symmetry

IMPACT FACTOR 2.2

CITESCORE 5.4

Article

Newton's First Law and the Grand Unification

Martin Tamm

MDPI

https://doi.org/10.3390/sym16121694

symmetry

MDPI

Article

Newton's First Law and the Grand Unification

Martin Tamm

Department of Mathematics, University of Stockholm, 106 91 Stockholm, Sweden; matamm@math.su.se;
Tel.: +46-708775876

Abstract: This paper is devoted to the study of stationary trajectories of free particles. From a classical point of view, this appears to be an almost trivial problem: Free particles should follow straight lines as predicted by Newton's first law, and straight lines are indeed the stationary trajectories of the standard action integrals in the classical theory. In the following, however, a general relativistic approach is studied, and in this situation it is much less evident what action integral should be used. As it turns out, using the traditional Einstein–Hilbert principle gives us stationary states very much in line with the classical theory. But it is suggested that a different action principle, and in fact one which is closer to quantum mechanics, gives stationary states with a much richer structure: Even if these states in a sense can represent particles which obey the first law, they are also inherently rotating. Although we may still be far from understanding how general relativity and quantum mechanics should be united, this may give an interesting clue to why rotation (or rather spin, which is a different but related concept) seems to be the natural state of motion for elementary particles.

Keywords: law of inertia; general relativity; quantum mechanics; Lagrangian; rotation; spin

1. Introduction

Newton's first law has, ever since it was formulated, been one of the most indisputable laws of physics (see [1]). Before Newton, it may have been a non-obvious statement, but to his followers it acquired an almost axiomatic status:

> "Every body continues in its state of rest, or of uniform motion in a straight line, unless it is compelled to change that state by an external force."

Later generations of physicists have tried to derive Newton's laws from more fundamental principles, most importantly the principle of least action: Bodies travel along straight lines because this is the simplest (or "most economical") way of developing. In a more mathematical language, and in the simplest possible situation in classical physics with just one particle, this can be expressed by saying that the straight line is the unique action-minimizing trajectory between the starting point and the endpoint.

When the big revolutions in physics came along at the beginning of the twentieth century, the first law survived remarkably well, both within general relativity and quantum mechanics. In general relativity, bodies should now move along geodesics instead of just straight lines. In quantum mechanics, Feynman's democracy of all histories approach has given us a sound motivation for identifying the trajectories which are stationary with respect to the action with the probability-maximizing trajectories, i.e., coarsely speaking the ones that have a reasonable chance of occurring. But none of these modifications is usually thought of as altering the essential meaning of the first law itself.

Could there possibly still be something to add to the first law after more than three centuries? In this paper, I will argue that there may actually be more to say. The point is not that the first law as we know it is wrong. But it may in fact be that, in a certain sense, the stationary solutions have a both more complex and more interesting structure than is usually thought. This opinion is based on a curious and perhaps unexpected property

Citation: Tamm, M. Newton's First Law and the Grand Unification. *Symmetry* **2024**, *16*, 1694. https://doi.org/10.3390/sym16121694

Academic Editor: Manuel Gadella

Received: 30 October 2024
Revised: 4 December 2024
Accepted: 16 December 2024
Published: 20 December 2024

of Lorentz geometry, a property which turns out to be particularly interesting when we attempt to unify general relativity and quantum mechanics.

The present paper is the result of an attempt to apply general relativity to the study of elementary particles. This is rather contrary to the historically dominating approach, where gravity is almost always neglected. Although there seems to be a general agreement among physicists that every particle should in principle give rise to a Schwarzschild metric surrounding it, the effects of this metric are in general considered to be so weak that they can be left out of the discussion.

Here, the idea will instead be the opposite one: Almost all properties of elementary particles will be neglected, except the influence that comes from the metric which they induce. It should be kept in mind that, as it is, we do not yet have access to any firm knowledge about how these metrics are induced. For this reason, the metrics in this paper should not be considered to be based on some fundamental physical theory. Rather, they should be viewed as simple mathematical attempts to illustrate some properties which could potentially be important, not only in Lorentz geometry, but also for the grand unification of general relativity and quantum mechanics.

Before going into details, let me also say that I agree with the general opinion that the effects of the Schwarzschild metric away from a particle are probably negligible for almost everything in particle physics. However, this is not the full story. The Schwarzschild metric is not defined in all of space-time because of the singularity at the origin. If we insist on a metric defined everywhere, which agrees with the Schwarzschild metric asymptotically, then inevitably it will have to have non-zero Ricci curvature somewhere, most likely close to or inside the particle.

The point is that this non-zero Ricci curvature may not be negligible at all. In fact, if we believe that the concept of mass–energy in general relativity in a unified theory should be essentially equivalent to the concept of mass–energy in quantum mechanics, then one can argue that this Ricci curvature could be as important as the usual quantum mechanical mass–energy.

