2 = \boldsymbol{r}^2 - c^2 t^2, \quad \boldsymbol{k}'^2 - \frac{\omega'^2}{c^2} = \boldsymbol{k}^2 - \frac{\omega^2}{c^2}, \quad \boldsymbol{k}'\boldsymbol{r}' - \omega' t' = \boldsymbol{kr} - \omega t. \tag{A.16}$$

Furthermore from (A.15) follows, in particular

$$\left[\boldsymbol{k}', \boldsymbol{V}\right] = \left[\boldsymbol{k}, \boldsymbol{V}\right], \quad \boldsymbol{k}'\boldsymbol{V} = \gamma\left(\boldsymbol{kV} - \omega\frac{\boldsymbol{V}^2}{c^2}\right) = \gamma\left(\boldsymbol{kV} - \omega\right) + \frac{\omega}{\gamma}. \tag{A.17}$$

As expected this shows that the components of the wave vector $\boldsymbol{k}$ perpendicular to the velocity $\boldsymbol{V}$ of the inertial system $\mathcal{I}'$ in $\mathcal{I}$ is not influenced by the transformation that, clearly, is known. The transformation formulae (A.7) and (A.8) in three-dimensional separation take on the form

$$P'_i\left(\boldsymbol{k}', \omega'\right) = \left\{\frac{V_i V_k}{\boldsymbol{V}^2} + \gamma\left(\delta_{ik} - \frac{V_i V_k}{\boldsymbol{V}^2} + \frac{\boldsymbol{k}'\boldsymbol{V}\delta_{ik} - V_i k'_k}{\omega'}\right)\right\} P_k\left(\boldsymbol{k}, \omega\right), \tag{A.18}$$

and using $\boldsymbol{B}'\left(\boldsymbol{k}', \omega'\right) = \frac{c}{\omega'}\left[\boldsymbol{k}', \boldsymbol{E}'\left(\boldsymbol{k}', \omega'\right)\right]$ (cf. also with formulae (8.2) in [42])

$$E_l\left(\boldsymbol{k}, \omega\right) = \left\{\frac{V_l V_j}{\boldsymbol{V}^2} + \gamma\left(\delta_{lj} - \frac{V_l V_j}{\boldsymbol{V}^2} + \frac{\boldsymbol{k}'\boldsymbol{V}\delta_{lj} - k'_l V_j}{\omega'}\right)\right\} E'_j\left(\boldsymbol{k}', \omega'\right), \tag{A.19}$$

where, in addition, the transformations (A.15) for the arguments have to be used. For the general susceptibilities defined by (A.9) using (A.18) and A.19) this leads finally to the transformation formula to the moving medium

$$\begin{aligned} \chi'_{ij}\left(\boldsymbol{k}', \omega'\right) = &\left\{\frac{V_i V_k}{\boldsymbol{V}^2} + \gamma\left(\delta_{ik} - \frac{V_i V_k}{\boldsymbol{V}^2} + \frac{\boldsymbol{k}'\boldsymbol{V}\delta_{ik} - V_i k'_k}{\omega'}\right)\right\} \\ &\cdot\left\{\frac{V_j V_l}{\boldsymbol{V}^2} + \gamma\left(\delta_{jl} - \frac{V_j V_l}{\boldsymbol{V}^2} + \frac{\boldsymbol{k}'\boldsymbol{V}\delta_{jl} - V_j k'_l}{\omega'}\right)\right\} \\ &\cdot\chi_{kl}\left(\underbrace{\boldsymbol{k}' + \left(\gamma - 1\right)\frac{\boldsymbol{k}'\boldsymbol{V}}{\boldsymbol{V}^2}\boldsymbol{V} + \gamma\frac{\omega'}{c^2}\boldsymbol{V}}_{=\boldsymbol{k}}, \underbrace{\gamma\left(\omega' + \boldsymbol{k}'\boldsymbol{V}\right)}_{=\omega}\right). \end{aligned} \tag{A.20}$$

The transformation of the corresponding permittivity tensors follow from

$$\varepsilon'_{ij}(\boldsymbol{k}',\omega') \equiv \delta_{ij} + 4\pi\chi'_{ij}(\boldsymbol{k}',\omega'), \quad \varepsilon_{kl}(\boldsymbol{k},\omega) \equiv \delta_{kl} + 4\pi\chi_{kl}(\boldsymbol{k},\omega). \tag{A.21}$$

If $\chi_{ij}(\boldsymbol{k},\omega)$ is a symmetric tensor then $\chi'_{ij}(\boldsymbol{k}',\omega')$ is also a symmetric tensor. Furthermore, we see that both indices of these tensors are independently transformed by exactly the same tensorial factors in (A.20) written on first and second lines.

