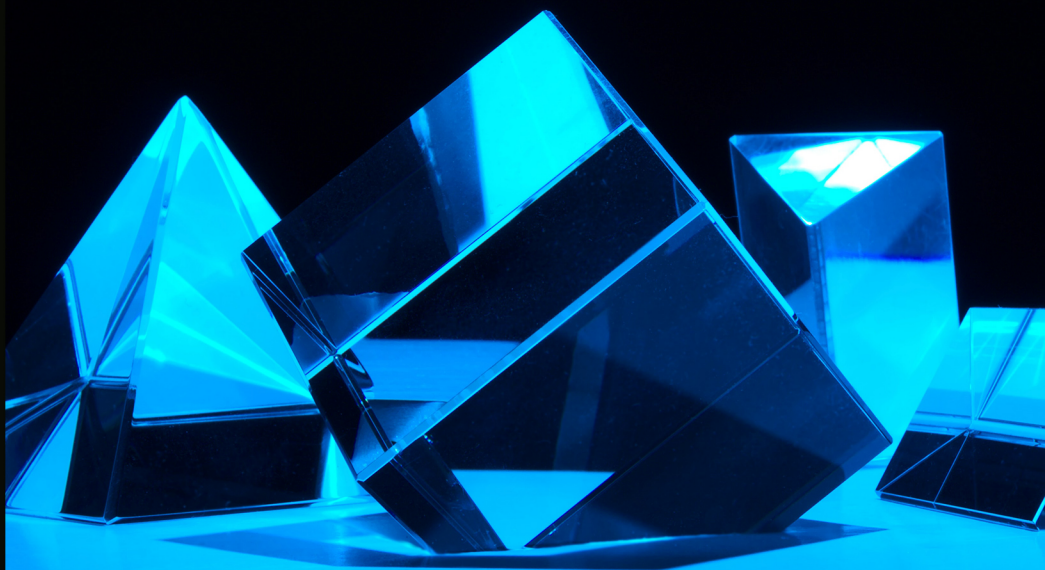
Realistic Interpretation of Quantum Mechanics



Emilio Santos Corchero

Realistic Interpretation of Quantum Mechanics

Realistic Interpretation of Quantum Mechanics

By

Emilio Santos Corchero

Cambridge Scholars Publishing

Realistic Interpretation of Quantum Mechanics

By Emilio Santos Corchero

This book first published 2022

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-7974-3
ISBN (13): 978-1-5275-7974-3

Contents

Preface

Quantum mechanics has been labeled magic because it combines an extremely good efficiency for the prediction of experimental results with a relatively simple formalism, but it does not provide a picture of the material world. Furthermore, when one attempts to get an intuitive picture logical contradictions appear. For instance, there are empirical facts suggesting that an electron is a point particle, or at least an object much smaller than an atom. However, interference effects and other phenomena seem to prove that it is an extended object (a wave). As a consequence, the interpretation of quantum mechanics has been the subject of continuous debate from the early period, almost one century ago, until today. The interpretations range from the pragmatic one supported by Bohr, with emphasis in the experiments, to those dazzled by the formalism, like the many-worlds that claims its universal validity, against common intuitions of human beings that do not believe to be simultaneously in several branches of a wave-function of the universe.

There is a dichotomy in current interpretations, namely they either reject a quantum picture of the world or offer a picture drastically departing from what we may derive from our everyday experience. In contrast, in this book I support 'realism', that is the view that science in general, and physics in particular, should explain how the world is, rather than just offering rules for the prediction of the results of the observations or experiments. That is, the book attempts a realistic interpretation that might provide a picture of the material world. Any picture must be free from contradictions and should be understood without sophisticated mathematical theories. In summary, realism includes the view that the world is made up of real stuff, existing in space and changing with time or, better stated, existing in a spacetime continuum.

The book may be seen as a kind of scientific memoir of the author. It is an organized survey of the attempts to understand quantum mechanics made along more than fifty years. Asides from revisiting relevant work of the author, the book contains many original contributions.

A fundamental hypothesis of the proposed interpretation of quantum theory is the reality of the vacuum fields. They appear as an unavoidable consequence of quantization although there is no agreement with respect to their nature. For some people they are an artifact of the formalism having observational consequences. The vacuum fields are labeled 'virtual', a word without a clear meaning that is used to avoid any commitment about their actual existence. This book strongly supports the reality of the vacuum fields as stochastic fields.

The interpretation of quantum theory resting on the reality of the vacuum fields opens a door for the solution of several problems in fundamental physics, some of them considered open and other allegedly closed but the common solution being still disputed by some people. I will mention six problems that will be discussed in this book: local realism, entanglement, quantum gravity, dark energy, dark matter and black holes. In the following I will comment briefly on each of them, stressing my personal opinion, that in most cases does not agree, or at least not fully, with the common view.

Local realism versus quantum mechanics has been the late stage of a debate that arose soon after the discovery of quantum mechanics in 1925. As is well known the main actors of the debate were Niels Bohr and Albert Einstein, who supported completeness and incompleteness of the theory, respectively. In 1935 Einstein, Podolsky and Rosen (EPR) introduced locality, or relativistic causality, as an argument *for* the incompleteness, but the debate did not end. In 1965, 10 years after Einstein's death, John Bell derived his celebrated inequalities that have been interpreted by most authors as a vindication of Bohr. I do not agree and the subject is treated extensively in this book, namely in chapter 1, section 1.1, chapter 2 section 2.3, the whole chapter 3 and chapter 6 section 6.6, asides from comments in other parts of the book.

Entanglement is a concept introduced by EPR and discussed in more detail by Schrödinger the same year 1935. There is a clear mathematical definition in terms of vectors in a Hilbert space (or wave-functions) and their consequences are extremely relevant, it being a crucial concept in the increasingly important field of quantum

information and quantum computation. However there is no clear physical interpretation of entanglement. It is closely related to the Bell inequalities and it is particularly studied in chapter 2 section 2.2.3, chapter 3, section 3.2.5, chapter 5 section 5.4 and chapter 6 section 6.6.

The remaining 4 problems belong to astrophysics or cosmology, and the whole chapter 7 is devoted to them. In section 2 the idea of 'quantizing the gravitational field' is revisited taking into account that gravity is not a force. The so-called gravity effects derive from the curvature of spacetime. Therefore general relativity is not a theory of gravity, but a theory of (curved) spacetime. Quantizing gravity is actually quantizing the spacetime curvature and this is understood in this book as an epistemological rather than ontological question. In fact, as said above 'quantization' means the need of studying everything in the material world as stochastic.

The rest of chapter 7 deals with the consequences of the hypothesis that the quantum vacuum fields may produce spacetime curvature and that curvature modifies those fields. In particular, it is necessary to take into account that the vacuum fields fluctuate. In section 7.4 it is argued that, as these fluctuations are incompatible with Minkowski space, it is worth to study the minimal modification produced by them and the result is that in the absence of matter they give rise to a cosmological constant term or, in other words, they are plausibly the origin of the 'dark energy'. On the other hand, in space containing baryonic matter the combination of that matter with the vacuum fluctuations may produce new effects or, in other words, modify the dark energy. It is also proposed that vacuum fluctuations might give rise to effects currently attributed to dark matter. This possibility is studied in section 7.5.

In astrophysical compact objects where the baryonic matter is able to produce a strong spacetime curvature, it is plausible that the effect on the quantum vacuum fields should be extremely big. In section 7.6 it is proposed that these changes might modify the evolution of such compact relativistic objects stopping collapse before a Schwarzschild singularity is produced.

Plan of the book. The book consists of seven chapters. In chapter 1 a number of nude observations, usually assumed specifically quantal, are analysed in order to show that they might be explained without departing from our proposed realistic view of nature. The chapter

starts with an epistemological introduction that supports the Einstein (realistic) against the Bohr-Heisenberg (positivistic and pragmatic) views of science. At the end of the chapter a sketch is presented of the view about the quantum world.

Chapter 2 is devoted to the standard, or canonical, Hilbert-space formalism of quantum theory. We start with the postulates of the theory followed by a critical analysis of the most popular interpretations of the formalism. A purpose of that analysis is to point out that we should not attempt to interpret the standard formalism, but rather the observations or experiments, that are independent of any theory. For the sake of completeness I also include a study of the proposed logical structure in terms of lattices of propositions and the comparison with classical logic, the Bell inequalities being a crucial test.

Chapter 3 deals in more detail with the Bell inequalities, that have been during half a century most relevant in discussions about the interpretation of quantum mechanics. The inequalities are assumed to be necessary conditions for a local realistic interpretation of nature. In fact, Bell's work seems to prove that there is a conflict between quantum mechanics and local realism, understanding locality as relativistic causality. I shall discuss to which extent this is true. In addition, a short survey is presented of the experiments performed or proposed in order to test the inequalities against the quantum predictions.

In chapter 4 alternative formulations of quantum theory are presented, namely de Broglie-Bohm, stochastic mechanics, Weyl-Wigner in phase space, and Feynman path integrals. These formulations either contradict some predictions of the standard formalism, and also experiments, or seem incompatible with a realistic interpretation. For instance, the Weyl-Wigner function seems to imply 'negative probabilities' and Feynman path integrals 'imaginary probabilities'. I shall argue that nevertheless both of them might be free from these shortcomings if correctly interpreted.

Chapter 5 is a survey of stochastic electrodynamics, a theory that studies within classical electrodynamics the motion of charged particles under the action of given forces but in the presence of a random electromagnetic radiation field. The theory agrees with quantum predictions in a limited domain, but disagrees in other cases. The relevance of stochastic electrodynamics is that it provides hints for a realistic interpretation of the whole quantum theory. In fact, it is a

theory that may be considered classical but makes predictions that fit in experiments allegedly quantal.

Chapter 6 provides a realistic analysis of several effects of the quantum vacuum radiation field that offers, for some experiments, a clear intuitive picture commonly claimed to be impossible. In particular, we discuss experiments showing wave-particle behaviour or violations of Bell inequalities.

Chapter 7 deals with quantum effects in astrophysics and cosmology. In the first part a personal view is presented of the meaning of general relativity, which in several respects differs from the current view. The rest of the chapter deals with the possibility of understanding dark energy and dark matter as effects of the quantum vacuum fluctuations. It is also suggested that the quantum vacuum fields might prevent collapse to black holes.

Acknowledgements. The book has been written in practical isolation after my retirement from the University of Cantabria (Spain). Therefore it is my sole responsibility. But the ideas exposed had been discussed along the years with many people. In particular, with all colleagues with whom I have published joint articles. Many of those authors appear in the references of the different chapters, but I want to mention here a few with whom I have had a more relevant interaction; namely, Trevor W. Marshall (Manchester), Ana María Cetto and Luis de la Peña (Mexico), Franco Selleri (Bari), Pierre Claverie (Paris), Kaled Dechoum and Humberto M. França (Sao Paulo), Jaroslaw Pykacz and Marek Zukowski (Gdansk), Michael Revzen (Haifa), Luis J. Boya (Zaragoza), Miguel Ferrero (Oviedo), Jose Luis Sánchez-Gómez and Antonio F. Rañada (Madrid), Xavier Barcons, Rafael Blanco, Ángel Mañanes, Saturnino Marcos, Luis Pesquera and Miguel A. Rodríguez (Santander), Ramón Lapiedra (Valencia), Manuel Gadella (Valladolid), Susana F. Huelga (Ulm), Alberto Casado and Ramón Risco (Sevilla), and Albert Bramón (Barcelona). I also want to thank Francisco Santos for his help with processing the manuscript in LaTeX and for a careful reading of it.

Last but not least, I acknowledge to my late wife Matilde Leal her moral support and patience.

I apologize to all those not mentioned.

CHAPTER 1

The quantum image of nature

1.1. The debate about the interpretation of quantum theory

1.1.1. Early interpretations: Schrödinger, Heisenberg and Bohr. Quantum theory began with Max Planck's formula for the blackbody spectrum, presented in December 14, 1900, a date later named by Sommerfeld the birthday of quantum theory. After years of slow progress without a definite theory, quantum mechanics appeared almost simultaneously during 1925-26 in two different forms: 'quantum mechanics' of Heisenberg and 'wave mechanics' of Schrödinger.

Soon afterwards Schrödinger himself and Dirac proved that both theories are equivalent; that is, they make the same predictions for the experiments. For the history of the early period see Jammer [**1**].

Quantum mechanics had a rapid and deep impact. It was soon applied to atoms, molecules and solids with great success. However the interpretation of the theory was not straightforward. Schrödinger suggested that his wave-function describes a continuous electric charge distribution. That picture was abandoned after correct criticisms by Bohr and other people. In particular a detailed calculation of the ionization energy of the helium atom proved that the electron should be seen as a particle much smaller than the atom.

In contrast, Heisenberg introduced his quantum theory as a set of calculational rules involving arrays of numbers (*matrices* in mathematical language) devoided of any intuitive picture. Furthermore physics without images should be considered a superior form of science because the only condition for the validity of a theory is the agreement of its predictions with the empirical evidence. That view was reinforced with Dirac's formulation in terms of an abstract vector space.

The approach of Heisenberg was supported by Bohr, who elaborated it introducing the 'complementary principle', with the aim of solving the particle (localized)-wave (extended) duality of quantum objects, and stressing the role of the Planck constant as an indivisible element of 'action'. This led to the *Copenhagen interpretation*, which became dominant for many years. This interpretation may be labeled as pragmatic because the referent of the theory is not the physical world but the experiment. However, it produced disconfort in some people, e.g. Einstein, and a long debate arose that lasts until today [**2**], [**3**], [**4**], [**5**], [**6**], [**7**], [**8**], [**9**], [**10**], [**11**]. For a recent review with extensive bibliography, see Drummond [**12**].

The interpretations usually refer to quantum mechanics as formulated in Hilbert space, a formalism to be treated in chapter 2. Therefore I postpone to that chapter a discussion of the most popular interpretations. In the following I comment on the two main classes of quantum interpretations, namely pragmatic and realistic.

1.1.2. The pragmatic approach to quantum mechanics. None of the interpretations proposed till now offer a clear intuitive picture of the quantum world. Nevertheless, most physicists do not worry for the lack of a picture and embrace a *pragmatic approach* close to the

Copenhagen interpretation. They accept a minimal interpretational framework with the following key features [**13**]:

1. Quantum theory is viewed as a scheme for predicting the probabilistic distribution of outcomes of measurements made on suitably prepared copies of a system.

2. The probabilities are interpreted in a statistical way as referring to relative frequencies.

Behind the pragmatic approach there is usually a philosophical position about physics (or science in general) that may be summarized as follows. It is common wisdom that a physical theory has at least two components [**14**]: (1) the formalism, or mathematical apparatus, of the theory, and (2) the rules of correspondence that establish a link between the formalism and the results of measurements. As an example let us consider the formalism of quantum mechanics based on the mathematical theory of Hilbert spaces (to be discussed in more detail in chapter 2). The formalism involves two kinds of operators: density operators, $\hat{\rho}$, that represent states, and self-adjoint operators, $\hat{A}$, that represent observables. The link with the measurement results is given by the postulate that the expectation value, $\mathrm{Tr}(\hat{\rho}\hat{A})$, corresponds to the statistical mean of the values obtained when one realizes several measurements on identically prepared systems (which determines $\hat{\rho}$) by means of an appropriate apparatus (that corresponds to $\hat{A}$).

If we assume that the formalism and the correspondence rules are the *only objects required to define a physical theory*, in the sense that the statistical regularities need not be further explained, then we get what has been called a *minimal instrumentalistic interpretation* of the theory [**15**], [**13**]. It may be identified with the purely pragmatic approach mentioned above.

Most people claiming to support that approach accept the following positions:

1. The notion of an individual physical system 'having' or 'possessing' values for all its physical quantities is *inappropriate* in the context of quantum theory.

2. The concept of 'measurement' is fundamental in the sense that the scope of quantum theory is *intrinsically* restricted to predicting the results of measurements.

3. The spread in the results of measurements on identically prepared systems must not be interpreted as reflecting a 'lack of knowledge' of some objectively existing state of affairs.

The *instrumentalistic* approach is quite different from, even opposite to, the *realistic* view traditional of classical physics. Between these two extremes there are a variety of approaches.

1.1.3. Realistic interpretations. The main opponent to a purely pragmatic approach to quantum mechanics was Albert Einstein. Indeed, his discussions with Niels Bohr are the paradigm of a scientific debate, hard in the scientific arguments but hearty from the personal point of view. One of the most celebrated moments of the debate was a 1935 article by Einstein, Podolsky and Rosen [**16**] (EPR) soon followed by Bohr's reply [**17**]. The former begins as follows: "Any serious consideration of a physical theory must take into account the distinction between the objective reality, which is independent of any theory, and the physical concepts with which the theory operates. These concepts are intended to correspond with *the objective reality*, and by means of these concepts *we picture this reality to ourselves*" (my emphasis).

It is true that in the years elapsed since the EPR paper the concept of 'objective reality' has been questioned as not clear. Due to the difficulties with the interpretation of quantum mechanics, many people working on foundations dismiss the 'realism' of EPR as 'naive'. Thus more sophisticated forms of realism have been proposed [**7**]. In any case, a deep discussion about the philosophical aspects of reality or realism is outside our scope.

