Quantum theory as a description of robust experiments: Derivation of the Pauli equation

CrossMark

Hans De Raedt [a], Mikhail I. Katsnelson [b], Hylke C. Donker [b], Kristel Michielsen [c,d,*]

[a] *Department of Applied Physics, Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, NL-9747AG Groningen, The Netherlands*

[b] *Radboud University Nijmegen, Institute for Molecules and Materials, Heyendaalseweg 135, NL-6525AJ Nijmegen, The Netherlands*

[c] *Institute for Advanced Simulation, Jülich Supercomputing Centre, Forschungszentrum Jülich, D-52425 Jülich, Germany*

[d] *RWTH Aachen University, D-52056 Aachen, Germany*

H I G H L I G H T S

- The Pauli equation is obtained through logical inference applied to robust experiments on a charged particle.
- The concept of spin appears as an inference resulting from the treatment of two-valued data.
- The same reasoning yields the quantum theoretical description of neutral magnetic particles.
- Logical inference provides a framework to establish a bridge between objective knowledge gathered through experiments and their description in terms of concepts.

A R T I C L E I N F O

A B S T R A C T

It is shown that the Pauli equation and the concept of spin naturally emerge from logical inference applied to experiments on a charged particle under the conditions that (i) space is homogeneous (ii) the observed events are logically independent, and (iii) the observed frequency distributions are robust with respect to small changes in the conditions under which the experiment is carried out. The derivation does not take recourse to concepts of quantum theory and is based on the same principles which have already been shown to lead to e.g. the Schrödinger equation and the probability distributions of pairs of particles in the singlet or triplet state. Application to Stern–Gerlach experiments with chargeless,

* Corresponding author at:Institute for Advanced Simulation, Jülich Supercomputing Centre, Forschungszentrum Jülich, D-52425 Jülich, Germany.

E-mail address: k.michielsen@fz-juelich.de (K. Michielsen).

magnetic particles, provides additional support for the thesis that quantum theory follows from logical inference applied to a well-defined class of experiments.

1. Introduction

In laboratory experiments, one **never** has complete knowledge about the mechanisms that affect the outcome of the measurements: there is always uncertainty. In addition, the outcomes of real experiments are always subject to uncertainties with respect to the conditions under which the experiments are carried out.

If there are uncertainties about the individual events and uncertainties about the conditions under which the experiment is carried out, it is often difficult or even impossible to establish relations between individual events. However, in the case that the frequencies of these events are robust (to be discussed in more detail later) it may still be possible to establish relations, not between the individual events, but between the frequency distributions of the observed events.

The algebra of logical inference provides a mathematical framework that facilitates rational reasoning when there is uncertainty [1–5]. A detailed discussion of the foundations of logical inference, its relation to Boolean logic and the derivation of its rules can be found in the papers [1,4] and books [2,3,5]. Logical inference is the foundation for powerful tools such as the maximum entropy method and Bayesian analysis [3,5]. To the best of our knowledge, the first derivation of a non-trivial theoretical description by this general methodology of scientific reasoning appears in Jaynes' papers on the relation between information and (quantum) statistical mechanics [6,7].

A recent paper [8] shows how some of the most basic equations of quantum theory, e.g. the Schrödinger equation and the probability distributions of pairs of particles in the singlet or triplet state emerge from the application of logical inference to (the abstraction of) robust experiments, without taking recourse to concepts of quantum theory. This logical-inference approach yields results that are unambiguous and independent of the individual subjective judgment. In addition, this approach provides a rational explanation for the extraordinary descriptive power of quantum theory [8]. As the introduction of the concept of intrinsic angular momentum, called spin, is a landmark in the development of quantum theory, it is natural to ask the question under which circumstances this concept appears in a logical-inference treatment.

A classical review of how the concept of spin has been introduced in quantum theory is given by van der Waerden [9]. The original motivation to introduce this new concept was the discovery of the anomalous Zeeman effect and its transition to the normal Zeeman effect with increasing magnetic field (the so-called Paschen-Back effect). Pauli introduced spin in a very formal way by attributing to the electron an additional intrinsic magnetic quantum number taking the values $\pm 1/2$ [10]. Although the picture of the spin in terms of a "rotating electron model" was quickly and widely accepted, Pauli was strongly against this picture because of its purely classical-mechanics character. A few years later he suggested the Pauli equation [11] in which this intrinsic degree of freedom was introduced by replacing the single-component wavefunction that appears in Schrödinger's equation by a two-component wavefunction and "Pauli matrices"; the most rigorous way to establish a relation with the idea of the rotating electron is just a formal observation that these Pauli matrices satisfy the same commutation relations as the generators of the rotation group in three-dimensional space and that the two-component wavefunctions (spinors) provide a double-valued representation of this group [9].

Bohm and his followers, in the spirit of their general approach to provide a causal interpretation of quantum mechanics, tried to construct a purely classical description of spin by analogy with the hydrodynamics of a rotating liquid [12,13]. Despite the beauty of the mathematical description, the interpretation of the spin as entity, a field, which is distributed over the whole space is rather exotic and can hardly be considered as a derivation and justification of the Pauli equation.

Bohr and Pauli suggested that spin and the related magnetic moment cannot be measured in experiments which can be interpreted in terms of classical trajectories (such as Stern–Gerlach experiments with a free-electron beam), see Ref. [14] and references therein. In an inhomogeneous magnetic field, spin effects cannot be separated from the effects of the Lorentz force due to the orbital motion of the charged particle. However, these difficulties are technical rather than conceptual as they do not consider the possibility that there are neutral particles (not subject to the Lorentz force) with magnetic moments, such as neutrons, for which a Stern–Gerlach experiment is not only possible in principle but has really been performed [15]. It is clear now that the naive way to demonstrate the "essentially non-classical" character of the spin degree of freedom was premature.

In this paper, we show how the Pauli equation and the concept of spin naturally emerge from the logical-inference analysis of experiments on a charged particle. We carefully analyze the additional assumptions (some of them having obvious analogs in Pauli's analysis of the anomalous Zeeman effect) which are required to pass, in a model-free way, to the Pauli equation.

Conceptually, we return to the roots by first introducing "spin" as some intrinsic degree of freedom characterized by a two-valued number. We will call this two-valued property "color" (e.g. red or blue) to make clear that we leave no room for (mis)interpretations in terms of models of a rotating particle and the like. This is in sharp contrast to the interpretation of Refs. [12,13]. Note that such a generalization of the concept of spin is very important in modern physics. For instance, the idea of isospin of elementary particles [16] was originally introduced [17] as a way to describe constituents of atomic nuclei in terms of the same particles (nucleons) with two subspecies (neutrons and protons). Another example is the pseudospin of the charge carriers in graphene [18] used to indicate that the carriers belong to sublattice A or B of the honeycomb crystal lattice. In both of these examples, there is nothing that is rotating!

We further illustrate the power of the approach by an application to Stern–Gerlach experiments with chargeless, magnetic particles, providing additional support to the idea that quantum theory directly follows from logical inference applied to a well-defined class of experiments [8].

To head off possible misunderstandings, it is important to mention that the underlying premise of our approach is that current scientific knowledge derives, through cognitive processes in the human brain, from the discrete events which are observed in laboratory experiments and from relations between those events that we, humans, discover. As a direct consequence of this underlying premise, the validity of the results obtained in our approach does not depend on the assumption that the observed events are signatures of an underlying objective reality which is mathematical in nature (for an overview of older and new work in this direction, see Ref. [19]). We take the point of view that the aim of physics is to provide a consistent description of relations between certain events that we perceive (usually with the help of some equipment) with our senses. Some of these relations express cause followed by an effect and others do not. A derivation of a quantum theoretical description from logical-inference principles does not prohibit the construction of cause-and-effect mechanisms that, when analyzed in the same manner as in real experiments, create the *impression* that the system behaves according to quantum theory [20–22]. Work in this direction has shown that it is indeed possible to build simulation models which reproduce, on an event-by-event basis, the results of interference/entanglement/uncertainty experiments with photons/neutrons [23–27].

The paper is organized as follows. In Section 2 we specify the measurement scenario and introduce the inference-probability that characterizes the observed detection events (all the elements of logical inference that are required for the purpose of the present paper are summarized in Appendix A). Then, we discuss and formalize the notion of a robust experiment. Although these three steps are similar to the ones taken in the logical-inference derivation of the Schrödinger equation [8], to make the presentation self-contained, we give a detailed account. The next three subsections address the problem of including additional knowledge about the motion of the particle in some limiting cases. In Section 2.8 we collect the results of the previous subsections and derive the Pauli equation. Section 3 shows that the same procedure leads to the quantum theoretical equation that describes the motion of an uncharged particle in a magnetic field. A discussion of the relation of the logical-inference derivation of the Pauli equation and earlier work on the hydrodynamic formulation of quantum theory is given in Section 4. A summary and discussion of more general aspects of the work presented in this paper can be found in Section 5.

2. Logical inference: derivation of the Pauli equation

2.1. Measurement scenario

We consider N repetitions of an experiment on a particle located in 3-dimensional space $\boldsymbol{\Omega}$. The experiment consists of sending a signal to the particle at discrete times labeled by the integer $\tau = 1, \ldots, M$. It is assumed that for each repetition, labeled by $n = 1, \ldots, N$, the particle is at the unknown position $\mathbf{X}_\tau \in \boldsymbol{\Omega}$. As the particle receives the signal, it responds by emitting another signal which is recorded by an array of detectors. For each signal emitted by a particle the data recorded by the detector system is used to determine the position $\mathbf{j}_{n,\tau} \in \mathcal{V}$ where $\mathcal{V}$ denotes the set of voxels with linear extent $[-\Delta, \Delta]/2$ that cover the 3-dimensional space $\boldsymbol{\Omega}$. The signal also contains additional information which is two-valued and encodes, so to speak, the "color" of the particle at the time when it responded to the signal emitted by the source. This color is represented by variables $k_{n,\tau} = \pm 1$. The frequency distribution of the $(\mathbf{j}, k)_{n,\tau}$'s changes with the applied electric and magnetic field from which we may infer that there is some form of interaction between the electromagnetic field and the particle.

The result of N repetitions of the experiment yields the data set

$$\Upsilon = \{(\mathbf{j}, k)_{n,\tau} | \mathbf{j}_{n,\tau} \in \mathcal{V}; \ k = \pm 1; \ n = 1, \ldots, N; \ \tau = 1, \ldots, M\}, \tag{1}$$

or, denoting the total counts of voxels $\mathbf{j}$ and color k at time τ by $0 \leq c_{\mathbf{j},k,\tau} \leq N$, the data set can be represented as

$$\mathcal{D} = \left\{ c_{\mathbf{j},k,\tau} \middle| \tau = 1, \ldots, M \ ; \ \sum_{k=\pm 1} \sum_{\mathbf{j} \in [-L^d, L^d]} c_{\mathbf{j},k,\tau} = N \right\}. \tag{2}$$

2.2. Inference-probability of the data produced by the experiment

The first step is to introduce a real number $0 \leq P(\mathbf{j}, k | \mathbf{X}_\tau, \tau, Z) \leq 1$ which represents the plausibility that we observe a detector click $(\mathbf{j}, k)$, conditional on $(\mathbf{X}_\tau, \tau, Z)$. For reasons explained in Appendix A, $P(\mathbf{j}, k | \mathbf{X}_\tau, \tau, Z)$ is called the inference-probability (or i-prob for short) and encodes the relation between the unknown location $\mathbf{X}_\tau$ and the location $\mathbf{j}$ and color k registered by the detector system at discrete time τ. Except for the unknown location $\mathbf{X}_\tau$, all other experimental conditions are represented by Z and are assumed to be fixed and identical for all experiments. Note that unlike in the case of parameter estimation, in the case at hand both $P(\mathbf{j}, k | \mathbf{X}_\tau, \tau, Z)$ and the parameters $\mathbf{X}_\tau$ are unknown.