From the transformation formulae we find then the formal Lorentz covariance of the expression

$$I_1 \equiv E_i(\pm k)E_i(k) - B_i(\pm k)B_i(k) = E_i(\pm k)\frac{c^2}{\omega^2}k_\mu k_\nu\left(\delta_{\mu i}\delta_{\nu j} - \delta_{\mu\nu}\delta_{ij}\right)E_j(k), \tag{A.22}$$

and the formal covariance of

$$I_2 \equiv 4\pi E_i(\pm k)\chi_{ij}(k)E_j(k). \tag{A.23}$$

The sum of both these covariant expressions can be written by means of the operator $L_{ij}(k)$ introduced in (2.7) as follows

$$I \equiv I_1 + I_2 = E_i(\pm k)L_{ij}(k)E_j(k). \tag{A.24}$$

Thus the basis of the presented concept of relativistic-covariant treatment of electrodynamics and optics of homogeneous dispersive media is formed by the vectorial wave equation for the electric field (5.4) in the form $L_{ij}(k)E_j(k) = 0$ together with one of the two equivalent representations

$$L_{ij}(k) \equiv \frac{c^2}{\omega^2}k_\mu k_\nu\left(\delta_{\mu i}\delta_{\nu j} - \delta_{\mu\nu}\delta_{ij}\right) + 4\pi\chi_{ij}(k) = \frac{c^2}{\omega^2}\left(k_i k_j - \boldsymbol{k}^2\delta_{ij}\right) + \varepsilon_{ij}(\boldsymbol{k},\omega), \tag{A.25}$$

in connection with the transformation formulae (A.8) for the electric field and for the tensor of susceptibility (A.10) as well as the use of the formal Lorentz invariance of the quantities considered in (A.22) and (A.24). The material properties are described by only one susceptibility tensor which depends on wave vector $\boldsymbol{k}$ and frequency ω before and after transformation from systems $\mathcal{I}$ to $\mathcal{I}'$.

One has to pay attention that for reason of consideration of the susceptibility tensor the transformation (A.18) is given in direction $\mathcal{I} \to \mathcal{I}'$ and the transformation (A.19) in inverse direction $\mathcal{I}' \to \mathcal{I}$ and that the transformation operators are almost but not fully equal where both are represented using $\boldsymbol{k}'$ and ω' in system $\mathcal{I}'$. These formulae describe the relativistic Doppler effect. The inverse formulae to (A.18), (A.19) and (A.15) and of transformation formulae in further text are obtained by substituting $\boldsymbol{V} \to -\boldsymbol{V}$ and by interchanging all quantities with and without primes. In contrast to (A.19) and (A.18) the well-known transformation formulae of Minkowski (e.g., [7], § 76 or, e.g., [5] [18]) are mixed transformations between electric and magnetic field which, furthermore, neglect the dispersion and are not made for the Fourier transforms of the fields.

Appendix B

Special Lorentz transformation of the energy-momentum tensor

In the transition from inertial system $\mathcal{I}$ to inertial system $\mathcal{I}'$ moving with velocity $\boldsymbol{V}$ in $\mathcal{I}$ the energy-momentum tensor transforms written in tensorial and in matrix form and as similarity transformation (upper index T means transposed)

$$T'_{\kappa\lambda} = \Lambda_{\kappa\mu}\Lambda_{\lambda\nu}T_{\mu\nu} = \Lambda_{\kappa\mu}T_{\mu\nu}\Lambda^{\mathrm{T}}_{\nu\lambda} = \Lambda_{\kappa\mu}T_{\mu\nu}\Lambda^{-1}_{\nu\lambda}, \quad \|\Lambda\| = 1, \tag{B.1}$$

that for the special Lorentz transformation (A.3) in space-time separation according to (8.1) is

$$\begin{aligned}
T'_{kl} &= \left(\delta_{km} + (\gamma-1)\frac{V_kV_m}{\boldsymbol{V}^2}\right)\left(\left(\delta_{ln} + (\gamma-1)\frac{V_lV_n}{\boldsymbol{V}^2}\right)T_{mn} - \gamma V_l g_m\right) \\
&\quad - \gamma\frac{V_k}{c^2}\left(\left(\delta_{ln} + (\gamma-1)\frac{V_lV_n}{\boldsymbol{V}^2}\right)S_n - \gamma V_l w\right), \\
g'_k &= \gamma\left\{-\left(\delta_{km} + (\gamma-1)\frac{V_kV_m}{\boldsymbol{V}^2}\right)\left(\frac{V_n}{c^2}T_{mn} - g_m\right) + \gamma\frac{V_k}{c^2}\left(\frac{V_n}{c^2}S_n - w\right)\right\}, \\
S'_l &= \gamma\left\{-\left(\delta_{ln} + (\gamma-1)\frac{V_lV_n}{\boldsymbol{V}^2}\right)\left(V_mT_{mn} - S_n\right) + \gamma V_l\left(V_m g_m - w\right)\right\}, \\
w' &= \gamma^2\left\{\frac{V_n}{c^2}\left(V_mT_{mn} - S_n\right) - \left(V_m g_m - w\right)\right\}.
\end{aligned} \tag{B.2}$$

From this follows

$$\begin{aligned}
T'_{kl} - T'_{lk} &= \left(\delta_{km} + (\gamma-1)\frac{V_kV_m}{\boldsymbol{V}^2}\right)\left(\delta_{ln} + (\gamma-1)\frac{V_lV_n}{\boldsymbol{V}^2}\right)\left(T_{mn} - T_{nm}\right) \\
&\quad + \gamma\left\{V_k\left(g_l - \frac{1}{c^2}S_l\right) - V_l\left(g_k - \frac{1}{c^2}S_k\right)\right\}, \\
g'_k - \frac{1}{c^2}S'_k &= \gamma\left\{\frac{V_m}{c^2}\left(T_{mk} - T_{km}\right) + \left(\delta_{km} - \frac{V_kV_m}{\boldsymbol{V}^2}\right)\left(g_m - \frac{1}{c^2}S_m\right)\right\},
\end{aligned} \tag{B.3}$$

showing that symmetric and anti-symmetric parts of the energy-momentum tensor transforms independently on each other.