In this book I strongly support Einstein's view. That is, I believe that a realistic interpretation is possible. The starting point is the claim that any physical theory should offer a *physical model* in addition to *the formalism and rules for the connection with the experiments.* The latter are obviously essential because they are required for the comparison of the theory with empirical evidence, which is *the test* for the validity of the theory. But in my opinion physical models are also necessary in order to *reach a coherent picture* of the world. Many quantum physicists apparently support the uselessness of pictures, but it is the case that when they attempt popular explanations of quantum phenomena they frequently propose actual pictures, many of them rather bizarre. For instance it has been claimed that quantum mechanics *compel us* to believe that there are a multiplicity of 'me' in parallel universes, or that an atom may be present in two distant places at the same time. This is an indication that the need for a 'picture the reality to ourselves' [**16**] cannot be easily dismissed.

Furthermore the existence of physical models might open the possibility for new developments and applications of quantum theory and therefore it is not only an academic question.

The contrast between the two great theories of the 20th century, quantum mechanics and relativity is interesting. The latter provides a beautiful physical model: There is a four-dimensional manifold with intrinsic curvature and all material objects (e.g. particles or fields) are defined in that continuum and the basic quantities of physics, like mass, energy or momentum, become geometric properties. But the calculational tool of general relativity (derived from the Riemann geometry) is rather involved, the fundamental (Einstein) equation being *nonlinear*. In quantum mechanics there is a relatively simple *linear* formalism involving vectors and operators in a Hilbert space. Indeed, the fundamental (Schrödinger) equation is linear. However, there is no coherent physical model behind. I would say that general relativity has *physical beauty*, the quantum formalism possesses *mathematical elegance.*

Historically the renunciation to physical models in quantum mechanics was a consequence of frustration caused by the failure of the models proposed during the first quarter of the 20th century. This was specially the case after Bohr's atom, consisting of point electrons moving in circular orbits around the nucleus. The model, generalized with the inclusion of elliptical orbits, certainly produced progress in the decade after 1913. However, it was obvious that the model mixed contradictory laws, namely classical electrodynamics and Bohr postulates. The success of quantum mechanics in the quantitative interpretation of experiments did not solve the problem, which became more acute. Thus the failure to find a good physical model of the microworld led to an almost universal acceptance of the current view that models may be unnecessary or even misleading.

I do not agree with that view, but this book is a defence of a realistic interpretation of the quantum phenomena. I am aware that the task is extremely difficult as is proved by the lack of such an interpretation after a century of quantum mechanics. However, I am convinced that many of the obstacles derive from some assumptions that are not necessary for the interpretation of the experiments. These assumptions have been introduced along the historical development of the theory and are now a part of the common view. Pointing out the main obstacles and how they might be removed is the purpose of this book. I do not pretend to provide a coherent and complete realistic

interpretation, but I hope that some of the ideas put forward might be useful in the progress towards a better understanding of quantum mechanics.

1.1.4. A note on the epistemology of physics. In order to practice science some previous philosophical questions should be answered. For instance, what is science? Or, what is the purpose of science? There are different philosophical positions about these questions that are closely connected with the different interpretations of quantum mechanics.

There is some agreement that the criterion to distinguish science from nonscientific knowledge is the proposal of Karl Popper [**18**], [**19**]: A claim is scientific if it may be refuted by observations or experiments. This definition is a consequence of a well known fact, namely *the possible existence of several different theories all of them predicting correctly the results of experiments in a given domain.* In other words the correctness of a theory is sufficient, *but not necessary*, for the appropriate prediction of empirical facts. For this reason a single experiment may refute a theory but a theory can never be fully confirmed empirically, and this is essentially the Popper thesis. As a consequence several different theories may exist that are able to predict correctly the empirical results, but suggesting quite different pictures of the microworld.

Popper's criterion is good enough as a matter of principle, but it is not so good in practice. In fact, it is the case that rarely an established theory breaks down as a consequence of a single experiment contradicting it. As Lakatos [**20**] has pointed out, *well tested theories are protected* in the sense that the empirical refutation of a single prediction may be interpreted without rejecting the theory, for instance assuming that the analysis of the experiment was incorrect. Indeed, it is a historical fact that established theories are only abandoned, or better superseded, when there is a new theory in agreement with the former one in its domain of validity but possessing a wider domain or other virtues.

Quantum mechanics is today a fully established theory and therefore it is very well protected in the sense of Lakatos. I do not only mean protection in the domain where the theory has been tested. What I want to stress is that over the years people have introduced a number of assumptions, today widely accepted, that are additions without possibility of empirical test. See section 2 below for several

examples. These unnecessary additions are also protected and, in my opinion, they are the main cause of the strong difficulties in reaching a realistic physical model of the quantum world.

Most working quantum physicists adhere to the pragmatic approach as described above. The support has its roots in a 'positivistic' attitude. Positivism is the philosophical doctrine that, in a broad sense, states that all knowledge should be founded on empirical evidence. In this sense it is accepted by everybody. But in a more strict sense it is a tendency to give value to the crude empirical data in detriment of the theoretical elaborations. For instance this was the opinion of Ernst Mach, who rejected the concept of atom because at that time (around 1900) atoms had not been directly observed.

Positivism was also behind Heisenberg's initial formulation of quantum mechanics resting upon the belief that only sets of numbers corresponding to the possible results of measurements should enter the theory. This led him to elaborate quantum mechanics as a calculational tool involving matrices (it was sometimes called 'matrix mechanics'). The combination of mathematical formalism and empirical results almost without further theoretical elaboration permeates the interpretation of quantum mechanics till now. An illuminating confrontation between the positivistic and realistic epistemologies is the conversation of Heisenberg with Einstein that took place in Berlin in 1926, as remembered by Heisenberg himself [**21**]. The most relevant part is the following.

> As soon as we were indoors, he [Einstein] opened the conversation with a question that bore on the philosophical background of my recent work. "What you have told us sounds extremely strange. You assume the existence of electrons inside the atom, and you are probably quite right to do so. But you refuse to consider their orbits, even though we can observe electron tracks in a cloud chamber. I should very much like to hear more about your reasons for making such strange assumptions."
>
> "We cannot observe electron orbits inside the atom," I must have replied, "but the radiation which an atom emits during discharges enables us to deduce the frequencies and corresponding amplitudes of its electrons. After all, even in the older physics wave numbers and amplitudes could be considered substitutes for electron orbits. Now, since *a good theory must be based*

on directly observable magnitudes, I thought it more fitting to restrict myself to these, treating them, as it were, as representatives of the electron orbits."

"But you don't seriously believe," Einstein protested, "that none but observable magnitudes must go into a physical theory?"

"Isn't that precisely what you have done with relativity?" I asked in some surprise. "After all, you did stress the fact that it is impermissible to speak of absolute time, simply because absolute time cannot be observed; that only clock readings, be it in the moving reference system or the system at rest, are relevant to the determination of time."

"Possibly I did use this kind of reasoning," Einstein admitted, "but it is nonsense all the same. Perhaps I could put it more diplomatically by saying that it may be heuristically useful to keep in mind what one has actually observed. But on principle, *it is quite wrong to try founding a theory on observable magnitudes alone.* In reality the very opposite happens, *it is the theory which decides what we can observe.* You must appreciate that observation is a very complicated process. The phenomenon under observation produces certain events in our measuring apparatus. As a result, further processes take place in the apparatus, which eventually and by complicated paths produce sense impressions and help us to fix the effects in our consciousness. Along this whole path – from the phenomenon to its fixation in our consciousness – we must be able to tell how nature functions, must know the natural laws at least in practical terms, before we can claim to have observed anything at all. Only theory, that is, knowledge of natural laws, enables us to deduce the underlying phenomena from our sense impressions. When we claim that we can observe something new, we ought really to be saying that, although we are about to formulate new natural laws that do not agree with the old ones, we nevertheless assume that the existing laws –covering the whole path from the phenomenon to our consciousness– function in such a way that we can rely upon them and hence speak of observations" (my emphasis).

The conversation continued for a while and at the end Einstein warned: "*You are moving on very thin ice. For you are suddenly speaking of what we know about nature and no longer about what nature really does. In science we ought to be concerned solely with what nature does.*" Einstein's arguments are a clear support to a realistic epistemology, and I fully agree with his views.

I believe that, as stated by Einstein "it is wrong to try founding a theory on observable magnitudes alone". But I believe that the following statement more close to Heisenberg's view is correct: *We should try to interpret as physically real just the observable magnitudes alone, but not the intermediate (mathematical) ones that appear in a calculation.* This will be our guide in this book for a realistic interpretation of quantum mechanics.

1.2. Specific features of quantum physics

I propose that the difficulties for a realistic interpretation of quantum phenomena do not derive from the empirical facts, or not only. Thus in the following I shall briefly revisit the most relevant of those phenomena in order to see whether the nude empirical facts do prevent any picture of the microworld. Actually, most textbooks of quantum mechanics emphasize the difficulty, or impossibility, to interpret typical quantum phenomena with a realistic view. The purpose of the following paragraphs is just the opposite. It will be shown that in fact those phenomena are compatible in most cases with a picture of the microworld. Of course, the picture is somewhat different from the one offered by classical physics, but not dramatically different.

Nevertheless, I shall confess that serious difficulties remain, so that neither this section nor the whole book will provide a systematic realistic interpretation free from difficulties.

1.2.1. The stability of atoms. Soon after Rutherford's experiment of 1911, that lead to the nuclear atom, Bohr proposed in 1913 a model which involved postulates contradicting classical electrodynamics. The common wisdom was, and it still is, that the contradiction cannot be avoided. That it appears even for the most basic empirical fact, the stability of the atom. But this is not true [**22**].

In fact, if studied with classical electrodynamics a hydrogen atom, consisting of one proton and one electron, cannot be stable *if isolated*. The reason is that an electron moving around the proton would radiate, and therefore it will lose energy until the atom collapses. But

the argument is not valid if there are many atoms in the universe because if all atoms radiate the hypothesis of isolation is not appropriate. It is more plausible to assume that there is some amount of radiation filling space. Then every atom would sometimes radiate but other times it would absorb energy from the radiation, eventually arriving at a dynamical equilibrium. This may explain, at least qualitatively, the stability of the atom. The picture that emerges is that *matter and radiation of the universe cannot be treated independently, and the complexity of the universe compel us to treat the radiation as a background stochastic field.* Therefore the electron of a hydrogen atom would move in a random way around the nucleus. That motion should be so complex that we cannot follow it in detail but only the probability distribution of positions can be determined. That distribution is what the Schrödinger wave-function provides via Born's rule. The assumption of background fields filling space fits in the quantum vacuum fields that appear in field quantization. They are assumed real stochastic fields throughout this book, and make up the basic hypothesis for the realistic interpretation of quantum theory as discussed in the following.

1.2.2. The connection between energy and frequency. A standard method to study the radiation field in free space is to expand it in plane waves (or in normal modes if it is enclosed in a cavity). In free space the number of modes, N, per unit volume and unit frequency interval is

$$N = \frac{\omega^2}{\pi^2 c^3}, \tag{1}$$

and the radiation energy is

$$E = \frac{1}{2}\hbar\omega \tag{2}$$

per normal mode of the radiation. That energy eq.(2) is just 1/2 the one postulated by Einstein in his 1905 article where he introduced the concept of quantum of radiation, later named photon. I will discuss the concept of photon in chapter 6. In the following I will derive some consequences via an heuristic approach. Firstly I propose to generalize eq.(2) for all possible vacuum fields, associated to the forces of nature.

If a hydrogen atom is in a dynamical equilibrium with radiation, it is plausible that the main interaction with the vacuum fields takes place with the normal modes of the field that have frequencies close to those of the electron motion. Also, in a dynamical equilibrium

it is plausible that the mean kinetic energy of the electron should be close to half the average energy of those normal modes having greater interaction with the atom (the other half would correspond to potential energy). Then if the electron moved around the nucleus in a circle with energy E (i.e. without emission or absorption of radiation), we might write the following equalities

$$|E| = \frac{1}{2}mv^2 = \frac{e^2}{2r}, \qquad v = r\omega, \qquad |E| \sim \frac{1}{2}\hbar\omega, \tag{3}$$

the latter corresponding to the condition of dynamical equilibrium with radiation. Of course, the motion is perturbed by the action of the vacuum fields, whence the electron motion would be irregular, not circular, but it is plausible that eqs.(3) would be roughly fulfilled. Hence the energy and the size of the atom may be got by eliminating the quantities v and ω amongst the 4 equalities, which leads to

$$E \sim -\frac{1}{2}\frac{me^4}{\hbar^2}, \qquad r \sim \frac{\hbar^2}{me^2}, \tag{4}$$

in rough agreement with the quantum prediction and with experiments.

1.2.3. Statistical character. The statistical character of measurements in the quantum domain is a consequence of existence of random vacuum fields as discussed above. However, it is appropriate to comment on it in more detail due to the great relevance attributed to it in books and articles about foundations of quantum physics.

In the classical domain typical experiments are affected by statistical errors. That is, the same experiment performed in similar conditions may give rise to (slightly) different results. For this reason it is a standard practice to report the results of measurements accompanied by an uncertainty interval. In the macroscopic domain the uncertainty is attributed to the difficulty of controlling a very large number of parameters (the environment), with the consequence that never (or rarely) an experiment may be repeated in exactly the same conditions. In any case, it is usual that the uncertainty is only a small fraction of the measured quantity. In contrast, in the microscopic domain it is frequent that the uncertainties are of the same order than the measured result. This is equivalent to saying that the same experiment may give rise to a number of different results, every one with some probability. However, in contrast with macroscopic (classical) physics, in quantum physics the probabilities are usually not atributted to lack of control in the experiment.

The current view is that quantum probabilities are radically different from the classical, ordinary life, probabilities. The latter are introduced when there is incomplete knowledge ('ignorance'), maybe unavoidable, about the truth of some assertion. For instance we may attach a probability 1/2 to the appearance of head when throwing a coin, because we cannot control all relevant variables in the experiment. In contrast, it is a common assumption that quantum probabilities are quite different, that they derive from a lack of strict causality of the natural laws, i.e. the fact that different effects may follow to the same cause. This is usually called the *fundamental or essential probabilistic* character of the physical laws. This is an example of a practical difficulty that has been (incorrectly in my opinion) raised to the rank of an ontological statement: *"Natural laws are not strictly causal"*.

Einstein disliked that assumption and strongly criticized it, as expressed in his celebrated sentence "*God does not play dice*". I understand very well Einstein's opinion. For him the rational understanding of nature was a kind of religion. The more loose (strict) the natural laws are, the smaller (greater) could be our rational understanding of nature. Accepting a weak causality is like accepting poor science. Nevertheless, some people are happy with the absence of determinism implied by the nonexistence of strict causality. For instance some claims have been made that the quantum lack of determinism may explain human free will. This question lies outside the scope of this book and shall not be further commented on.

But I do not support determinism in the mechanistic view of Laplace. As said above quantum mechanics is a stochastic theory. I believe that strictly causal laws might perhaps exist, but there is also a universal noise which permeates everything and prevents any practical determinism. Strict causality combined with stochasticity (randomness) is in practice indistinguishable from essential probability, and the former is more plausible. In order to clarify this matter let us think about Brownian motion. Under macroscopic observations the random motion of a Brownian particle may appear as lacking causality; but we assume that, taking into account the molecules of the liquid where the particle is immersed, the whole motion is governed by Newtonian dynamics, which is causal.

It may be argued that the uncertainties in the measurements in the quantum domain do not look like typical uncertainties derived

from noise. Indeed, the latter may be usually approximated by continuous probability distributions. In the quantum domain there are instances, e.g. the Stern-Gerlach experiment, where there is uncertainty between just two values (for more about that experiment see the subsection about discrete states, below). I shall not discuss further this difficulty here, but we will discuss it in more detail in later chapters.

1.2.4. Heisenberg uncertainty relations. The Heisenberg 'uncertainty principle' is the most frequently quoted evidence for the dramatic splitting between classical and quantum physics. In fact, the principle appears in popular writings like a kind of mysterious property of our world. However, the arguments given below strongly suggest that the Heisenberg inequalities are consequences of the stochasticity inherent to the microworld rather than a fundamental principle. I shall not discuss here the general relation dealing with conjugate dynamical variables, but restrict attention to the experimentally proved impossibility of determining simultaneously the position and the velocity (or momentum) of a particle. This implies that it is not possible to prepare a particle with both position and velocity sharply defined, and also that no measurement may provide the values of both these quantities at the same time. Hence it is impossible to determine the path of a particle.

In any motion under the action of a random force some constraints may appear for the simultaneous determination of position and velocity, that might be stated in the form of inequalities. As an illustrative example this is shown to be the case in Brownian motion. A Brownian particle possesses a highly irregular path whose instantaneous velocity cannot be measured (with ordinary, macroscopic set-ups). Only the mean velocity, $\overline{\mathbf{v}}$, during some time interval may be measured, that is,

$$\overline{\mathbf{v}} = \frac{|\Delta \mathbf{r}|}{\Delta t},$$

where $|\Delta \mathbf{r}|$ is the distance between the initial and final positions in the time interval Δt. On the other hand there is a relation, derived by Einstein in 1905, between the expected value of the square of the distance, $|\Delta \mathbf{r}|^2$, and the time interval, Δt. Namely,

$$\left\langle |\Delta \mathbf{r}|^2 \right\rangle = D \Delta t,$$

where D is called the diffusion constant and $\langle\rangle$ means ensemble average, that is, the average over many measurements involving the same time interval. If we eliminate Δt amongst the two equalities we get

$$\left\langle |\Delta \mathbf{r}|^2 \right\rangle = \langle \overline{v}^2 \rangle \, \Delta t^2 \qquad \Rightarrow \qquad \left\langle |\Delta \mathbf{r}|^2 \right\rangle \langle \overline{v}^2 \rangle \simeq D^2,$$

a relation having some similarity with the Heisenberg uncertainty relation. We conclude that a plausible interpretation of the Heisenberg principle is that the quantum motion possesses a random component having some similarity (not identity!) with Brownian motion. This similarity has been the basis for the development of stochastic mechanics, which provides an intuitive picture of some typically quantum phenomena. However, this theory presents difficulties as will be discussed in chapter 4. Actually, the Brownian motion inequality derived above is different from the Heisenberg inequalities because $\overline{v}^2$ is the mean squared velocity during a time interval, rather than the uncertainty of the velocity. In chapter 5 I will show that Heisenberg inequalities may be derived as a consequence of the radiation spectrum.