We make the following, seemingly reasonable assumptions:

1. Each repetition of the experiment represents an event of which the outcome is logically independent of any other such event. By application of the product rule (see Appendix A), a direct consequence of this assumption is that

$$P(\Upsilon | \mathbf{X}_1, \ldots, \mathbf{X}_M, N, Z) = \prod_{\tau=1}^{M} \prod_{n=1}^{N} P(\mathbf{j}_{n,\tau}, k_{n,\tau} | \mathbf{X}_\tau, \tau, Z), \tag{3}$$

and hence

$$P(\mathcal{D} | \mathbf{X}_1, \ldots, \mathbf{X}_M, N, Z) = N! \prod_{\tau=1}^{M} \prod_{\mathbf{j} \in \mathcal{V}} \prod_{k=\pm 1} \frac{P(\mathbf{j}, k | \mathbf{X}_\tau, \tau, Z)^{c_{\mathbf{j},k,\tau}}}{c_{\mathbf{j},k,\tau}!}. \tag{4}$$

2. It is assumed that it does not matter where the experiment is carried out. This implies that the i-prob should have the property

$$P(\mathbf{j}, k | \mathbf{X}_\tau, \tau, Z) = P(\mathbf{j} + \boldsymbol{\zeta}, k | \mathbf{X}_\tau + \boldsymbol{\zeta}, \tau, Z), \tag{5}$$

where $\boldsymbol{\zeta}$ is an arbitrary 3-dimensional vector. The relation Eq. (5) expresses the assumption that space is homogeneous.

2.3. Condition for reproducibility and robustness

If the frequencies with which the detectors fire vary erratically with $\{\mathbf{X}_\tau\}$, the experiment would most likely be called "irreproducible". Excluding such experiments, it is desirable that frequency distributions of the data exhibit some kind of robustness, smoothness with respect to small changes of the unknown values of $\{\mathbf{X}_\tau\}$. Unless the experimental setup is sufficiently "robust" in the sense just explained, repeating the run with slightly different values of $\{\mathbf{X}_\tau\}$ would often produce results that are very different from those of other runs and it is common practice to discard such experimental data. Therefore, a "good" experiment must be a robust experiment.

The robustness with respect to small variations of the conditions under which the experiment is carried out should be reflected in the expression of the i-prob to observe data sets which yield reproducible averages and correlations (with the usual statistical fluctuations). The next step therefore is to determine the expression for $P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)$ which is most insensitive to small changes in $\mathbf{X}_\tau$. It is expedient to formulate this problem as an hypothesis test. Let H_0 and H_1 be the hypothesis that the same data $\mathcal{D}$ is observed for the unknown locations $\{\mathbf{X}_\tau\}$ and $\{\mathbf{X}_\tau + \boldsymbol{\varepsilon}_\tau\}$, respectively. The evidence Ev of hypothesis H_1, relative to hypothesis H_0, is defined by [3,5]

$$
\begin{aligned}
\text{Ev} &= \ln \frac{P(\mathcal{D}|\mathbf{X}_\tau + \boldsymbol{\varepsilon}_\tau, \tau, N, Z)}{P(\mathcal{D}|\mathbf{X}_\tau, \tau, N, Z)} \\
&= \sum_{\mathbf{j},k,\tau} c_{\mathbf{j},k,\tau} \ln \frac{P(\mathbf{j}, k|\mathbf{X}_\tau + \boldsymbol{\varepsilon}_\tau, \tau, Z)}{P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)},
\end{aligned}
\tag{6}
$$

where the logarithm serves to facilitate algebraic manipulations. If H_1 is more (less) plausible than H_0 then Ev > 0 (Ev < 0). In statistics, the r.h.s. of Eq. (6) is known as the log-likelihood function and used for parameter estimation. In contrast, in the present context, the function Eq. (6) is *not* used to estimate $\mathbf{X}_\tau$ but is a vehicle to express the robustness with respect to the coordinates $\mathbf{X}_\tau$.

Writing Eq. (6) as a Taylor series in $\boldsymbol{\varepsilon}$ we have

$$
\begin{aligned}
\text{Ev} &= \sum_{\mathbf{j},k,\tau} c_{\mathbf{j},k,\tau} \ln \left[1 + \frac{\boldsymbol{\varepsilon}_\tau \cdot \nabla_\tau P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)}{P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)} + \frac{1}{2} \frac{(\boldsymbol{\varepsilon}_\tau \cdot \nabla_\tau)^2 P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)}{P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)} + \mathcal{O}(\varepsilon_\tau^3) \right] \\
&= \sum_{\mathbf{j},k,\tau} c_{\mathbf{j},k,\tau} \left[\frac{\boldsymbol{\varepsilon}_\tau \cdot \nabla_\tau P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)}{P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)} - \frac{1}{2} \left[\frac{\boldsymbol{\varepsilon}_\tau \cdot \nabla_\tau P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)}{P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)} \right]^2 \right. \\
&\quad \left. + \frac{1}{2} \frac{(\boldsymbol{\varepsilon}_\tau \cdot \nabla_\tau)^2 P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)}{P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)} \right] + \mathcal{O}(\varepsilon_\tau^3),
\end{aligned}
\tag{7}
$$

where ∇_τ differentiates with respect to $\mathbf{X}_\tau$. Here and in the following we assume that $\boldsymbol{\varepsilon}_\tau$ is sufficiently small such that the third and higher order terms in the $\boldsymbol{\varepsilon}$'s can be ignored. According to our criterion of robustness, the evidence Eq. (7) should change as little as possible as $\mathbf{X}_\tau$ varies. This can be accomplished by minimizing, in absolute value, all the coefficients of the polynomial in $\boldsymbol{\varepsilon}_\tau$, *for all allowed $\boldsymbol{\varepsilon}_\tau$ and $\mathbf{X}_\tau$*. The clause "for all allowed $\boldsymbol{\varepsilon}_\tau$ and $\mathbf{X}_\tau$" implies that we are dealing here with an instance of a robust optimization problem [28].

The first and third sum in Eq. (7) vanish identically if we choose $c_{\mathbf{j},k,\tau}/N = P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)$. Indeed, we have

$$
\begin{aligned}
\sum_{\mathbf{j},k,\tau} c_{\mathbf{j},k,\tau} \frac{(\boldsymbol{\varepsilon}_\tau \cdot \nabla_\tau)^\alpha P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)}{P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)} &= N \sum_{\mathbf{j},k,\tau} (\boldsymbol{\varepsilon}_\tau \cdot \nabla_\tau)^\alpha P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z) \\
&= N \sum_\tau (\boldsymbol{\varepsilon}_\tau \cdot \nabla_\tau)^\alpha \sum_{\mathbf{j},k} P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z) \\
&= N \sum_\tau (\boldsymbol{\varepsilon}_\tau \cdot \nabla_\tau)^\alpha 1 = 0,
\end{aligned}
\tag{8}
$$

for $\alpha = 1, 2, \ldots$. Although this choice is motivated by the desire to eliminate contributions of order $\boldsymbol{\varepsilon}_\tau$, it follows that our criterion of robustness automatically suggests the intuitively obvious procedure to assign to $P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)$ the value of the observed frequencies of occurrences $c_{\mathbf{j},k,\tau}/N$ [3,5].

Dropping irrelevant numerical factors and terms of $\mathcal{O}(\boldsymbol{\varepsilon}_\tau^3)$, the remaining contribution to the evidence

$$\text{Ev} = \sum_{\mathbf{j},k,\tau} \frac{1}{P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)} \left[\boldsymbol{\varepsilon} \cdot \nabla_\tau P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)\right]^2, \tag{9}$$

vanishes identically (for all $\boldsymbol{\varepsilon}_\tau$) if and only if $\nabla_\tau P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z) = 0$ in which case it is clear that we can only describe experiments for which the data does not exhibit any dependence on $\mathbf{X}_\tau$.

> Experiments which produce frequency distributions that do not depend on the conditions do not increase our knowledge about the relation between the conditions and the observed data. Therefore, we explicitly exclude such non-informative experiments.

Thus, from now on, we explicitly exclude the class of experiments for which $\nabla_\tau P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z) = 0$.

The clause "for all allowed $\boldsymbol{\varepsilon}_\tau$" can be eliminated using the Cauchy–Schwarz inequality. We have

$$\text{Ev} = \sum_{\mathbf{j},k,\tau} \left[\frac{\boldsymbol{\varepsilon}_\tau \cdot \nabla_\tau P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)}{P^{1/2}(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)}\right]^2 \leq \widehat{\varepsilon}^2 \sum_{\mathbf{j},k,\tau} \frac{1}{P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)} \left[\nabla_\tau P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)\right]^2, \tag{10}$$

where $\widehat{\varepsilon}^2 = \max_\tau \boldsymbol{\varepsilon}_\tau^2$. As the $\boldsymbol{\varepsilon}_\tau$'s are arbitrary (but small), it follows from Eq. (10) that we find the robust solution(s) by searching for the global minimum of

$$I_F = \sum_{\mathbf{j},k,\tau} \frac{1}{P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)} \left[\nabla_\tau P(\mathbf{j}, k|\mathbf{X}_\tau, \tau, Z)\right]^2, \tag{11}$$

which is the Fisher information of the measurement scenario described above.

2.4. Continuum limit

Propositions such as "detector $(\mathbf{j}, k)$ has clicked at time τ" are ultimately related to sensory experience and are therefore discrete in nature. On the other hand, the basic equations of quantum theory such as the Schrödinger, Pauli and Dirac equations are formulated in continuum space. Taking the continuum limit of the discrete formulation connects the two modes of description. Here and in the following, we use the symbols for (partial) derivatives for both the case that the continuum approximation is meaningful and the case that it is not. In the latter, operator symbols such as $\partial/\partial t$ should be read as the corresponding finite-difference operators.

Assuming that the continuum limit is well-defined, we have $\mathcal{V} \to \boldsymbol{\Omega}$ and the Fisher information reads

$$\begin{aligned} I_F &= \int d\mathbf{x}\, dt \sum_{i=1}^{3} \sum_{k=\pm 1} \frac{1}{P(\mathbf{x}, k|\mathbf{X}, t, Z)} \left[\frac{\partial P(\mathbf{x}, k|\mathbf{X}, t, Z)}{\partial X_i}\right]^2 \\ &= \int d\mathbf{x}\, dt \sum_{i=1}^{3} \sum_{k=\pm 1} \frac{1}{P(\mathbf{x}, k|\mathbf{X}, t, Z)} \left[\frac{\partial P(\mathbf{x}, k|\mathbf{X}, t, Z)}{\partial x_i}\right]^2 \\ &= \int d\mathbf{x}\, dt \sum_{k=\pm 1} \frac{1}{P(\mathbf{x}, k|\mathbf{X}, t, Z)} \left[\nabla P(\mathbf{x}, k|\mathbf{X}, t, Z)\right]^2, \end{aligned} \tag{12}$$

where ∇ denotes derivatives with respect to $\mathbf{x}$ and we have simplified the notation somewhat by writing $\mathbf{X} = \mathbf{X}_t$. We have changed derivatives with respect to $\mathbf{X}$ to derivatives with respect to $\mathbf{x}$ by assuming that $(P(\mathbf{x}, k|\mathbf{X}, t, Z) = P(\mathbf{x}+\mathbf{y}, k|\mathbf{X}+\mathbf{y}, t, Z)$ holds for all $\mathbf{y}$ (see assumption 2 in Section 2.2).