The key tool in the following will be the principle of least action. In a certain sense, this may be the closest we have to a universal law of physics. On the other hand, it is not at all clear how this principle should be interpreted and what the action should look like, in particular in general relativity. There, the most commonly used action principle is the Einstein–Hilbert principle (see [2,3]):

$$I = \int R \, dV. \tag{1}$$

Arguably, this is the simplest choice. But there is no generally accepted deduction of (1) from fundamental principles, and it is also well known (see [4]) that there are many other action principles which can be used to generate the field equations of general relativity. In the following, (1) will be compared to another (non-standard) action principle:

$$I = \int R^2 \, dV. \tag{2}$$

In particular, it will be seen that the stationary states of (2) have a more complicated, and also more interesting, structure than those of (1), especially when it comes to properties related to rotation and spin.

Remark 1. *The word non-standard here refers to the fact that (2) does not reproduce the field equation in the usual way using the classical calculus of variation (as in the case of (1)). Only when we apply a multiple-histories perspective to general relativity, can the ordinary field equations be obtained. A complete treatment of this approach would use methods from statistical mechanics, and would also be very difficult to make rigorous. However, in the Appendix A a description of the main ideas is included (see also [5]).*

In a sense, this is all much closer to quantum mechanics than is the Einstein–Hilbert principle. It is in fact very much at the heart of this paper that the classical Einstein–Hilbert principle, even if it was a great step forwards at the time when it was introduced, may nowadays be a road blocker on our way to a unification between general relativity and quantum physics.

Remark 2. *The action principle (2) is by no means the only attempt to use classical differential geometry to bridge the gap from general relativity to quantum mechanics in general and spin in particular. Particularly interesting examples have been discussed within the framework of Einstein–Cartan theory, using an extended geometric framework in order to include, e.g., spinors. See, e.g., [6,7].*

This paper can be seen as a continuation of my paper [8]. The main theme is the same; namely, that in Lorentz geometry, as viewed from the perspective of (2), rotation is, coarsely speaking, more natural than non-rotation. In particular, the ground states of particles, i.e., the states with the lowest energy, in a certain sense seem to be rotating under rather general circumstances, even if the concept of rotation itself may not be well defined in Lorentz geometry (in particular when the Ricci curvature is non-zero). In the previous paper, it was merely argued that curvature, as measured by (2), goes down when we start to rotate the metric. In this paper, the purpose is to show how the metrics, which are actually stationary with respect to the action principle (2), can be computed. (See the Mathematica file in the Supplementary Materials).

This is considerably more precise information. But it can only be reached at the price of much heavier computations. In this paper, I have decided to work with just two very simple metrics, where everything can be computed. Nevertheless, the goal should be to prove theorems for general classes of metrics, at least as general as in [8,9] to start with. Hence, this paper should only be considered as a first attempt to attack the problem in extremely simple special cases.

2. Stationary States

In this section, I investigate the consequences of the action principles (1) and (2) in Section 1, and in particular take a closer look at their stationary metrics, which from now on will be referred to as stationary states. It would be far beyond the reach of this paper to try to analyze all such states, so I will limit myself to very simple metrics in what can loosely be described as circular orbits (see below). It should also be remembered that rotation is a very non-trivial concept in Lorentz geometry. This is especially so when the Ricci tensor is non-zero, and the simple model here is by no means the final and perfect one. At best, it can perhaps be viewed as a model which concentrates on the connection between the metric of a hypothetical orbiting particle and the surrounding space-time, but neglects the effects of the rotation on the particle itself. However, it will hopefully be sufficient to demonstrate the remarkable difference between (1) and (2).

To make this discussion more precise, let us start by considering the following simple model for a particle of radius 1 in a circular orbit of radius d, and which moves around the z-axis with velocity b.

First, consider the function

$$h_b(x, y, z, t) = 1 + \phi((x - d\cos(bt/d))^2 + (y - d\sin(bt/d))^2 + z^2), \tag{3}$$

where $d > 0$ and $\phi(s)$ is a sufficiently differentiable one-variable function with support in $[0, 1]$. Then, let

$$\mathfrak{g}(b) = h_b dx^2 + h_b dy^2 + h_b dz^2 - \frac{1}{h_b} dt^2. \tag{4}$$

Clearly, this gives a metric which coincides with the usual Minkowski metric, except close to the curve

$$\Gamma_0 : \quad (x, y, z, t) = (d\cos(bt/d), d\sin(bt/d), 0, t). \tag{5}$$

Remark 3. *The reader may wonder why this particular method of parametrization is used. This will hopefully become clearer later on (see Remark 5). In fact, it will be seen that b, with this parametrization, acquires a more or less universal value, independent of d.*

A very schematic picture of such a metric is given in Figure 1.

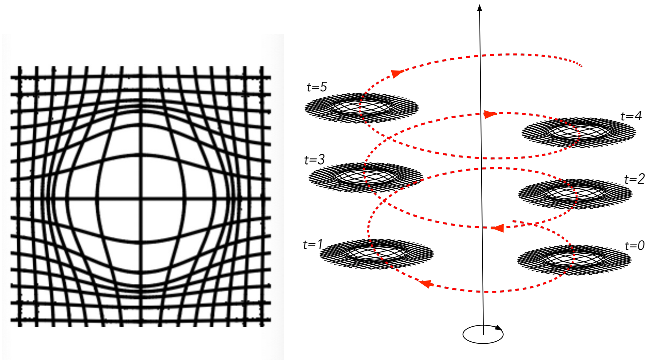

Figure 1. To the left, a very schematic illustration of a perturbation of the standard metric on Euclidean three-space (although by necessity drawn in two dimensions), which circulates with time around an axis in space (to the right).

