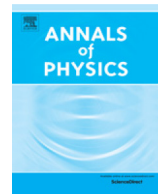




Contents lists available at ScienceDirect

Annals of Physics

journal homepage: www.elsevier.com/locate/aop



CrossMark

Logical inference approach to relativistic quantum mechanics: Derivation of the Klein–Gordon equation

H.C. Donker^{a,*}, M.I. Katsnelson^a, H. De Raedt^b, K. Michielsens^c

^a Radboud University, Institute for Molecules and Materials, Heyendaalseweg 135, NL-6525AJ Nijmegen, The Netherlands

^b Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, NL-9747 AG Groningen, The Netherlands

^c Institute for Advanced Simulation, Jülich Supercomputing Centre, Forschungszentrum Jülich, D-52425 Jülich, Germany

HIGHLIGHTS

- Logical inference applied to relativistic, massive, charged, and spinless particle experiments leads to the Klein–Gordon equation.
- The relativistic Hamilton–Jacobi is scrutinized by employing a field description for the four-velocity.
- Logical inference allows analysis of experiments with uncertainty in detection events and experimental conditions.

ARTICLE INFO

Article history:

Received 8 December 2015

Accepted 22 April 2016

Available online 27 April 2016

Keywords:

Logical inference

Quantum theory

Klein–Gordon equation

ABSTRACT

The logical inference approach to quantum theory, proposed earlier De Raedt et al. (2014), is considered in a relativistic setting. It is shown that the Klein–Gordon equation for a massive, charged, and spinless particle derives from the combination of the requirements that the space–time data collected by probing the particle is obtained from the most robust experiment and that on average, the classical relativistic equation of motion of a particle holds.

© 2016 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

* Correspondence to: Radboud University, Institute for Molecules and Materials, Theory of Condensed Matter, P.O. box 9010/60, 6500 GL Nijmegen, The Netherlands.

E-mail address: h.donker@science.ru.nl (H.C. Donker).

<http://dx.doi.org/10.1016/j.aop.2016.04.018>

0003-4916/© 2016 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The inception of quantum theory was one of taking leaps. This is illustrated by e.g. Schrödinger's paper [1] in which he proposes his celebrated wave equation. In this article [1] Hamilton's principal function S is postulated to take the form $S = k \ln \psi$ with k a constant and is then substituted in the Hamilton–Jacobi equation (HJE). Upon variation of the resulting quadratic functional with respect to ψ (which Schrödinger later justifies using Huygens' principle [2]) an equation linear in ψ , now known as the Schrödinger equation, was obtained. The derivation of the Klein–Gordon equation [3–7] is essentially identical to that of the Schrödinger equation namely, an action Ansatz is substituted in the relativistic Hamilton–Jacobi equation, and after variation of the resulting quadratic functional with respect to ψ , the relativistic analogue of the Schrödinger equation is obtained [3–7].

Because of the ad hoc assumptions involved in obtaining these equations, standard quantum mechanics textbooks usually present the formalism of quantum theory as a set of postulates (see e.g. Refs. [8–11]) and considerable activity focuses on eliminating some of these postulates [12–19]. Instead of starting from a set of postulates, the current work presents an alternative derivation of the relativistic wave equation based on the principles of logical inference (LI) [20–23]. Specifically, we demonstrate how the Klein–Gordon equation for a massive, charged and spinless particle follows from LI based on the analysis of data recorded by a detector, thereby extending earlier work [24–27] to the relativistic domain.

The key concept in LI is the plausibility [23], a mental construct which quantifies e.g. the chance that a detection event occurs. In general, the degree of plausibility is expressed by a real number in the range of 0 and 1 [23]. The algebra of LI facilitates plausible reasoning in the presence of uncertainty in a mathematically well-defined manner [20–23]. In real experiments there is not only uncertainty about the individual detection events but there obviously is also uncertainty in the conditions under which the experiments are carried out. Inevitably, the conditions of the experiment will vary whenever the experiment is repeated. But if the experimental data is to be reproducible, the experiment must be robust (to be quantified later) with respect to small changes in the conditions under which the experiment is being performed. Earlier work has shown that the equations of non-relativistic quantum theory can be obtained by analyzing such robust experiments [24–27]; most notable are the Schrödinger [25] and the Pauli equation [26]. Importantly, the requirement that the experiment is to be robust implies that the plausibility must be viewed as an objective assignment (i.e. conditional probability) rather than a subjective one [25]. The present work extends this approach to the relativistic domain: it shows how the Klein–Gordon equation [3,7] for a massive, charged, and spinless relativistic particle emerges by an analysis similar to the one employed in Refs. [25,26].