The Heisenberg inequalities become an obstacle for a realistic interpretation of quantum mechanics when the practical difficulty (or impossibility) of simultaneous knowledge of position and velocity is elevated to the category of an ontological statement: "*Trajectories of quantum particles do not exist*". Of course, the Heisenberg inequalities are reinforced by the fact that they are predicted by the quantum formalism, but the analogy with the Brownian motion inequality suggests that the quantum formalism may be a disguised form of specifying a stochastic theory.

1.2.5. Discrete energy states. As is well known, the first quantum hypothesis, introduced by Planck in 1900, was that material systems may possess only energies belonging to a discrete set. The assumption was extended by the Einstein 1905 proposal that light consists of discrete pieces of energy (photons) and the successful application of this principle to the photoelectric effect. In 1913 Bohr incorporated the idea to his atomic model postulating that atoms can only exist in states having energies within a discrete set, E_0, E_1, E_2,... The model also assumed that the absorption and emission of light takes place with transitions between these states, the frequency, ω_{jk}, of the light being related to the difference of atomic energies by

$$\hbar \, \omega_{jk} = E_j - E_k. \tag{5}$$

The mere existence of spectra (e.g. of atoms) with a discrete set of frequencies does not put any problem for a realistic picture of the atoms. Indeed, discrete frequencies are an essential feature of musical instruments. Thus it is not strange that Erwin Schrödinger had this analogy in his mind when he derived his celebrated equation. In fact, the title of his pioneer paper on wave mechanics was "Quantization as an eigenvalue problem". Eigenvalue problems appear in the mathematical treatment of linear differential equations with boundary conditions, like in the study of vibrations of a string (say of a piano) fixed by both ends. It is the connection of frequencies with energy differences, as in Bohr eq.(5), which is difficult to understand. I suggest that it may be related to the connection energy-frequency of the normal modes of the vacuum fields, eq.(2), but I do not hold a fully satisfactory explanation.

We might believe that the frequencies are real but the discrete states are just an extra assumption introduced via eq.(5) because in practice the frequencies are observed but the energies usually not. However, the connection was confirmed by the experiment of Frank and Hertz in 1914. It consists of the scattering of electrons on mercury atoms in vapour state with the result that, for high enough electron energies, inelastic scattering was observed with a decrease of the electron energy by 4.9 eV. This quantity precisely corresponds to a frequency of the mercury spectrum via the relation eq.(5), whence the said quantity was interpreted as the energy difference between the ground state and the first excited state of the atom. As a consequence of these facts, and others, it has been fully accepted that the set of energy states of atoms is discrete. Furthermore, the discontinuities appear naturally in the quantum formalism where physical quantities correspond to operators in a Hilbert space having a discrete spectrum (in the mathematical sense of spectrum of an operator). Discreteness has also been ascribed to other dynamical quantities like angular momentum as mentioned below.

It is true that the quantum discontinuities give rise to difficulties for an intuitive understanding of quantum physics. I am not able to offer a plausible picture of the phenomenon, but the difficulty of doing it increases by the way the fact is presented at an elementary level. Indeed, the usual statement that material systems (e.g. atoms) may transition between two different energy states never possessing any intermediate energy is wrong. Quantum electrodynamics predicts that spectral lines are *not* sharp, but possess some width. Thus Bohr's

eq.(5) should be taken as an approximation, the possible atomic energies actually consisting of a continuous set. Eq.(5) simply recalls that the probabilities of some energy states are strongly concentrated near some discrete values. This remark solves one paradox which appears when the emission or absorption of light is presented at a popular level, namely the contradiction between assuming that atomic transitions are instantaneous and assuming that the emitted light has a sharp frequency. The fact is that the transition has a finite duration, Δt, and the emitted light has a finite linewidth, $\Delta\omega$, fulfilling the inequality

$$\Delta\omega\Delta t \gtrsim 1, \tag{6}$$

which is well known from classical optics. The inequality holds true for any periodic motion and the quantum formalism also predicts it. Indeed, sharp energies of atoms appear only in a calculation to lowest order of approximation (i.e. in the limit when the electron charge goes to zero, $e \to 0$). However, when radiative corrections of quantum electrodynamics are taken into account the calculation leads to spectral lines with a finite width. The *corrections* are small and *may be neglected in elementary calculations* but *they are essential for a realistic interpretation.* This is a typical example of how the emphasis on the simplicity of the calculations, rather than the clarity of the concepts, has the consequence that quantum mechanics appears as counterintuitive. A realistic picture of the atomic emission is possible assuming that light is emitted in a continuous process lasting a time Δt that fulfils the inequality (6), the total energy of atom plus light being conserved at all times. Indeed, this fits with the quantum evolution equation of the atom coupled to the electromagnetic field (that is, in quantum electrodynamics). Of course, the picture does not explain why the emitted light has large intensity for some frequencies within a discrete set but very low intensity for other frequencies.

Similar arguments may be used in order to understand the quantization of angular momentum, as shown for instance in the Stern-Gerlach experiment. In popular expositions the experimental results are presented as if all atoms arrive at one amongst two sharp lines in a screen. Then it is difficult to reach a picture of what is taking place in the interaction between the atom and the inhomogeneous magnetic field. However, the truth is that what appears in the screen are two wide spots, something that is less counterintuitive. It is the

case that an accurate quantum mechanical treatment of the experiment precisely predicts that fact [**23**]. A picture of the phenomenon may be reached assuming that during the interaction of the atom with the magnetic field some fluctuation and dissipation takes place which tends to align (approximately) the atomic magnetic moment with the (main) field. After the atomic magnetic moment and the magnetic field are aligned either parallel or antiparallel, the magnetic force determines the atomic motion towards the screen.

1.2.6. The apparent lack of objective properties. In classical physics it is assumed that any observation or measurement just reveals ('removes a veil') a property which exists objectively with independence of any observation. In quantum mechanics this seems to be untrue. Let us clarify the motivation for that belief with an example. We consider a physical system possessing three observable properties which I shall label A, B and C.

I will assume that the observables A and C may be measured in the same experiment, and similarly for B and C, but for some reasons A and B cannot be measured with the same experimental arrangement. Then with repeated measurements in identically prepared systems it is possible to obtain the joint probability distribution for the results of the former measurement, which I will represent by the density, $\rho(a,c)$, that the observable A takes the value a, and the observable C the value c. Similarly we may obtain $\rho(b,c)$, but it is not possible to obtain empirically a joint probability density $\rho(a,b)$ because we cannot measure A and B simultaneously. Up to here no problem arises, everything agrees with the intuition.

If we now think that the measurement just reveals preexisting values of the observable quantities, we are compelled to assume that, in every state, the system possesses the values a, b and c, independently of any observation or measurement. More generally, the preparation procedure should lead to a state with a joint probability distribution, $\rho(a,b,c)$, for the three observables. If this is the case the joint probabilities for two observables should be the marginals of the joint distribution, that is,

$$\rho(a,c) = \int \rho(a,b,c)\, db, \qquad \rho(b,c) = \int \rho(a,b,c)\, da. \tag{7}$$

However, it has been shown, in some experiments, that there are particular cases of states and observables where no (positive) joint probability density $\rho(a,b,c)$ exists such that the marginals eqs.(7) agree

with the empirical results. The non-existence of a joint probability fulfilling eqs.(7) in general is predicted from the quantum formalism, and it is the essential content of the Kochen-Specker theorem (see chapter 3 section 3.2.4). As in the case of the Heisenberg principle, the practical impossibility has been raised to the rank of an ontological statement: *"Physical systems do not possess properties independently of measurements."*

Is that statement justified? It is not. What the experiments have shown is that the observed properties depend not only on the state of the system but on the whole experimental set-up. In fact, we may assume that physical systems possess some properties which in specific experimental set-ups give rise to observable quantities, but the observables may not exist independently of the experiment. Actually, a similar situation also happens in classical physics, as for instance when we play dice. If we get a number, say 2, we cannot claim that the value 2 was preexistent to our experiment. The result 2 is actually *created* by the experiment of throwing the dice. Returning to quantum physics, there is a simple explanation for the frequent inexistence of properties independent of measurements (some particular properties do exist, for instance the rest mass of particles). We may assume that the measured properties are *contextual*, that is, they depend not only on the state of the system but on the whole experimental context. This point was correctly emphasized by Bohr and, in my opinion, solves all problems of interpretation which might follow from the Kochen-Specker theorem. (Of course, the theorem provides some quantitative statements which should be explained, but here I am addressing the question whether the practical impossibility of getting joint probabilities prevents a realistic interpretation.)

The real difficulty arises when people attempt to reach conclusions which go beyond what follows from the facts. Indeed, we can state that *some properties* do not exist independently of measurements *in some particular instances*, but we should not extrapolate and say that in nature *there are no properties* independent of the observation. This absurd extrapolation was correctly criticized by Einstein with his celebrated rhetorical question *"Is the moon there when nobody looks?"*

One might ask why in the microscopic domain it is frequent that values of the observables are created by the experiments whilst this situation is rare at the macroscopic level. An explanation may be as follows. In the macroscopic world we may study systems with instruments more fine than the object to be studied. E.g., we may

look at the interior of an orange using a knife. In the microscopic domain any macroscopic equipment used for the study of atoms will consist of atoms. This makes our knowledge less direct in the micro than in the macroscopic domain, and more dependent on the context.

The fact that the measurement cannot be understood as simply revealing the values of preexisting properties has led to the introduction of some 'postulates of the measurement' in the Hilbert space quantum formalism as discussed in the next chapter.

1.2.7. Wave-particle duality. The assumption that all quantum entities have a dual nature, particle and wave, is the source of most difficulties for an intuitive understanding of quantum mechanics. But if we do not want to destroy the basic properties of space, the wave-particle duality really involves a contradiction. In fact *particle* means something localized, *wave* means something extended. More precisely, particle (wave) means much smaller (much greater) than some reference length, e.g. a few times the size of an atom in the nonrelativistic quantum mechanics of atoms and molecules. For this reason it is bizarre to say that an atom (with radius about 10 nm) crosses simultaneously through two slits (distant about 1μm). Thus it is not strange that for some people the interference experiments contain all the mysteries of quantum mechanics (in Feynman's words). The problem for a realistic interpretation of the quantum phenomena posed by the wave-particle duality is certainly hard. It will be considered for some particular cases in later chapters of the book, but in the following I sketch a possible general solution. We may assume that in nature there are both particles and fields (waves), the particle behaviour of fields deriving from the interaction with particles and the wave behaviour of particles from the interaction with fields. The difference with the macroscopic world, where there are also particles and fields, is that interactions are more relevant and complex in the microscopic domain.

I think that electrons (or protons, neutrons, atoms, molecules) are particles, whilst radiation consists of waves. 'Photons' are not particles but mathematical constructs useful for the description of some phenomena [**6**]. Then, how can we interpret the interference experiments where we observe fringes typical of waves, but these fringes appear as sets of localized events which are typical of particles? In the case of radiation the interference may be easily understood in classical terms, and the problem is the particle behaviour in detection. The

opposite is true for particles like atoms. Its localized detection is easy to understand but their interference puts the problem. Let us study the two cases separately.

The detection of 'individual photons' in a photographic plate is due to the atomic nature of the plate. In this interpretation, saying that radiation are particles because they give rise to individual blackened grains is like saying that wind is corpuscular because the number of trees falling in the forest is an integer. Of course, in both cases, the photo and the forest, there is a random element. This is obvious for the wind but there is also a random element in the radiation: the quantum noise or quantum fluctuations of the *vacuum fields*. The assumption that vacuum fields are real is an essential hypothesis in this book. I claim that this hypothesis is crucial for a realistic interpretation of quantum mechanics. The detection process in a photon counter may be explained as a transfer of energy from the field to individual atoms or molecules, this energy amplified appropriately by the detector giving rise to one count. In this process the vacuum fields would play a relevant role.

The wave behaviour of neutrons, atoms or molecules, for instance in the two-slits experiments, is more difficult to understand. We might assume that it is caused by the vacuum fields, consisting of a random electromagnetic radiation, fluctuations of the spacetime metric and other components. That noise is what interferes, producing a pattern which guides the particles. The picture has some similarity with the old proposal by L. de Broglie (the pilot wave theory) or the picture offered by 'Bohmian mechanics', to be discussed in chapter 4, but there are two important differences. Firstly in our view there is a clear physical entity different from the particles, namely the vacuum field fluctuations (or quantum noise), whilst the particle remains localized all the time. Secondly there is a random element which is not present in Bohmian mechanics.

In any case, the wave-particle duality is a big problem for an intuitive picture of the microscopic world. I will make a more detailed discussion of some examples in chapters 4 and 6.

1.2.8. Quantum wholeness and nonlocality. Maybe the phenomena that put the greatest difficulties for a realistic interpretation is the existence of a kind of wholeness in some quantum phenomena. Indeed, there exist correlations between distant bodies that apparently cannot be explained with an intuitive picture. Usually these

phenomena are associated to *entanglement*. I propose that entanglement is a correlation mediated by the vacuum fields. In chapters 5 and 6 we will present examples that fit with this assumption.

More dramatic is the existence, proved by John Bell in 1964, of inequalities that are necessary for realistic local models. The subject requires a detailed discussion and the whole chapter 3 is devoted to it.

1.2.9. Conclusions. The analysis of the most characteristic quantum phenomena leads us to emphasize a point that is crucial for the attempt of reaching a picture of the quantum world. *The difficulties for a realistic interpretation of quantum mechanics may derive from a number of unneeded assumptions, adhered to the quantum theory for historical reasons. In some cases the difficulties are caused by an excess of idealization in the interpretation of the experiments.* The use of idealizations is supposed to contribute to clarity, but in my opinion it has the opposite effect; it contributes to misunderstanding. It is true that it may simplify the teaching of *how to use* quantum mechanics, but it puts a strong obstacle for an intuitive picture of the quantum world. A typical example, discussed above, is the use of first order perturbation theory in the study of emission or absorption of light, which suggests sudden jumps in the atom thus hiding the fact that a more accurate treatment (via quantum electrodynamics) predicts a continuous evolution of atom plus field.

The conclusion of the analysis made in this section is that *a realistic interpretation of quantum mechanics should refer to the actual observations or experiments, but not to the predictions obtained from the quantum formalism.*

1.3. The assumption that the vacuum fields are real

For the realistic interpretation of quantum mechanics proposed in this book, the hypothesis that the quantum vacuum fields are real stochastic fields is crucial. Therefore I devote this section to give arguments supporting that assumption.

1.3.1. The quantum vacuum in field theory. The existence of random radiation in vacuum, even at zero Kelvin, appears for the first time in Planck's second radiation theory of 1912. This zero-point energy of the electromagnetic field (ZPF) was rejected because it is divergent, although the consequences of its possible reality were soon explored by several authors, including Einstein and Nernst [**26**].

The ZPF reappeared in 1927 when Dirac quantized the electromagnetic field via an expansion in normal modes, that is, plane waves in the case of free space. In the standard (Hilbert-space) formalism the Hamiltonian operator of the field may be written (see chapter 2) as

$$(8)\quad H = \frac{1}{2}\hbar \sum_{js} \omega_j \left(a_{js} + a_{js}^{\dagger}\right)^2 = \frac{1}{2}\hbar \sum_{js} \omega_j (a_{js}^2 + a_{js}^{\dagger 2} + 2a_{js}^{\dagger} a_{js} + 1),$$

where ω_j is the (angular) frequency and s the polarization of a normal mode, a_{js} and $a_{js}^{\dagger}$ being the 'annihilation and creation operators of photons' in that mode, and we have taken the commutation rules into account in the last equality. Energy is given by the vacuum expectation of the Hamiltonian, that is,

$$\langle 0\,|H|\,0\rangle = \frac{1}{2}\hbar \sum_{j} \omega_j \left\langle 0 \left| a_{js}^2 + a_{js}^{\dagger 2} + 2a_{js}^{\dagger} a_{js} \right| 0 \right\rangle + \frac{1}{2}\hbar \sum_{j} \omega_j \langle 0\,|1|\,0\rangle$$

$$(9)\qquad = \sum_{js} (\frac{1}{2}\hbar\omega_j),$$

the former expectation being null. The result corresponds to a mean energy $\frac{1}{2}\hbar\omega_j$ per mode, that fits in the arguments of section 1.2.2 above.

A problem is that the total energy density in space diverges when we sum over all (infinitely many) modes. The standard solution to the divergence problem is to remove the term that contributes in eq.(9), a procedure which is known as 'normal ordering'. It consists of writing the annihilation operators to the right; that is, assuming that the correct Hamiltonian is not the first expression in eq.(8), but the last one with unity removed. It may be realized that the normal ordering is equivalent to choosing the zero of energies at the level of the vacuum. It provides a practical procedure useful in quantum-mechanical calculations, but it is not a good solution for many authors. They see it as an 'ad hoc' assumption, just aimed at removing unpleasant divergences. For these authors the ZPF is a logical consequence of quantization, and the solution of the divergence problem should come from a more natural mechanism. Furthermore it has been shown that the assumption of reality of the ZPF combined with the classical laws of electrodynamics allows explaining some phenomena usually taken as purely quantum, an approach known as stochastic electrodynamics that will be reviewed in chapter 5.