As has already been said, the choice of this kind of metric is not based on any particular physical theory. For the moment it just serves as an example of a specific mathematical phenomenon, and the main reason for this choice is simplicity. In fact, everything we need can be computed explicitly using Mathematica, at least for sufficiently simple choices of ϕ. I will come back to the question of possible physical implications in Section 3.

As a particularly simple starting point, let us start with the specific choices

$$\phi(s) = \frac{1}{25}(s-1)^2 \qquad \text{and} \qquad \phi(s) = -\frac{1}{25}(s-1)^2 \tag{6}$$

for $0 \leq s < 1$ and $\phi(s) = 0$ otherwise. It is important to note that these choices of $\phi(s)$ are regular enough to avoid singular contributions to the curvature.

Remark 4. *It can be argued that the most interesting case is when $\phi(s)$ is, at least predominantly, non-negative. A physical motivation for this assumption is that it can be interpreted as saying that time should run more slowly in the presence of matter (compare with (3) and (4)). However, although the two cases in (6) are geometrically rather different, the resulting stationary states are remarkably similar. Here, only the case with the plus sign is presented, but the computations in the minus sign case are included in the Mathematica file of the Supplementary Materials.*

It does not make sense to compute the integrals in (1) and (2) over all of space-time, since the result would obviously be infinite. What makes sense however is to compute the time average over a given finite time interval, e.g., $I_T = [0, T]$, i.e.,

$$E_1 = \frac{1}{T}\int_{\mathbb{R}^3 \times I_T} R_{\mathfrak{g}} \sqrt{-\det(\mathfrak{g})}\, dxdydzdt, \tag{7}$$

and

$$E_2 = \frac{1}{T}\int_{\mathbb{R}^3 \times I_T} R_{\mathfrak{g}}^2 \sqrt{-\det(\mathfrak{g})}\, dxdydzdt. \tag{8}$$

In the case of the given metric $\mathfrak{g}$ above, it is easy to see, by a simple rotation invariance argument in $\mathbb{R}^3$, that if we compute these integrals only over the x, y, z-variables, then the results will be independent of t. From this, it follows that the integrals can be further simplified to

$$E_1 = \int_{\mathbb{R}^3} R_{\mathfrak{g}} \sqrt{-\det(\mathfrak{g})}\, dxdydz = \int_{D_b(t)} R_{\mathfrak{g}} h_b \, dxdydz, \tag{9}$$

and

$$E_2 = \int_{\mathbb{R}^3} R_{\mathfrak{g}}^2 \sqrt{-\det(\mathfrak{g})}\, dxdydz = \int_{D_b(t)} R_{\mathfrak{g}}^2 h_b \, dxdydz, \tag{10}$$

where

$$D_b(t) = \{(x,y,z) : (x - d\cos(bt/d))^2 + (y - d\sin(bt/d))^2 + z^2 \leq 1\}. \tag{11}$$

Here, $\sqrt{-\det(\mathfrak{g})} = h_b$ comes from (4), and the integrals are in fact independent of the choice of t.

Note that these integrals can be interpreted as the integrals (1) and (2) per unit of time. In other words, E_1 and E_2 represent action per unit of time, which means, according to the classical interpretation of action, that they represent some kind of energy. I do not claim that it is obvious what kind of energy they would represent. Nevertheless, it may be interesting to keep this interpretation in mind.

To compute integrals like these is not difficult in principle but, as is often the case when the Riemann tensor is involved, extremely time and labor consuming. In fact, almost all computations in this paper were carried out on a computer; see the Mathematica file in the Supplementary Materials.

Let us now determine the stationary points of (1) and (2) in the following sense: In principle, we want to compare the values of E_1 and E_2 along the circular orbit in (5) with their values along close-by curves, but different choices of spaces of such curves are not obviously equivalent. Here, I have chosen to work with the perhaps simplest choice, but it is worth keeping in mind that there may be better alternative formulations, although it is the opinion of the author that the main results would not be affected.

In the following, the space $\mathfrak{B}$ will be the space of all sufficiently differentiable curves which are periodic with period $2\pi d/b$, i.e., with the same period in t as $\mathfrak{g}_b$ itself.

It is important to note that every curve which is close to Γ_0 (in the sense that not only are the curves close, but also their direction vectors are close) can be parametrized by t. It follows that all small perturbations can be written in the form $(f_0(t) + \epsilon f(t), t)$, where f_0 and ϵf denote

$$f_0(t) : \begin{cases} x = & d\cos(bt/d), \\ y = & d\sin(bt/d), \\ z = & 0, \end{cases} \qquad \epsilon f(t) : \begin{cases} x = & \epsilon f_1(t), \\ y = & \epsilon f_2(t), \\ z = & \epsilon f_3(t). \end{cases} \tag{12}$$

In other words, we will use the same time parameter t for all curves involved. There is of course a case for using invariant methods, but for reasons of technical complexity in the computations, this is currently out of reach.