2. Logical inference approach

2.1. Particle detection experiment

Consider an experiment in which a particle source and detectors are located at fixed positions relative to the laboratory reference frame. The source emits a particle that interacts with one of the detectors and triggers a detection event that yields data in the form of three spatial coordinates $\mathbf{r} = (x, y, z)$ of the detector and the clock time t at which the event occurred. The experiment is considered to be ideal in the sense that every emitted particle triggers one and only one detector.

The experiment is repeated N times, meaning that we let N particles pass through the detector. Each time a particle is created, the (laboratory) clock time is reset. We label the particles and the corresponding data by the index $n = 1 \dots N$ and denote the spatial and temporal resolution by Δ_s and Δ_t , respectively. As particle n passes through the detector, the latter produces a time stamp t_n and a vector of spatial coordinates $\mathbf{r}_n = (x_n, y_n, z_n)$, which because of the limited resolution, correspond to the time-bin $j_n = \text{ceiling}(t_n/\Delta_t)$ and space-bin $\mathbf{k}_n = \text{ceiling}(\mathbf{r}_n/\Delta_s)$ where, element-wise, the function $\text{ceiling}(x)$ returns the smallest integer not smaller than x . In practice the number of time-bins and space-bins is necessarily finite. Therefore we must have $0 \leq j_n \leq J$ and $(0, 0, 0) \leq \mathbf{k}_n \leq \mathbf{K} = (K_x, K_y, K_z)$, where J, K_x, K_y and K_z are (large) integer numbers.

The data collected after N repetitions of the experiment is given by the set of quadruples

$$\mathcal{Y} = \{(\mathbf{j}_n, \mathbf{k}_n) \mid 0 \leq j_n \leq J; \mathbf{0} \leq \mathbf{k}_n \leq \mathbf{K}; n = 1 \dots N\}, \quad (1)$$

or, denoting the total amount of clicks in bin $\mathbf{j} = (j_n, \mathbf{k}_n)$ by c_j , by the equivalent data set

$$\mathcal{D} = \left\{ c_j \mid \sum_j c_j = N \right\}. \quad (2)$$

Note that at this stage, we have not yet assumed that there is a relation between the space–time coordinates of the particle and the data set \mathcal{D} .

2.2. Inference probability and Fisher information

Having specified the measurement scenario, the next step in the LI approach is to encode the relation between the space–time coordinates of the particle and the n th detection event $\mathbf{j}_n = (j_n, \mathbf{k}_n)$ through the inference-probability (i-prob) $P(\mathbf{j}_n|\theta, Z)$ where θ and Z specify the conditions under which the experiment is being performed [25]. The i-prob is, at this stage, a necessarily subjective number between zero and one that expresses the uncertainty with which the n th particle produces the data \mathbf{j}_n . The particle is assumed to be characterized by its own (unknown) clock time θ measured in a reference frame attached to the particle. The proposition Z represents all other experimental conditions (e.g. applied electromagnetic potentials) which are considered fixed for the duration of the experiment but are deemed irrelevant for the problem at hand.

It is common practice to assume, as a first step, that events are independent, meaning that knowing all earlier and future events, it is impossible to say with certainty what the event will be. Following this practice, we assume that the N detection events are independent. Then, according to the algebra of LI [20–23], it follows immediately that the i-prob $P(\mathcal{Y}|\theta, N, Z)$ to observe data set \mathcal{Y} factorizes as

$$P(\mathcal{Y}|\theta, N, Z) = \prod_{n=1}^N P(\mathbf{j}_n|\theta, Z), \quad (3)$$

or, equivalently,

$$P(\mathcal{D}|\theta, N, Z) = N! \prod_j \frac{P(\mathbf{j}|\theta, Z)^{c_j}}{c_j!}. \quad (4)$$

The salient feature of the experiment considered here is that there is uncertainty about the individual detection events, that there is uncertainty in the mapping from θ to the spatial coordinates and the time of the detection events. However, if the experimental data is to increase our capability to uncover relations among the observed events at different space–time points, the experiment must be robust [25]. In the case at hand this means that small changes in the unknown clock time θ do not lead to erratic changes in the observed data \mathcal{D} , even though there is no reproducibility on the level of individual events.

It is convenient to express the requirement of robustness as an hypothesis test [25]. The evidence [22,23] Ev for the hypothesis that $\theta + \epsilon$ produces the data \mathcal{D} relative to the hypothesis that θ produces the same data is given by [22,23,25]

$$\text{Ev} = \ln \frac{P(\mathcal{D}|\theta + \epsilon, N, Z)}{P(\mathcal{D}|\theta, N, Z)}. \quad (5)$$

The notion of a robust experiment then translates to the statement that for all θ and arbitrary but small ϵ , the evidence $|\text{Ev}|$ should be as small as possible. In searching for the solution of the global optimization problem, we exclude the trivial, non-informative experiment for which $P(\mathcal{D}|\theta, N, Z)$ does not depend on θ [25]. Making use of Eq. (4) and expanding Eq. (5) to second order in ϵ yields

$$\text{Ev} = \sum_j c_j \ln \frac{P(\mathbf{j}|\theta + \epsilon, Z)}{P(\mathbf{j}|\theta, Z)} = \sum_j c_j \left\{ \epsilon \frac{P'(\mathbf{j}|\theta, Z)}{P(\mathbf{j}|\theta, Z)} - \frac{\epsilon^2}{2} \left[\left(\frac{P'(\mathbf{j}|\theta, Z)}{P(\mathbf{j}|\theta, Z)} \right)^2 - \frac{P''(\mathbf{j}|\theta, Z)}{P(\mathbf{j}|\theta, Z)} \right] \right\}, \quad (6)$$

where the primes indicate partial derivatives with respect to θ .