Around 1947 two discoveries reinforced the hypothesis that the quantum vacuum fields are real, namely the Lamb shift and the Casimir effect. Lamb and Retherford observed an unexpected absorption of microwave radiation by atomic hydrogen, that was soon explained in terms of the interaction of the atom with the quantized electromagnetic field, that involves the vacuum radiation (ZPF). Indeed, Willis Lamb claimed to be the discoverer of the ZPF by experiment. Furthermore he wrote that "photons are the quanta of the electromagnetic field, but they are not particles" [**6**]. The discovery led in a few years to the development of quantum electrodynamics (QED), a theory that allows predictions in spectacular agreement with experiments, and it was the starting point for the whole theory of relativistic quantum fields. The success of QED rests on renormalization techniques, that is, assuming that physical particles, e.g. electrons, are dressed with 'virtual fields' making their physical mass and charge different from the bare quantities. In my opinion the assumptions behind renormalization are actually a reinforcement of the reality of the quantum vacuum fields, although people avoid commitment with that conclusion using the word 'virtual' as an alternative to 'really existing'.

The Casimir effect consists of the attraction between two parallel perfectly conducting plates in vacuum. The force F per unit area depends on the distance d between the plates,

$$F = -\frac{\pi^2 \hbar c}{240 d^4}, \tag{10}$$

a force confirmed empirically [**26**]. The reason for the attraction may be understood qualitatively as follows. In equilibrium the electric field of the zero-point radiation, ZPF, should vanish on or be normal to any plate surface, otherwise a current would be produced. This fact constrains the possible low frequency normal modes of the radiation, that is, those having wavelengths of the order the distance between plates or larger, although the distribution of high frequency modes is barely modified by the presence of the plates. Therefore, ascribing an energy $\frac{1}{2}\hbar\omega_j$ to every mode, the total energy of the ZPF in space becomes a function of the distance between plates and the derivative of that function with respect to the distance leads to eq.(10). Actually, the calculated energy of the ZPF diverges if we sum over all radiation modes, including those with arbitrarily high frequency, but there are regularization procedures that give the correct result [**26**].

They essentially subtract the field energy with the plates present minus the energy with the plates removed. The physical picture of the phenomenon is that the radiation pressures in both faces of each plate are different and this is the reason for a net force on the plate. The Casimir effect is currently considered the most strong argument for the reality of the quantum vacuum fields. For us it is specially relevant because it provides an example of the fact that the difference between the radiation arriving at the two faces of a plate is what matters for the radiation-plate interaction, rather than the radiation acting on one side of the plate. We will make a similar assumption in the model of detector to be proposed in chapter 6 section 6.5.

1.3.2. The quantum vacuum in astrophysics and cosmology. In laboratory physics, where gravity usually plays no role, the reality of the quantum vacuum fields is not too relevant a question. In fact, their possibly huge, or divergent, energy may be usually ignored choosing the zero energy at the vacuum level as said above. However, this choice is no longer innocuous in the presence of gravity because, according to relativity theory, energy gravitates; whence the huge vacuum energy should produce a huge gravitational field. Therefore the possible existence of a vacuum energy is a relevant question in astrophysics and cosmology. A more extended study of the question will be made in chapter 7, summed up as follows.

From long ago the quantum vacuum has been related to the cosmological constant, a term that Einstein introduced in general relativity in order to make possible his 1917 model of universe. The reason for the relation is that the vacuum, even if it possesses some energy density, ρ, should be Lorentz invariant. Then it should have also pressure, P, with the vacuum equation of state $P = -\rho$, that is, equivalent to a cosmological constant term. Actually, there was no empirical evidence for a cosmological constant until 1999 but, even before that date, many authors speculated about the possibility that the quantum vacuum fields give rise to a cosmological term. However, there was a big problem, namely the vacuum energy density appears to be infinite if no cut-off exists, and the only natural cut-off seems to be at the Planck scale. This choice would give a cosmological term about 10^{123} times the value derived from observations, a huge discrepancy known as the 'cosmological constant problem'. The most simple solution is to assume a cancelation between positive and negative terms of the vacuum energy. However, it seems difficult to believe

that the cancelation is not exact but it is fine tuned to reduce the disagreement by 123 orders of magnitude. This problem might be an argument against the reality of the vacuum fields. Nevertheless, the difficulty with the fine tuning is solved in a plausible way if we assume that the *mean vacuum energy and pressure* are exactly balanced but the cosmological constant derives from the vacuum *fluctuations* (see chapter 7 section 7.4 for details).

As a conclusion there are strong arguments for the reality of the quantum vacuum fields possessing energy and pressure. Firstly the vacuum fields are a logical consequence of quantization and should not be removed artificially. Secondly the action of the vacuum fields in laboratory experiments may be weak, usually not observable, due to an almost complete cancelation of the effects of radiation traveling in opposite directions, but it may be measured in some delicate breakings of balance like the Casimir effect. In astrophysics and cosmology the alleged huge energy and pressure of the vacuum fields may not hold, because a cancelation might exists between positive and negative contributions.

1.4. Sketch of a physical model of the quantum world

A realistic interpretation, providing a physical model of the world, would make quantum mechanics more palatable to lovers of *theory* in the sense of contemplation. It would allow 'understanding quantum mechanics'. The picture supported here is sketched in the following. Nevertheless, as said above, there are real difficulties for a realistic interpretation that this book does not eliminate completely.

The framework of the physical world is a manifold with four dimensions and intrinsic curvature: the spacetime. In physics we use names like energy, momentum or angular momentum for some properties of matter that I propose to be types of spacetime curvature. I believe that the Einstein equation of general relativity is just the 'dictionary' that translates the physical concepts to geometrical properties, the latter summarized in the Einstein curvature tensor and the former enclosed in the stress-energy tensor (see chapter 7 section 7.2 for details). In spacetime there are fields, that is, scalar, vector or spinor functions of the spacetime points. The first two kinds (scalar and vector fields) are named Bose fields and the last Fermi fields. Maybe the Standard Model of particle physics contains all existing fields in nature, but we cannot exclude that there is physics beyond that model.

The theory stating the appropriate formalism for the study of the fields and their interactions is relativistic quantum field theory. Quantization is a name for the process of describing the fields as (peculiar) stochastic processes. To begin with, the quantization should be applied to spacetime, which should be treated as stochastic. That is, a probability distribution over possible samples of spacetime or, in other words, a distribution of metrics. For a more extended discussion see chapter 7 section 2.

I believe that Bose fields are actually continuous, that is, waves in popular language. In contrast, Fermi fields describe small, or point-like, particles. The stochastic characterization of fields is commonly made using a Hilbert space (more properly C*-algebras). For a simple presentation the field is written as a linear combination of normal modes (e.g. plane waves in free space) and the field amplitudes are represented by operators on the Hilbert space. This provides a formalism very useful for calculations but lacking a realistic interpretation, that is, a picture of reality. For the electromagnetic field a picture may be obtained via the Weyl-Wigner transform, which will be studied in some detail in chapter 6 section 6.3. In summary it characterizes the vacuum state of the field as a Gaussian stochastic field with a mean energy per mode $\frac{1}{2}\hbar\omega$ whence states with additional energy (field excitations in quantum language) may be got using the creation operators for 'photons'. I believe that this picture might be extended to all Bose fields. The picture of Fermi fields is less clear, but I believe that they consist of particles, as said above. All this refers to fields interacting via electromagnetic or weak forces alone, because strongly interacting fields, that is, hadrons, probably cannot be studied in isolation; whence getting a physical picture is more difficult.

The objects that we may study in the laboratory, for instance an atom or a piece of bulk matter, are built from the fundamental fields. They are usually quite complex, involving many interacting fields, in particular the vacuum fields. In any case, an important consequence is that no system may be fully isolated from the rest of the universe. Indeed, every system should be effectively interacting with many other systems via the vacuum fields. But in order to be able to make physics we should assume that microscopic systems, even if not isolated, may be treated with a theory that, in some form, takes the interaction with the environment into account. I believe that this theory is quantum mechanics. For instance, if we represent the state of an atom by a

state vector it is plausible to assume that this representation corresponds to the atom 'dressed' by all fields that interact with it. This is consistent with the fact that in quantum electrodynamics the physical electrons are never 'bare' but 'dressed with virtual photons and electron-positron pairs'. The word 'virtual' is just a name for something that we know to have observable effects, but we do not want to be committed with its real existence. In my view the representation of an atom by a wave-function takes the (approximate) action of the vacuum fields into account as is shown by the use of the physical, rather than bare, mass and charge of the electrons in calculations (say when solving Schrödinger equation). Thus the magic of quantum theory is that it allows studying complex objects via a quite simple formalism.

As a consequence it is a daring attitude to pretend that a state vector represents faithfully the *actual state of an individual system*. But it is more plausible to assume that the state vector represents the *relevant* information available about the system. The conclusion is that the quantum-mechanical description is necessarily incomplete and this is the real cause of the claimed 'irreducible probabilistic character of the physical laws'. For instance the fact that an atom decays at a time that cannot be predicted derives from the fluctuations of the vacuum fields that actually stimulate the decay.

The concept of isolated system is the cornerstone of classical physics and, therefore, it is not strange that it was also introduced in quantum physics. It is true that early authors dealing with quantum theory, like Planck, Einstein and Nernst, studied the possible existence and influence of vacuum (i.e. not thermal) fluctuations. However, the success of the Bohr atomic model, where the concept of fluctuation was absent, reinforced the idea that quantum systems may be treated as isolated. These fluctuations reappeared in modern quantum mechanics associated to the zero-point energy of bounded quantum systems. However, in the alternative of either rejecting the assumption of isolated system or dismissing the reality of the quantum fluctuations, the mainstream of the community chose the latter. In my opinion that choice has been the source of most difficulties for a realistic interpretation of the quantum formalism.

The existence of real vacuum fluctuations gives rise to two characteristic traits of quantum physics. Firstly quantum theory should be *probabilistic*. Secondly it should present a kind of *wholeness*, quite strange to classical physics where the concept of isolated system is crucial. The fact that the vacuum fluctuations at different points may be

correlated is the origin of the wholeness, which manifests specially in the phenomenon of *entanglement*.

Bibliography

[1] M. Jammer: *The conceptual development of quantum mechanics.* McGraw-Hill, New York, 1966.

[2] M. Jammer: *The philosophy of quantum mechanics: The interpretations of quantum mechanics in historical perspective.* John Wiley and Sons, New York, 1974.

[3] J. A. Wheeler, W. H. Zurek: *Quantum theory and measurement.* Princeton University Press, Princeton, 1983. This book contains reprints of many early articles on the foundations of quantum physics.

[4] A. Fine: *The Shaky Game: Einstein realism and the quantum theory. Science and its conceptual foundations.* Univ. of Chicago Press,1986

[5] G. Auletta: *Foundations and interpretation of quantum mechanics.* World Scientific, Singapore, 2000.

[6] W. Lamb: *The interpretation of quantum mechanics.* Rinton Press, Princeton, 2001.

[7] B. d'Espagnat: *On physics and philosophy.* Princeton U. P. (2006).

[8] P. Mittelstaedt: The problem of interpretation of modern physics. *Found. Phys.* **41**, 1667-1676 (2011).

[9] R. Beneduci, F. E. Schroeck: On the unavoidability of the interpretations of quantum mechanics. *Am. J. Phys.* **82**, 80-82 (2014).

[10] T. Norsen: *Foundations of quantum mechanics.* Springer. 2017.

[11] F. Laloë: *Do we really understand quantum mechanics? Strange correlations, paradoxes, and theorems.* 2nd. edition, Cambridge Univ. Press, 2019.

[12] B. Drummond: Understanding quantum mechanics: a review and synthesis in precise language. *Open Phys.* **17**, 390-437 (2019).

[13] C. J. Isham: *Lectures on quantum theory. Mathematical and structural foundations.* Imperial College Press, London, 1995.

[14] F. Suppe: *The structure of scientific theories.* University of Illinois Press, Urbana, 1977.

[15] M. Redhead: *Incompleteness, nonlocality and realism.* Clarendon Press. Oxford, 1990.

[16] A. Einstein, B. Podolsky, N. Rosen: Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47**, 777-780 (1935). (Reprinted in Wheeler and Zurek (1983)).

[17] N. Bohr: Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **48**, 696 (1935).

[18] K. R. Popper: *The logic of scientific discovery.* Hutchinson, London, 1980. German original, 1934.

[19] K. R. Popper: *Conjectures and refutations. The growth of scientific knowledge.* Basic books. New York, 1962.

[20] I. Lakatos: *The methodology of scientific research programmes.* Cambridge University Press, Cambridge, 1980

[21] W. Heisenberg: *Physics and beyond : encounters and conversations.* Harper & Row, New York, 1971. Chapter 5: Quantum Mechanics and a talk with Einstein (1925-1926) p. 59-69.

[22] E. Santos: Is there an electromagnetic background radiation underlying the quantum phenomena? *Anales de Física (Madrid)*, **64**, 317-320 (1968).

[23] T. W. Marshall, E. Santos: Can the Stern-Gerlach effect exhibit quantum nonlocality? In P. Bush, P. Lahti and P. Mittelstaedt (Eds.), *Symposium on the Foundations of Modern Physics*. Singapore: World Scientific, pp. 361-368 (1993).

[24] A. Einstein: Remarks concering the essays brought toghether in this cooperative volume. In P. A. Schilpp (Ed.), *Albert Einstein: Philosopher-Scientist* (pp. 665-688). La Salle Illinois. Open Court. 1949.

[25] N. Bohr: The quantum postulate and the recent development of atomic theory. *Nature*, **121**, 580-90 (1928). (Reprinted in Wheeler and Zurek (1983)).

[26] P. W. Milonni: *The quantum vacuum. An introduction to quantum electrodynamics.* Academic Press. New York. 1994.

CHAPTER 2

The Hilbert space formulation of quantum mechanics

2.1. The origin of the formalism

The standard formalism of quantum mechanics rests on the mathematical theory of Hilbert spaces. Historically the formulation in terms of abstract mathematical objects is due to Paul Dirac [**1**], who proposed it soon after the formulation in terms of matrices by Werner Heisenberg in the summer 1925. Dirac's formalism provides a unification with Erwin Schrödinger's wave mechanics published in early 1926. In the following I comment on the origin of the standard quantum formalism which explains the difficulty for the physical interpretation of the theory. For the history of this period see the books by Jammer quoted in chapter 1.

2.1.1. Heisenberg, Dirac, and von Neumann. Heisenberg formulated quantum mechanics substituting matrices for the dynamical variables of classical mechanics. The original article is hard to read and several aspects were clarified in collaboration with Max Born and Pascual Jordan. The starting point is the observation that some observable quantities in atomic spectra depend on two indices, whence the set of possible values may be arranged as a square matrix. In particular the frequencies of the emitted or absorbed radiation by an atom, ω_{jk}, which are related to energy differences via Bohr equation $\hbar\omega_{jk} = E_j - E_k$. This led Heisenberg to represent all dynamical quantities by square matrices. In the attempt to use the classical laws of motion for the new theory he was confronted with the novelty that, unlike classical variables, the product of (square) matrices is not commutative in general. Actually, this proved not to be a problem but an advantage that Heisenberg exploited postulating that the matrices associated to the position and momentum of a particle fulfil the commutation rule (in one dimension)

$$\sum_k X_{jk}P_{kl} - \sum_k P_{jk}X_{kl} = i\hbar\delta_{jl}, \tag{11}$$

where X_{jk} is the (j,k) element of the matrix associated to position, and similar for the momentum, i is the imaginary unit, $\hbar$ is the Planck constant and δ_{jk} is the Kronecker delta (i.e. $\delta_{jj} = 1$ and $\delta_{jk} = 0$ if $k \neq j$). From eq.(11) and the classical dynamical laws Heisenberg devised a method for the calculation of atomic spectra and other observable properties of atoms. A more practical method was the equation postulated by Schrödinger a few months later. But in the meantime Dirac made an important contribution that I revisit in the following.

Dirac started from the fact that square matrices might be interpreted as transformations of vectors. Indeed, one of the applications of matrices is to represent changes in position of bodies. For instance, a rotation in space may be described via the change in position of any particle, say $\mathbf{r} \equiv (x_1, x_2, x_3)$, to another position, $\mathbf{r}' \equiv (x'_1, x'_2, x'_3)$, that is,

$$x'_j = \sum_k R_{jk} x_k \qquad \Rightarrow \qquad \mathbf{r} \to \mathbf{r}' = \hat{R}\mathbf{r},$$

where R_{jk} is the (j,k) element of the rotation matrix and $\hat{R}$ an operator that transforms any vector in another one. Mathematically the operator is a mapping of the space V of vectors onto itself. Rotations in two dimensions commute, that is, the order of two mappings is irrelevant, which may be represented $\hat{R}_1\hat{R}_2 = \hat{R}_2\hat{R}_1$, but this is not generally true in three dimensions. The same happens with the associated matrices.

By analogy Dirac assumed that the matrices of Heisenberg correspond to operators acting on the vectors of some linear space. In particular the evolution of a quantum system might be viewed as a mapping of the form

$$| \psi(t)\rangle = \hat{U}(t, t_0) \mid \psi(t_0)\rangle, \tag{12}$$

where $\mid \psi(t)\rangle$ is a vector, written with the notation introduced by Dirac which is now standard. The vector $\mid \psi\rangle$ might be represented by a column matrix and the operator by a square matrix as those used by Heisenberg in his quantum mechanics. The vector is supposed to correspond to a state of a physical system, whence the vector space becomes the space of states.