Furthermore, it is understood that integrations are over the domain defined by the measurement scenario. Technically speaking, after passing to the continuum limit, $P(\mathbf{x}|\mathbf{X}, t, Z)$ denotes the probability density, not the probability itself. However, as mentioned above, we write integration and derivation symbols for both the discrete case and its continuum limit and as there can be no confusion about which case we are considering, we use the same symbol for the probability density and the probability.

For later use, it is expedient to write Eq. (12) in a different form which separates the data about the position of the clicks and the associated color k as much as possible. According to the product rule, we have

$$P(\mathbf{x}, k|\mathbf{X}, t, Z) = P(k|\mathbf{x}, \mathbf{X}, t, Z)P(\mathbf{x}|\mathbf{X}, t, Z), \tag{13}$$

which we may, without loss of generality, represent as

$$P(\mathbf{x}, k = +1|\mathbf{X}, t, Z) = P(\mathbf{x}|\mathbf{X}, t, Z)\cos^2\frac{\theta(\mathbf{x}, \mathbf{X}, t, Z)}{2}$$

$$P(\mathbf{x}, k = -1|\mathbf{X}, t, Z) = P(\mathbf{x}|\mathbf{X}, t, Z)\sin^2\frac{\theta(\mathbf{x}, \mathbf{X}, t, Z)}{2}. \tag{14}$$

Substituting Eq. (14) into Eq. (12) we obtain

$$I_F = \int d\mathbf{x}\, dt \left\{ \frac{1}{P(\mathbf{x}|\mathbf{X}, t, Z)} [\nabla P(\mathbf{x}|\mathbf{X}, t, Z)]^2 + [\nabla\theta(\mathbf{x}, \mathbf{X}, t, Z)]^2 P(\mathbf{x}|\mathbf{X}, t, Z) \right\}, \tag{15}$$

which is the Fisher information for the measurement scenario described earlier. Note that up to this point, we have not assumed that the particle moves or carries a magnetic moment nor did we assign any particular meaning to $\theta(\mathbf{x}, \mathbf{X}, t, Z)$.

According to the principle laid out earlier, our task is to search for the global minimum of Eq. (15), the Fisher information of the measurement scenario described above, thereby excluding the uninformative class of solutions.

2.5. Including knowledge

It is instructive to first search for the global minimum of Eq. (15) in the case that we do not know whether the particle moves or not and do not know about the effect of the applied electromagnetic field on the frequency distribution of the $(\mathbf{j}, k)_{n,\tau}$'s. In this situation, we may discard the time dependence altogether and search for the non-trivial global minimum of

$$\widetilde{I}_F = \int d\mathbf{x} \left\{ \frac{1}{P(\mathbf{x}|\mathbf{X}, Z)} [\nabla P(\mathbf{x}|\mathbf{X}, Z)]^2 + [\nabla\theta(\mathbf{x}, \mathbf{X}, Z)]^2 P(\mathbf{x}|\mathbf{X}, Z) \right\}. \tag{16}$$

For pedagogical purposes, we first focus on the case of one spatial dimension and discard the color dependence, that is we set $\nabla\theta(\mathbf{x}, \mathbf{X}, Z) = 0$ and assume that $\Omega \to [0, L]$ where $[0, L]$ is the range covered by the detection system. With the latter assumption $P(x|X, Z) = 0$ for $x \leq 0$ or $x \geq L$.

Recalling the assumption that space is homogeneous (see Eq. (5)), we search for solutions of the form $P(x|X, Z) = f(x - X, Z)$. As $f(x - X, Z) \geq 0$, we may substitute $P(x|X, Z) = f(x - X, Z) = \psi^2(x - X, Z)$ in Eq. (16) and we obtain

$$\widetilde{I}_F = 4 \int_0^L dx \left(\frac{\partial\psi(x - X, Z)}{\partial x} \right)^2. \tag{17}$$

Recall that the requirement of a global minimum entails that $\widetilde{I}_F$ is constant, independent of the unknown position $\mathbf{X}$ of the particle.

The extrema of Eq. (17) are easily found by a standard variational calculation. Introducing the Lagrange multiplier μ to account for the constraint $\int_0^L dx\, \psi^2(x - X, Z) = \int_0^L dx P(x|X, Z) = 1$, the

extrema are the solutions of

$$\frac{\partial^2 \psi(x - X, Z)}{\partial x^2} - \frac{\mu}{4} \psi(x - X, Z) = 0. \tag{18}$$

For $\mu > 0$, the solutions of Eq. (18) are hyperbolic functions, a family of solutions that is not compatible with the constraint $P(x|X, Z) = 0$ for $x = 0, L$ and can therefore be ruled out. Writing $\mu = -4v^2$, the general solution of Eq. (18) reads

$$\psi(x - X, Z) = c_1(Z) \sin v(x - X) + c_2(Z) \cos v(x - X)$$
$$= [c_1(Z) \cos vX + c_2(Z) \sin vX] \sin vx - [c_1(Z) \sin vX - c_2(Z) \cos vX] \cos vx, \tag{19}$$

where $c_1(Z)$ and $c_2(Z)$ are integration constants. Imposing the boundary condition $\psi(x - X, Z) = 0$ for $x = 0$ we must have $c_1(Z) \sin vX = c_2(Z) \cos vX$. Hence the second term in Eq. (19) vanishes for all x. In addition, imposing the boundary condition $\psi(x - X, Z) = 0$ for $x = L$, we must have either $c_1(Z) \cos vX + c_2(Z) \sin vX = 0$ in which case $\psi(x - X, Z) = 0$ for all x or $v = n\pi/L$ for $n = 1, 2, \ldots$ in which case the non-trivial solutions read

$$\psi(x - X, Z) = \left[c_1(Z) \cos \frac{n\pi X}{L} + c_2(Z) \sin \frac{n\pi X}{L} \right] \sin \frac{n\pi x}{L}, \quad n = 1, 2, \ldots. \tag{20}$$

Using $c_1(Z) \sin vX = c_2(Z) \cos vX$ with $v = n\pi/L$ we find that

$$\psi^2(x - X, Z) = [c_1^2(Z) + c_2^2(Z)] \sin^2 \frac{n\pi x}{L}, \quad n = 1, 2, \ldots, \tag{21}$$

and from $\int_0^L dx \, \psi^2(x - X, Z) = 1$ we find that $L[c_1^2(Z) + c_2^2(Z)]/2 = 1$. Hence

$$P(x|X, Z) = \frac{2}{L} \sin^2 \frac{n\pi x}{L}, \quad n = 1, 2, \ldots, \tag{22}$$

which are nothing but the solutions of the Schrödinger equation of a free particle in a one-dimensional box [29]. Note that the r.h.s of Eq. (22) does not depend on X. In other words, from the measured data we cannot infer anything about the unknown position X, in concert with the notion that the particle is "free". From Eq. (20) it follows that $\widetilde{I_F} = (2n\pi/L)^2$, independent of X as it should be. Clearly, the solution for non-trivial global minimum of $\widetilde{I_F}$ is given by Eq. (22) with $n = 1$.

Returning to the case that the frequency distribution of the $(\mathbf{j}, k)_{n,\tau}$'s indicates that the motion of the particle depends on the applied electric or magnetic field, we can incorporate this additional knowledge as a constraint on the global minimization problem. In general, the global minimization problems that we will consider take the form $\lambda I_F + \Lambda$ where λ is a parameter (not a Lagrange multiplier) that "weights" the uncertainty in the conditions (represented by I_F) relative to the knowledge represented by the functional

$$\Lambda = \int d\mathbf{x} \, dt \sum_{k=\pm 1} F(\mathbf{x}, k, t, Z) P(\mathbf{x}, k|\mathbf{X}, t, Z), \tag{23}$$

where $F(\mathbf{x}, k, t, Z)$ is a function which encodes the additional knowledge and which does not depend on the unknown position $\mathbf{X}$.

The assumption that space is homogeneous allows us to replace derivatives with respect to $\mathbf{X}$ by derivatives with respect to $\mathbf{x}$. This helps in searching for the global minimum of $\lambda I_F + \Lambda$ because it can be found by searching for the extrema of

$$\lambda I_F + \Lambda = \int d\mathbf{x} \, dt \sum_{k=\pm 1} \left\{ \frac{\lambda}{P(\mathbf{x}, k|\mathbf{X}, t, Z)} [\nabla P(\mathbf{x}, k|\mathbf{X}, t, Z)]^2 + F(\mathbf{x}, k, t, Z) P(\mathbf{x}, k|\mathbf{X}, t, Z) \right\}, \tag{24}$$

as a functional of the $P(\mathbf{x}, k|\mathbf{X}, t, Z)$'s. By the standard variational procedure, the extrema of $\lambda I_F + \Lambda$ are the solutions of

$$\frac{\lambda [\nabla P(\mathbf{x}, k|\mathbf{X}, t, Z)]^2}{P^2(\mathbf{x}, k|\mathbf{X}, t, Z)} + 2\lambda \nabla \left[\frac{\nabla P(\mathbf{x}, k|\mathbf{X}, t, Z)}{P(\mathbf{x}, k|\mathbf{X}, t, Z)} \right] - F(\mathbf{x}, k, t, Z) = 0, \quad k = 1, 2. \tag{25}$$

On the other hand, the global minimum of $\lambda I_F + \Lambda$ should not depend on unknown $\mathbf{X}$ because if it did, it was not a global minimum and in addition, the values of $\lambda I_F + \Lambda$ would tell us something about $\mathbf{X}$, a contradiction to the assumption that $\mathbf{X}$ is unknown.

Taking the derivative of Eq. (24) with respect to $\mathbf{X}$ (recall $\mathbf{X} = \mathbf{X}_t$) yields

$$\nabla_t(\lambda I_F + \Lambda) = -\int d\mathbf{x}\, dt \sum_{k=\pm 1} \left[\lambda \frac{[\nabla P(\mathbf{x}, k|\mathbf{X}, t, Z)]^2}{P^2(\mathbf{x}, k|\mathbf{X}, t, Z)} \right.$$

$$\left. + 2\lambda \nabla \left[\frac{\nabla P(\mathbf{x}, k|\mathbf{X}, t, Z)}{P(\mathbf{x}, k|\mathbf{X}, t, Z)} \right] - F(\mathbf{x}, k, t, Z) \right] \nabla_t P(\mathbf{x}, k|\mathbf{X}, t, Z). \tag{26}$$

Comparing Eqs. (25) and (26) and recalling the constraint $\nabla_\tau P(\mathbf{j}, k|\mathbf{X}_\tau, N, Z) \neq 0$ used to eliminate uninformative solutions, we conclude that the extrema (and therefore also the global minimum) of Eq. (24) are (is) independent of $\mathbf{X}_t$, as required.

2.6. Motion of the particle

We consider the limiting case that there is no uncertainty on the position of the particle, that is $x = X$ for all clicks. Then the motion of the particle and the motion of the positions of the detector clicks map one-to-one, for each repetition of the experiment (by assumption).