Our goal is now to minimize $|\text{Ev}|$ for all θ simultaneously. First note that as $\sum_j P(\mathbf{j}|\theta, Z) = 1$, all partial derivatives of $\sum_j P(\mathbf{j}|\theta, Z)$ with respect to θ are zero. Therefore the first and the third term in Eq. (6) vanish if we make the assignment $c_j = NP(\mathbf{j}|\theta, Z)$. This is an important result: the criterion of robustness not only enforces the intuitively obvious assignment $P(\mathbf{j}|\theta, Z) = c_j/N$ but by doing so, it changes the subjective nature of $P(\mathbf{j}|\theta, Z)$ into an objective, physically measurable quantity (the relative frequency of outcomes). Thus, it is at this point that the possibility to view the i-prob as a subjective assignment is eliminated [25].

With this assignment, the expression for the evidence becomes

$$\text{Ev} = -\frac{\epsilon^2 N}{2} \sum_j \frac{1}{P(\mathbf{j}|\theta, Z)} \left(\frac{\partial P(\mathbf{j}|\theta, Z)}{\partial \theta} \right)^2, \quad (7)$$

and as ϵ is arbitrary, we can find the solution of the optimization problem by minimizing the Fisher information

$$I_F = \sum_j \frac{1}{P(\mathbf{j}|\theta, Z)} \left(\frac{\partial P(\mathbf{j}|\theta, Z)}{\partial \theta} \right)^2, \quad (8)$$

for all θ simultaneously.

The basic equations of (relativistic) quantum theory are formulated in terms of continuous space and time. Therefore, to derive such equations from a LI approach, it is necessary to take the continuum limit of Eq. (8). This is readily accomplished in the standard manner by letting the temporal resolution Δ_t and spatial resolution Δ_s approach zero while keeping the four dimensions of the four-dimensional volume fixed. Taking the continuum limit and ignoring irrelevant prefactors, Eq. (8) becomes

$$I_F = c \int d^3\mathbf{r} \int dt \frac{1}{P(t, \mathbf{r}|\theta, Z)} \left(\frac{\partial P(t, \mathbf{r}|\theta, Z)}{\partial \theta} \right)^2 \equiv \int d^4\mathbf{x} \frac{1}{P(\mathbf{x}|\theta, Z)} \left(\frac{\partial P(\mathbf{x}|\theta, Z)}{\partial \theta} \right)^2, \quad (9)$$

where $\mathbf{x} = [ct, x, y, z] = [x^0, x^1, x^2, x^3]$ denotes the four-vector of a location in space–time and c is the speed of light in vacuum. Strictly speaking, Eq. (9) makes a slight abuse of notation: in the continuum limit $P(\mathbf{x}|\theta, Z)$ is a probability density whilst $P(\mathbf{x}|\theta, Z)d\mathbf{x}$ is the corresponding (dimensionless) i-prob. Henceforth it is assumed that this change of notation is implicitly understood.

2.3. Special relativity

The above discussion focused on the relation between a robust experiment and the observed data but does not refer to any physical theory yet. The knowledge or expectation about the physical behavior enters the LI approach by imposing constraints on the minimization of Eq. (9). Generally speaking, in the absence of uncertainty, we may expect to observe data that complies with the classical mechanical description. Thus, in the case at hand, we require that in the absence of uncertainty, the LI approach yields the results of the special theory of relativity (STR).

Proper time, that is the time measured by a clock at rest, is a central notion in the STR. In the measurement scenario described above, the i-prob $P(\mathbf{x}|\theta, Z)d\mathbf{x}$ was already assumed to depend on the proper time θ of the particle. In the spirit of the STR, we assume that

$$P(\mathbf{x}|\theta, Z) = P(\tau|\theta, Z), \quad (10)$$

where τ ($c^2\tau^2 \equiv c^2t^2 - x^2 - y^2 - z^2$) is the proper time of the detection event. In words, we assume that the i-prob to observe a space–time event depends on Lorentz scalars only. Note that because the clock is being reset with each repetition of the experiment, the proper times that enter our description are proper time intervals. In addition, we assume that space–time is homogeneous, meaning that