After that, it was necessary to characterize the evolution operator that should be derived from the Heisenberg equations for the time evolution of the matrices associated to de dynamical variables representing dynamical variables. The fundamental result is that the time derivative of $\hat{U}(t, t_0)$ is related to the Hamiltonian operator so that eq.(12) leads to

$$i\hbar \frac{d}{dt} \mid \psi(t)\rangle = \hat{H} \mid \psi(t)\rangle,$$

where $\mid \psi\rangle$ is a vector of the assumed space. This is the basic quantum equation for the evolution, that bears the name of Schrödinger although he did not write it in this general form (i.e. with abstract operators and vectors).

As an example let us apply it to a particle in a potential V which is (in one dimension)

$$i\hbar\frac{d}{dt}\mid\psi(t)\rangle=\left(\frac{1}{2m}\hat{p}^2+V(\hat{x})\right)\mid\psi(t_0)\rangle. \tag{13}$$

For practical calculations it is convenient to consider the vectors $\mid\psi\rangle$ to be elements of a linear space of functions $\{\psi(x,t)\}$. Then we shall search for appropriate operators $\hat{p}$ and $\hat{x}$ acting on functions. It may be realized that the following operators are consistent with the commutation rules in eq.(11)

$$\hat{x}\mid\psi\rangle\rightarrow x\psi(x,t),\qquad \hat{p}\mid\psi\rangle\rightarrow-i\hbar\frac{\partial}{\partial x}\psi(x,t).$$

In fact, for any vector $\mid\psi\rangle$ we have

$$(\hat{p}\hat{x}-\hat{x}\hat{p})\mid\psi\rangle=-i\hbar\mid\psi\rangle\quad\leftrightarrow\quad\left[-i\hbar\frac{\partial}{\partial x}(x\psi)+xi\hbar\frac{\partial}{\partial x}\psi=-i\hbar\psi\right].$$

Hence we get Schrödinger equation in the form proposed by him, that is (now in 3 dimensions),

$$i\hbar\frac{\partial}{\partial t}\psi(x,y,z,t)=-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\psi+V(x,y,z)\psi.$$

In summary, taking advantage of the analogy between matrices and functions of one or several variables, Dirac showed that both formulations of quantum mechanics (Heisenberg's and Schrödinger's) were physically equivalent, something that the second author also established independently. The vector spaces of functions require some mathematical restrictions. In particular, topological postulates are needed when the spaces have infinite dimension. For instance, the solutions of Schrödinger equation, usually named wave-functions, should be twice derivable and normalizable.

In 1932 John von Neumann [**2**] established the quantum formalism on a rigorous mathematical basis defining the Dirac linear spaces to be Hilbert spaces. Hilbert spaces became the standard, or canonical, formulation of quantum mechanics. For practical calculations in systems of electrons (or atoms, molecules and solids) it proves convenient to use Schrödinger equation or the relativistic generalization due to Dirac in 1928 [**1**].

2.1.2. The lack of a physical picture. Hilbert spaces supply an elegant mathematical formalism able to establish general properties like symmetries. They have also allowed to extend the theory from

atoms and molecules to nuclei and later to fundamental particles. It is not strange that the combination of beautiful formalism and practical success has dazzled physicists, making the Hilbert space formalism the standard one for quantum theory. In particular most proposed interpretations of the theory refer to that formalism as we will discuss in section 2.4 below. To this fact has contributed also the failure of alternative formulations like Bohmian or stochastic mechanics and the difficulties of interpretation for other formalisms like the Weyl-Wigner (but see chapter 6) or the Feynman path integrals. These alternative formulations will be revisited in chapter 4.

The abstract character of the formalism darkens the physical interpretation, that is, it makes it difficult to get a picture of the natural world. In contrast, in classical physics the abstract formulations like Lagrange's or Hamilton's came after a slow process from observations of particular phenomena, in a route similar to the process of learning, going from simple cases to general formulations. Indeed, the development began with the study of motion and the causes of it, forces. The naive observation suggests that motion is more rapid if the force applied is greater, a common experience when we push a body. After a long process, which included accepting the relevance of friction and later the existence of forces at a distance, Newton was able to state the foundations of classical mechanics with his well known laws of dynamics. The interpretation is not difficult in the sense that these laws provide an intuitive picture of motion and its relation with forces. From this point on, a process of increasing abstraction led to the concepts of energy and momentum and later to the elegant mathematical formulations due to Lagrange and Hamilton. But it seems that no physical picture could be easily obtained if we started from a formulation in terms of generalized coordinates and momenta.

In sharp contrast, quantum mechanics appeared in abstract form at the very beginning, in an attempt to have a simple calculational tool to deal with rather complex phenomena like the emission and absorption of light by atomic gases, that is atomic spectra. The empirical evidence was initially interpreted introducing simple assumptions that evaded a deep understanding of the detailed interaction of atoms with light. This was the case for the Bohr atomic model, able to predict atomic spectra. The elaboration of these preliminary theories did not pursued in the direction of getting a deeper picture of the structure of atoms but rather with the aim of finding a good calculational tool free from internal contradictions. At the end the standard formalism

did not lead just to modifications of the existing picture of the natural world (i.e. classical physics), but to a change in the very concept of what is science. This was indeed the case with the quantum mechanics of Heisenberg, reinforced with the philosophical contributions of this author and Bohr, as discussed in chapter 1.

Schrödinger wave mechanics attempted a less drastic change of the old classical laws, more in line with tradition, not questioning that the purpose of physics is to provide a picture of the natural world. Indeed, Schrödinger proposed to understand line spectra of atoms by analogy with the well known sound spectra of musical instrument. This approach was dismissed because Schrödinger interpreted his equation as showing that the electron is a continuous charge distribution, which is untenable as discussed in chapter 1. At present the Schrödinger equation is fully integrated in the standard formalism, interpreting that his wave-function does not refer to actual fields or waves but to probability amplitudes. In my view 'probability amplitude' is just the juxtaposition of two contradictory words. I believe that the wave-function may be a kind of average of a set of field amplitudes. That is, I think that the analogy of Schrödinger's equation with equations for the vibrations of musical instruments is valid; air in a flute is similar to vibrations of fields, including vacuum fields, in an atom. Consequently I think that the correct route to understand quantum theory is not to start with elementary (nonrelativistic) quantum mechanics, extending later the interpretation to field theory, which has been the historical path. We should try to understand first quantum field theory, as attempted in chapter 1 section 1.4.

2.2. The standard postulates of quantum mechanics

Here the formalism of quantum mechanics will be summarized presenting it as a set of postulates, although I will comment on these standard postulates critically in section 2.2.2. For the sake of clarity, I will not present the postulates in full generality, but make the fiction that all operators corresponding to observables possess discrete (numerable) spectrum. For a more rigorous formulation and explanation of the postulates see any book on quantum mechanics, e.g. the one by Galindo and Pascual [**3**].

In classical physics one distinguishes between mechanics and field theory. The former deals with systems consisting of a finite set of point particles (or bodies that may be approximated as pointlike) and the latter with continuous distributions of matter. However, in

practice the distinction is rather made between systems possessing either finite or infinite number of degrees of freedom, so that rigid bodies are usually studied in books of mechanics, rather than field theory. On the other hand it is frequent that systems consisting of a very large (finite) number of particles are more easily treated as continuous distributions and therefore studied as fields. This is for instance the case of elasticity theory.

In quantum physics there is also a distinction between mechanics (of finite sets of point particles) and field theory. However, the standard view is that the latter is the fundamental theory whilst the former is an approximation valid in a limited domain, when there is no creation and annihilation of particles, which requires that the velocities involved are small in comparison with that of light. Indeed, a characteristic trait of quantum field theory is the prediction of the possible creation or annihilation of particles. Therefore for high velocities (special) relativistic *field* theory must be used, but no quantum relativistic *mechanics* exists. In fact, kinetic energies of order the rest mass of particles would allow possible creation, whence a field theoretical treatment is necessary. For these reasons it appears that we should formulate the quantum postulates for field theory and derive from them those of mechanics. However, it is the case that all experiments would involve measurements with instruments at rest in the laboratory or moving slowly and, on the other hand, measurement is a crucial process according to the quantum canonical formalism, that requires specific postulates. Therefore it is appropriate to establish the postulates for quantum mechanics and later on to extend them to field theory. This is what I will do here.

2.2.1. States and observables. *1. To every physical system it is associated a Hilbert space, H, complex and separable. A pure state of the physical system at a time t is represented by a unitary ray belonging to the Hilbert space. A normalized element of the ray, $|\,\psi(t)\rangle$, is called state vector.*

A pure state contains the maximal possible information about the system. The opposite to pure is 'mixed state', which would correspond to a probability distribution of pure states. That is, if $\{|\,\psi_j\rangle\}$ is a set of pure states, a mixed state is defined by a set of probabilities $\{p_j\}$. Although it is possible to study more general cases, here I will assume that every two vectors of the set $\{|\,\psi_j\rangle\}$ used to define the mixed state are normalized and orthogonal, that is, $\langle\psi_j \mid \psi_l\rangle = \delta_{jk}$. Then

any mixed state may be represented by a 'density operator' defined in terms of the probabilities of pure states as

$$\hat{\rho} = \sum_j p_j \mid \psi_j \rangle \langle \psi_j \mid . \tag{14}$$

Density operators fulfil the conditions of being self-adjoint, positive, and having unit trace, that is,

$$\hat{\rho} = \hat{\rho}^\dagger, \qquad \hat{\rho} \geq \hat{\rho}^2, \qquad \mathrm{Tr}\,(\hat{\rho}) = 1. \tag{15}$$

Pure states may be also defined as those corresponding to density operators having only one term in the sum eq.(14) . In this case that operator is idempotent, that is, it fulfils $\hat{\rho} = \hat{\rho}^2$. This definition is equivalent to saying that it corresponds to a ray, as stated above.

2. Every observable, A, of a physical system is represented by a self-adjoint, linear operator, $\hat{A}$, acting on the Hilbert space of the system.

Here, and in the following, I will distinguish the *physical* observable, A, from the *operator* which represents it, $\hat{A}$. I shall avoid the name 'observable' for an operator in spite of this being a common practice in books and articles.

Any self-adjoint operator, $\hat{A}$, (with discrete spectrum) may be represented in the form

$$\hat{A} = \sum_j a_j \hat{P}_j, \tag{16}$$

where $\{a_j\}$ is the set of eigenvalues of $\hat{A}$ and $\hat{P}_j$ is a projector (fulfilling $\hat{P}_j^2 = \hat{P}_j$) onto the subspace associated to a_j, that is, the subspace spanned by the eigenvectors of $\hat{A}$ pertaining to eigenvalue a_j. Recently a more general type of observable is used, which is represented by an operator defined by an expression like eq.(16) with $\hat{P}_j$ not necessarily being a projector but only a positive operator (fulfilling $\hat{P}_j \geq \hat{P}_j^2$). Operators like eq.(16) are known as 'positive operator valued measures (POVM)'. The generalization is useful in some cases, but it is not needed for a realistic interpretation of quantum mechanics and will not be used in this book.

After these postulates (Schrödinger's) wave mechanics of a single particle may be seen as a realization of the Hilbert space by means of (square integrable, therefore normalizable) functions $\psi\,(x, y, z)$ and

(Heisenberg's) quantum mechanics as a realization by means of (infinite) matrices. Some authors, although not all, introduce the linear superposition principle in the form of a postulate, as follows:

3. If $\mid \psi_j\rangle$ *and* $\mid \psi_l\rangle$ *represent two possible pure states of the system, then any linear combination*

$$\alpha \mid \psi_j\rangle + \beta \mid \psi_l\rangle,$$

where α *and* β *are complex numbers satisfying* $|\alpha|^2 + |\beta|^2 = 1$*, is also a possible state of the system, except for superselection rules.* A superselection rule occurs if $\mid \psi_j\rangle$ and $\mid \psi_l\rangle$ are eigenvectors, with different eigenvalue, of an operator representing an absolutely conserved quantity (e.g. the electric charge).

2.2.2. Measurement. *4. If a physical system is in a pure state represented by the normalized vector* $\mid \psi\rangle$*, then in a measurement of the observable* A *the probability of getting the value* a_j *is* $P_j = |\langle a_j \mid \psi\rangle|^2$*, where* $\langle a_j \mid$ *is the eigenvector of the operator* $\hat{A}$ *corresponding to the eigenvalue* a_j*.*

A consequence of this postulate is that the expectation value of the observable A when the system is in the pure state $\mid \psi\rangle$ is

$$\langle A\rangle = \sum_{j=1}^{n} P_j a_j = \sum_{j=1}^{n} a_j \left|\langle a_j \mid \psi\rangle\right|^2 .$$

For a mixed state the expectation value becomes

$$\langle A\rangle = \sum_{k=1}^{n} p_k \langle \psi_k \mid \hat{A} \mid \psi_k\rangle = \mathrm{Tr}\left(\hat{\rho}\hat{A}\right),$$

where $\mathrm{Tr}\,()$ means the trace and the density operator $\hat{\rho}$ was defined in eq.(14). This may be extended to integer powers of the observable, that is,

$$\langle A^n\rangle = \sum_{k=1}^{n} p_k \langle \psi_k \mid \hat{A}^n \mid \psi_k\rangle = \mathrm{Tr}\left(\hat{\rho}\hat{A}^n\right).$$

An important application of the postulate is the measurement of position of a particle. We may get the probability distribution of positions, $\rho\,(\mathbf{x})$, via its moments. We have

$$\langle \mathbf{x}^n\rangle = \langle \psi \mid \hat{\mathbf{x}}^n \mid \psi\rangle.$$

When we pass to the Schrödinger representation by wave-functions, rather than vectors in the Hilbert space, the position operator is just

multiplication times the wave-function. Thus we have

$$\langle \mathbf{x}^n \rangle = \int d^3\mathbf{x}\psi^* (\mathbf{x})\, \mathbf{x}^n \psi (\mathbf{x}) = \int d^3\mathbf{x}\mathbf{x}^n \left|\psi (\mathbf{x})\right|^2,$$

which shows that

$$\rho (\mathbf{x}) = \left|\psi (\mathbf{x})\right|^2. \tag{17}$$

This is known as Born's rule, and it is fundamental for the interpretation of the wave-function. It was introduced as a postulate by Max Born [**4**].

An important property is commutativity. Two operators $\hat{A}$ and $\hat{B}$ commute if

$$\left[\hat{A}, \hat{B}\right] = \hat{A}\hat{B} - \hat{B}\hat{A} = 0.$$

Commuting self-adjoint operators possess a complete set of common eigenvectors and it is said that they are compatible. In fact, it may be proved from the postulates that if the operators associated to two observables do not commute, we cannot expect that the observables can be measured simultaneously because this leads to contradiction with the postulates. On the other hand it is usually assumed that compatible observables can be measured simultaneously.

5. In an ideal measurement of the observable $\hat{A}$, the state after a measurement with the result λ will be one of the eigenstates of $\hat{A}$ belonging to the eigenvalue λ.

In the Hilbert space it is possible to define 'complete sets of commuting (self-adjoint linear) operators' (CSCO), as a set fulfilling: 1) Every two operators commute, 2) There is a common basis of eigenvectors, 3) No operator may be removed without violating condition 2. The most relevant property is that a pure state is determined by the eigenvalues of all operators of a CSCO. Thus Postulate 5 implies that a pure state might be manufactured by the simultaneous measurement of a CSCO on the system.

2.2.3. Evolution. *6. In the time interval between two measurements, pure states of the physical system remain pure, and the representative vector evolves according to the Schrödinger equation*

$$i\hbar \frac{d}{dt} \mid \psi (t)\rangle = \hat{H} \mid \psi (t)\rangle,$$

where $\hat{H}$ is a self-adjoint linear operator called Hamiltonian of the system.

The Schrödinger equation allows getting the time variation of the expectation value of any observable. In fact, we have

$$\frac{d}{dt}\langle\psi(t) \mid \hat{A} \mid \psi(t)\rangle = \frac{i}{\hbar}\langle\psi(t) \mid \left[\hat{A}, \hat{H}\right] \mid \psi(t)\rangle.$$

This is called the Schrödinger picture. The same result may be obtained in the Heisenberg picture. It consists of using always the same wave-function, say $\psi(0)$, but a time dependent operator, $\hat{A}_H(t)$, whose evolution is given by

$$\frac{d}{dt}\hat{A}(t) = \frac{i}{\hbar}\left[\hat{A}_H(t), \hat{H}\right] \tag{18}$$

$$\Rightarrow \qquad \langle\psi(t) \mid \hat{A} \mid \psi(t)\rangle = \langle\psi(0) \mid \hat{A}_H(t) \mid \psi(0)\rangle$$

which is called Heisenberg equation.

2.2.4. Commutation rules. *7. If in a physical system the Cartesian coordinates of position are $x_1, x_2, \ldots, x_n$ and their conjugate momenta $p_1, p_2, \ldots, p_n$, then the operators representing these observables should fulfil the commutation rules*

$$[\hat{x}_j, \hat{x}_k] = 0, \qquad [\hat{p}_j, \hat{p}_k] = 0, \qquad [\hat{x}_j, \hat{p}_k] = i\hbar\delta_{jk}\hat{I},$$

where $\hat{I}$ is the identity operator in the Hilbert space and $[\hat{x}_j, \hat{x}_k] \equiv \hat{x}_j\hat{x}_k - \hat{x}_k\hat{x}_j$.