From the data $\mathbf{x}(t)$ we can compute the vector field $\mathbf{U}(\mathbf{x}, t)$ defined by

$$\frac{d\mathbf{x}}{dt} = \mathbf{U}(\mathbf{x}, t). \tag{27}$$

In principle, $\mathbf{U}(\mathbf{x}, t)$ is fully determined by the data obtained by repeating the experiment under different (initial) conditions. In practice, however, it is unlikely that we have enough data to compute $\mathbf{U}(\mathbf{x}, t)$ for all $(\mathbf{x}, t)$.

We only consider the case in which the position of the clicks is encoded by its (x, y, z)-coordinates in an orthogonal frame of reference attached to the observer. Under the usual assumptions of differentiability etc., we can use the Helmholtz-like decomposition of a vector field $\mathbf{U}(\mathbf{x}, t) = \nabla S(\mathbf{x}, t) - \nabla \times \mathbf{W}(\mathbf{x}, t)$. We will not use this form but write [30]

$$\mathbf{U}(\mathbf{x}, t) = \nabla S(\mathbf{x}, t) - \mathbf{A}(\mathbf{x}, t), \tag{28}$$

where $S(\mathbf{x}, t)$ is a scalar function and $\mathbf{A}(\mathbf{x}, t)$ a vector field. Clearly Eq. (28) has some extra freedom which we can remove by requiring that $\mathbf{A}(\mathbf{x}, t) = \nabla \times \mathbf{W}(\mathbf{x}, t)$. This amounts to requiring that $\nabla \cdot \mathbf{A} = 0$. It is convenient not do this at this stage so we take Eq. (28) and will impose $\nabla \cdot \mathbf{A} = 0$ later. As mentioned earlier, if differentiability is an issue we should use the finite-difference form of the ∇ operators.

For convenience, we drop the $(\mathbf{x}, t)$ arguments and switch to a component-wise notation in the few paragraphs that follow. From Eqs. (27) and (28) it directly follows that [30]

$$\frac{d^2 x_i}{dt^2} = \frac{\partial U_i}{\partial t} + \sum_{j=1}^{3} \frac{\partial U_i}{\partial x_j} U_j$$

$$= \frac{\partial^2 S}{\partial x_i \partial t} - \frac{\partial A_i}{\partial t} + \sum_{j=1}^{3} \left(\frac{\partial^2 S}{\partial x_i \partial x_j} - \frac{\partial A_i}{\partial x_j} \right) \left(\frac{\partial S}{\partial x_j} - A_j \right)$$

$$= \frac{\partial}{\partial x_i} \left[\frac{\partial S}{\partial t} + \frac{1}{2} \sum_{j=1}^{3} \left(\frac{\partial S}{\partial x_j} - A_j \right)^2 \right] + \sum_{j=1}^{3} \left(\frac{\partial A_j}{\partial x_i} - \frac{\partial A_i}{\partial x_j} \right) \left(\frac{\partial S}{\partial x_j} - A_j \right) - \frac{\partial A_i}{\partial t}, \tag{29}$$

where $i = 1, 2, 3$ labels the coordinate of the detector clicks.

Introducing the vector field $\mathbf{B} = \nabla \times \mathbf{A}$ the second term in Eq. (29) can we written as

$$\sum_{j=1}^{3} \left(\frac{\partial A_j}{\partial x_i} - \frac{\partial A_i}{\partial x_j} \right) \left(\frac{\partial S}{\partial x_j} - A_j \right) = \left(\frac{d\mathbf{x}}{dt} \times \mathbf{B} \right)_i. \tag{30}$$

It is important to note that in order to derive Eq. (30), it is essential that the position is represented by three coordinates. Switching back to the vector notation we have

$$\frac{d^2\mathbf{x}}{dt^2} = \nabla\left[\frac{\partial S}{\partial t} + \frac{1}{2}(\nabla S - \mathbf{A})^2\right] + \frac{d\mathbf{x}}{dt} \times \mathbf{B} - \frac{\partial \mathbf{A}}{\partial t}. \tag{31}$$

Up to now, we have not made any assumption other than that space is three-dimensional. Next comes a crucial step in the reasoning. Let us hypothesize that there exists a scalar field $\phi = \phi(\mathbf{x}, t)$ such that

$$\frac{\partial S}{\partial t} + \frac{1}{2}(\nabla S - \mathbf{A})^2 = -\phi. \tag{32}$$

Then, upon introducing the vector field $\mathbf{E} = -\nabla\phi - \partial\mathbf{A}/\partial t$, Eq. (31) becomes

$$\frac{d^2\mathbf{x}}{dt^2} = \mathbf{E} + \frac{d\mathbf{x}}{dt} \times \mathbf{B}. \tag{33}$$

Although Eq. (33) has the same the structure as the equation of motion of a charged particle in an electromagnetic field ($\mathbf{E}$, $\mathbf{B}$), our derivation of Eq. (33) is solely based on the elementary observation that the data yields the vector field $\mathbf{U}(\mathbf{x}, t)$ (see Eq. (28)), some standard vector-field identities and the hypothesis that there exist a scalar field ϕ such that Eq. (32) holds. No reference to charged particles or electromagnetic fields enters the derivation. Put differently (and putting aside technicalities related to differentiability), if there exist a scalar field ϕ such that Eq. (32) holds, then mathematics alone dictates that the equation of motion must have the structure Eq. (33), with $\mathbf{E}$ and $\mathbf{B}$ having no relation to the electromagnetic field acting on a charged particle. The latter relation is established when the data shows that there is indeed an effect of electromagnetic field on the motion of the particle, an effect from which it is inferred that the particle carries charge. This relation can be made explicit by introducing the symbols m for the mass and q for the charge of the particle and by replacing $\mathbf{A}$ by $q\mathbf{A}/m$ (we work with MKS units throughout this paper) and ϕ by $(q\phi + u)/m$ where u represent all potentials that are not of electromagnetic origin. Then we have

$$m\frac{d^2\mathbf{x}}{dt^2} = -\nabla u + q\mathbf{E} + q\frac{d\mathbf{x}}{dt} \times \mathbf{B}, \tag{34}$$

and upon replacing S by S/m and $V = q\phi + u$

$$\frac{\partial S}{\partial t} + \frac{1}{2m}(\nabla S - q\mathbf{A})^2 + V = 0. \tag{35}$$

Note that we have obtained the Hamilton–Jacobi equation (35) without making any reference to a Hamiltonian, the action, contact transformations and the like. In essence, Eqs. (28)–(35) follow from Eq. (27), some mathematical identities and the crucial assumption that there exists a V such that Eq. (35) holds. Summarizing:

> If we can find scalar fields S and V and a vector field $\mathbf{A}(\mathbf{x}, t)$ such that Eq. (35) holds for all $(\mathbf{x}, t)$ then the clicks of the detectors will carve out a trajectory that is completely determined by the classical equation of motion Eq. (34) of a particle in a potential and subject to electromagnetic potentials.

Of course, there is nothing really new in this statement: it is just telling us what we know from classical mechanics but there is a slight twist.

First, given the data $\mathbf{x}(t)$ of the detector clicks, this data will not comply with the equations of classical mechanics unless we can find scalar fields S (the action) and V (the potential) and a vector field $\mathbf{A}(\mathbf{x}, t)$ (vector potential) such that Eq. (35) holds. Second, in the case of interest to us here, there is uncertainty on the mapping between the particle position $\mathbf{X}(t)$ and the position of the corresponding

clicks $\mathbf{x}(t)$ and there is no reason to expect that Eq. (35) will hold. Instead of requiring that Eq. (35) holds, we will require that there exists two scalar fields $V_k(\mathbf{x}, t)$ for $k = \pm 1$ such that

$$\int d\mathbf{x}\, dt \sum_{k=\pm 1} \left[\frac{\partial S_k(\mathbf{x}, t)}{\partial t} + \frac{1}{2m} (\nabla S_k(\mathbf{x}, t) - q\mathbf{A}(\mathbf{x}, t))^2 + V_k(\mathbf{x}, t) \right] P(\mathbf{x}, k|\mathbf{X}, t, Z) = 0, \quad (36)$$

where we regard the particles that respond with $k = +1$ or $k = -1$ as two different objects, the clicks generated by each object being described by its own Hamilton–Jacobi equation with potentials $V_k(\mathbf{x}, t)$.

The next step is to disentangle as much as possible the motion of the positions of the clicks from their k-values. We introduce $S_k(\mathbf{x}, t) = S(\mathbf{x}, t) - kR(\mathbf{x}, t)$ for $k = \pm 1$ and after some rearrangements we obtain

$$
\begin{aligned}
\Lambda &= \int d\mathbf{x}\, dt \sum_{k=\pm 1} \left[\frac{\partial S_k(\mathbf{x}, t)}{\partial t} + \frac{1}{2m} (\nabla S_k(\mathbf{x}, t) - q\mathbf{A}(\mathbf{x}, t))^2 + V_k(\mathbf{x}, t) \right] P(\mathbf{x}, k|\mathbf{X}, t, Z) \\
&= \int d\mathbf{x}\, dt \left\{ \frac{1}{2m} \left[(\nabla S(\mathbf{x}, t) - q\mathbf{A}(\mathbf{x}, t))^2 + \left(\nabla R(\mathbf{x}, t) \right)^2 \right. \right. \\
&\quad \left. - 2\cos\theta(\mathbf{x}, \mathbf{X}, t, Z)\nabla R(\mathbf{x}, t) (\nabla S(\mathbf{x}, t) - q\mathbf{A}(\mathbf{x}, t)) \right] \\
&\quad + \left[\frac{\partial S(\mathbf{x}, t)}{\partial t} - \cos\theta(\mathbf{x}, \mathbf{X}, t, Z) \frac{\partial R(\mathbf{x}, t)}{\partial t} \right] + V_0(\mathbf{x}, t) \\
&\quad + V_1(\mathbf{x}, t)\cos\theta(\mathbf{x}, \mathbf{X}, t, Z) \Bigg\} P(\mathbf{x}|\mathbf{X}, t, Z),
\end{aligned}
\quad (37)
$$

where $V_0(\mathbf{x}, t) = [V_{+1}(\mathbf{x}, t) + V_{-1}(\mathbf{x}, t)]/2$, $V_1(\mathbf{x}, t) = [V_{+1}(\mathbf{x}, t) - V_{-1}(\mathbf{x}, t)]/2$ and we made use of $\sum_{k=\pm 1} kP(\mathbf{x}, k|\mathbf{X}, t, Z) = \cos\theta(\mathbf{x}, \mathbf{X}, t, Z)P(\mathbf{x}|\mathbf{X}, t, Z)$. Omitting the terms involving $\cos\theta(\mathbf{x}, \mathbf{X}, t, Z)$ and $R(\mathbf{x}, t)$, Eq. (37) reduces to the expression of the averaged Hamilton–Jacobi equation which entered the derivation of the time-dependent Schrödinger equation [8].

2.7. Including the motion of the magnetic moment

The function $\cos\theta(\mathbf{x}, \mathbf{X}, t, Z)$ determines the ratio of $k = \pm 1$ clicks and $R(\mathbf{x}, t) = (S_{-1}(\mathbf{x}, t) - S_{+1}(\mathbf{x}, t))/2$ is half of the difference between the actions of the $k = -1$ and $k = +1$ clicks. We can relate these two functions to the direction of a classical magnetic moment by imposing the constraint that when the positions of the clicks (=particle position in this case) do not move, we recover the classical–mechanical equation of motion of a magnetic moment in a magnetic field, for every $\mathbf{x}$.