$$P(\tau + \delta|\theta + \delta, Z) = P(\tau|\theta, Z), \quad (11)$$

where δ is an arbitrary shift of proper time. A remark about the last assumption may be in order. In the LI approach the measurement scenario and the observed data are key to formulate the notion of

a robust experiment. Physics on the other hand is about connecting mental pictures, concepts about the nature of the world around us, to the observed data. Viewed in this light, Eq. (11) expresses our expectation that carrying out the experiment at another point in space–time does not change our mental picture. From the definition of the derivative, it follows directly from Eq. (11) that

$$\frac{\partial P(\tau|\theta, Z)}{\partial \tau} = -\frac{\partial P(\tau|\theta, Z)}{\partial \theta}, \quad (12)$$

and using this identity, Eq. (9) becomes

$$I_F = \int d^4\mathbf{x} \frac{1}{P(\mathbf{x}|\theta, Z)} \left(\frac{\partial P(\mathbf{x}|\theta, Z)}{\partial \tau} \right)^2 = \int d^4\mathbf{x} \frac{1}{P(\tau|\theta, Z)} \left(\frac{\partial P(\tau|\theta, Z)}{\partial \tau} \right)^2. \quad (13)$$

Recall that the objective of the LI approach is to find $P(\mathbf{x}|\theta, Z)$ that minimizes I_F for all θ simultaneously, subject to constraints that we discuss next.

2.4. Motion of the particle

In the absence of uncertainty, successive observed detection events map one-to-one on the relativistic motion of the particle, described by the laws of the STR. In the LI approach, this limiting case enters through a “correspondence principle” in terms of the HJE [25,26]. This is not a surprise: as mentioned in the introduction, the HJE is one of the key ingredients in the derivation of the Schrödinger [1] and the Klein–Gordon equation [3–7,28] and it plays a similar role in the LI derivation of these equations. In the present paper, we do not postulate a HJE but, in analogy with the derivation of the non-relativistic HJE [29,26], we follow an alternative path and derive the relativistic HJE for a massive and charged particle from a field description of the four-velocity $d\mathbf{x}/d\tau$.

We start by assuming that there exists a (four-)vector field $\mathbf{U}(\mathbf{x})$ such that

$$\frac{d\mathbf{x}^\mu}{d\tau} = -U^\mu(\mathbf{x}). \quad (14)$$

Here and in the following, we use the standard co/contra-variant notation and the summation convention and denote the Minkowski metric by $\eta = \text{diag}(1, -1, -1, -1)$. Taking the derivative of Eq. (14) with respect to τ yields

$$\frac{d^2\mathbf{x}^\mu}{d\tau^2} = -\frac{\partial U^\mu}{\partial(ct)} \frac{d(ct)}{d\tau} - \sum_{i=1}^3 \frac{\partial U^\mu}{\partial x^i} \frac{dx^i}{d\tau} = -\partial^\nu U^\mu \frac{dx_\nu}{d\tau}, \quad (15)$$

where ∂^μ is the shorthand for $\partial/\partial x_\mu$. As the norm of the four-velocity is c , we have $\mathbf{U}^2 \equiv U^\alpha U_\alpha = c^2$ is a constant and hence any derivative thereof is zero. Therefore we have

$$\partial^\mu \mathbf{U}^2 = 2 \left(U^0 \frac{\partial U^0}{\partial x_\mu} - \sum_{i=1}^3 U^i \frac{\partial U^i}{\partial x_\mu} \right) = 2U^\nu \partial^\mu U_\nu = 0. \quad (16)$$

Substitution of Eq. (14) into Eq. (16) yields

$$(\partial^\mu U^\nu) \frac{dx_\nu}{d\tau} = 0. \quad (17)$$

From Eqs. (15) and (17) it then follows that

$$\frac{d^2\mathbf{x}^\mu}{d\tau^2} = (\partial^\mu U^\nu - \partial^\nu U^\mu) \frac{dx_\nu}{d\tau} = F^{\mu\nu} \frac{dx_\nu}{d\tau}. \quad (18)$$

Eq. (18) has the same form as the Lorentz force equation of a particle moving in an electrodynamic field [30] if we identify $F^{\mu\nu}$ with the field-strength tensor of electrodynamics. In order that this identification makes sense, it is necessary to assume that the particles in the particle detection experiment are massive and charged.