2.2.5. Quantization. *8. If a system possesses an observable with classical expression $A(x_j, p_k; t)$, then the operator corresponding to this observable is a similar function of the operators $\hat{x}_j, \hat{p}_k$ and the time, t, if there is no ambiguity in the ordering of the operators.*

9. If a system consists of several subsystems, the Hilbert space is the direct product of the spaces of the subsystems.

In the case of a system with two subsystems, this means that if $\{\mid \psi_k\rangle\}$ form a basis in the space of the first subsystem and $\{\mid \chi_l\rangle\}$ a basis in the space of the second one, then $\{\mid \psi_k\rangle \otimes \mid \chi_l\rangle\}$ is a basis in the complete system. For instance, if the states of a particle are represented by the (square integrable) functions of the position variable $\mathbf{r}_1$, and the states of another particle by functions of the variable $\mathbf{r}_2$, then the states of the two-particle system would be functions of both variables, $\mathbf{r}_1$ and $\mathbf{r}_2$.

2.2.6. Quantum field theory. As said above a simple generalization of quantum mechanics to relativistic systems is not possible due to the phenomenon of creation and annihilation of particles. It should be replaced by quantum field theory. From the mathematical point of view passing from a finite to an infinite number of degrees of freedom requires going beyond the Hilbert space and to work with noncommutative algebras, having some mathematical properties not to be specified here [**5**]. But for the purposes of this book the Hilbert space formalism will be sufficient, provided that we use the following approximation. As shown below a quantum field may be expanded in (an infinite number of) normal modes. The states of each mode may be represented by vectors in a Hilbert space, but the states of the whole field could not. However, we might truncate the set of modes and deal only with a finite number of them. In this form the states of the field may be represented by vectors in a Hilbert space.

Postulates 1 to 6 apply to quantum field theory with small modifications. Postulates 7 and 8 should be replaced because in quantum field theory the position coordinates are parameters (as is time in quantum mechanics) and the basic operators represent fields, that is, functions of the coordinates and time. The typical example is the electromagnetic field, which may be formulated by means of a four-vector, whence it is possible to obtain the electric and magnetic fields which together form an antisymmetric tensor in four-dimensional spacetime.

The quantum fields may be studied by an expansion in normal modes, the coefficients being operators. For instance, for an scalar field the expansion would be

$$\hat{\phi}\left(\mathbf{r},t\right)=\sum_{j}\left[\hat{a}_{j}F_{j}\left(\mathbf{r},t\right)+\hat{a}_{j}^{\dagger}F^{*}\left(\mathbf{r},t\right)\right],$$

where $F_j\left(\mathbf{r},t\right)$ and its complex conjugated $F_j^*\left(\mathbf{r},t\right)$ are functions for the mode j, in particular plane waves in the case of free fields. $\hat{a}_j$ and $\hat{a}_j^\dagger$ are called annihilation and creation operators, respectively. Fields with integer spin (e.g. the electromagnetic field which has spin 1) fulfil the commutation rules

$$[\hat{a}_j,\hat{a}_k]=0,\qquad\left[\hat{a}_j^\dagger,\hat{a}_k^\dagger\right]=0,\qquad\left[\hat{a}_j,\hat{a}_k^\dagger\right]=\delta_{jk}.$$

They are called Bose fields.

Fields with half-odd spin, like the Dirac field of electrons and positrons, are called Fermi fields. They may also be expanded in normal modes, but the coefficients fulfil anticommutation rules. For instance,

the Dirac field consists of spinors with four components and there are two kinds of annihilation operators, $\hat{b}_j$ and $\hat{d}_j$, for electrons and positrons respectively, and similarly two kinds of creation operators, $\hat{b}_j^\dagger$ and $\hat{d}_j^\dagger$, fulfilling the anticommutation rules

$$\left\{\hat{b}_j, \hat{b}_k\right\} = \left\{\hat{b}_j^\dagger, \hat{b}_k^\dagger\right\} = \left\{\hat{d}_j, \hat{d}_k\right\} = \left\{\hat{d}_j^\dagger, \hat{d}_k^\dagger\right\} = 0,$$
$$\left\{\hat{b}_j, \hat{d}_k\right\} = \left\{\hat{b}_j^\dagger, \hat{d}_k^\dagger\right\} = \left\{\hat{b}_j, \hat{d}_k^\dagger\right\} = \left\{\hat{d}_j, \hat{b}_k^\dagger\right\} = 0,$$
$$\left\{\hat{b}_j, \hat{b}_k^\dagger\right\} = \delta_{jk}, \qquad \left\{\hat{d}_j, \hat{d}_k^\dagger\right\} = \delta_{jk},$$

where

$$\left\{\hat{b}_j, \hat{b}_k\right\} \equiv \hat{b}_j\hat{b}_k + \hat{b}_k\hat{b}_j.$$

For details see any book on quantum field theory, e.g. [**6**].

2.3. Difficulties for a realistic interpretation

There are two criticisms that I would make to the standard postulates of quantum theory as presented in section 2.2. Firstly, postulates 1 to 3 about states are stronger than needed; secondly, the measurement postulates 4 and 5 possess an instrumentalist flavor, that is, they appear as practical rules rather than fundamental hypotheses and therefore should not be included in a theory. As a consequence I believe that the usual postulates prevent, or make difficult, finding a realistic interpretation of quantum theory. In the following I discuss the question in more detail.

2.3.1. States with nodal surfaces. Postulates 1 to 3 above are too strong because the set of states of a physical system that can be found in nature or manufactured in the laboratory may correspond to a small subset of the vectors of the Hilbert space. This fact is more or less explicitly recognized in careful textbooks, which assume (as in our postulate 1) that not all vectors or density operators correspond to states. However, the weak statement compels us to *reject the one to one correspondence between pure states and vectors*, whence the definition of pure state cannot be derived from the fact that vectors (or, equivalently, idempotent density operators) correspond to the boundary of the whole set of density matrices. It seems more appropriate to postulate that physical states correspond to density operators without any explicit mention to vectors, as in our postulate $1'$ in section 2.3.5 below. Thus postulate 3 should be removed as many textbooks do. Indeed, the postulate, when introduced, has the purpose of making

the explanation of interference simple. However, this is not necessary because interference follows from the Schrödinger equation even without the postulate.

The interpretation of some vector states as corresponding to real physical states leads to a counterintuitive picture. This is the case when the associated wave-function ψ presents nodal surfaces, that is, surfaces where $\psi = 0$, for example the excited states of the hydrogen atom. In fact, the associated wave-functions are products of a radial function $R(r)$ times a spherical harmonic function, the latter carrying the angular dependence. The function $R(r)$ is the product of an exponential times a polynomial and, except in the ground state, the polynomial possesses zeroes that give rise to spherical surfaces where $\psi = 0$. This implies, via Born's rule, that the probability to find the electron is zero in these surfaces but finite outside them, which is very strange. I propose that such wave-functions do not correspond to actual physical states.

An argument for the one to one correspondence between vectors and pure states is that the complete set of vectors is required for some mathematical derivations. However, this is not a valid argument as shown by the well known example of vector spaces in classical field theory, involving Fourier analysis. It is the case that, for the solution of some linear partial differential equations, a complete set of basis functions is needed in order to be able to write any solution as a Fourier expansion in that basis. However, this does not imply that all the basis functions represent physical states. For instance, let us consider the solution of the diffusion equation, say for the cooling of a plate with boundaries $x = -L$ and $x = L$. Assuming that the temperature, T, does not depend on the coordinates y and z, the distribution is governed by the diffusion equation

$$\frac{\partial T(x,t)}{\partial t} = \sigma \frac{\partial^2 T(x,t)}{\partial x^2},$$

where t is the time and σ is a constant related to the conductivity of the medium. The solution may be easily found via a Fourier series expansion. The analogy is reinforced by the fact that functions in the Fourier expansion are also vectors in a vector space. For instance, if the initial and boundary conditions are

$$T(x,0) = T_0, \qquad T(\pm L, t) = T_L < T_0,$$

then the result is

$$T(x,t) = T_L + \sum_n \frac{4(T_0 - T_L)}{(2n+1)\pi}(-1)^n \cos\left[\frac{(2n+1)\pi x}{2L}\right] \exp\left[-\frac{(2n+1)^2 \pi^2 \sigma t}{4L^2}\right].$$

It is obvious that the functions $\cos[(2n+1)\pi x/(2L)]$ *do not* correspond to actual temperature distributions, they are just auxiliary mathematical functions. I propose that the same is true for the quantum vectors or wave-functions.

In order to see the argument more clearly I reproduce a calculation that appears in almost every text of quantum mechanics: The absorption of light by an atom from a plane electromagnetic wave using time-dependent perturbation theory. Here the derivation is made without introducing concepts like photon absorption or transition probability between atomic states, quite common in text-books but devoid of a clear (realistic) physical interpretation. Let us consider an atom initially in the ground state with wave-function $\mid \psi_0\rangle$, interacting with a linearly polarized electromagnetic plane wave. We will represent the electric field of the wave by $F(t)$, choosing the propagation direction as Z axis and the direction of the field as X axis. The total Hamiltonian of the atom, say with a single electron, interacting with the radiation may be written as

$$H = H_0 - exF(t), \tag{19}$$

where e is the electron charge. The Schrödinger equation may be solved by expanding the atomic state in terms of eigenvectors of the unperturbed Hamiltonian H_0, that is,

$$\begin{aligned} i\hbar\frac{d}{dt} \mid \psi(t)\rangle &= (H_0 - exF(t)) \mid \psi(t)\rangle \\ \Rightarrow \quad & \sum_n i\hbar\dot{c}_n(t) \mid \psi_n\rangle \exp(-iE_n t/\hbar) \\ & = -\sum_n c_n(t)\, exF(t) \mid \psi_n\rangle \exp(-iE_n t/\hbar). \end{aligned}$$

The perturbation approximation consists of neglecting all c_n on the right side except c_0 which is put equal to unity. Then multiplication of both terms on the left by $\langle\psi_k \mid$ leads to

$$\dot{c}_k(t) = \frac{i}{\hbar} eF(t) \langle\psi_k \mid x \mid \psi_0\rangle \exp(i\omega_{0k} t), \qquad \omega_{0k} = \frac{E_0 - E_k}{\hbar}.$$

The integration is trivial and I omit it. The measurable quantity is the energy absorbed by the atom from the field in the time interval $(0,T)$, that may be got by subtracting the final atomic energy from the initial one, that is,

$$\begin{aligned}\Delta E &= \langle \psi(T) \mid H_0 \mid \psi(T)\rangle - \langle \psi_0 \mid H_0 \mid \psi_0\rangle \\ &= \sum_{kl} c_l(T)\, c_k^*(T)\, \langle \psi_k \mid H_0 \mid \psi_l\rangle - E_0 = \sum_{k\neq 0} |c_k(T)|^2 (E_k - E_0) \\ &= e^2\hbar^{-2} \sum_{k\neq 0} |\langle \psi_k \mid x \mid \psi_0\rangle|^2 \left| \int_0^T F(t) \exp(i\omega_{0k}t)\, dt \right|^2 \hbar\omega_{0k}.\end{aligned}$$

The result is quite plausible and easy to understand, namely there is absorption when the spectrum of the incoming electromagnetic wave possesses frequencies close to one of the characteristic frequencies of the atom, ω_{0k}, and the intensity absorbed depends on some properties of the atom (basically a matrix element). In most books the derivation goes further by taking the final result as a sum of products, each one corresponding to a 'photon energy' $\hbar\omega_{0k}$ times a 'transition probability' from the ground state ψ_0 to an excited state ψ_k. However, this gives rise to a lot of interpretation problems. For instance: Was the photon inside the wave before the absorption? Or, if the energy of one photon is greater than the energy carried by the wave during the time interval $(0,T)$, where did the photon energy come from? These problems disappear if we respect the rule stated in chapter 1 at the end of section 1.1.4, that is: We should try to interpret as physically real just the observed results, but neither the intermediate (mathematical) steps in the calculation nor additional non-testable results.

2.3.2. Entangled states. Entanglement is a quantum property of systems with several degrees of freedom that appears when the total state vector cannot be written as a product of vectors associated to one degree of freedom each. In formal terms a typical entangled state is the following

$$| \psi(1,2)\rangle = \sum_{m,n} c_{mn} \mid \psi_m(1)\rangle \mid \psi_n(2)\rangle, \tag{20}$$

where 1 and 2 correspond to two different degrees of freedom, usually belonging to different subsystems that may be placed far from each other. The essential condition is that the state eq.(20) cannot be written as a single product. That is, the sum cannot be reduced to just one term via a change of basis in the Hilbert space. As will

be discussed in section 2.4 below, entanglement appears as a specifically quantum form of correlation, which is claimed to be dramatically different from the correlations in classical physics.

The relevance of entanglement was stressed by Schrödinger [**7**] in 1935. He wrote that entanglement is not one but *the* characteristic trait of quantum mechanics. It is a fact that entanglement is extremely relevant for the interpretation of quantum mechanics. But, in the ensemble interpretation to be studied in section 2.4.5 below it poses no difficulty for a realistic interpretation of quantum mechanics, it is just a way to write some correlations that arise frequently in the microscopic world. However, if one assumes that quantum mechanics is complete — that is, a state vector like eq.(20) represents a pure state — then I believe that a realistic interpretation is impossible.

In fact, if we adhere to the completeness assumption then we are confronted with consequences in sharp contradiction with the intuition. One should believe that a state vector like eq.(20) represents complete information about the state of the system but incomplete information about each of the two subsystems. In fact, the state of the first subsystem should be obtained by taking the partial trace with respect to the second subsystem, leading to the following density matrix (assuming all state vectors normalized)

(21)
$$\rho(1) = \mathrm{Tr}_2\left[\mid \psi(1,2)\rangle\langle\psi(1,2) \mid\right] = \sum_{m,n} |c_{mn}|^2 \mid \psi_m(1)\rangle\langle\psi_m(1) \mid .$$

The density matrix represents a mixed state, where the information is incomplete, that is, we only know the probability, $P_m = \sum_n |c_{mn}|^2$, that the first subsystem is in state $\mid \psi_m(1)\rangle$. It is as if a student claims to know *completely* the subject matter of a given book, but she/he is admittedly *ignorant about every chapter*. This contradicts the usual meaning of complete information about the whole as being information about every part. This property of the entangled state may be stated more formally in terms of information entropy. The completeness assumption leads to the following von Neumann entropy, S, [**2**]

$$S = -\mathrm{Tr}\left[\rho \log \rho\right].$$

Hence the entropy of the state eq.(20) is 0 whilst the entropy of eq.(21) is

$$S_1 = -\sum_m P_m \log P_m > 0.$$

Another important property of entanglement is that it is a necessary and sufficient condition for the violation of the Bell inequalities, as discussed in more detail in chapter 3 section 3.2.5.

2.3.3. What is a quantum state? After presenting a few problems for the interpretation of quantum states we see that in order to understand quantum mechanics it is crucial to know what a quantum state really is. In classical physics the concept of state is clear, it rests on the concepts of isolation for either a particle or a wave or any combination of them. It is common view that neither the concept of particle nor the concept of wave may be transferred easily to quantum physics. Thus the standard answer to the question whether the electron is a particle or a wave is that it is *neither*. However, the answer involves a contradiction: anything is either localized (particle) or extended (wave), of course with respect so some reference size, say for an electron as compared with an atom. I believe that a more correct answer is that the electron is *both*. In fact, an electron cannot be seen as an isolated point particle. The physical electron corresponds to a cloud of electrons and positrons, electromagnetic radiation and other fields like neutrino-antineutrino pairs with a total charge e. The cloud has a size at least as large as the electron Compton wavelength, but possibly much larger.

This statement may be put in a different form as follows. The vacuum consists of a set of real fluctuating fields that are modified by the presence of the electron. In this book we claim that the fields are real, in contrast with the common opinion that they are virtual. I believe that *virtual* is a word without any clear meaning, used in order to avoid commitment with either the assertion that the fields are real or they are not. In summary, in contrast with the classical view, an electron cannot be seen as a particle whose state is defined by just its position and its velocity, which is the definition of pure state in classical mechanics.

Then we may ask what is the physical interpretation of state in a more complex system like an atom. It is not just a system of $Z+1$ point (or small) particles, the nucleus plus Z electrons. In the ground state of the atom the nucleus might perhaps be treated as a small particle localized in a region far smaller than the atom, but this is not the case for the electrons. What exists is a large number of electrons and positrons that are created (maybe with emission of radiation) or annihilated (with absorption of radiation) in pairs, with a conservation

of the total electric charge, that is Ze. Many other quantum fields are likely involved that correspond to a modification of the vacuum fields. In summary, I believe that the quantum state of any physical system is a quite complex structure consisting of many interacting fields evolving in time.

Sometimes it is argued that in a nonrelativistic treatment the possible creation or annihilation of electron-positron pairs should not be taken into account because the energies required are far larger than typical atomic energies. However, the argument is flawed. The total mass-energy of say two electrons plus a positron at extremely small distances may not be greater than the mass of a single electron due to a possibly strong electrostatic negative energy of interaction. In other words, the total mass-energy of the atom may be only slightly smaller than the mass of the nucleus plus the electrons, that mass defined when they are at a large distance from each other. In summary the internal structure of quantum systems like atoms should be always treated taking relativistic quantum fields of the vacuum into account. Of course, this is actually accepted when it is recognized that in renormalization calculations the bare mass or charge are quite different from the physical ones. The relevant fact is that the quantum formalism has the virtue that quite complex structures like atoms may be treated via simple equations like Schrödinger's. Of course, that equation is just a (fairly good) approximation, but there are small corrections that may be calculated using not so simple techniques like renormalization. My view is quite different from the standard one. It is usual to assume that quantum equations are exact when we ignore the interaction with the vacuum fields and corrections appear when the interaction is switched on. I believe that the interaction with the vacuum fields is precisely what makes quantum theory different from classical theories. I hope that the point will be more clear in chapter 5 where it is shown that, taking the vacuum electromagnetic field into account, classical electrodynamics predicts phenomena currently assumed as purely quantum.