In the limit that $m \to \infty$ (corresponding to the situation that the positions of the clicks hardly change with time) we have

$$
\lim_{m \to \infty} \Lambda = \int d\mathbf{x}\, dt \left\{ \left[\frac{\partial S(\mathbf{x}, t)}{\partial t} - \cos\theta(\mathbf{x}, \mathbf{X}, t, Z) \frac{\partial R(\mathbf{x}, t)}{\partial t} \right] + V_0(\mathbf{x}, t) \right.
$$
$$
\left. + V_1(\mathbf{x}, t)\cos\theta(\mathbf{x}, \mathbf{X}, t, Z) \right\} P(\mathbf{x}|\mathbf{X}, t, Z).
\quad (38)
$$

Without loss of generality, we may assume that $V_0(\mathbf{x}, t) = \widetilde{V}_0(\mathbf{x}, t) + \widehat{V}_0(\mathbf{x}, t)$ where $\widetilde{V}_0(\mathbf{x}, t)$ does not depend on $\theta(\mathbf{x}, \mathbf{X}, t, Z)$ and $R(\mathbf{x}, t)$ while $\widehat{V}_0(\mathbf{x}, t)$ may. Writing $\widehat{V}_1(\mathbf{x}, t) = \widehat{V}_0(\mathbf{x}, t) + V_1(\mathbf{x}, t)\cos\theta(\mathbf{x}, \mathbf{X}, t, Z)$, searching for the extrema of Eq. (38) through variation with respect to $\cos\theta(\mathbf{x}, \mathbf{X}, t, Z)$, $R(\mathbf{x}, t)$, $S(\mathbf{x}, t)$ and $P(\mathbf{x}, t)$ yields

$$\frac{\partial R(\mathbf{x}, t)}{\partial t} = \frac{\partial \widehat{V}_1(\mathbf{x}, t)}{\partial \cos\theta(\mathbf{x}, \mathbf{X}, t, Z)} \quad (39)$$

$$\frac{\partial \cos\theta(\mathbf{x}, \mathbf{X}, t, Z)}{\partial t} = -\frac{\partial \widehat{V}_1(\mathbf{x}, t)}{\partial R(\mathbf{x}, t)} \quad (40)$$

$$S(\mathbf{x}, t) \frac{\partial P(\mathbf{x}|\mathbf{X}, t, Z)}{\partial t} = 0 \tag{41}$$

$$\frac{\partial S(\mathbf{x}, t)}{\partial t} = \cos \theta(\mathbf{x}, \mathbf{X}, t, Z) \frac{\partial R(\mathbf{x}, t)}{\partial t} - \widehat{V}_0(\mathbf{x}, t) - \widehat{V}_1(\mathbf{x}, t). \tag{42}$$

From Eq. (41) it follows that $P(\mathbf{x}|\mathbf{X}, t, Z)$ does not change with time, in concert with the assumption that the positions of the clicks are stationary. Comparing Eqs. (39) and (40) with Eq. (C.7), it is clear that we will recover the classical equations of motion of the magnetic moment if (i) we set $\widehat{V}_1(\mathbf{x}, t) = -\gamma \mathbf{m}(\mathbf{x}, t) \cdot \mathbf{B}(\mathbf{x}, t)$ where $\mathbf{m}(\mathbf{x}, t)$ is a unit vector, and (ii) make the symbolic identification $z = \cos \theta(\mathbf{x}, \mathbf{X}, t, Z)$ and $\varphi(\mathbf{x}, t) = R(\mathbf{x}, t)/a$ where a needs to be introduced to give $a\varphi(\mathbf{x}, t)$ the dimension of $S(\mathbf{x}, t)$. Substituting the inferred expression for $\widehat{V}_1(\mathbf{x}, t)$ in Eq. (37) yields

$$\begin{aligned}
\Lambda = \int d\mathbf{x}\, dt \, \bigg\{ &\frac{1}{2m} \Big[(\nabla S(\mathbf{x}, t) - q\mathbf{A}(\mathbf{x}, t))^2 + a^2 \big(\nabla\varphi(\mathbf{x}, t)\big)^2 \\
&- 2a \cos\theta(\mathbf{x}, \mathbf{X}, t, Z) \nabla\varphi(\mathbf{x}, t) \, (\nabla S(\mathbf{x}, t) - q\mathbf{A}(\mathbf{x}, t)) \Big] \\
&+ \bigg[\frac{\partial S(\mathbf{x}, t)}{\partial t} - a \cos\theta(\mathbf{x}, \mathbf{X}, t, Z) \frac{\partial\varphi(\mathbf{x}, t)}{\partial t} \bigg] + V_0(\mathbf{x}, t) \\
&- a\gamma \mathbf{m}(\mathbf{x}, t) \cdot \mathbf{B}(\mathbf{x}, t) \bigg\} P(\mathbf{x}|\mathbf{X}, t, Z).
\end{aligned} \tag{43}$$

2.8. Derivation of the Pauli equation

We now have all ingredients to derive the Pauli equation from the principle that logical inference applied to the most robust experiment yields a quantum theoretical description [8]. According to this principle, we should search for the global minimum of the Fisher information for the experiment, subject to the condition that when the uncertainty vanishes, we recover the equations of motion of classical mechanics [8]. Thus, we should search for the global minimum of

$$F = \lambda I_F + \Lambda, \tag{44}$$

where I_F and Λ are given by Eqs. (15) and (43), respectively.

In Appendix B, it is shown that the quadratic functional Q which yields the Pauli equation is identical to Eq. (44) if we make the identification $V_0(\mathbf{x}, t) = q\phi(\mathbf{x}, t)$, $a = \hbar/2$, $\gamma = q/m$ and $\lambda = \hbar^2/8m$ and

$$\Phi(\mathbf{x}, t) = \begin{pmatrix} P^{1/2}(\mathbf{x}, k = +1|\mathbf{X}, t, Z) e^{iS_1(\mathbf{x}, t)/\hbar} \\ P^{1/2}(\mathbf{x}, k = -1|\mathbf{X}, t, Z) e^{iS_2(\mathbf{x}, t)/\hbar} \end{pmatrix}. \tag{45}$$

This then completes the derivation of the Pauli equation from logical inference principles.

2.9. Discussion

In Section 2.6, we showed how to include the knowledge that in the absence of uncertainty the particle's motion is described by Newtonian mechanics. Obviously, this treatment requires the particle to have a nonzero mass. On the other hand, in our logical inference treatment of the free particle in Section 2.5, the notion of mass does not enter in the derivation of Eq. (22) but neither does the concept of motion. This raises the interesting question how to inject into the logical inference treatment the notion of moving massless particles with spin. We believe that the analogy with the pseudo-spin in graphene mentioned in the introduction may provide a fruitful route to explore this issue.

The carbon atoms of ideal single-layer graphene form a hexagonal lattice with the π-band (originating from p_z-orbitals of carbon atoms) well separated from other bands [18]. The electronic

band structure of graphene has the remarkable feature that in the continuum limit, the low-energy excitations are described by the two-dimensional Dirac equation for two species of massless fermions (corresponding to two valleys, K and K'). The fact that the wave function of each of these two species is a two-component "spinor" is not related to the intrinsic spin of the electron but is a manifestation of the two sub-lattice and bipartite structure of the hexagonal lattice [18]. This feature (Dirac-like spectrum) is present already in the simplest model in which only the nearest-neighbor hopping is taken into account [31] but, actually, it is robust and follows just from discrete symmetries, namely, time-reversal and inversion symmetries [18]. A generalization to a 4-dimensional lattice, retaining the property that the continuum limit yields the Dirac equation, is given in Ref. [32]. This is a nice illustration of the fact that the model of a rotating electron is not the only way to arrive at the concept of spin. In our derivation of the Pauli equation, we have to make the additional assumption (based on experimental observations such as the anomalous Zeeman effect) that the interaction of this intrinsic degree of freedom with an external magnetic field is described by the standard classical expression for the energy of a magnetic moment.

The next important step might be the derivation of the Dirac equation. The Creutz model [32] suggests that we should consider incorporating into the logical inference treatment, the additional knowledge that one has objects hopping on a lattice instead of particles moving in a space-time continuum. Recall that up to Section 2.4, the description of the measurement scenario, robustness etc. is explicitly discrete. In Section 2.4, the continuum limit was taken only because our aim was to derive the Pauli equation, which is formulated in continuum space-time. Of course, the description of the motion of the particle in Section 2.6 is entirely within a continuum description but there is no fundamental obstacle to replace this treatment by a proper treatment of objects hopping on a lattice. Therefore it seems plausible that the logical inference approach can be extended to describe massless spin-1/2 particles moving in continuum space-time by considering the continuum limit of the corresponding lattice model. An in-depth, general treatment of this problem is beyond the scope of the present paper and we therefore leave this interesting problem for future research.

A comment on the appearance of $\hbar$ is in order. First of all, it should be noted that recent work has shown that $\hbar$ may be eliminated from the basic equations of (low-energy) physics by a re-definition of the units of mass, time, etc. [33,34]. This is also clear from the way $\hbar$ appears in the identification that we used to show that the quadratic functional Q which yields the Pauli equation (see Eq. (B.4)) is the same as Eq. (44). With the MKS units adopted in the present paper, Planck's constant $\hbar$ enters because of dimensional reasons ($a = \hbar/2$) and also controls the importance of the term that expresses the robustness of the experimental procedure ($\lambda = \hbar^2/8m$). The actual value of λ can only be determined by laboratory experiments. Note that the logical-inference derivation of the canonical ensemble of statistical mechanics [6,7] employs the same reasoning to relate the inverse temperature $\beta = 1/k_BT$ to the average thermal energy.

We end this section by addressing a technicality. Mappings such as Eq. (45) are not one-to-one. This is clear: we can always add a multiple of $2\pi\hbar$ to $S_1(\mathbf{x}, t)$ or $S_2(\mathbf{x}, t)$, for instance. In the hydrodynamic form of the Schrödinger equation [35], the ambiguity that ensues has implications for the interpretation of the gradient of action as a velocity field [36,37]. As pointed out by Novikov, similar ambiguities appear in classical mechanics proper if the local equations of motion (Hamilton equations) are not sufficient to characterize the system completely and the global structure of the phase space has to be taken into consideration [38]. However, for the present purpose, this ambiguity has no effect on the minimization of F because Eq. (44) does not change if we add to $S_1(\mathbf{x}, t)$ or $S_2(\mathbf{x}, t)$ a real number which does not depend on $(\mathbf{x}, t)$ (as is evident from Eq. (37)) or, equivalently, if we multiply $\Phi(x|\mathbf{X}, t, Z)$ by a global phase factor and add a constant to $\varphi(\mathbf{x}, t)$.

3. Stern–Gerlach experiment: neutral magnetic particle

The Stern–Gerlach experiment with silver atoms [39] and neutrons [15] demonstrates that a magnetic field affects the motion of a neutral particle, suggesting that a minimalist theoretical description should account for the interaction of the magnetic moment of the particle and the applied magnetic field. As is clear from the definition of the Pauli Hamiltonian equation (B.2), in the Pauli

equation the magnetic field is directly linked to the charge q of the particle. Therefore, in this form the Pauli equation cannot be used to describe the motion of a neutral magnetic particle in a magnetic field.

In quantum theory, this problem is solved by the ad-hoc introduction of the intrinsic magnetic moment which is proportional to the spin and by replacing $q\hbar/2m$ by the gyromagnetic ratio γ, the value of which is particle-specific.