If $S = S(\mathbf{x})$ represents a scalar field, the transformation $A^\mu = U^\mu + \partial^\mu S$ yields

$$(\partial S - \mathbf{A})^2 = c^2 \quad (19)$$

where we introduced the shorthand notation $(\partial S)^2 = (\partial_\alpha S)(\partial^\alpha S)$. As it is the four-velocity $d\mathbf{x}/d\tau$ which corresponds to a physically relevant quantity, imposing gauge invariance enforces introducing a non-vanishing canonical momentum $p^\mu = \partial^\mu S$ in order to keep the norm of the four-velocity fixed to c . Eq. (19) is the relativistic HJE in disguise. Indeed, making use of $\partial S = [\partial S/\partial x_0, \partial S/\partial x_1, \partial S/\partial x_2, \partial S/\partial x_3] = [\partial S/\partial x^0, -\partial S/\partial x^1, -\partial S/\partial x^2, -\partial S/\partial x^3]$ and $\mathbf{A} = [A^0, A^1, A^2, A^3] = [A_0, -A_1, -A_2, -A_3]$, introducing the symbols m and q for the mass and charge of the particle, respectively, and changing in Eq. (19) symbols according to $\partial S \rightarrow (\partial S/\partial ct, \partial S/\partial x, \partial S/\partial y, \partial S/\partial z)/m$ and $\mathbf{A} \rightarrow q(\Phi, A_x, A_y, A_z)/m$ (where Φ and (A_x, A_y, A_z) are the usual scalar and vector potential, respectively [30]), we find

$$\begin{aligned} \left(\frac{\partial S(\mathbf{x})}{\partial ct} - \frac{q}{c} \Phi(\mathbf{x}) \right)^2 &= \left(\frac{\partial S(\mathbf{x})}{\partial x} + \frac{q}{c} A_x(\mathbf{x}) \right)^2 + \left(\frac{\partial S(\mathbf{x})}{\partial y} + \frac{q}{c} A_y(\mathbf{x}) \right)^2 \\ &+ \left(\frac{\partial S(\mathbf{x}|\theta, Z)}{\partial z} + \frac{q}{c} A_z(\mathbf{x}) \right)^2 + m^2 c^2, \end{aligned} \quad (20)$$

which is the relativistic HJE for a charged, massive particle in an electromagnetic field [3,7].

2.5. Derivation of the Klein–Gordon equation

As a first step, it is expedient to write Eq. (13) in an alternative form by noting that $c\tau = \sqrt{\eta_{\mu\nu} x^\mu x^\nu}$ implies $\partial^\alpha \tau = x^\alpha / (c^2 \tau)$ such that

$$\begin{aligned} \eta_{\mu\nu} (\partial^\mu P(\tau|\theta, Z)) (\partial^\nu P(\tau|\theta, Z)) &= \left(\frac{\partial P(\tau|\theta, Z)}{\partial \tau} \right)^2 \eta_{\mu\nu} (\partial^\mu \tau) (\partial^\nu \tau) \\ &= \frac{1}{c^2} \left(\frac{\partial P(\tau|\theta, Z)}{\partial \tau} \right)^2, \end{aligned} \quad (21)$$

and hence Eq. (13) can be written as

$$I_F = c^2 \int d^4 \mathbf{x} \frac{1}{P(\tau|\theta, Z)} (\partial P(\tau|\theta, Z))^2. \quad (22)$$

The general guiding principle of the LI approach is that the experiment that yields the most robust data is described by the probability density $P(\mathbf{x}|\theta, Z)$ that minimizes I_F for all θ simultaneously, subject to additional constraints that are deemed relevant to the experiment at hand [25]. In the present case, we require that the description is compatible with the special theory of relativity. For a massive, charged particle and in the absence of uncertainty, the latter requirement implies that the classical, relativistic HJE (19) should hold. We can inject this requirement into the LI approach by considering the functional

$$F = c^2 \int d^4 \mathbf{x} \left\{ \frac{(\partial P(\tau|\theta, Z))^2}{P(\tau|\theta, Z)} + \lambda [(\partial S - \mathbf{A})^2 - c^2] P(\tau|\theta, Z) \right\}, \quad (23)$$

where λ is a weighting factor that reflects the importance of the uncertainty and robustness relative to the contribution of the classical dynamics. It is straightforward to show that the expression Eq. (23) is invariant under Lorentz transformations.

Extremization of Eq. (23) can be carried out by the standard variational calculus and yields a set of nonlinear partial differential equations for $P(\mathbf{x}|\theta, Z)$ and $S(\mathbf{x})$. It is not difficult to show that at an extremum, (i) the value of F does not depend on the value of the unknown proper time θ and that (ii) the value of F is zero, independent of λ . The latter result implies that the extrema describe situations in which the uncertainty about the detection events is perfectly balanced by the certainty that the classical HJE describes the motion of the observed detection events.