If we denote the energy-momentum tensor by a four-dimensional matrix T and the Lorentz transformation by a matrix Λ then the transformation (B.1) with $\Lambda^{\mathrm{T}} = \Lambda^{-1}$ (Minkowski metrics $g_{\mu\nu} = \delta_{\mu\nu}$) can be written as a similarity transformation as follows

$$T' = \Lambda T \Lambda^{-1}. \tag{B.4}$$

This shows that the real-valued independent invariants of the energy-momentum tensor with respect to Lorentz transformations are the invariants of the similarity transformations (B.4) which can be chosen as the coefficients in the four-dimensional Hamilton-Cayley identity for T

$$T^4 - \langle T\rangle T^3 + [T]T^2 - |T|T + \|T\|I = 0, \tag{B.5}$$

where the first three invariants with respect to similarity transformations $\langle T\rangle, [T], |T|$ are formally given by the same formulae as for three-dimensional operators but, e.g., with $\langle T\rangle$ the four-dimensional trace of T and, in addition, we have to consider the four-dimensional determinant $\|T\|$ that means

$$\begin{aligned}&\langle T\rangle \equiv T^{\mu}_{\mu}, \quad [T]=\frac{1}{2}\left(\langle T\rangle^{2}-\langle T^{2}\rangle\right), \quad |T|=\frac{1}{6}\left(\langle T\rangle^{3}-3\langle T\rangle\langle T^{2}\rangle+2\langle T^{3}\rangle\right)\\ &\|T\|=\frac{1}{24}\left(\langle T\rangle^{4}-6\langle T\rangle^{2}\langle T^{2}\rangle+3\langle T^{2}\rangle^{2}+8\langle T\rangle\langle T^{3}\rangle-6\langle T^{4}\rangle\right).\end{aligned} \tag{B.6}$$

We now use an index-less representation of the energy-momentum tensor T of the form[17]

$$T \equiv \begin{pmatrix} \mathrm{T} & \mathrm{i}c\boldsymbol{g} \\ \frac{\mathrm{i}}{c}\boldsymbol{S} & -w \end{pmatrix}, \tag{B.7}$$

with T the stress tensor T (or its matrix) as a three-dimensional operator and find for the four-dimensional invariants of T expressed by the three-dimensional invariants formed from the three-dimensional operator T, from the vectors $\boldsymbol{g}$ and $\boldsymbol{S}$ and from the scalar w in (B.7)

$$\begin{aligned}&\langle T\rangle=\langle \mathrm{T}\rangle-w,\\ &[T]=[\mathrm{T}]-\langle \mathrm{T}\rangle w+\boldsymbol{Sg},\\ &|T|=|\mathrm{T}|-[\mathrm{T}]w+\langle \mathrm{T}\rangle \boldsymbol{Sg}-\boldsymbol{S}\mathrm{T}\boldsymbol{g},\\ &\|T\|=-|\mathrm{T}|w+[\mathrm{T}]\boldsymbol{Sg}-\langle \mathrm{T}\rangle \boldsymbol{S}\mathrm{T}\boldsymbol{g}+\boldsymbol{S}\mathrm{T}^{2}\boldsymbol{g},\end{aligned} \tag{B.8}$$

where $\langle \mathrm{A}\rangle$ denotes the trace of an arbitrary three-dimensional operator A, $[\mathrm{A}]$ its second invariant and $|\mathrm{A}|$ its determinant according to

$$\begin{aligned}&\langle \mathrm{A}\rangle \equiv \frac{1}{2}\varepsilon_{ijk}\varepsilon_{ljk}A_{il}=A_{kk}, \quad [\mathrm{A}]\equiv \frac{1}{2}\varepsilon_{ijk}\varepsilon_{lmk}A_{il}A_{jm}=\frac{1}{2}\left(\langle \mathrm{A}\rangle^{2}-\langle \mathrm{A}^{2}\rangle\right),\\ &|\mathrm{A}|\equiv \frac{1}{6}\varepsilon_{ijk}\varepsilon_{lmn}A_{il}A_{jm}A_{kn}=\frac{1}{6}\left(\langle \mathrm{A}\rangle^{3}-3\langle \mathrm{A}\rangle\langle \mathrm{A}^{2}\rangle+2\langle \mathrm{A}^{3}\rangle\right),\end{aligned} \tag{B.9}$$

and where $\boldsymbol{xy}\equiv x_n y_n$ is the scalar product of three-dimensional vectors $\boldsymbol{x}$ with $\boldsymbol{y}$ and $\boldsymbol{x}\mathrm{A}\boldsymbol{y}\equiv x_m A_{mn} y_n$ a three-dimensional bilinear form. The first invariant $\langle T\rangle$ is the already mentioned trace of T and the invariant $\|T\|$ the determinant of the 4-dimensional energy-momentum tensor. All are also relativistic invariants due to (B.4). Among the invariants (B.8) one has only one invariant which is linear in the components of the energy-momentum tensor T, namely, the trace $\langle T\rangle=\langle \mathrm{T}\rangle-w$.