In the logical-inference treatment, no such ad-hoc procedure is necessary. We simply set $q = 0$ in Eq. (43) and use Eq. (45) to find the equivalent quadratic form. The Hamiltonian that appears in this quadratic form reads

$$H = -\frac{\hbar^2}{2m}\nabla^2 - \gamma\boldsymbol{\sigma} \cdot \mathbf{B}(\mathbf{x}, t), \tag{46}$$

where γ is the gyromagnetic ratio which, in general, is not given by q/m. As mentioned earlier, the appearance in Eq. (46) of the Pauli-matrices is a direct consequence of logical inference applied to robust experiments that yield data in the form of the position and one of the two kinds of detector clicks.

4. Relation to earlier work

Readers familiar with the hydrodynamic formulation of quantum theory [35] and its interpretation in terms of Bohmian mechanics [40,41] undoubtedly recognize the steps which transform quadratic functional equation (B.4) yielding the Pauli equation (B.2) and the functional Q given by Eq. (B.27). In fact, the functional Q has been used as the starting point for the hydrodynamic representation [42] and a causal interpretation [12,43,44] of the Pauli equation. In this formulation, the two-component spinor can be given a classical–mechanical interpretation in terms of an assembly of very small rotating bodies which are distributed continuously in space. Within this interpretation spins of different bodies interact.

Clearly, the logical-inference treatment does not support this interpretation: the functional equation (B.27) is the result of analyzing a robust experiment that yields data in the form of $(\mathbf{x}, k)$ where $\mathbf{x}$ is a 3-dimensional coordinate and $k = \pm1$ denotes the two-valued "color", together with the requirement that on average and in special cases, the data should comply with the classical–mechanical motion.

An expression of Eq. (B.27) in which the separation of the contribution of the Fisher information and the classical–mechanical fields is explicit has been given by Reginatto [45]. This expression is different from ours. Comparing Eq. (15) with Eq. (6, 7) in Ref. [45], we find that the expressions are fundamentally different due to the fact that the representation (7), when substituted in (6), does not yield Eq. (B.27).

5. Conclusion

It is somewhat discomforting that it takes a considerable amount of symbolic manipulations to derive the Pauli equation from the combination of the measurement scenario, the notion of a robust experiment and the behavior expected in some limiting cases. Therefore, it may be worthwhile to recapitulate what has been done in simple words, without worrying too much about the technicalities.

The first step is to describe the measurement scenario. It is assumed that the object (particle) we are interested in responds to the signal that we send to probe it. The response of the object triggers a detection event. In the case at hand, the data representing the detector clicks consist of spatial coordinates and two-valued "color" indices. We assign an i-prob to the whole data set. To make progress, it is necessary to make assumptions about the data-collection procedure. We assume that each time we probe the object, the data produced by the detection system is logically independent of all other data produced by previous/subsequent probing. With this assumption, together with the assumption that is does not matter where we carry out the experiment, the notion of a robust experiment is found to be equivalent to the global minimum of the Fisher information for the corresponding measurement scenario (see Eq. (15)).

The next step is to bring in the knowledge that in the extreme case that there is no uncertainty about the outcome of each detection event, we expect to observe data that is compliant with classical,

Newtonian mechanics both for the motion of a particle as well as for the motion of its magnetic moment in the case that the particle does not move (see Eq. (43)).

The third step is to find the balance between the uncertainty in the detection events represented by Eq. (15) and the "classical mechanics" knowledge represented by Eq. (43) by searching for the global minimum of Eq. (44) for all possible unknown positions of the particle. The result of this calculation is a fairly complicated non-linear set of equations for the i-prob to observe a click.

The final step is to observe that by transformation equation (45), this non-linear set of equations and the Pauli equation are equivalent. The latter, being a set of linear equations, is much easier to solve than its non-linear equivalent.

In the logical inference approach, the assumption that each time we probe the object, the detection system reports a two-valued "color" index and our requirement that in the extreme case mentioned earlier we expect to see the motion of a classical magnetic moment automatically leads to the notion of a "quantized" (i.e. two-valued) intrinsic magnetic moment. The notion of spin appears as an inference, forced upon us by the (two-valued) data and our assumptions (which do not make reference to concepts of quantum theory) that the experiment is robust, etc.

From a more general perspective, it is remarkable that the logic inference approach introduces the concept of "spin" in a way which is not much different from the way real numbers are introduced. Indeed, the latter appear as a necessity to provide an answer to questions such as "what new kind of number do we have to introduce such that the square of it yields the integer n". If $n = m^2$ where m is an integer, no new concept has to be introduced but if say $n = 2$, the answer to the question is given the symbolic name $\sqrt{2}$.

Similarly, in our logical-inference treatment the concept of spin naturally appears as a result of describing situations in which there is two-valued data and the requirement that in a limiting case we recover the classical equation of motion. This concept of spin only exists in our mind, in complete agreement with the fact that this concept maybe put to very good use whenever there are two-valued variables that may or may not relate to (intrinsic) angular momentum, as in the theory of the electronic properties of graphene, for example [18].

It will not have escaped the reader that in the logical-inference derivation of the Pauli equation as well as in earlier work along this line [8,46] there are no postulates regarding "wavefunctions", "observables", "quantization rules", "quantum measurements" [47], "Born's rule", etc. This is a direct consequence of the basic premise of this approach, namely that current scientific knowledge derives, through cognitive processes in the human brain, from the discrete events which are observed in laboratory experiments and from relations between those events that we, humans, discover. These discrete events are not "generated" according to certain quantum laws: instead these laws appear as the result of (the best) inference based on data available in the form of discrete events. In essence, for all the basic but fundamental cases treated so far, the machinery of quantum theory appears as a result of transforming a set of non-linear equations into a set on linear ones. The wavefunction, spinor, spin, …, are all mathematical concepts, vehicles that render a class of complicated nonlinear minimization problems into the minimization of quadratic forms. As products of our collective imagination, these concepts are extraordinarily useful but have no tangible existence, just like numbers themselves. Of course, it remains to be seen whether the logical-inference approach can be extended to e.g. many-body and relativistic quantum physics.

In summary: the Pauli equation derives from logical inference applied to robust experiments in which there is uncertainty about individual detection events which yield information about the particle position and its two-valued "color". This derivation adds another, new instance to the list of examples [8,46] for which the logical-inference approach establishes a bridge between objective knowledge gathered through experiments and their description in terms of concepts.

Acknowledgments

Appendix A. The algebra of logical inference

This appendix does not contain any original material but is provided to render the present paper self-contained.

If we are only concerned about quantifying the truth of a proposition given the truth of another proposition, it is possible to construct a mathematical framework, an extension of Boolean logic, that allows us to reason in a manner which is unambiguous and independent of the individual, in particular if there are elements of uncertainty in the description [1–5].

The algebra of logical inference can be derived [2–5] from three so-called "desiderata". The formulation which follows is taken from Ref. [4].

Desideratum 1. *Plausibilities are represented by real numbers.* The plausibility that a proposition A is true conditional on proposition B being true will be denoted by $P(A|B)$.

Desideratum 2. *Plausibilities must exhibit agreement with rationality.* As more and more evidence supporting the truth of a proposition becomes available, the plausibility should increase monotonically and continuously and the plausibility of the negation of the proposition should decrease monotonically and continuously. Moreover, in the limiting case that proposition A is known to be either true or false, the plausibility $P(A|B)$ should conform to the rules of deductive reasoning. In other words, plausibilities must be in qualitative agreement with the patterns of plausible reasoning uncovered by Pólya [48].

Desideratum 3. *All rules relating plausibilities must be consistent.* Consistency of rational reasoning demands that if the rules of logical inference allow a plausibility to be obtained in more than one way, the result should not depend on the particular sequence of operations.

These three desiderata only describe the essential features of the plausibilities and definitely do not constitute a set of axioms which plausibilities have to satisfy.

It is a most remarkable fact that these three desiderata suffice to uniquely determine the set of rules by which plausibilities may be manipulated [2–5]. Omitting the derivation, it follows that plausibilities may be chosen to take numerical values in the range [0, 1] and obey the rules [2–5]:

1. $P(A|Z) + P(\bar{A}|Z) = 1$ where $\bar{A}$ denotes the negation of proposition A and Z is a proposition assumed to be true.
2. $P(AB|Z) = P(A|BZ)P(B|Z) = P(B|AZ)P(A|Z)$ where the "product" BZ denotes the logical product (conjunction) of the propositions B and Z, that is the proposition BZ is true if both B and Z are true. This rule will be referred to as "product rule". It should be mentioned here that it is not allowed to define a plausibility for a proposition conditional on the conjunction of mutual exclusive propositions. Reasoning on the basis of two or more contradictory premises is out of the scope of the present paper.
3. $P(A\bar{A}|Z) = 0$ and $P(A + \bar{A}|Z) = 1$ where the "sum" $A + B$ denotes the logical sum (inclusive disjunction) of the propositions A and B, that is the proposition $A + B$ is true if either A or B or both are true. These two rules show that Boolean algebra is contained in the algebra of plausibilities.

The algebra of logical inference, as defined by the rules (1–3), is the foundation for powerful tools such as the maximum entropy method and Bayesian analysis [3,5]. The rules (1–3) are unique [3–5]. Any other rule which applies to plausibilities represented by real numbers and is in conflict with rules (1–3) will be at odds with rational reasoning and consistency, as embodied by the desiderata 1–3.

The rules (1–3) are identical to the rules by which we manipulate probabilities [5,49–51]. However, the rules (1–3) were not postulated. They were derived from general considerations about rational reasoning and consistency only. Moreover, concepts such as sample spaces, probability measures etc., which are an essential part of the mathematical foundation of probability theory [50,51], play no role in the derivation of rules (1–3). Perhaps most important in the context of quantum theory is that in the logical inference approach uncertainty about an event does not imply that this event can be represented by a random variable as defined in probability theory [51].

There is a significant conceptual difference between "mathematical probabilities" and plausibilities. Mathematical probabilities are elements of an axiomatic framework which complies with the algebra of logical inference. Plausibilities are elements of a language which also complies with the

algebra of logical inference and serve to facilitate communication, in an unambiguous and consistent manner, about phenomena in which there is uncertainty.

The plausibility $P(A|B)$ is an intermediate mental construct that serves to carry out inductive logic, that is rational reasoning, in a mathematically well-defined manner [3]. In general, $P(A|B)$ may express the degree of believe of an individual that proposition A is true, given that proposition B is true. However, in the present paper, we explicitly exclude applications of this kind because they do not comply with our main goal, namely to describe phenomena "in a manner independent of individual subjective judgment".

> To take away this subjective connotation of the word "plausibility", we call $P(A|B)$ the "inference-probability" or "i-prob" for short.

A comment on the notation used throughout this paper is in order. To simplify the presentation, we make no distinction between an event such as "detector D has fired" and the corresponding proposition "D = detector D has fired". If we have two detectors, say D_x where $x = \pm 1$, we write $P(x|Z)$ to denote the i-prob of the proposition that detector D_x fires, given that proposition Z is true. Similarly, the i-prob of the proposition that two detectors D_x and D_y fire, given that proposition Z is true, is denoted by $P(x, y|Z)$. Obviously, this notation generalizes to more than two propositions.

Appendix B. Pauli equation: quantum theory

In this appendix, we show that the quadratic form, the minimization of which yields the Pauli equation, is identical to the one derived in Section 2 through logical inference.