It is now expedient to write Eq. (23) more explicitly as

$$\begin{aligned}
 F = c^2 \int d^4\mathbf{x} & \left\{ \frac{1}{P(\mathbf{x}|\theta, Z)} \left[\left(\frac{\partial P(\mathbf{x}|\theta, Z)}{\partial ct} \right)^2 - \left(\frac{\partial P(\mathbf{x}|\theta, Z)}{\partial x} \right)^2 \right. \right. \\
 & \left. \left. - \left(\frac{\partial P(\mathbf{x}|\theta, Z)}{\partial y} \right)^2 - \left(\frac{\partial P(\mathbf{x}|\theta, Z)}{\partial z} \right)^2 \right] \right. \\
 & + \lambda \left[\left(\frac{\partial S(\mathbf{x})}{\partial ct} - A^0(\mathbf{x}) \right)^2 - \left(\frac{\partial S(\mathbf{x})}{\partial x} + A^1(\mathbf{x}) \right)^2 - \left(\frac{\partial S(\mathbf{x})}{\partial y} + A^2(\mathbf{x}) \right)^2 \right. \\
 & \left. \left. - \left(\frac{\partial S(\mathbf{x})}{\partial z} + A^3(\mathbf{x}) \right)^2 - c^2 \right] P(\mathbf{x}|\theta, Z) \right\}. \quad (24)
 \end{aligned}$$

We do not know of any direct method to solve the nonlinear set of equations that results from searching for the extrema of Eq. (24) but, by analogy with the non-relativistic case, we may consider a quadratic functional of a complex-valued field $\varphi(\mathbf{x})$ and use the polar representation of this field to construct the corresponding functional in terms of this representation [25,26,31].

To this end, considering the quadratic functional

$$\begin{aligned}
 Q = 4c^2 \int d^4\mathbf{x} & \left\{ \left[\frac{\partial \varphi^*(\mathbf{x})}{\partial ct} + \frac{i\sqrt{\lambda}}{2c} A^0(\mathbf{x}) \varphi^*(\mathbf{x}) \right] \left[\frac{\partial \varphi(\mathbf{x})}{\partial ct} - \frac{i\sqrt{\lambda}}{2c} A^0(\mathbf{x}) \varphi(\mathbf{x}) \right] \right. \\
 & - \left[\frac{\partial \varphi^*(\mathbf{x})}{\partial x} - \frac{i\sqrt{\lambda}}{2c} A^1(\mathbf{x}) \varphi^*(\mathbf{x}) \right] \left[\frac{\partial \varphi(\mathbf{x})}{\partial x} + \frac{i\sqrt{\lambda}}{2c} A^1(\mathbf{x}) \varphi(\mathbf{x}) \right] \\
 & - \left[\frac{\partial \varphi^*(\mathbf{x})}{\partial y} - \frac{i\sqrt{\lambda}}{2c} A^2(\mathbf{x}) \varphi^*(\mathbf{x}) \right] \left[\frac{\partial \varphi(\mathbf{x})}{\partial y} + \frac{i\sqrt{\lambda}}{2c} A^2(\mathbf{x}) \varphi(\mathbf{x}) \right] \\
 & \left. - \left[\frac{\partial \varphi^*(\mathbf{x})}{\partial z} - \frac{i\sqrt{\lambda}}{2c} A^3(\mathbf{x}) \varphi^*(\mathbf{x}) \right] \left[\frac{\partial \varphi(\mathbf{x})}{\partial z} + \frac{i\sqrt{\lambda}}{2c} A^3(\mathbf{x}) \varphi(\mathbf{x}) \right] - \frac{\lambda c^2}{4} \varphi^*(\mathbf{x}) \varphi(\mathbf{x}) \right\}. \quad (25)
 \end{aligned}$$

Substituting the polar representation

$$\varphi(\mathbf{x}) = \sqrt{P(\mathbf{x}|\theta, Z)} e^{i\sqrt{\lambda} S(\mathbf{x})/2}, \quad (26)$$

in Eq. (25) yields $Q = F$. Equations for the extrema of the functional Q can be found by variation with respect to $\varphi^*(\mathbf{x})$, yielding the linear partial differential equation

$$\begin{aligned}
 \left[\frac{\partial}{\partial ct} + i\frac{\sqrt{\lambda}}{2c} A^0(\mathbf{x}) \right]^2 \varphi(\mathbf{x}) & = \left\{ \left[\frac{\partial}{\partial x} - i\frac{\sqrt{\lambda}}{2c} A^1(\mathbf{x}) \right]^2 + \left[\frac{\partial}{\partial y} - i\frac{\sqrt{\lambda}}{2c} A^2(\mathbf{x}) \right]^2 \right. \\
 & \left. + \left[\frac{\partial}{\partial z} - i\frac{\sqrt{\lambda}}{2c} A^3(\mathbf{x}) \right]^2 - \frac{\lambda c^2}{4} \right\} \varphi(\mathbf{x}), \quad (27)
 \end{aligned}$$

which has the same mathematical structure as the KG equation [32]. This can be made more explicit by changing symbols according to $\mathbf{A} \rightarrow q(\Phi, A_x, A_y, A_z)/m$ and $\lambda = 4m^2/\hbar^2$, yielding