When computing the variations with respect to perturbations in the space $\mathfrak{B}$, the integrals can no longer be reduced to three dimensions (as in (9) and (10)), since the variations in general do not share the rotational invariance properties of the metric we start with. Hence, the integration in the time direction must also be taken into account. Keeping the periodicity in mind, this is accomplished by making T in (7) and (8) equal to the period $2\pi d/b$.

After some very long computations (see the Mathematica file in the Supplementary Materials) of the curvature, some even longer integrations, and finally some partial integrations in the t-variable (as is standard when deriving Euler–Lagrange equations), we finally arrive at the following (note that $f_3(t)$ does not contribute to the first-order term for any b, d):

$$E_1(f_0 + \epsilon f) = \tag{13}$$

$$E_1(f_0) + \epsilon \int_0^{2\pi d/b} c^*(b,d)\left(\cos\left(\frac{bt}{d}\right) f_1(t) + \sin\left(\frac{bt}{d}\right) f_2(t)\right) dt + O(\epsilon^2),$$

where

$$c^*(b,d) = \frac{b}{2\pi d}\frac{157376\pi b^2}{2165625d} = \frac{78688b^3}{2165625d^2}, \tag{14}$$

and

$$E_2(f_0 + \epsilon f) = \tag{15}$$
$$E_2(f_0) + \epsilon \int_0^{2\pi d/b} c^{**}(b,d)\left(\cos\left(\frac{bt}{d}\right) f_1(t) + \sin\left(\frac{bt}{d}\right) f_2(t)\right) dt + O(\epsilon^2),$$

where

$$c^{**}(b,d) = \frac{8b^3}{78203125d^4}\times \tag{16}$$
$$\Big[(-625625 + 625625i)d^2\Big((-10248 - 10248i) + 5\sqrt{7872271 - 9841045i}\cot^{-1}\left(\sqrt{-1-5i}\right) +$$
$$(3320 + 7125i)\sqrt{-1+5i}\cot^{-1}\left(\sqrt{-1+5i}\right)\Big) +$$
$$64b^2\Big(547 + 33d^2\Big(-3605978 + 34125\sqrt{-22006 - 4210i}\cot^{-1}\left(\sqrt{-1-5i}\right) +$$
$$34125\sqrt{-22006 + 4210i}\cot^{-1}\left(\sqrt{-1+5i}\right)\Big)\Big)\Big].$$

From the first formula (13), it is very easy to see that no orbiting metric (with $d > 0$) could ever be stationary in the Einstein–Hilbert case, since we can always choose f_1, f_2 so that the first-order term is non-zero. In fact, in this case, the only stationary solution is the straight line corresponding to $d = 0$, which is formally not included in (13). This is actually exactly what we should expect from a classical point of view.

More interesting, however, is the second formula (15). In this case, it is actually equally easy to see that there are no stationary solutions as long as $c^{**}(b,d) \neq 0$. But the point is that in this case there are non-trivial solutions in addition to the straight line. In fact, the equation

$$c^{**}(b,d) = 0 \tag{17}$$

has non-trivial real solutions for b (for positive values of d).

Due to its simple rational form, (17) can be solved exactly. As could be suspected from the equation itself, however, the expressions are quite complicated. A numerical plot of the zeros is shown in Figure 2.

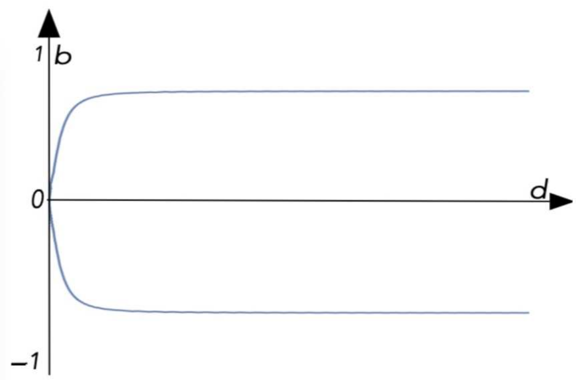

Figure 2. The solutions to the equation $c^{**}(b,d) = 0$ for $d > 0$.

Remark 5. *What Figure 2 tells us is that the speed parameter b is essentially independent of the radius d of the orbit, except for small values of this radius (as compared with the size of the 3-dimensional region where the metric deviates from the flat standard metric).*

Summarizing the above discussion, we have the following theorems:

Theorem 1. *Periodic metrics of the type* $\mathfrak{g}(b)$*, as in (3) and (4), where* $d > 0$*, are never stationary with respect to the minimizing principle (1).*

Theorem 2. *Periodic metrics of the type* $\mathfrak{g}(b)$*, as in (3) and (4), where* $d > 0$*, are stationary with respect to the minimizing principle (2) if and only if b and d satisfy the equation* $c^{**}(b, d) = 0$.