The Pauli equation for a particle with mass m and charge q can be written as

$$i\hbar \frac{\partial}{\partial t} \Phi = H\Phi, \tag{B.1}$$

where

$$\Phi = \Phi(\mathbf{x}, t) = \begin{pmatrix} \Phi_1(\mathbf{x}, t) \\ \Phi_2(\mathbf{x}, t) \end{pmatrix}, \tag{B.2}$$

is a two-component wavefunction and the Hamiltonian is given by

$$H = \frac{1}{2m} \{\boldsymbol{\sigma} \cdot [-i\hbar\boldsymbol{\nabla} - q\mathbf{A}(\mathbf{x}, t)]\}^2 + q\phi(\mathbf{x}, t)$$

$$= \frac{1}{2m} [-i\hbar\boldsymbol{\nabla} - q\mathbf{A}(\mathbf{x}, t)]^2 + q\phi(\mathbf{x}, t) - \frac{q\hbar}{2m}\boldsymbol{\sigma} \cdot \mathbf{B}(\mathbf{x}, t), \tag{B.3}$$

where $\boldsymbol{\sigma} = (\sigma^x, \sigma^y, \sigma^z)^T$ denote the Pauli-spin matrices.

By the standard variational argument, it follows that the Pauli equation is an extremum of the quadratic form (functional)

$$Q = \int d\mathbf{x}\, dt \left[\frac{i\hbar}{2} \left(\frac{\partial \Phi^\dagger}{\partial t} \Phi - \Phi^\dagger \frac{\partial \Phi}{\partial t} \right) + \frac{1}{2m} \left(+i\hbar\boldsymbol{\nabla}\Phi^\dagger - q\mathbf{A}(\mathbf{x}, t)\Phi^\dagger \right) \right.$$

$$\left. \times \left(-i\hbar\boldsymbol{\nabla}\Phi - q\mathbf{A}(\mathbf{x}, t)\Phi \right) + q\phi(\mathbf{x}, t)\Phi^\dagger\Phi - \frac{q\hbar}{2m}\Phi^\dagger\boldsymbol{\sigma} \cdot \mathbf{B}(\mathbf{x}, t)\Phi \right], \tag{B.4}$$

with respect to variations in $\Phi^\dagger$. We want to know how Eq. (B.4) looks like when we substitute the polar representation

$$\Phi = \begin{pmatrix} \sqrt{P_1(\mathbf{x}, t)}e^{iS_1(\mathbf{x}, t)/\hbar} \\ \sqrt{P_2(\mathbf{x}, t)}e^{iS_2(\mathbf{x}, t)/\hbar} \end{pmatrix}, \tag{B.5}$$

for both components of the spinor. We have

$$\frac{\partial \Phi_k(\mathbf{x}, t)}{\partial t} = \left[\frac{1}{2P_k(\mathbf{x}, t)} \frac{\partial P_k(\mathbf{x}, t)}{\partial t} + \frac{i}{\hbar} \frac{\partial S_k(\mathbf{x}, t)}{\partial t} \right] \sqrt{P_k(\mathbf{x}, t)}e^{iS_k(\mathbf{x}, t)/\hbar}, \tag{B.6}$$

for $k = 1, 2$ and

$$\Phi^\dagger \frac{\partial \Phi}{\partial t} = \sum_{k=1}^{2} \left[\frac{1}{2P_k(\mathbf{x}, t)} \frac{\partial P_k(\mathbf{x}, t)}{\partial t} + \frac{i}{\hbar} \frac{\partial S_k(\mathbf{x}, t)}{\partial t} \right] P_k, \tag{B.7}$$

from which it directly follows that

$$\frac{i\hbar}{2} \left(\frac{\partial \Phi^\dagger}{\partial t} \Phi - \Phi^\dagger \frac{\partial \Phi}{\partial t} \right) = \frac{\partial S_1(\mathbf{x}, t)}{\partial t} P_1(\mathbf{x}, t) + \frac{\partial S_2(\mathbf{x}, t)}{\partial t} P_2(\mathbf{x}, t). \tag{B.8}$$

Likewise we have

$$- i\hbar \left(\nabla \Phi_k - \frac{iq}{\hbar} \mathbf{A}(\mathbf{x}, t) \Phi_k \right) = -i\hbar \left[\frac{1}{2P_k(\mathbf{x}, t)} \nabla P_k(\mathbf{x}, t) + \frac{i}{\hbar} \nabla S_k(\mathbf{x}, t) - \frac{iq}{\hbar} \mathbf{A}(\mathbf{x}, t) \right]$$
$$\times \sqrt{P_k(\mathbf{x}, t)} e^{iS_k(\mathbf{x}, t)/\hbar}, \tag{B.9}$$

and

$$\frac{\hbar^2}{2m} \left(\nabla \Phi^\dagger + \frac{iq}{\hbar} \mathbf{A}(\mathbf{x}, t) \Phi^\dagger \right) \left(\nabla \Phi - \frac{iq}{\hbar} \mathbf{A}(\mathbf{x}, t) \Phi \right)$$

$$= \frac{\hbar^2}{2m} \sum_{k=1}^{2} \left\{ \frac{1}{4P_k^2(\mathbf{x}, t)} \left[\nabla P_k(\mathbf{x}, t) \right]^2 + \frac{1}{\hbar^2} \left[\nabla S_k(\mathbf{x}, t) - q\mathbf{A}(\mathbf{x}, t) \right]^2 \right\} P_k(\mathbf{x}, t). \tag{B.10}$$

Furthermore, it follows that

$$\phi(\mathbf{x}, t) \Phi^\dagger \Phi = \phi(\mathbf{x}, t) \left[P_1(\mathbf{x}, t) + P_2(\mathbf{x}, t) \right], \tag{B.11}$$

$$\Phi^\dagger \sigma^x \Phi = 2\sqrt{P_1(\mathbf{x}, t) P_2(\mathbf{x}, t)} \cos \frac{S_2(\mathbf{x}, t) - S_1(\mathbf{x}, t)}{\hbar}, \tag{B.12}$$

$$\Phi^\dagger \sigma^y \Phi = 2\sqrt{P_1(\mathbf{x}, t) P_2(\mathbf{x}, t)} \sin \frac{S_2(\mathbf{x}, t) - S_1(\mathbf{x}, t)}{\hbar}, \tag{B.13}$$

$$\Phi^\dagger \sigma^z \Phi = P_1(\mathbf{x}, t) - P_2(\mathbf{x}, t). \tag{B.14}$$

Thus, we have all the expressions to write Eq. (B.4) in terms of $P_1(\mathbf{x}, t)$, $P_2(\mathbf{x}, t)$, $S_1(\mathbf{x}, t)$, and $S_2(\mathbf{x}, t)$.

Without loss of generality and without assigning a particular meaning to the new symbols yet, we write

$$P_1(\mathbf{x}, t) = P(\mathbf{x}, t) \cos^2 \frac{\theta(\mathbf{x}, t)}{2}$$
$$P_2(\mathbf{x}, t) = P(\mathbf{x}, t) \sin^2 \frac{\theta(\mathbf{x}, t)}{2}. \tag{B.15}$$

Then we have

$$\sum_{k=1}^{2} \frac{1}{4P_k(\mathbf{x}, t)} (\nabla P_k(\mathbf{x}, t))^2$$

$$= \frac{1}{4} \left\{ \frac{1}{P_1(\mathbf{x}, t)} \left[\nabla P(\mathbf{x}, t) \cos^2 \frac{\theta(\mathbf{x}, t)}{2} - P(\mathbf{x}, t) \cos \frac{\theta(\mathbf{x}, t)}{2} \sin \frac{\theta(\mathbf{x}, t)}{2} \nabla \theta(\mathbf{x}, t) \right]^2 \right.$$

$$+ \frac{1}{P_2(\mathbf{x}, t)} \left[\nabla P(\mathbf{x}, t) \sin^2 \frac{\theta(\mathbf{x}, t)}{2} + P(\mathbf{x}, t) \cos \frac{\theta(\mathbf{x}, t)}{2} \sin \frac{\theta(\mathbf{x}, t)}{2} \nabla \theta(\mathbf{x}, t) \right]^2 \right\}$$

$$= \frac{1}{4} \left\{ \frac{1}{P(\mathbf{x}, t)} \left[\nabla P(\mathbf{x}, t) \right]^2 + P(\mathbf{x}, t) \left[\nabla \theta(\mathbf{x}, t) \right]^2 \right\}. \tag{B.16}$$

Similarly,

$$\phi(\mathbf{x}, t)\Phi^\dagger\Phi = \phi(\mathbf{x}, t)P(\mathbf{x}, t), \tag{B.17}$$

$$\Phi^\dagger\sigma^x\Phi = P(\mathbf{x}, t)\sin\theta(\mathbf{x}, t)\cos\frac{S_2(\mathbf{x}, t) - S_1(\mathbf{x}, t)}{\hbar}, \tag{B.18}$$

$$\Phi^\dagger\sigma^y\Phi = P(\mathbf{x}, t)\sin\theta(\mathbf{x}, t)\sin\frac{S_2(\mathbf{x}, t) - S_1(\mathbf{x}, t)}{\hbar}, \tag{B.19}$$

$$\Phi^\dagger\sigma^z\Phi = P(\mathbf{x}, t)\cos\theta(\mathbf{x}, t). \tag{B.20}$$

Next, we introduce

$$S_1(\mathbf{x}, t) = S(\mathbf{x}, t) - \frac{\hbar\varphi(\mathbf{x}, t)}{2},$$

$$S_2(\mathbf{x}, t) = S(\mathbf{x}, t) + \frac{\hbar\varphi(\mathbf{x}, t)}{2}, \tag{B.21}$$

and obtain

$$\frac{i\hbar}{2}\left(\frac{\partial\Phi^\dagger}{\partial t}\Phi - \Phi^\dagger\frac{\partial\Phi}{\partial t}\right) = \frac{\partial S(\mathbf{x}, t)}{\partial t}[P_1(\mathbf{x}, t) + P_2(\mathbf{x}, t)] - \frac{\hbar}{2}\frac{\partial\varphi(\mathbf{x}, t)}{\partial t}[P_1(\mathbf{x}, t) - P_2(\mathbf{x}, t)]$$

$$= \left[\frac{\partial S(\mathbf{x}, t)}{\partial t} - \frac{\hbar}{2}\cos\theta(\mathbf{x}, t)\frac{\partial\varphi(\mathbf{x}, t)}{\partial t}\right]P(\mathbf{x}, t), \tag{B.22}$$

$$\sum_{k=1}^2\left[\nabla S_k(\mathbf{x}, t) - q\mathbf{A}(\mathbf{x}, t)\right]^2 P_k(\mathbf{x}, t)$$

$$= P(\mathbf{x}, t)\sum_{k=1}^2\frac{1 - (-1)^k\cos\theta(\mathbf{x}, t)}{2}\left[\nabla S(\mathbf{x}, t) - q\mathbf{A}(\mathbf{x}, t) + \frac{(-\hbar)^k}{2}\nabla\varphi(\mathbf{x}, t)\right]^2$$

$$= \left\{\left[\nabla S(\mathbf{x}, t) - q\mathbf{A}(\mathbf{x}, t)\right]^2 + \frac{\hbar^2}{4}\left[\nabla\varphi(\mathbf{x}, t)\right]^2\right\}P(\mathbf{x}, t)$$

$$- \hbar\cos\theta(\mathbf{x}, t)\nabla\varphi(\mathbf{x}, t)\left[\nabla S(\mathbf{x}, t) - q\mathbf{A}(\mathbf{x}, t)\right]P(\mathbf{x}, t), \tag{B.23}$$

$$\Phi^\dagger\sigma^x\Phi = P(\mathbf{x}, t)\sin\theta(\mathbf{x}, t)\cos\varphi(\mathbf{x}, t), \tag{B.24}$$

$$\Phi^\dagger\sigma^y\Phi = P(\mathbf{x}, t)\sin\theta(\mathbf{x}, t)\sin\varphi(\mathbf{x}, t), \tag{B.25}$$

$$\Phi^\dagger\sigma^z\Phi = P(\mathbf{x}, t)\cos\theta(\mathbf{x}, t). \tag{B.26}$$