According to (9.9) the trace $\langle T\rangle$ does not vanish in general but it vanishes in case of neglected dispersion. The second invariant $[T]$ and the higher invariants vanish for energy-momentum tensors which factorize in the form of dyadic products of two vectors. This means that one has only one non-vanishing eigenvalue of this tensor also if one takes into account the dispersion as it is shown for the second invariant in (12.15). One has to expect that the second and the higher invariants of the energy-momentum tensor do not vanish in higher approximations of the slowly varying amplitudes as it is suggested by considerations in Section 9.

[17]We remind that the spatial part T and T itself are not uniquely defined in literature with respect to sign and also to their notation.

Journal of Modern Physics, 2021, 12, 1922-1929
https://www.scirp.org/journal/jmp
ISSN Online: 2153-120X
ISSN Print: 2153-1196

Matter, Dark Matter and Quartic Potential Generated by Unstable Confinement of Quarks

Leif Matsson

Department of Physics, University of Gothenburg, Gothenburg, Sweden
Email: leif.matsson@telia.com

How to cite this paper: Matsson, L. (2021) Matter, Dark Matter and Quartic Potential Generated by Unstable Confinement of Quarks. *Journal of Modern Physics*, **12**, 1922-1929.
https://doi.org/10.4236/jmp.2021.1213109

Received: October 21, 2021
Accepted: November 27, 2021
Published: November 30, 2021

Abstract

The Standard Model for particle physics is here extended by making a non-equilibrium filling of the empty vacuum after the start of Big Bang. The process is described as an unstable binding of massless quarks to massless antiquarks. When the filled part of vacuum condenses, the system becomes stabilized, quarks acquire mass and become confined and a quartic potential is induced, which hence need not be introduced ad hoc. The coupling and scale parameters in this potential have become asymmetric microscopic functions of the quark and antiquark densities. The so obtained dynamics can explain how the matter-antimatter asymmetry in the Universe and dark matter emerged. Quantum corrections are included and the model then gives ordinary matter, dark matter and dark energy contents at correct orders of magnitude.

Keywords

Dark Matter, Dark Energy, Cosmology, Quartic Potential

1. Introduction

It has been thought that Electroweak (EW) baryogenesis and QCD confinement could help to explain how the Standard Model (SM) emerged [1]-[7]. However, many questions have remained open and many new problems have emerged. For instance, the lambda cold dark matter (LCDM) model postulates that the expansion of the Universe is driven by dark energy (DE) and dark matter (DM), and that the galaxy structures we see today are due to density variations in the very early Universe. Such variations are expected to produce gravitational waves and thus a signal in the cosmic microwave background. However, such a signal has not been observed [8]. This is not a problem in the model suggested here, because it starts from a completely empty vacuum.

The generation of matter and dark matter is in this model induced by a non-equilibrium filling of vacuum which is here assumed to have been totally empty at the Big Bang. Big Bang is assumed to have started at infinite energy conditions from a completely empty vacuum and that the Universe then contained equal and isotropic densities of free massless quarks and antiquarks, leptons and antileptons. At lower energies, the QCD model appears to remain intact with a stable vacuum even if a finite number of quark-antiquark pairs are generated. However, above a finite critical energy (temperature) E_c, surplus quarks (and surplus leptons) became generated, implying that vacuum could then not be expected to have been stable enough for quantum field theories to have been valid. The non-equilibrium filling of vacuum has another advantage. It generates the quartic (so called Ginsburg Landau or Higgs) potential, which thus need not be assumed in an *ad hoc* manner as in the SM.

The Dirac equation [9] and QED did not have any vacuum problems, because these models were not used at very high energies. As demonstrated before [10], surplus quarks were generated by a non-equilibrium filling of the empty QCD vacuum, a process which had to be described by classical fields. The non-equilibrium filling of vacuum, followed by condensation (assumed to have occurred when the energy decreased after Big Bang), generated a quartic potential, in which the coupling $\lambda = g^2$ and the scale a became asymmetric microscopic functions of the quark and antiquark densities.

The density of surplus quarks in the numerator of the coupling g in this model is proportional to the density of nucleons and hence approximately to the density of ordinary mass in the Universe. Except for this density of surplus quarks, the denominator of g also contains the two equal densities of non-surplus quarks and antiquarks, which are here interpreted as building blocks of dark matter (DM). It is assumed here that top quarks dominated at higher energies already from the start of Big Bang, but all types of quarks should give the same result. It is not known how fast the maximal density of surplus quarks, corresponding to the actual density of protons in the Universe, is attained. However, the nonquantal vacuum filling process should be normalized such that $\lambda = g^2 = 0.11$ at $E_c = 10^5$ GeV (**Figure 1**), where vacuum is sufficiently filled for the standard model to start working. At energies lower than 10^5 GeV, the matter-DM ratio then approaches the value observed.

Leptons are assumed to contribute similarly, but much less, and are therefore neglected here. Under assumption that confinement started as a non-equilibrium binding of massless quarks, color by color, the suggested model yields observable matter, DM and DE contents at correct orders of magnitude. The results obtained show that the critical energy level E_c is much below 10^{10} GeV at which $\lambda_{quartic}$ in the SM becomes negative [11].