Remark 6. *A natural question to ask next is: What is the character of these stationary metrics discussed above? Preliminary computer computations of the second variations seem to show that all stationary metrics, both rotating and non-rotating, are (local) minima. However, the value of the integral in (2) tends to be smaller for the rotating ones, as compared to the non-rotating ones, indicating that the rotating ones are the natural "ground states".*

For the numerical computations, see the Mathematica file in the Supplementary Materials. There it is shown that the value of the "energy" in (10) for the non-rotating stationary metric is approximately 0.092*, whereas the corresponding value for the rotating ones (with d not too small) is* 0.051*, that is, approximately* 55% *of the value in the non-rotating case.*

3. Possible Areas of Application

Can the results in the previous section lead to new physics? First of all it should be remembered that everything proved here is based on classical differential geometry without any trace of ordinary quantum mechanics in the methodology. Such an approach could hardly be expected to give any final solutions to the fundamental problems in particle physics. Having said this, however, it is my belief that studying the implications of classical general relativity within the realm of quantum physics may be precisely what is needed today in order to achieve the grand unification, since, historically, there has been a very definite bias towards studying implications in the other direction, or even towards attempting to somehow deduce general relativity from quantum mechanics.

Keeping this in mind, I will here give two examples of situations where differential geometry may have something to add to our understanding. None of these examples should be considered as more than possible shadows of the concepts of a final theory. But still, these shadows may be easier for us to understand in a context where we have access to the machinery of classical differential geometry.

The first example of a situation in which the stationary states in this paper may have relevance comes from string theory. A possible (classical) model for an open string, which has been discussed in more detail by the author in [8,9], may look as follows: We simply define a string by the region in space which it occupies. And for an open string, the simplest such region is the convex hull of two spheres.

In particular, studying such rotating strings, where mass is concentrated in the two spheres at the ends, can mathematically be viewed as considering two orbiting metrics, as in Section 2, rotating around a common axis (see Figure 3). In this case, the mathematics used to derive the stationary metrics in Section 2 carries over essentially without any changes at all, at least if we assume that the string is long enough so that the metrical disturbances at the ends do not interfere with each other. This is, for example, the situation for the schematic rotating string in Figure 3. And in general, strings in string theory are usually thought of as being thin objects, or perhaps even as being simply one-dimensional.

Summarizing, the stationary states of the classical strings as described here with the least energy (as measured by the integral in (10)) are rotating. This, in a sense, mimics what happens in usual string theory. But in the present paper, the underlying mechanism is purely classical and is based on the concept of curvature.

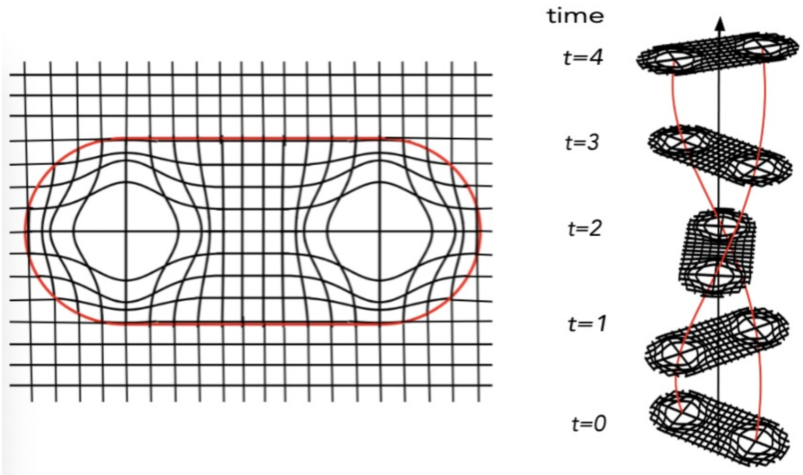

Figure 3. A schematic illustration of a classical string (to the left). To the right, the same string rotating around an axis in space (not in time).

This kind of classical strings does not seem very controversial from the point of view of Newton's first law: Their main property of interest consists in the observation that the states of such strings with the lowest energy are rotating. But if we consider their centers of mass, these are still expected to move along straight lines (or more generally, geodesics), just as in the classical theory.

But what happens if we just consider the stationary states in Section 3 themselves, without pairing them together to strings as above? Clearly, they would, if interpreted as free particles traveling in space-time, be traveling in closed orbits or along helices. Such behavior is certainly not what we expect from free particles.

There could however be exceptions: I would actually suggest thinking about such stationary states, as in Section 2, as candidates for a kind of classical theory of photons. It is not my intention here to advance some new speculative theory about what a photon actually is. I only want to draw attention to the fact that the minimizing principle in Section 2 can actually give rise to something with both wave-like and particle-like properties (see Figure 4).

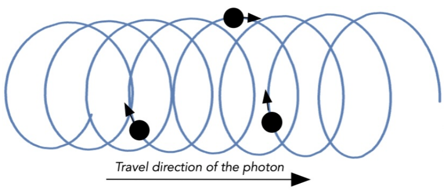

Figure 4. A schematic illustration of such a classical photon, as in the text.

Obviously, the kind of classical stationary states considered here may be far from the usual quantum mechanical picture, and it is by no means obvious how to bridge this gap at this point. Nevertheless, the dichotomy between the two aspects of light, as waves and as particles, has been one of the defining factors in the development of modern quantum physics, most clearly manifested in Bohr's complementarity principle (see [10]). So perhaps it could be fruitful for our understanding to study this kind of two-sidedness in a completely classical situation, where we have full access to our geometric intuition? There have been many situations in the history of physics where the key to new quantum phenomena has been to start from a corresponding classical analogue.