Collecting all terms we find

$$Q = \int d\mathbf{x}\,dt\left(\frac{\hbar^2}{8m}\left\{\frac{1}{P(\mathbf{x}, t)}[\nabla P(\mathbf{x}, t)]^2 + [\nabla\theta(\mathbf{x}, t)]^2 P(\mathbf{x}, t)\right\}\right.$$

$$+ \left[\frac{\partial S(\mathbf{x}, t)}{\partial t} - \frac{\hbar}{2}\cos\theta(\mathbf{x}, t)\frac{\partial\varphi(\mathbf{x}, t)}{\partial t}\right]P(\mathbf{x}, t)$$

$$+ \frac{1}{2m}\left\{\left[\nabla S(\mathbf{x}, t) - q\mathbf{A}(\mathbf{x}, t)\right]^2 + \frac{\hbar^2}{4}[\nabla\varphi(\mathbf{x}, t)]^2\right.$$

$$\left.- \hbar\cos\theta(\mathbf{x}, t)\nabla\varphi(\mathbf{x}, t)\left[\nabla S(\mathbf{x}, t) - q\mathbf{A}(\mathbf{x}, t)\right]\right\}P(\mathbf{x}, t)$$

$$+ q\phi(\mathbf{x}, t)P(\mathbf{x}, t) - \frac{q\hbar}{2m}\left[B_x\sin\theta(\mathbf{x}, t)\cos\varphi(\mathbf{x}, t)\right.$$

$$\left.+ B_y\sin\theta(\mathbf{x}, t)\sin\varphi(\mathbf{x}, t) + B_z\cos\theta(\mathbf{x}, t)\right]P(\mathbf{x}, t)\right). \tag{B.27}$$

Appendix C. Classical mechanics of a magnetic moment

For completeness, we collect some well-known facts about the classical–mechanical description of the rotational motion of a magnetic moment $\mathbf{M} = \mathbf{M}(t) = (M_x, M_y, M_z)^T$ which does not move and interacts with a magnetic field $\mathbf{B} = \mathbf{B}(t) = (B_x, B_y, B_z)^T$ [44]. The motion of the magnetic moment is completely determined by the torque equation

$$\frac{d\mathbf{M}}{dt} = \gamma \mathbf{M} \times \mathbf{B}, \tag{C.1}$$

where γ is the gyromagnetic ratio. In terms of components we have

$$\frac{dM_x}{dt} = \gamma (M_y B_z - M_z B_y)$$

$$\frac{dM_y}{dt} = \gamma (M_z B_x - M_x B_z)$$

$$\frac{dM_z}{dt} = \gamma (M_x B_y - M_y B_x). \tag{C.2}$$

The total magnetic moment $M_0 = (M_x^2 + M_y^2 + M_z^2)^{1/2}$ is constant in time. Therefore we can write $\mathbf{M} = M_0 \mathbf{m}$ where

$$\mathbf{m} = (\cos \varphi \sin \theta, \sin \varphi \sin \theta, \cos \theta)^T, \tag{C.3}$$

is a unit vector and (θ, φ) are its spherical coordinates. The equations of motion of these coordinates read

$$\begin{aligned}
\frac{d\varphi}{dt} &= \frac{d \arctan(m_y/m_x)}{dt} \\
&= \frac{\gamma}{m_x^2 + m_y^2} \left(m_x \frac{dm_y}{dt} - m_y \frac{dm_x}{dt} \right) \\
&= \gamma \left[-B_z + \frac{m_z}{m_x^2 + m_y^2} (m_x B_x + m_y B_y) \right] \\
&= \gamma \left[-B_z + \frac{z}{\sqrt{1 - z^2}} (B_x \cos \varphi + B_y \sin \varphi) \right],
\end{aligned} \tag{C.4}$$

where $z = \cos \theta$ and

$$\begin{aligned}
\frac{dz}{dt} &= \gamma (m_x B_y - m_y B_x) \\
&= \gamma \left[\sqrt{1 - z^2} (-B_x \sin \varphi + B_y \cos \varphi) \right].
\end{aligned} \tag{C.5}$$

If we define a Hamiltonian H_M by

$$\begin{aligned}
H_M &= -\gamma \left(B_x \sin \theta \cos \varphi + B_y \sin \theta \sin \varphi + B_z \cos \theta \right) \\
&= -\gamma \left[z B_z + \sqrt{1 - z^2} (B_x \cos \varphi + B_y \sin \varphi) \right],
\end{aligned} \tag{C.6}$$

it follows that

$$\begin{aligned}
\frac{d\varphi}{dt} &= +\frac{\partial H_M}{\partial z} \\
\frac{dz}{dt} &= -\frac{\partial H_M}{\partial \varphi}.
\end{aligned} \tag{C.7}$$

From Eq. (C.7), it follows that the pair (φ, z) are conjugate variables.

Finally, it is easy to check that the equations of motion Eq. (C.7) follow by searching for the extremum of the functional

$$\mathcal{M} = \int dt \ \left(-z \frac{d\varphi}{dt} + H_M \right).$$
(C.8)

References

[1] R.T. Cox, Amer. J. Phys. 14 (1946) 1–13.
[2] R.T. Cox, The Algebra of Probable Inference, Johns Hopkins University Press, Baltimore, 1961.
[3] M. Tribus, Rational Descriptions, Decisions and Designs, Expira Press, Stockholm, 1999.
[4] C.R. Smith, G. Erickson, in: J. Skilling (Ed.), Maximum Entropy and Bayesian Methods, Kluwer Academic Publishers, Dordrecht, 1989, pp. 29–44.
[5] E.T. Jaynes, Probability Theory: The Logic of Science, Cambridge University Press, Cambridge, 2003.
[6] E.T. Jaynes, Phys. Rev. 106 (1957) 620–640.
[7] E.T. Jaynes, Phys. Rev. 108 (1957) 171–190.
[8] H. De Raedt, M.I. Katsnelson, K. Michielsen, Ann. Phys. 347 (2014) 45–73.
[9] B.L. van der Waerden, in: M. Fierz, V.W. Weisskopf (Eds.), Theoretical Physics in the Twentieth Century: A Memorial Volume to Wolfgang Pauli, Interscience, New York, 1960.
[10] W. Pauli, Z. Phys. 31 (1924) 373–385.
[11] W. Pauli, Z. Phys. 43 (1927) 601.
[12] D. Bohm, R. Schiller, J. Tiomno, Nuovo Cimento Series 10 (1) (1955) 48–66.
[13] T. Takabayasi, Progr. Theoret. Phys. 14 (1955) 232.
[14] H. Batelaan, T.J. Gay, J.J. Schwendiman, Phys. Rev. Lett. 79 (1997) 4517–4521.
[15] J.E. Sherwood, T.K. Stephenson, S. Bernstein, Phys. Rev. 96 (1954) 1546–1548.
[16] D. Griffiths, Introduction to Elementary Particles, Wiley, New York, 1987.
[17] W. Heisenberg, Z. Phys. 77 (1932) 1–11.
[18] M.I. Katsnelson, Graphene: Carbon in Two Dimensions, Cambridge Univ. Press, Cambridge, 2012.
[19] A. Khrennikov, Beyond Quantum, Pan Stanford Publishing, Singapore, 2014.
[20] G. 't Hooft, Found. Phys. Lett. 10 (1997) 105–111.
[21] K. De Raedt, H. De Raedt, K. Michielsen, Comp. Phys. Comm. 171 (2005) 19–39.
[22] G. 't Hooft, in: T.M. Nieuwenhuizen, B. Mehmani, V. Špička, M.J. Aghdami, A. Yu Khrennikov (Eds.), Beyond the Quantum, World Scientific, Singapore, 2007, pp. 3–19.
[23] K. Michielsen, F. Jin, H. De Raedt, J. Comput. Theor. Nanosci. 8 (2011) 1052–1080.
[24] H. De Raedt, K. Michielsen, Ann. Phys. (Berlin) 524 (2012) 393–410.
[25] H. De Raedt, F. Jin, K. Michielsen, Quantum Matter 1 (2012) 1–21.
[26] Hans De Raedt, Kristel Michielsen, Front. Phys. 2 (2014).
[27] K. Michielsen, H. De Raedt, Internat. J. Modern Phys. C 25 (2014) 01430003.
[28] Website, For additional information see, http://en.wikipedia.org/wiki/Robust_optimization.
[29] L.E. Ballentine, Quantum Mechanics: A Modern Development, World Scientific, Singapore, 2003.
[30] J.P. Ralston, Emergent mechanics, quantum and un-quantum, in: Proc. SPIE 8832, 88320W1–25, 2013.
[31] P.R. Wallace, Phys. Rev. 71 (1947) 622–634.
[32] M. Creutz, J. High Energy Phys. 0804 (2008) 17.
[33] G.E. Volovik, JETP Lett. 90 (2010) 697–704.
[34] J.P. Ralston, Revising your world-view of the fundamental constants, in: Proc. SPIE 8832, 883216–1–16, 2013.
[35] E. Madelung, Z. Phys. 40 (1927) 322–326.
[36] T.C. Wallstrom, Phys. Rev. A 49 (1994) 1613–1617.
[37] G.E. Volovik, in: R.T. Jantzen, H. Kleinert, R. Ruffini (Eds.), Proceedings of the Eleventh Marcel Grossmann Meeting on General Relativity, World Scientific, Singapore, 2009, p. 350.
[38] S.P. Novikov, Uspekhi Mat. Nauk. 37 (1982) 3–49.
[39] W. Gerlach, O. Stern, Z. Phys. 9 (1922) 349–352.
[40] D. Bohm, Phys. Rev. 85 (1952) 166–179.
[41] T. Takabayasi, Progr. Theoret. Phys. 8 (1952) 143.
[42] T. Takabayasi, Progr. Theoret. Phys. 2 (1954) 810.
[43] D. Bohm, R. Schiller, Nuovo Cimento Series 10 1 (1955) 67–91.
[44] R. Schiller, Phys. Rev. 125 (1962) 1116–1123.
[45] M. Reginatto, Phys. Lett. A 249 (1998) 355–357.
[46] H. De Raedt, M.I. Katsnelson, K. Michielsen, Quantum theory as the most robust description of reproducible experiments: application to a rigid linear rotator, in: Proc. SPIE 8832, 883212–1–11, 2013.
[47] A.E. Allahverdyan, R. Balian, Th. M. Nieuwenhuizen, Phys. Rep. 525 (2013) 1–166.
[48] G. Pólya, Mathematics and Plausible Reasoning, Princeton University Press, Princeton, 1954.
[49] J.M. Keynes, A Treatise on Probability, Macmillan, London, 1921.
[50] W. Feller, An Introduction to Probability Theory and its Applications, vol. 1, Wiley & Sons, New York, 1968.
[51] G.R. Grimmet, D.R. Stirzaker, Probability and Random Processes, Clarendon Press, Oxford, 1995.