2.3.4. Preparation and measurement. Many authors, in particular those supporting the Copenhagen interpretation (see section 2.4 below), propose to connect theory and experiments ascribing an operational meaning to the postulates. Thus states are devoid of an *ontological* status and are identified with *operational* 'preparation procedures' and observable properties with 'measurements'. These

identifications may be useful in practice, but I think that it is flawed to introduce physical operations, like 'preparation' or 'measurement', as a part of the postulates. In fact, it is not appropriate to establish a direct correspondence between 'preparation' and 'density matrix' or between 'measurement' and 'self-adjoint operator' (or more generally POVM, positive operator valued measure). Preparation is a rather complex set of physical manipulations, whence the density matrix appropriate for a microscopic system may be guessed, rather than derived, most times after a process of trial and error on the part of the experimental physicists performing the particular preparation. Similarly for measurement. Indeed, an empirical result is taken as relevant only after the experiment has been critically analyzed and repeated by different groups of researchers. This view agrees with Bell's, who wrote: "I am now convinced that the word 'measurement' has now been so abused that the field would be significantly advanced by banning its use altogether, in favor for example of the word 'experiment' "[**8**, page 166].

The existence of measurement postulates is an anomaly of the quantum formalism and it has been strongly criticized by philosophers of science like Karl Popper [**9**] or Mario Bunge [**10**]. In particular, the latter stresses that a physical theory should not include any *general theory of measurement*, but particular recipes or protocols for every specific measurement. This is most clear in chemistry. There are recipes for, say, the preparation of pure alcohol or the analysis of water from a river. However, it would be absurd to search for a 'general theory of preparation or analysis' in chemistry. In my opinion the same is true in physics, including quantum physics. The existence of a 'theory of measurement' is peculiar and it does not exist in any other theory in physics (or, more generally, in natural science). It is true that from a philosophical (epistemological) point of view any theory requires some assumptions for the connection with the results of observations or experiments. For instance, in classical mechanics we use the concepts of time, space, particle, isolated system, etc., and there are rules telling us how these concepts should be related to the (mathematical) formalism. However, it would be absurd to search for a 'general theory of preparation or measurement' in any branch of science, including quantum physics.

Thus I propose to remove the measurement postulates. In a realistic interpretation of the theory, they should be replaced by ontological statements about the properties possessed by systems, rather

than about the results of measurements. Measurement is a physical process that ends at the macroscopic level where the results are recorded, e.g. by a computer. Thus the measurement process should be studied using the (ontological) postulates of the theory. In the macroscopic domain no measurement postulates are needed, just the assumption that the measurement reveals properties of the system, and this should be the case for quantum theory. I believe that in the quantum domain the measurement postulates should be replaced by a weak form of Born's rule as stated in the following.

2.3.5. Proposed weak postulates. I propose to replace the standard postulates 1 and 3 in section 2 by the following one:

1'. To every physical system it is associated a Hilbert space, H, complex and separable. A state of the physical system at a time t is represented by a density operator, that is, an operator fulfilling the properties in eq.(15).

From a mathematical point of view it may be appealing to postulate also the reciprocal, namely that every density operator represents a possible state. Of course, if all density operators can be associated to possible states, it would be plausible to give a special status to the boundary of the set, that is the density operators of the form

$$\hat{\rho} = \mid \psi\rangle\langle\psi \mid,$$

$\mid \psi\rangle$ being a normalized vector. These (idempotent) density operators are associated to rays in the Hilbert space, representable by normalized vectors, and the associated states might be named 'pure states'. Therefore if one assumes that all rays may represent physical states it is natural to introduce Postulate 1 in terms of state vectors rather than density matrices and to reinforce the argument introducing the superposition principle, postulate 3, which nevertheless must be restricted with the superselection rules.

I propose that the postulates should be as weak as possible and therefore substitute Postulate 1' for the standard Postulate 1, also removing the superposition principle Postulate 3. It is interesting that the assumption that any system, except the whole universe, should be represented by a density matrix is unavoidable in the popular many-worlds interpretation of quantum mechanics (see section 4 below).

The proposed postulate $1'$ does not contradict the standard postulates 1 and 3, but it allows some freedom, which will be used in later chapters to remove obstacles for a realistic interpretation of some quantum phenomena. On the other hand the superposition principle,

when it is really required, is a consequence of Schrödinger equation (e.g. in the study of interference experiments) as said above. The replacement of postulates 1 and 3 by $1'$ eliminates the concept of pure states, represented by state vectors, as those with maximal information. The almost unavoidable identification, in the standard formulation, of maximal information with complete information has been the source of the most polemic, and dangerous in my opinion, statements in quantum mechanics, as I will discuss below. In particular it leads to a sharp distinction between the probabilities in classical physics (and other sciences) and the probabilities associated to quantum measurements. It also leads to the assumption that quantum mechanics is complete thus putting strong difficulties for a realistic interpretation.

If the states of systems are represented by density matrices, rather than state vectors, then the evolution (Schrödinger) equation should be more adequately formulated as the evolution of the density operator. That is, postulate 6 should be replaced by the following one.

6'. The density operator representative of the state of the system evolves according to the equation

$$\frac{d}{dt}\hat{\rho}(t) = -\frac{i}{\hbar}\left[\hat{H}, \hat{\rho}\,(t)\right],$$

where $\hat{H}$ is a self-adjoint operator called Hamiltonian of the system. It is easy to see that from postulate $6'$ it is possible to derive eq.(18) as well as from postulate 6.

Although I believe that postulates of measurement should not appear in the theory, a connection between the formalism and the experiments is needed. But I propose to replace postulate 4 by Born's rule eq.(17) as follows:

4' The probability distribution of positions of a particle in a state represented by the density operator $\hat{\rho}$, eq.(14) *is given, in terms of wave-functions, by*

$$\rho\,(\mathbf{x}) = \sum_j p_j \langle \mathbf{x} \mid \psi_j \rangle \langle \psi_j \mid \mathbf{x} \rangle = \sum_j p_j \mid \psi_j\,(\mathbf{x}) \mid^2, \qquad p_j > 0.$$

This postulate is an ontological statement ("the particle position is...") and not a measurement statement as is frequently stated it textbooks ("if we perform a position measurement the result will be..."). I assume (tentatively) that this postulate is sufficient because at the end all measurements reduce to position measurements (e.g. of a pointer).

In summary, I propose to remove the postulates 3 and 5, and substitute the new postulates 1', 4' and 6' for the standard ones 1, 4

and 6, maintaining unchanged the remaining postulates, that is, 2, 7 and 9. 'Postulate' 8 is not really a postulate but a rule to get (or guess) quantum equations by analogy with classical equations. This set of postulates may be named *restricted quantum formalism*, as opposed to the standard one. An attempt will be made below to get a realistic interpretation of this restricted formalism.

2.4. Critical comments on current interpretations

As said above, Heisenberg's quantum mechanics was proposed as an abstract formalism without any physical picture behind. Bohr justified the absence of a model and, on this basis, he elaborated the 'Copenhagen interpretation' [**11**]. Several modifications or novel interpretations have appeared later. Relevant papers up to 1983 are reprinted in a book by Wheeler and Zurek [**12**]. I shall comment briefly on the most popular interpretations in the following. I stress that these interpretations refer to the standard postulates as stated in section 2.1.1. Our interpretation of the postulates modified as in section 2.1.3 will be considered later.

For the sake of clarity I will illustrate my comments with the celebrated 'Schrödinger cat' *gedanken* (imaginary) experiment [**7**]. It is a relevant example because it involves the phenomenon of entanglement, crucial for any attempt of interpretation of quantum mechanics. The fanciful experiment consists of a box containing a radioactive atom and a cat together with a device that kills the cat, say instantaneously, when the atom decays. I will assume that both the atom in the excited state and the live cat are put inside the box at time t_1. The question is what may be said about the atom and the cat at times $t > t_1$. In particular, what is the prediction of quantum mechanics for the states of both the cat and the atom when the box is open at time t_2. Any person with knowledge of the law of radioactive decay, but ignorant of quantum mechanics, would claim that the probability of being both the cat alive and the atom excited at time $t \in [t_1, t_2]$ is

$$P(t) = \exp\left[-\lambda\,(t - t_1)\right], \tag{22}$$

λ^{-1} being the mean lifetime of the atom. In particular, the probability at the moment of opening the box will be $P(t_2)$, eq.(22). This may be named the response of a 'naive realist'. In contrast, the answer of an educated quantum physicist will depend on the interpretation that she/he supports, as I discuss in the following.

2.4.1. Copenhagen (Bohr) interpretation. According to the Copenhagen interpretation (CI) the referent of quantum mechanics is not the material world but the experiments. That is, the theory deals with the relations between the world and the observers. As Bohr put it "the finite magnitude of the quantum of action prevents altogether a sharp distinction being made between a phenomenon and the agency by which it is observed" [**11**]. Thus CI is close to the pragmatic approach as described in chapter 1. According to this approach we should not make assertions about the bodies, but about the results of possible observations or measurements. Thus a sentence like "the probability that the atom *is* in the excited state at time t" is considered meaningless. A meaningful assertion should be something like "if we perform a measurement of the state of the atom at time t, the probability that we get the result 'excited' is given by eq.(22)". The approach is instrumentalist and it might be called a 'protocol for the use of the quantum formalism' rather than an interpretation.

Bohr elaborated a philosophical background with the introduction of the 'complementarity principle' and the 'correspondence principle', in order to solve two theoretical difficulties of the formalism. Firstly it is unsatisfactory that quantum mechanics applies only to the microscopic world whilst the macroscopic one is governed by classical theories. Bohr's solution to this problem was to assume that there is a smooth transition, quantum laws approaching the classical ones in the limit when Planck constant becomes negligible, formally when $\hbar \to 0$. This is the essential content of the *correspondence principle,* that Bohr also applied to several instances deriving some relevant results. The second theoretical problem was the existence of apparent contradictions, in particular those derived from the fact that quantum entities behave sometimes like particles and other times like waves. In order to solve that problem Bohr proposed the *complementarity principle*, which stresses the incompatibility between causal laws and spacetime description, due to the finite (nonzero) value of the quantum of action. After that he showed that there is no contradiction in practice because the behavior of the quantum entities does not derive from the microscopic system alone, but it also depends on the full context [**11**], including macroscopic measuring devices.

The rules of CI for the use of quantum mechanics are to some extent independent on the two mentioned Bohr's principles, and I will comment only on the rules. CI assumes (or at least it does not

reject the assumption) that macroscopic bodies have objective properties (that is, independent of any measurement) and their evolution is governed by the laws of classical physics. Thus it is meaningful to ask whether a cat is either alive or dead at any time. A more difficult question is whether we are allowed to assign a probability to each of these possibilities. The application of quantum mechanics, with the CI rules, to the 'cat experiment' is that for $t \in (t_1, t_2)$ the atomic state should be represented by the state vector

$$(23) \qquad | \psi(t)\rangle = c_g(t) \mid g\rangle + c_e(t) \mid e\rangle, \quad \text{with}$$
$$c_e(t) = \sqrt{exp(-\lambda(t-t_0))}, \qquad c_g(t) = \sqrt{1 - exp(-\lambda(t-t_0))},$$

where $| g\rangle$ ($| e\rangle$) is the state vector of the atom in the ground (excited) state. Now there are two possibilities, depending on what is supposed to be a measurement:

1) If we assume that the actual measurement takes place when the box is open, then quantum mechanics says nothing about the atom and the cat for times $t \in (t_1, t_2)$. At time t_2 it predicts that the probability of both the cat being alive and the atom being excited is given by the squared modulus of the amplitude $c_g(t)$, eq.(23), which precisely agrees with the naive prediction $P(t_2)$, eq.(22).

2) We might assume that the cat, being a macroscopic system, can act as a measuring device. In this case, CI tells us that, for any time $t \in (t_1, t_2)$, the probability of both the cat being alive and the atom excited is eq.(22). This interpretation (the cat as a measuring device) is consistent with the fact that, if the cat is found dead at time t_2, a careful study of the corpse (involving macroscopic manipulations) might determine the time of death, say t_d. This would allow reconstructing the whole history: The cat was alive and the atom excited until t_d. We must assume that, if a similar experiment is performed many times, the distribution of times t_d would converge to an agreement with the probability eq.(22).

Bohr's approach to the problem of the 'state vector (or wave function) collapse' is interesting. This is the discontinuous change of the state vector when a measurement is made, e.g. a change from eq.(23) to $| \psi\rangle = | g\rangle$ at the time of opening the box. In our example we may naively believe that the collapse is just a change of our information as a result of the observation. However, Bohr strongly opposed to the belief that the *wave function just represents our information about the system*, with the implicit consequence that this information may

be incomplete. See section 2.4.5 for a more detailed discussion of the completeness question.

2.4.2. John von Neumann. CI is very good from the practical point of view and avoids any bizarre assumption (which is not the case in more elaborated interpretations discussed below). The problem with the CI is that it creates what has been called an 'infamous boundary' [**13**], that is, a discontinuity between micro and macrophysics. The former should be studied within quantum mechanics, the latter using classical physics. In order to remove the boundary and get an interpretation where quantum mechanics is valid also for macroscopic systems, John von Neumann [**2**] introduced a theory of measurement and he even gave a model for it. His approach has been also currently named Copenhagen interpretation, but this is misleading because von Neumann's interpretation is different from Bohr's, as a matter of principle. However, it was an elaboration of the Copenhagen interpretation rather than an alternative, which may justify the name. For short, I shall label it MCI with M standing for modified or measurement. MCI has been supported in most papers and books of quantum mechanics until around 1980.

The modification introduced by von Neumann with respect to Bohr was to take seriously the assumption that quantum mechanics is the universal theory and that classical theories are just approximations. Thus he proposed studying the measurement within quantum mechanics and made a model involving the coupling of the microscopic system with the measuring apparatus. However, this gave rise to a number of difficulties that will be discussed below, but previously we clarify the matter studying the application of von Neumann's ideas to the cat experiment.

In MCI both the cat and the atom should be treated as quantum objects. Therefore eq.(23) is no longer appropriate and we should represent the state of the whole system, atom plus cat, by

$$(24) \qquad |\,\psi\,(t)\rangle = c_g(t)\,|\,g\rangle\,|\,deadcat\rangle + c_e(t)\,|\,e\rangle\,|\,livecat\rangle.$$

Of course, one may point out that a dead cat does not correspond to a pure state to be represented by the vector $|\,deadcat\rangle$. Indeed, there may be very many quantum states corresponding to a dead cat and similarly for a live cat. However, this is not a real problem because MCI assumes that any physical system is associated to a well defined state vector. (When the appropriate state vector is not known we should use a probability distribution over those vectors, which may

be formalized via a density matrix. But for ease of understanding let us use a single state vector as in eq.(24)).

Eq.(24) represents a typical entangled state. If CI had been modified with the assumption that state vectors actually represent statistical ensembles, this would have lead to the ensemble interpretation, to be discussed below. However, the mainstream of the scientific community rejected it and supported the 'completeness' of quantum mechanics, in the sense that *the state vector represents the actual state of an individual physical system, as opposed to a statistical ensemble.* With this assumption the MCI leads to bizarre consequences, which was the point that Schrödinger [**7**] attempted to stress with his cat example. Indeed, for many people it is impossible to understand the meaning of a state represented by a superposition of alive and dead cat. Is it something intermediate between life and death?

The problem is not only the highly counterintuitive character of superpositions of macroscopic systems, it is the disagreement with empirical evidence. Indeed, those macroscopic superpositions cannot be manufactured in practice (there is a lot of literature about the actual preparation of 'Schrödinger cats', but they always involve mesoscopic rather than truly macroscopic systems). Thus it seems that the quantum evolution (the Schrödinger equation) is violated at the macroscopic level. This has been called the problem of the *objectification or individuation* [**14**], [**15**]. That is, the fact that *a particular value is obtained in the measurement amongst several possible values,* something not predicted by the quantum formalism except if an explicit postulate is included (postulate 5 of section 2.2.1 above). This postulate forces us to change the state vector at the time of measurement, a change usually called the state vector, or wave function, 'collapse'. That change is not predicted by the Schrödinger equation and in fact it precludes the validity of that equation during measurements. In the CI the collapse was just a change of the mathematical representation needed for the analysis of the experiment. In MCI, however, it becomes a real physical change because it is assumed that the state vector corresponds to an individual system (rather than a statistical ensemble). In the next subsection I discuss possible solutions that have been proposed.

A problem related to objectification is the existence of quantum jumps [**16**], the typical example being the decay of a radioactive atom. For instance, an atom of uranium 238 may remain as such during a

million years but, at some unexpected time, it decays with the emission of an alpha particle (a nucleus of helium 4). The sudden decay (within a small fraction of a second) apparently contradicts the Schrödinger equation. People like to say that the observation of the (spontaneous) decay is a particular case of measurement and therefore the problem of quantum jumps becomes an example of objectification after a measurement. In my opinion, however, there is a clear difference between objectification and quantum jump. The former is more properly the 'disentanglement' of an entangled state involving macroscopic bodies. For instance, the fact that we see the cat either alive or dead at the time of opening the box in the Schrödinger cat example. In contrast, a quantum jump refers to a discontinuous change of a microscopic system. In any case the difficulties with both objectification and jump may be solved simultaneously, for instance in either hidden variables or collapse theories. The latter is discussed in the next subsection and the former later on.