4. Discussion and Conclusions

What could the calculations in this paper possibly have to say about physics in general and the unification of general relativity and quantum physics in particular? Although the metrics used in Section 2 should not be considered to represent any kind of realistic physical model, the fact that rotating metrics appear to generate less scalar curvature than non-rotating ones seems to be a very general one. This has previously been investigated by the author in [8,9], but from a slightly different point of view. It is the author's belief that the kind of structure of the stationary states that has been studied in this paper may be a very general phenomenon. But so far, the only firm information that we have is based on very heavy computer computations, which can only be carried out for very simple examples.

The problem with the grand unification is the problem of building a bridge between two very different conceptual frameworks: General relativistic concepts, like curvature, have no natural interpretation in quantum mechanics; and likewise, quantum mechanical concepts, like wave functions, have no natural interpretation in general relativity. It is the belief of the author that the best road towards a unification is to try to study concepts which have meaningful interpretations in both worlds.

This paper is part of an attempt to show that the idea of rotation/spin could be such a concept. Although it is of course well known that quantum mechanical spin is something different from classical rotation, it is equally well known that there must be some kind of connection. The way spin was discovered, essentially through the work of Pauli [11] and Dirac [12], shows that it is an indispensable part of quantum mechanics. On the other hand, this approach, ingenious as it was, gives no hint to what really underlies the concept.

What the results in this paper are intended to show is that a conceptual reformulation of general relativity could give us such a hint to how we may investigate this problem on a deeper level.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/sym16121694/s1, In fact, almost all computations in this paper has been carried out using Mathematica on an ordinary mac. The corresponding Mathematica file can be downloaded and run on any computer running a sufficiently modern version of Mathematica. For the particular method for computing curvature on a computer, the reader is referred to [13].

Funding: This research received no external funding.

Data Availability Statement: Data is contained within the article or Supplementary Materials.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A. Multiple Histories as a Possible Foundation for General Relativity

This approach has emerged from a very simple way of treating multiple histories, with roots in classical statistical mechanics. It is not at all claimed that this approach covers all or even the most central problems of multiple-history theory (which obviously need a quantum mechanical treatment), but it may still be that some problems can most easily be understood using classical probabilities. In any case, making the following heuristic motivation of the field equations in the case of the action principle based on (2) into completely rigorous mathematics is a huge project. Here, it is included only to show the possibilities of this approach, and it should be kept in mind that certainly a lot of work remains.

The starting point is to consider all possible geometries, defined by different metrics, in some region. Together, these are said to constitute an ensemble. Although there is an enormous number of such geometries on a microscopic scale, it can very well be that on a macroscopic scale, under suitable circumstances, there is only one geometry (or perhaps a few) that has a non-negligible probability of occurring. This is perfectly analogous to what happens, e.g., in the classical theory of gases (see [14]).

So, how do we define the ensemble of all such metrics g in a given macroscopic region U in 4-space, subject to certain boundary conditions and, in addition, to the condition that the four-volume of U with respect to g is a fixed number? In other words, how do we assign probabilities to all these geometries? From a general relativistic point of view, it seems natural to assume that these probabilities should depend on the curvature, and by far the most natural real measure of curvature is given by the scalar curvature R.

But even if we restrict ourselves to probabilities which just depend on R, there are still many ways to assign a statistical weight to each geometry, and we know very little about how this should be performed: When looking at the geometry at shorter and shorter distances, the geometric fluctuations are expected to become larger and larger for quantum mechanical reasons, but that is essentially as much as we can say.

Fortunately, however, it may not be necessary to know all the details about these probabilities. In fact, assume more or less any probability distribution (with mean-value zero) for the total scalar curvature

$$\int_{U_\alpha} R\, dV, \tag{A1}$$

on a generic small set U_α, and assume also that the probability distributions on disjoint sets behave more or less as independent stochastic variables. Using the central limit theorem (see Fischer [15]), and the additivity of the integral in (A1) in disjoint regions, we can conclude that the statistical weight of a larger set U, which can be written as a disjoint union $U = \cup_\alpha U_\alpha$, should behave as

$$\sim \exp\{-\mu \int_U R^2\, dV\}, \qquad \text{for some positive constant } \mu. \tag{A2}$$

(see Tamm 2021 [5] for more details). Observe that the scalar curvature R in this setting is not something which is defined at each point from the beginning. Rather, it should be viewed as a kind of average:

$$R = \frac{1}{m(U)} \int_U R'\, dV, \qquad (\text{where } m(U) \text{ denotes space-time volume}). \tag{A3}$$

Note the perhaps somewhat confusing notation: Strictly speaking, R and R' are both measures of curvature, but are defined on different length scales. Only on a macroscopic scale do these averages converge to the smooth classical function we are used to, and what happens at very short distances, where the concept of curvature is bound to lose its meaning, is something which we know very little about.