2. Unstable Confinement after Big Bang

After the start of Big Bang, an infinite number of the empty negative energy vacuum states below the critical vacuum level $-E_c$ must have been filled before quantum effects could be expected to have had any impact. The non-equilibrium filling of this part of the QCD vacuum, which started at infinite energy conditions, could be viewed as a strong but unstable binding of massless quarks to antiquarks

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = k \cdot \rho(x,t) \cdot \rho_\alpha(x,t) - k'\psi(x,t), \tag{1}$$

where ρ, ρ_a and ψ are classical fields (densities) of quarks, antiquarks and quark-antiquark pairs, and k and k' are the association and dissociation constants. This form of confinement process is here assumed to have been equal for all flavours and colours. The filling of the remaining finite number of vacuum states above $-E_c$ (energies below E_c), is sufficiently stable to be described as usual in the SM.

By insertion of the initial constraints, $\rho(\boldsymbol{x},t) = \rho_0 - \psi(\boldsymbol{x},t)$ and $\rho_a(\boldsymbol{x},t) = \rho_{a0} - \psi(\boldsymbol{x},t)$, where ρ_0 and ρ_{a0} are the two equal initial densities of quarks and antiquarks, here assumed to be equal, Equation (1) becomes

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = k \cdot \left(\left(a - \psi(\boldsymbol{x},t) \right)^2 - \left(a^2 - b^2 \right) \right), \tag{2}$$

in which $a = (\rho_0 + \rho_{a0} + K)/2$, $b^2 = \rho_0 \rho_{a0}$, and $K = k'/k$. Integration of Equation (2) then yields

$$\ln\left(\frac{a + \sqrt{a^2 - b^2} - \psi}{a - \sqrt{a^2 - b^2} - \psi} \frac{a(1-g)}{a(1+g)} \right) = \ln\left(\frac{\rho_K(\boldsymbol{x},t)}{\rho_{\alpha K}(\boldsymbol{x},t)} \frac{\rho_{\alpha K}}{\rho_K} \right) = 2kag \cdot (t - t_0) \tag{3}$$

where $\rho_K(\boldsymbol{x},t) = \rho_K - \psi(\boldsymbol{x},t)$ and $\rho_{aK}(\boldsymbol{x},t) = \rho_{aK} - \psi(\boldsymbol{x},t)$ are the time dependent quark and antiquark densities. With $\rho_K + \rho_{aK} = 2a$ and $g = \sqrt{(a^2 - b^2)/a^2} = (\rho_K - \rho_{aK})/(\rho_K + \rho_{aK}) > 0$ the corresponding after start 'new initial' densities ρ_K and ρ_{aK} can be written as $\rho_K = a(1+g) \geq \rho_0$ and $\rho_{aK} = a(1-g) \leq \rho_{a0}$. For $k' > 0$ the coupling g thus becomes an asymmetric microscopic function of the quark and antiquark densities, which shows that the suggested model works like a seesaw.

At the start of Big Bang, the Universe contained equal densities ρ_0 and ρ_{a0} of free massless quarks and antiquarks, which might then have condensed pairwise, $(\rho\rho_a)(\rho\rho_a)(\rho\rho_a)\ldots$, without generating any surplus quarks. But as described in Equation (3), the quarks confined asymmetrically, without leaving any antiquarks behind, into a system, which apart from pairs of non-surplus quarks ρ_{NS} and antiquarks ρ_{aK}, contained surplus quarks ρ_s, $\rho_s(\rho_{aK}\rho_{Ns})(\rho_{aK}\rho_{Ns})(\rho_{aK}\rho_{Ns})\ldots$ This still implies that the non-surplus quarks were equally many as the antiquarks, which they were also before the start of Big Bang. However, it should be observed that at infinite energy (temperature), before the confinement had started, all pairs $(\rho\rho_a)$ were free and there was no difference between the two systems.

The numerator of the coupling g equals the density of surplus quarks which can also be separated out in the denominator of g. Apart from that the denominator of $g = \rho_s/(\rho_s + \rho_{Ns} + \rho_{aK})$ also contains the two equal densities of non-surplus quarks ρ_{Ns} and antiquarks ρ_{aK}, which are here interpreted as building blocks of DM and black holes. When the Universe cooled down to the critical energy level E_o, the strong binding between non-surplus quarks and antiquarks started to become stable. The coupling g then increased above $1/(1+1+1) \approx 0.33$. Surplus leptons are expected to have emerged similarly.

3. A Microscopic Form of Potential

After binding into quark-antiquark pairs, this system condensed. This process can be described by a change of variable

$$\varphi(\boldsymbol{x},t) = \frac{\sqrt{a^2 - b^2}}{1 - \psi(\boldsymbol{x},t)/a} \tag{4}$$

which can also be viewed as a partition function where ψ/a plays a role as non-equilibrium "fugacity" driven by Equation (2).

The derivative of Equation (4) combined with Equation (2) gives

$$\frac{\mathrm{d}\varphi}{k\mathrm{d}t} = \frac{\mathrm{d}\varphi}{\mathrm{d}x} = g\left(a^2 - \varphi^2\right), \tag{5}$$

which describes the condensation of the strongly bound $\rho\rho_a$ - pairs as a travelling wave $\varphi(x) = a\tanh(agx)$ that propagates at a velocity $k = x/t$.