$$\begin{aligned}
 \frac{1}{c^2} \left[i\hbar \frac{\partial}{\partial t} - q\Phi(\mathbf{x}) \right]^2 \varphi(\mathbf{x}) & = \left\{ \left[\frac{\hbar}{i} \frac{\partial}{\partial x} - \frac{q}{c} A_x(\mathbf{x}) \right]^2 + \left[\frac{\hbar}{i} \frac{\partial}{\partial y} - \frac{q}{c} A_y(\mathbf{x}) \right]^2 \right. \\
 & \left. + \left[\frac{\hbar}{i} \frac{\partial}{\partial z} - \frac{q}{c} A_z(\mathbf{x}) \right]^2 + m^2 c^2 \right\} \varphi(\mathbf{x}). \quad (28)
 \end{aligned}$$

Obviously, the weighting factor $\lambda = 4m^2/\hbar^2$ cannot be determined on the basis of logic only but has to follow from a comparison of the outcome of calculations based on Eq. (27) with experimental data. It is of interest to inquire to what extent Eq. (28) allows us to infer from the observed data properties of the massive charged particles. The speed of light in vacuum c certainly does not depend on the properties of the massive charged particle. Then, from Eq. (28), it is immediately clear that its solutions are invariant under the transformation $\hbar \rightarrow \hbar\xi$, $q \rightarrow q\xi$, and $m \rightarrow m\xi$. Hence, from the observed data we may be able to determine two but not three of the constants that appear in Eq. (28). For instance, by a suitable redefinition of the units of mass and charge, \hbar can be eliminated from Eq. (28) [33].

In practice, instead of solving the set of nonlinear equations in terms of $P(\mathbf{x}|\theta, Z)$ and $S(\mathbf{x})$ that result from minimizing Eq. (24), it is much easier to first solve Eq. (27) and then use Eq. (26) to find $P(\mathbf{x}|\theta, Z) = \varphi(\mathbf{x})^* \varphi(\mathbf{x})$. It is important to recognize that the LI approach gives us the probability for observing a space–time event \mathbf{x} but does not yield an estimate of the proper time of the particle θ . The latter was and remains unknown. The LI approach suggests that the wave function $\varphi(\mathbf{x})$ is only a mathematical vehicle, be it an extraordinarily useful one, to transform a set of nonlinear partial differential equations into a linear set of partial differential equations. The interplay of the two real quantities $S(\mathbf{x})$ and $P(\mathbf{x}|\theta, Z)$ which account for respectively, the classical relativistic physics and the uncertainty on the collected data, can be disentangled through the use of single complex wave function. But, as a mathematical tool, the wave function does not need an interpretation: it is $P(\mathbf{x}|\theta, Z)$ that is directly linked to the observed events.

3. Discussion

We have shown how the Klein–Gordon equation for massive, charged particles derives from logical inference applied to experiments for which the observed events are independent and for which the frequency distribution of these events is robust with respect to small changes of the conditions under which experiments are carried out. The present derivation is a logical generalization of earlier work [24–27] to the relativistic domain, the fundamental difference being that the measured time is subject to uncertainty.

Obviously, the transition from non-relativistic to relativistic quantum theory is expected to bring in some radically new features. Landau and Peierls [34] pointed out that in relativistic quantum theory the particle position cannot be measured with an accuracy higher than its Compton wavelength. Measuring the position of an electron with an accuracy higher than its Compton wavelength requires an energy that exceeds the threshold for the creation of electron–positron pairs [32], rendering meaningless the question which of the electrons is the original one. Therefore, there is a common belief that relativistic quantum theory cannot be a theory of individual particles but it must be a *field theory* for a non-constant number of particles [35,36]. The requirement of a field theory description is also linked to the fact that the charge density of the Klein–Gordon equation is not positive definite, as mentioned by Dirac [37] and also stressed by Feshbach and Villars [32]. This is due to the second order time derivative in the Klein–Gordon equation and indicates that the wave function describes in fact two degrees of freedom instead of one [32].

It is worth noting that there is no mention of the direction of time in the logical inference approach. Indeed, when we derived Eq. (13) we allowed both $\theta < \tau$ and $\theta > \tau$. This looks unusual from the point of view of single-particle quantum mechanics (see, however, a discussion of time (a)symmetry in Refs. [38,39]). Within relativistic quantum mechanics it seems more natural. As was suggested by Wheeler one might interpret anti-particles as particles with the sign of the proper time reversed, i.e. as if the particles are moving backward in time [40], and, at least, for non-interacting particles this interpretation seems to be possible. In the measurement scenario analyzed here, there is no way to discern the absorption of particles from the emission of anti-particles. Both types of events contribute equally well to the detection counts. This relates to the measurement scenario where we make no distinction between detection events for which $\tau > \theta$ and detection events for which $\tau < \theta$; causality is not a prerequisite in the derivation presented here.