It is also worth recalling that ensembles in statistical mechanics are notoriously difficult to define and treat in a rigorous way. But on the other hand, they are usually very insensitive to the exact details of these definitions. Instead of going deeper into the nature of the ensemble, I will here try to explain how this is all connected to the action

$$L = \int_U R^2\, dV, \tag{A4}$$

and why the macroscopic metrics which dominate the ensemble should satisfy the field equations in vacuum, $R_{ij} = 0$.

A standard procedure in statistical mechanics (see [14]) looks as follows: First, compute the "state sum" (summing over all possible metrics):

$$\Xi = \sum_g \exp\{-\mu \int_U R_g^2\, dV\}. \tag{A5}$$

The negative of the logarithm of the state sum, $L = -\log \Xi$, gives what is usually called the "Helmholtz free energy". Theory now tells us that the macrostates which minimize L (among all states with a given volume) are the by far most probable ones, i.e., the ones which are observed in nature.

Remark A1. *Note that we are here working in four dimensions, not in three as in the usual theory. In particular, the Helmholtz free energy here is not directly related to ordinary energy in the usual sense.*

To compute the state sum and the free energy exactly can be a huge task. However, usually the sum is dominated by a single term (or at most a few), together with the corresponding "density of states Ω_g", i.e., the number of close-by metrics with more or less the same value of the weight function:

$$\Xi \sim \exp\{-\mu \int_U R_g^2 \, dV\} \cdot \Omega_g, \tag{A6}$$

$$L = -\log \Xi \approx \mu \int_U R_g^2 \, dV - \log \Omega_g. \tag{A7}$$

Finding the states which minimize the free energy can be very difficult, since they are determined by a complicated interplay between the sizes of the two terms in (A7) above. It is in fact a very common situation in statistical mechanics that the two terms in (A7) compete with each other: Sometimes the states with the largest weight win, and sometimes states with lower probability win (because there are many more of them). Compare, e.g., with the Ising model ([14,16]).

A possible method to minimize L could be to try to compute and solve the Euler–Lagrange equation, but in this geometric situation, this would be a very difficult problem.

However, the case of the vacuum equations is a kind of exception: *If* there is a solution to $R_{ij} = 0$ in the region U, which furthermore fulfills the given boundary conditions, then it can be argued that this solution should also minimize L. And, conversely, that a metric which minimizes L should satisfy $R_{ij} = 0$.

In fact, in this case we can try to simultaneously minimize both terms in (A7) by using methods from statistical mechanics instead of the Euler–Lagrange equation.

In the present situation, it is obvious that a metric with $R_{ij} = 0$ will also have $R = 0$, which of course minimizes the R^2 integral. But it is also clear that there are many other metrics with $R = 0$ with the same property. So the question is: Why will the second term $-\log \Omega$ be smaller ($\log \Omega$ will be bigger) for metrics with $R_{ij} = 0$?

The reason for this is a subtle one, and the answer that statistical mechanics gives is that it has to do with the character of the minima: Given a metric g with $R_g = 0$ and an $\epsilon > 0$, to compute the density of states Ω_g basically means to figure out how many variations δg there are such that $R^2_{g+\delta g} < \epsilon$. In addition, it must be noted that in the macroscopic case, it is really the number of such variations when $\epsilon \to 0$ which matters. And it is here that higher-order terms become important, not just the first-order terms as in the Euler–Lagrange equation: A minimum where the second variation vanishes is much flatter than a minimum where it does not, which means that in the first case there will be many more variations contributing to Ω_g.

In our case, we obtain for the orders of magnitude of the variations (writing R instead of R_g):

$$\text{if} \quad R_{ij} \neq 0, \quad \text{then} \quad \delta R \sim |\delta g| \quad \text{and} \quad \delta(R^2) \sim |\delta g|^2, \tag{A8}$$

$$\text{if} \quad R_{ij} = 0, \quad \text{then} \quad \delta R \sim |\delta g|^2 \quad \text{and} \quad \delta(R^2) \sim |\delta g|^4. \tag{A9}$$

The difference between the second-order terms in (A8) and the fourth-order terms in (A9) is exactly what makes metrics with $R_{ij} = 0$ overwhelmingly more probable than other metrics with just $R = 0$.

The first statement (A8) is a rather trivial one: Its essence is just the same as the fact from elementary analysis that if $\gamma(s)$ is differentiable and $\gamma(0) = 0$, then for small Δs, $\gamma(\Delta s) \sim \Delta s$, and $\gamma^2(\Delta s) \sim (\Delta s)^2$. As for the second statement (A9), the corresponding idea could similarly be expressed by saying that if $\gamma(0) = \gamma'(0) = 0$, then for small Δs, $\gamma(\Delta s) \sim (\Delta s)^2$, and $\gamma^2(\Delta s) \sim (\Delta s)^4$. But the problem here is that for this comparison to make sense, we somehow need to identify the derivative of R in the direction δg_{ij} with the component R_{ij} of the Ricci tensor. As it turns out, this is in a sense exactly what can be achieved, but to establish this demands non-trivial differential geometry. Below is included an argument, but it is a short and non-technical explanation, rather than a technically complete proof.