The square of Equation (5) yields a microscopic form of the quartic potential $V(\varphi)$

$$\frac{1}{2}\left(\frac{\mathrm{d}\varphi}{\mathrm{d}x}\right)^2 = \frac{g^2}{2}\left(a^2 - \varphi^2\right)^2 \equiv V(\varphi), \tag{6}$$

which hence need not be inserted *ad hoc* like in electroweak theory, and in which both the scale a and the coupling g have become microscopic functions of the quark and antiquark densities. As will be shown, after inclusion of quantum corrections below E_o, the microscopic potential $V(\varphi)$ approaches the quartic potential in the EW theory when the energy decreases further.

By breakdown of symmetry, $\varphi \to -a + \varphi$, Equation (6) then gives the equation of motion

$$\frac{1}{2}\left(\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right)^2 - \frac{1}{2}\left(\frac{\mathrm{d}\varphi}{\mathrm{d}x}\right)^2 + (2ga)^2\varphi = 6\lambda\varphi^2\left(a - \varphi/3\right) \tag{7}$$

where $2ga = m_B$ is the mass of a boson, a precursor to the Higgs boson, and $\lambda = g^2$. It is not known exactly when the condensation started and at what rate, *i.e.* if the symmetry breakdown occurred exactly at 10^5 GeV.

4. Quantum Corrections

In the previous "classical" approach [10], $2ga$ was identified as the mass of the Higgs boson $m_H \approx 125$ GeV. With $a = \rho_0 \approx 174$ GeV and $g \approx m_H/348 = 0.36$, the coupling $\lambda = g^2 = 0.13$ was then interpreted as the matter to all matter ratio (except dark energy), as if quantum effects were already included. However, in this work $2ga = m_B$ is identified as the mass of the becoming Higgs particle, a precursor boson that gets the mass, $m_H = 125$ GeV, at a certain critical energy below E_c at which quantum effects started to contribute. However, before the energy has decreased to E_c, the density of surplus quarks must have increased from zero to the value at which quantum effects start to contribute.

The problem is then to combine the result of the nonequilibrium filling of vacuum, which had to be described in the terms of classical fields, with the finite energy result of QCD quantum field theory [11], and to determine the critical energy level E_c at which this happens. But in order to simplify this task, the scale a is first modified from 174 GeV to the level of the top quark mass $m_{top} = 173$ GeV.

The mass m_H in the RG improved m_H/m_{top} - ratio at different energies in Fig. 2 in ref. [11] is here interpreted as the mass of the precursor boson m_B in Equation (7). This mass changes with energy such that the m_H/m_{top} - ratio attains 125/173 ≈ 0.72, and the numerator thus attains the mass of the Higgs boson, at about (or somewhat below) 10^3 GeV in Fig 2 in ref. [11]. By inclusion of quantum effects below 10^5 GeV in the model suggested here the m_B/m_{top} - ratio thus first increases from 0 to 0.66, and according to Fig. 2 in ref. [11], it then increases to 0.68, 0.72, 0.77 when the energy decreases further to 10^4, 10^3, 10^2 GeV. Given these numbers of the m_B/m_{top} - ratio and $m_{top} = 173$ GeV, the boson mass m_B then correspondingly becomes 114.18, 117.64, 124.56, 133.21 GeV at the actual energy levels.

The coupling constant $g = m_B/2a$ thus first increases from 0 to 0.33 = 114.18/346 and hence $\lambda = g^2 = 0.11$ at 10^5 GeV in **Figure 1** is thus reached from below when the energy decreases from higher energies.

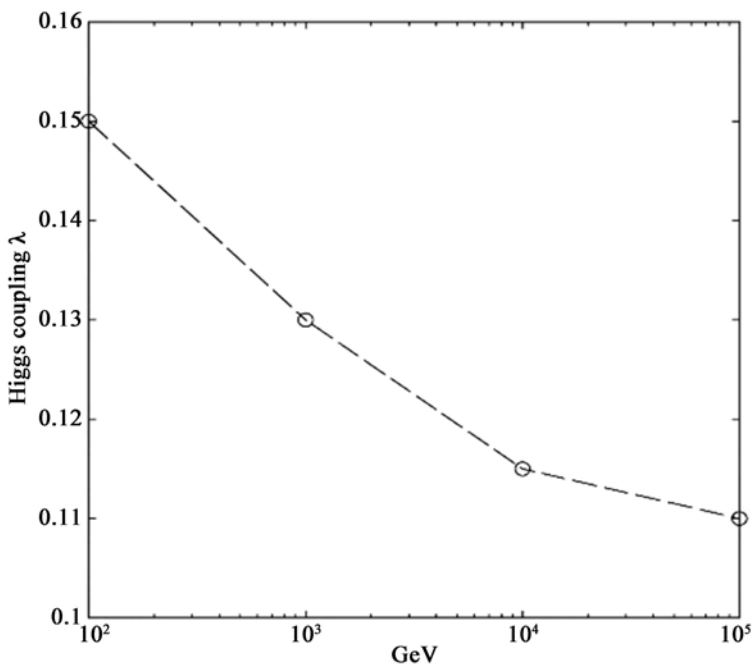


Figure 1. The coupling $\lambda = g^2$ as a function of energy.