Naturally one might ask how to extend this approach to particles with non-zero spin (e.g., Dirac equation). We leave this challenging program for future research.

Acknowledgment

MIK and HCD acknowledge financial support by the European Research Council, project 338957 FEMTO/NANO.

References

- [1] E. Schrödinger, *Ann. Phys.* 384 (1926) 361–376.
- [2] E. Schrödinger, *Ann. Phys.* 384 (1926) 489–527.
- [3] W. Gordon, *Z. Phys.* 40 (1926) 117–133.
- [4] V. Fock, *Z. Phys.* 38 (1926) 242–250.
- [5] V. Fock, *Z. Phys.* 39 (1926) 226–232.
- [6] J. Kudar, *Ann. Phys.* 386 (1926) 632–636.
- [7] O. Klein, *Z. Phys.* 41 (1927) 407–442.
- [8] P.A.M. Dirac, *The Principles of Quantum Mechanics*, in: *International Series of Monographs on Physics*, Oxford University Press, 1957.
- [9] L.D. Landau, E.M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory*, in: *Course of Theoretical Physics*, vol. 3, Pergamon Press, 1959.
- [10] D.J. Griffiths, *Introduction to Quantum Mechanics*, second ed., Pearson international edition, Pearson Prentice Hall, 2005.
- [11] L.E. Ballentine, *Quantum Mechanics: A Modern Development*, second ed., World Scientific, 2015.
- [12] P. Busch, *Phys. Rev. Lett.* 91 (2003) 120403.
- [13] C.M. Caves, C.A. Fuchs, K.K. Manne, J.M. Renes, *Found. Phys.* 34 (2004) 193–209.
- [14] S. Rieder, K. Svozil, *Foundations of Probability and Physics*, vol. 889, 2007, pp. 235–242.
- [15] W.H. Zurek, *Rev. Modern Phys.* 75 (2003) 715.
- [16] L. Hardy, *Quantum theory from five reasonable axioms*, arXiv:quant-ph/0101012.
- [17] G. Chiribella, G.M. D'Ariano, P. Perinotti, *Phys. Rev. A* 81 (2010) 062348.
- [18] Č. Brukner, *Physics* 4 (2011) 55.
- [19] L. Masanes, P. Müller, *New J. Phys.* 13 (2011) 063001.
- [20] R.T. Cox, *Amer. J. Phys.* 14 (1946) 1–13.
- [21] R.T. Cox, *The Algebra of Probable Inference*, Johns Hopkins University Press, Baltimore, 1961.
- [22] M. Tribus, *Rational Descriptions, Decisions and Designs*, Expira Press, Stockholm, 1999.
- [23] E. Jaynes, G. Bretthorst, *Probability Theory: The Logic of Science*, Cambridge University Press, 2003.
- [24] H. De Raedt, M.I. Katsnelson, K. Michielsen, *Proc. SPIE* 8832 (2013) 1. 883212–1–11.
- [25] H. De Raedt, M.I. Katsnelson, K. Michielsen, *Ann. Phys.* 347 (2014) 45–73.
- [26] H. De Raedt, M.I. Katsnelson, H.C. Donker, K. Michielsen, *Ann. Phys.* 359 (2015) 166–186.
- [27] H. De Raedt, M.I. Katsnelson, H.C. Donker, K. Michielsen, *Proc. SPIE* 9570 (2015) 957002–1–14.
- [28] L. Motz, A. Selzer, *Phys. Rev.* 133 (1964) 1622–1624.
- [29] J.P. Ralston, *Proc. SPIE - Int. Soc. Opt. Eng.* 8832 (2013) 88320W–1–25.
- [30] J.D. Jackson, *Classical Electrodynamics*, third ed., John Wiley & Sons, Inc., 1993.
- [31] E. Madelung, *Z. Phys.* 40 (1927) 322–326.
- [32] H. Feshbach, F. Villars, *Rev. Modern Phys.* 30 (1958) 24–45.
- [33] J.P. Ralston, *Proc. SPIE* 8832 (2013) 883216–1–16.
- [34] L.D. Landau, R. Peierls, *Z. Phys.* 69 (1931) 56–69.
- [35] M.E. Peskin, D.V. Schroeder, *An Introduction to Quantum Field Theory*, Westview, 1995.
- [36] A. Zee, *Quantum Field Theory in a Nutshell*, second ed., Princeton University Press, 2010.
- [37] P.A.M. Dirac, *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* (1928) 610–624.
- [38] J.G. Cramer, *Rev. Modern Phys.* 58 (1986) 647–687.
- [39] Y. Aharonov, S. Popescu, J. Tollaksen, *Phys. Today* (2010) 27–32.
- [40] R.P. Feynman, *Physics, 1963–1970, Nobel Lectures in Physics*, World Scientific, 1998, Ch. The development of the space–time view of quantum electrodynamics.