Writing h and h_{ij} for δg and δg_{ij}, we can compute the s-derivative of the scalar curvature $R(s)$ along the differentiable, volume-preserving one-parameter family $g_s = g_0 + s \cdot h + \dots$ of metrics passing through a given extremal metric $g = g_0$. The result is that at $s = 0$,

$$\frac{dR}{ds} = -\sum_{i,j} h^{ij} R_{ij} + \text{divergence terms,} \quad \text{where} \quad h^{ij} = \sum_{k,l} g^{ik} g^{jl} h_{kl}. \tag{A10}$$

(For this calculation, see, e.g., [17].)

Remembering from (A3) that the macroscopic curvature R should be interpreted as a mean value, we obtain for a volume-preserving microscopic variation h with support in U that the divergence terms disappear (by the divergence theorem), and that we are left with

$$\frac{d}{ds} R = -\frac{1}{m(U)} \int_U \left(\sum_{i,j} R_{ij} h^{ij} \right) dV \approx -\sum_{i,j} R_{ij} \frac{1}{m(U)} \int_U h^{ij} \, dV. \tag{A11}$$

Here, we use that the macroscopic R_{ij}s vary very slowly on the microscopic set U. Also, the calculation above must include the contribution from the variation in dV, which, however, vanishes because of the condition of fixed volume. From this formula it is clear that if $R_{ij} = 0$, then R is flat, in the sense that the left-hand side in (A11) vanishes in all directions.

If, on the other hand, the left-hand side in (A11) vanishes in all directions, then it is a standard exercise in linear algebra to check that $R_{ij} = 0$, using the fact that the h_{ij}s can be chosen arbitrarily, only subject to the condition that the variation is volume-preserving, which can be expressed as follows (see [17]):

$$\sum_{i,j} g_{ij} \int_U h^{ij} \, dV = 0. \tag{A12}$$

In fact, the only case in which the R_{ij}s can be chosen $\neq 0$ so that the right-hand side of (A11) is zero for all h_{ij} satisfying (A12), is to let $R_{ij} = k \cdot g_{ij}$ for some $k \neq 0$. But this is not compatible with the condition that R, which is the trace of R_{ij}, is zero.

Summing up, $-\log \Omega_g$ should then be minimized precisely when the vacuum field equations are fulfilled.

References

1. Newton, I. *Philosophiae Naturalis Principia Mathematica*; Royal Society of London: London, UK, 1687.
2. Hilbert, D. Die Grundlagen der Physik. (Erste Mitteilung.) Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen. *Math.-Phys. Kl.* **1915**, *1915*, 395–407.
3. Palatini, A. Deduzione invariantiva delle equazioni gravitazionali dal principio di Hamilton. *Rend. Circ. Mat. Palermo* **1919**, *43*, 203–212. [CrossRef]
4. Eddington, A.S. *Relativitätstheorie in Mathematischer Behandlung*; Springer: Berlin/Heidelberg, Germany, 1925.
5. Tamm, M. Natural Lagrangians. *Universe* **2021**, *7*, 74. [CrossRef]
6. Trautman, A. *Einstein Cartan Theory, Encyclopedia of Mathematical Physics*; Francoise, J.-P., Naber, G.L., Tsou S.T., Eds.; Elsevier: Oxford, UK, 2006; Volume 2.
7. Battista, E.; De Falco, V.; Usseglio, D. First post-Newtonian N-body problem in Einstein-Cartan theory with the Weyssenhoff fluid: Lagrangian and first integrals. *Eur. Phys. J.* **2023**, *83*, 112. [CrossRef] [PubMed]
8. Tamm, M. Different Aspects of Spin in Quantum Mechanics and General Relativity. *Symmetry* **2023**, *15*, 2016. 10.3390/sym15112016. [CrossRef]
9. Tamm, M. Minimizing Curvature in Euclidean and Lorentz Geometry. *Symmetry* **2021**, *13*, 1433. 1433. [CrossRef]
10. Significance of Complementarity in Physics: Dialectica. *Nature* **1949**, *163*, 435. [CrossRef]
11. Pauli, W. Zur Quantenmechanik des magnetischen Elektrons. *Z. Phys.* **1927**, *43*, 601–632. [CrossRef]
12. Dirac, P.M. The quantum theory of the electron. *Proc. R. Soc. Lond. A* **1928**, *117*, 610–624. [CrossRef]

13. Parker, L.; Christensen, S.M. *MathTensor: A System for Doing Tensor Analysis by Computer*; Addison-Wesley: Boston, MA, USA, 1994.
14. Huang, K. *Statistical Mechanics*, 2nd ed.; John Wiley & Sons, Inc.: New York, NY, USA, 1987.
15. Fischer, H. A *History of the Central Limit Theorem: From Classical to Modern Probability Theory, Sources and Studies in the History of Mathematics and Physical Sciences*; Springer: New York, NY, USA, 2011.
16. Onsager, L. Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition. *Phys. Rev.* **1944**, *65*, 117. [CrossRef]
17. Schoen, R.M. Variational theory for the total scalar curvature functional for riemannian metrics and related topics. In *Topics in Calculus of Variations*; Giaquinta, M., Ed.; Lecture Notes in Mathematics; Springer: Berlin/Heidelberg, Germany, 1989; Volume 1365. [CrossRef]