The coupling g then increases from 0.330 to 0.340, 0.360, 0.385, and $\lambda = g^2$ correspondingly becomes 0.109, 0.116, 0.130, 0.148 (see **Figure 1** here) at the actual energy levels. Accordingly, $\lambda = (125/346)^2 = m_H/2a = 0.13$ at 10^3 GeV and $\lambda = 0.148 \approx 0.15$ at 10^2 GeV. However, the m_H/m_W - ratio in Fig 2 in ref. [11] shows that $m_H/m_W = 1.55 \approx 124.56/80.40$, hence, $m_H \approx 125$ GeV is reached already at 3.46×10^2 GeV. The start of impact of quantum corrections should thus be correspondingly delayed from 10^5 GeV to $3.46 \cdot 10^4$ GeV.

It seems that the Universe has always remembered its quantum origin [12 - 14], from the start of Big Bang at infinite energy when vacuum was totally empty, until it became sufficiently filled for quantum field theories (QFTs) in the SM to start working. But as described here, for the density of quarks $\rho_s = (\rho_K - \rho_{aK})$ (and correspondingly for leptons) to increase from zero to a critical level corresponding to the amount of observable matter in the Universe, the filling of vacuum must have been a non-equilibrium process. This cannot be explained by the SM, which only allows the creation or annihilation of a finite number of particle-antiparticle pairs.

On the contrary, in the model suggested here the numerator of the coupling g describes the gap between the densities of quarks ρ_s and antiquarks ρ_{aK}. After separation of the density of quarks ρ_K in the denominator of g into surplus quarks ρ_s and non-surplus quarks ρ_{Ns}, the density of non-surplus quarks ρ_{Ns} becomes equal to the density of antiquarks ρ_{aK}

$$g = (\rho_K - \rho_{aK}) / (\rho_K + \rho_{aK}) = \rho_s / (\rho_s + \rho_{Ns} + \rho_{aK}) > 0 . \tag{8}$$

Quantum effects are here assumed to have started to contribute at 10^5 GeV where g was first equal to $1/(1+1+1) = 1/(1+2) = 1/3$ and $\lambda = g^2 \approx 0.11$ (**Figure 1**). The massive non-surplus quarks ($\rho_{Ns} \approx 1$) and antiquarks ($\rho_{aK} \approx 1$) then formed more strongly bound states of dark matter, and the denominator of g then decreased below 3. As already explained, the coupling λ then increased successively from $0.109 \approx 0.11$ to $0.148 \approx 0.15$, see **Figure 1**.

As is clear from eq. (8), the coupling g and hence also $\lambda = g^2$ define the relationship between matter and DM. Accordingly, $0.148 = 0.049/0.33$ corresponds to the ratio between matter and the sum of matter and DM. The corresponding relative content of DM in the Universe thus becomes $0.33 - 0.049 = 0.28$, and the remaining 67% corresponds to the density of dark energy, which also corresponds to the value of the cosmological constant. The cosmological constant has thus decreased from 1 to 0.67 during the generation of 28% DM.

5. Summary

Infrared divergence is but one reason why quark confinement cannot be described perturbatively in QCD [15]. However, as demonstrated here, in Equation (1), the confinement can be described non-perturbatively, and since the density of quarks must increase over that of antiquarks, this vacuum filling must be a non-equilibrium process. A second reason is that quantum fields lose their definition due to vacuum instabilities above the critical energy E_c, after emptying of a finite number of vacuum states (creation of a finite number of quark-antiquark pairs). The suggested model thus described the non-equilibrium dynamics above

E_c that generated surplus quarks, how this is connected with the QCD part of the SM, and it also explained the generation of the quartic potential, the separation between matter and DM and DE.

The non-equilibrium dynamics described here emerged by filling of the infinite number vacuum states that became emptied at energies above E_c. When this part of the filled vacuum condensed it generated a quartic potential in which the coupling g and scale a became microscopic asymmetric functions of the quark and antiquark densities. By inclusion of quantum effects, the coupling g then gave a matter to DM ratio and DE density at correct orders of magnitude. The model also gave a realistic value of the cosmological "constant", $\Lambda = 0.67$, corresponding to the DE density that determines the rate of inflation in the Universe [16]-[22]. At lower energies, the obtained microscopic quartic potential agreed with the quartic potential in EW theory. The model thus provides a form of direct link between the Universe immediately after the start of Big Bang at infinite high energy and high energy particle physics described by the SM.

The masses of neutrinos are usually described by a mixing of flavours in different seesaw models [23]-[30]. However, also neutrino quantum fields require a sufficiently filled vacuum. It would be interesting to see if the suggested model, which should also be able to explain the generation of surplus neutrinos, could help to discriminate between the various seesaw models.

Conflicts of Interest

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

References

[1] Kuzmin, V.A., Rubakov, V.A. and Shaposhnikov, M.E. (1985) *Physics Letters B*, **155**, 36-42. https://doi.org/10.1016/0370-2693(85)91028-7

[2] Cohen, A.G., Kaplan, D.B. and Nelson, A.E. (1993) *Annual Review of Nuclear and Particle Science*, **43**, 27-70. https://doi.org/10.1146/annurev.ns.43.120193.000331

[3] Carena, M., Quiros, M., Riotto, A., Vilja, I. and Wagner, C.E.M. (1997) *Nuclear Physics B*, **503**, 387-404. https://doi.org/10.1016/S0550-3213(97)00412-4

[4] Kubo, J., Lim, K.S. and Lindner, M. (2014) *Physical Review Letters*, **113**, Article ID: 091604. https://doi.org/10.1103/PhysRevLett.113.091604

[5] Ballesteros, G., Redondo, J., Ringwald, A. and Tamarit, C. (2017) *Physical Review Letters*, **118**, Article ID: 071802. https://doi.org/10.1103/PhysRevLett.118.071802

[6] Ipek, S. and Tait, T.M.P. (2019) *Physical Review Letters*, **122**, Article ID: 11200. https://doi.org/10.1103/PhysRevLett.122.112001

[7] Carena, M., Quiros, M. and Zhang, Y. (2019) *Physical Review Letters*, **122**, Article ID: 201802. https://doi.org/10.1103/PhysRevLett.122.201802

[8] Ade, P.A.R., *et al.* 2021) *Physical Review Letters*, **127**, Article ID: 151301. https://doi.org/10.1103/PhysRevLett.127.151301

[9] Dirac, P.A.M. (1930) *Proceedings of the Royal Society of London*, **126**, 360. https://doi.org/10.1098/rspa.1930.0013

[10] Matsson, L. (2016) *World Journal of Mechanics*, **6**, 441-445. https://doi.org/10.4236/wjm.2016.611031

[11] Buttazo, D., Degrassi, G., Giardino, P.P., Giudice, G.F., Sala, F., Salvio, A. and Strumia, A. (2013) *JHEP*, **12**, 89. https://doi.org/10.1007/JHEP12(2013)089

[12] Green, D. and Porto, R.A. (2020) *Physical Review Letters*, **124**, Article ID: 251302. https://doi.org/10.1103/PhysRevLett.124.251302

[13] Hu, W., Spergel, D.N. and White, M.J. (1997) *Physical Review D*, **55**, 3288. https://doi.org/10.1103/PhysRevD.55.3288

[14] Maldacena, J. (2016) *Fortschritte der Physik*, **64**, 10-23. https://doi.org/10.1002/prop.201500097

[15] Matsson, L. and Meuldermans, R. (1977) *Physics Letters B*, **70**, 309-312. https://doi.org/10.1016/0370-2693(77)90665-7

[16] Guth, A. (1981) *Physical Review D*, **23**, 347-356. https://doi.org/10.1103/PhysRevD.23.347

[17] Linde, A.D. (1982) *Physics Letters B*, **108**, 389-393. https://doi.org/10.1016/0370-2693(82)91219-9

[18] Albrect, A. and Steinhardt, P.J. (1982) *Physical Review Letters*, **48**, 1220-1223. https://doi.org/10.1103/PhysRevLett.48.1220

[19] Bezrukov, F. and Shaposhnikov, M. (2008) *Physics Letters B*, **659**, 703-706. https://doi.org/10.1016/j.physletb.2007.11.072

[20] Bezrukov, F., Kalmykov, M.Y., Kniehl, B.A. and Shaposhnikov, M. (2012) *JHEP*, **10**, 140. https://doi.org/10.1007/JHEP10(2012)140

[21] Kanemura, S., Matsui, T. and Nabeshima, T. (2013) *Physics Letters B*, **723**, 126-131. https://doi.org/10.1016/j.physletb.2013.05.002

[22] Okada, N. and Shafi, Q. (2015) *Physics Letters B*, **747**, 223-228. https://doi.org/10.1016/j.physletb.2015.06.001

[23] Fukuda, Y., *et al.* (1998) *Physical Review Letters*, **81**, 1562-1567.

[24] Minkowski, P. (1977) *Physics Letters B*, **67**, 421-428. https://doi.org/10.1016/0370-2693(77)90435-X

[25] Gell-Mann, M., Ramond, P. and Slansky, R. (1979) *Conf. Proc. C*, **315**, Article ID: 790927.

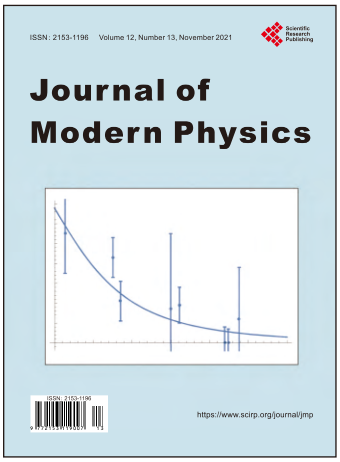
[26] Yanagida, T. (1980) *Progress of Theoretical Physics*, **64**, 1103-1105. https://doi.org/10.1143/PTP.64.1103

[27] Fukugita, M. and Yanagida, T. (1986) *Physics Letters B*, **174**, 45-47. https://doi.org/10.1016/0370-2693(86)91126-3

[28] Ipek, S., Plascencia, A.D. and Turner, J. (2018) *JHEP*, **12**, 111. https://doi.org/10.1007/JHEP12(2018)111

[29] Dror, J.A., Hiramatsu, T., Kohri, K., Murayama, H. and White, G. (2020) *Physical Review Letters*, **124**, Article ID: 041804. https://doi.org/10.1103/PhysRevLett.124.041804

[30] Greenberg, O.W. (1964) *Physical Review Letters*, **13**, 598-602. https://doi.org/10.1103/PhysRevLett.13.598



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**