2.4.3. The objectification and the quantum jumps problems. There are a variety of proposed solutions to the objectification problem. In the original (Bohr) Copenhagen interpretation there is no real problem: the Schrödinger equation is just a mathematical tool able to relate the *preparation* of a (microscopic) system to *measurements* made on it. The wave function (or state vector) is just a convenient form of dealing with the probabilities involved. That is, a preparation gives rise, after some time, to a set of probabilities when the system is placed in an appropriate experimental context. The objectification is a change due to the measurement. But the change must be *postulated* because the interaction between the microscopic system and the macroscopic apparatus can be described neither by quantum nor by classical theories in CI. The objectification problem does not exist either in the 'many worlds interpretation' (MWI) that will be discussed in the next subsection.

From the time of von Neumann's book [**2**] (1932) until around 1980's, and for a fraction of the scientific community until today, the MCI has been the most popular interpretation. For this reason a very large number of papers and books have been devoted to propose possible solutions to the objectification problem.

John von Neumann pointed out that a measurement only finishes when a (human) observer is conscious of the result of the experiment. This would solve the objectification problem if we assume that the

mind is not governed by quantum mechanics. This proposition is also supported by London and Bauer [**17**] and described by Wigner [**18**]. In the cat experiment, this seems to imply that the cat really dies when we look at the box after opening it, or even when we are informed by another person of the result of the experiment (which leads to the 'Wigner's friend' paradox). Many people dislike this solution.

In practice, many authors (maybe not too fond of the subtleties of foundational questions) accepted a kind of peaceful coexistence of the two (contradictory) postulates: the Schrödinger equation and the quantum theory of measurement. However, the mere existence of a measurement theory is an anomaly, as discussed above.

Collapse theories. A proposal that became popular around 1985 is to modify the Schrödinger equation in such a way that the change fulfil two consistency requirements: 1) For microscopic systems it produces an extremely weak, practically undetectable, modification in the evolution of the wave function, and 2) For macroscopic systems it gives rise to a rapid disentanglement, that is, an evolution from any superposition to a single term. There have been several explicit models of this type, called 'collapse theories'. The most satisfactory one has been proposed by Ghirardi, Rimini, and Weber, and is usually called the GRW theory [**19**], [**20**]. It involves phenomenological parameters that, if the theory is taken seriously, acquire the status of new constants of nature. There have been also attempts at deriving the parameters from fundamental arguments, like the action of gravitational forces (effects of general relativity).

In spite of their phenomenological character, the collapse theories have relevance since they have made clear that there are new ways to overcome the difficulties of the quantum formalism. Moreover, they have allowed a clear identification of the formal features which should characterize any unified theory of micro and macro processes. Last but not least, collapse theories qualify themselves as rival theories of quantum mechanics and one can identify some of their physical implications which, in principle, would allow crucial tests discriminating between the two. I shall not review here the collapse theory, which would lead far from the main purpose of this book. The interested reader may look at a review by Ghirardi [**21**].

2.4.4. Many-worlds. The many worlds interpretation (MWI) offers a radical solution to the objectification problem; it assumes that objectification never takes place. That is, the evolution of an

isolated system is always governed by the Schrödinger equation. Now, no system involving a macroscopic body may be completely isolated, so in the study of its evolution we should consider the wave vector of the whole universe. In particular, in the cat experiment we should include, in addition to the atom and the cat, also the box, the human observer and everything else. Thus eq.(24) should be replaced by

$$\begin{aligned} \mid \psi\,(t)\rangle = c_g(t) \mid g\rangle \mid deadcat\rangle \mid world_g\rangle \\ + c_e(t) \mid e\rangle \mid livecat\rangle \mid world_e\rangle, \end{aligned} \tag{25}$$

where $\mid world_g\rangle$ represents the rest of the world associated to the atom in the ground state and the cat dead, and similarly for $\mid world_e\rangle$. Eq.(25) seems to say that there are two copies of the human observer and of the whole world. In the last copy the observer sees the cat alive and the atom excited, in the first she/he sees the cat dead and the atom in the ground state. The state vector of the universe is a linear combination of these copies.

MWI is the unavoidable end of the logical path if we believe that quantum mechanics (as defined by the standard postulates excluded those of measurement) is universally valid. It was initially proposed by Everett [**22**] with the name 'relative states interpretation' and elaborated later by de Witt [**23**], who introduced the name 'many worlds'. The aims of MWI are: 1) to retain the unrestricted validity of the quantum formalism, 2) to remove the need for the state vector collapse, 3) to remove the need for an external observer, and 4) derive the Born rule [**24**]. The latter is the rule for finding the probabilities of the different possible outcomes as a result of a measurement.

Apart from the difficulty of understanding the real meaning of 'multiplicity of worlds', the main problem of MWI is to reproduce the Born rule without introducing any explicit probabilistic postulate (like our postulate 4' of the previous section). The standard approach for that is the theory of decoherence [**25**], [**26**]. Decoherence is the evolution predicted by quantum mechanics when a system possesses very many degrees of freedom, as is the case for a measuring device in contact with the environment. It involves a loss of coherence which leads from a state vector (representing a pure quantum state) to a density matrix, as a result of the interaction with the environment. That density matrix is approximately diagonal in an appropriate basis (the preferred basis), so that it looks like a probability distribution defined on a set of pure states, as is exhibited in the example of eq.(26) below. In the context of MWI the density matrix may be

seen as coming from taking the partial trace over those degrees of freedom that are not of interest. For instance, if we take the partial trace, with respect to the world states, of the (idempotent) density matrix associated to the state vector eq.(25) we get with very good approximation

$$\mathrm{Tr}_{world} \mid \psi\rangle\langle\psi \mid \simeq |c_g(t)|^2 \mid gd\rangle\langle gd \mid + |c_e(t)|^2 \mid el\rangle\langle el \mid, \tag{26}$$

where $\mid gd\rangle$ is short for $\mid g\rangle \mid deadcat\rangle$ and $\mid el\rangle$ for $\mid e\rangle \mid livecat\rangle$ and the orthogonality of the state vectors $\mid world_g\rangle$ and $\mid world_e\rangle$ has been taken into account. Eq.(26) is mathematically identical to the representation of the quantum state of the atom plus the cat that we should use when we do not know its actual state, and consequently we attribute the probability $|c_g(t)|^2$ to the atom being in the ground state and the cat dead, and $|c_e(t)|^2$ the probability of the alternative. The question, to be discussed below, is whether eq.(26) actually corresponds to a mixture or not.

Decoherence theory is actually more involved than it may appear from our example. Firstly, we should consider very many terms in the sum which represents the quantum state of the world, rather than only two as in our simplified example eq.(25). There is also an ambiguity in the world state vector because, it being a linear combination of (tensor) products of state vectors, it could be written in many different forms depending on the choice of basis in the Hilbert space. This leads to the problem of the preferred basis, whose solution is one of the achievements of decoherence theory. I shall not discuss here the different approaches and the technical issues of the MWI and decoherence in more detail, and refer to the vast literature on the subject (see, e.g., [**24**] and references therein). Related to decoherence is the 'consistent histories' approach [**27**], which will not be discussed here.

MWI has the virtue that it makes quantum mechanics a selfconsistent theory resting upon a simple hypothesis: its universal validity. In this respect it is superior to the old-fashioned CI and MCI. However, as usually understood it leads to a rather bizarre picture of the world. For the sake of clarity I will consider a measurement with reference to eq.(26), although now 'cat' means the macroscopic measuring device able to suffer an irreversible evolution. Once MWI plus decoherence theory leads to a reduced density matrix like eq.(26), it seems plausible to interpret it as representing a mixture, $|c_g(t)|^2$ and $|c_e(t)|^2$ being probabilities in the usual sense of mathematical measures of

information. However, this interpretation is not compatible with the assumption that quantum mechanics is complete. That is, the hypothesis that eq.(26) represents a mixture is not compatible with the assumption that the state vector of the universe corresponds to an individual world (although with many branches), rather than a statistical ensemble of possible worlds. However, it is irrelevant in practice whether we assume that the world state vector represents complete or incomplete information. In fact, a *detailed knowledge* of that state vector would always lie beyond the human capabilities. Therefore the assumption that eq.(26) represents an actual mixture, and quantum mechanics is incomplete, is in my opinion the most plausible.

In contrast, the conjunction of assuming universal validity (i.e. MWI) and completeness of quantum mechanics leads to the extravagant view that there are many parallel worlds [**24**]. I think that this belief is unnecessary. Actually, the view rests on what has been termed *a Platonic paradigm* by M. Tegmark [**28**], who defines it as follows: "The outside view (the mathematical structure) is physically real, and the inside view and all the human language we use to describe it is purely a useful approximation for describing our subjective perceptions." The mathematical structure referred to by Tegmark is the formalism of quantum mechanics. Thus the Platonic paradigm is equivalent to assuming that standard quantum mechanics is the absolute truth and everything else are shadows.

In my opinion scientific theories, quantum mechanics in particular, are something more modest. They are attempts at describing, rather imperfectly, "the objective reality, which is independent of any theory" [**29**]. In consequence I prefer to retain as much as possible of the MWI, logically superior to CI or MCI, but without adhering to the Platonic paradigm. The choice is obvious to me: we should reject the completeness of quantum mechanics, that rejection leading to the ensemble interpretation to be discussed in the next subsection. (But most people assume that MWI is not compatible with an ensemble interpretation. Even if it is compatible the relation is not trivial and will not be discussed here).

We have seen that in CI and MCI, discussed above, it is necessary to introduce a probabilistic postulate, which is substituted for Schrödinger evolution equation during measurement. That postulate (Born's rule) allows calculating the probabilities of the different possible outcomes of a measurement and the corresponding state vector

after the collapse. In MWI it is controversial whether a probabilistic postulate is introduced. Many authors consider that this is not the case, that the quantum probabilities may be got from the formalism. Actually, Everett introduced in his original formulation a measure given by the squares of the amplitudes in the sum (of normalized state vectors) which the world state vector consists of. In our example, eq.(25), that measure may allow to assume that $|c_g(t)|^2$ and $|c_e(t)|^2$ are probabilities. Therefore it is my opinion that *MWI does introduce a probabilistic postulate*, even if it is most natural, as Everett emphasized [**22**].

An interesting consequence of the MWI is that a state vector is only appropriate for the whole world. In contrast, the states of the systems which we may actually study (subsystems of the universe) should be represented by density matrices. This leads to the *conjecture that only a subset of the whole set of possible density matrices represents physical states*. This conjecture strongly limits the validity of the superposition principle and gives rise to the problem of determining what is the subset of the whole set of density matrices which correspond to physical (realizable) states.

2.4.5. Ensemble interpretation. Is quantum theory complete? The Copenhagen, von Neumann and many worlds interpretations have in common the assumption that the description offered by quantum mechanics is complete. They may be grouped within the class of 'orthodox' interpretations. An alternative to completeness is the assumption that the wave function just represents our knowledge about the actual state of a system. In other words, that the state vector (or the wave-function) represents a statistical ensemble of systems rather than a single system. This hypothesis has been named the 'ensemble interpretation' and it was supported by Einstein [**30**], and by a few authors in more recent times [**31**].

The dichotomy between completeness and incompleteness of quantum mechanics or, in modern language, between epistemological and ontological treatment of the wave-function, has been the subject of a controversy lasting for the whole existence of quantum mechanics. As is well known, in the early period the most famous debate took place between Bohr and Einstein (see references in chapter 1). Maybe the most important contribution to the debate was the 1935 article by Einstein, Podolsky and Rosen (EPR) [**29**]. Due to its relevance

for the (ensemble) interpretation of quantum mechanics supported in this book, I discuss it in some detail.

The article introduces, or clarifies, four relevant concepts:

(1) It stresses the need of realism in physical theories as we discussed in chapter 1 section 1.1.3.
(2) It provides an example of entanglement, a crucial concept in quantum theory that we have discussed in section 2.3.2 and will be studied again in section 2.5 and in chapters 3 and 6.
(3) It stated locality as a necessary property of any theory, as we will discuss in chapter 3.
(4) Finally it shows that either locality or completeness of quantum mechanics does not hold true, this being the declared purpose of the article.

The authors considered a system consisting of two particles placed at a distance in a quantum state such that the particles are correlated in both position and momentum. The state is possible according to the standard quantum formalism and the authors in fact wrote explicitly the wave-function of the composite system. According to the Heisenberg uncertainty principle it is not possible to determine simultaneously the position and the momentum of one of the particles, but nothing forbids measuring only one of the two observables with good accuracy. Due to the correlation, if the position of one particle, say number 1, is measured then we will know the position of particle number 2 without interacting with it in any way. Thus, after the measurement, we may attribute to the second particle a wave-function representing a state with definite position (but indefinite momentum). On the other hand a measurement of the momentum on the first particle allows attaching to the second particle a wave-function corresponding to a definite momentum (but indefinite position). The point of the argument is that the actual state of the second particle should be the same in both cases, because the measurement performed on its partner has perturbed nothing in its state. Therefore we can describe that state by means of two different wave functions. Hence the authors concluded that the wave-function just specifies our knowledge about the particle. In other words the wave-function should be treated as epistemological rather than ontological.

Crucial for the EPR [**29**] argument is the assumption that no influence could affect a particle due to a measurement performed on another distant particle, a hypothesis known as 'locality' (see chapter 3 section 3.3.2 for the definition of locality in the sense of relativity

theory). Bohr rebutted [**32**] the EPR argument claiming that in quantum mechanics there is a kind of wholeness that makes the assumed locality not to hold true. The common wisdom, resting upon Bell's theorem to be discussed in chapter 3, is that Bohr was right and EPR were wrong, but I do not agree.

In addition to nonlocality, a bizarre consequence of assuming an ontological status for the wave-function is exhibited in the EPR example. According to the orthodox interpretation of the quantum formalism, the wave function of the two-particle system represents a pure state but the state of each particle is not pure. In fact, the state of the two-particle system is represented by a wave-function, which in standard quantum mechanics means that our knowledge of the two-particle state is complete. However, the state of one of the particles cannot be represented by a wave-function, but by a density operator obtained by taking the partial trace of the density operator associated with the two-particle wave-function. That density operator represents a statistical mixture, meaning incomplete knowledge. (I must point out that this fact does not contradict the representation by a wave function made by EPR as discussed above. In the EPR argument the quantum state attributed to particle 2 *follows* from a measurement performed on particle 1, but now we are considering the state *before* any measurement is made). The conclusion is that we have complete information about a composite system, but incomplete about each part, contrary to the usual definition of 'complete'. This contradiction will be discussed more formally in section 2.3.3. In my opinion this behaviour of entangled quantum systems is another argument for the epistemological character of the wave-function. If that character is assumed, our knowledge will be incomplete for both the composite system and each one of its parts, whence no paradox would arise.

In spite of the above arguments, during the whole history of quantum mechanics the 'orthodox view' has been that the theory is *complete*, as stressed by Bohr and his followers. However, Bohr's completeness may be seen as a support to the 'instrumentalistic approach' rather than a statement about the relation between the wave-function and reality. In contrast, for Einstein the relevant question was whether "the ψ-function corresponds to a single system or to a (statistical) ensemble of systems" [**30**] (Einstein carefully avoided the name wave function — in order not to commit himself to the existence of waves associated to particles — and used the name ψ-function instead). He

clearly supported the second assumption, which may be stated saying that he adhered to the interpretation of the wave function as information. This interpretation has been vindicated by more recent authors, for instance Chris Fuchs who has written "quantum states are states of information, knowledge, belief, pragmatic gambling commitments, not states of nature." [**33**]. See also Englert and references therein [**34**].

At this moment it is appropriate to emphasize that an epistemological interpretation of the wave function does not imply for it a purely subjective character. In many cases the available information is such that everybody would attribute the same wave-function to the physical system, whence it acquires some objective character. A related question is whether the wave-function collapse after a measurement is a physical change or just a change in our information (see section 2.3.3). In my opinion, it is wrong to adhere exclusively to one of the possibilities. Actually both, or a combination of both, may appear in measurements. The EPR argument provides an example of a pure change of information, but an atom crossing a Stern-Gerlach apparatus may suffer an actual physical change in the direction of the spin. In any case the quantum postulate that "in a measurement the state of the system goes to an eigenstate of the measured observable" may be appraised as an elegant formal statement, but in actual experiments things are more involved.

The assumption that the quantum wave-function (or state vector) represents information is, in some respect, a vindication of some of Einstein's views on the interpretation of quantum mechanics. This vindication does not refer to his opinions (it is generally assumed that he was wrong in his beliefs on locality, allegedly refuted by Bell's theorem, see below). He is rather vindicated as having pointed out what are the relevant questions to be answered. In fact, a close scrutiny of Einstein's letters to different authors shows that his main interest was not the question whether the wave-function ψ represents an ensemble of possible systems — or, what is almost equivalent, if it only represents our information — but on whether a given real (ontic) state may correspond to different quantum-mechanical states ψ. Einstein clarified the point in a letter to Schrödinger [**35**], [**36**]. An extended discussion about Einstein's opinions, with many references, appears in a paper by Harrigan and Spekkens [**37**].

2.4.6. Hidden variables theories. The ensemble interpretation poses a question: What is the ensemble of the real (ontic) states corresponding to a given quantum state (a wave-function), and what is the probability distribution on the ensemble? That is, we should assume that there is an unknown subquantum theory such that quantum mechanics is derived from it in the same way that we study classical statistical mechanics of particles. The statistical approach is usually, but not necessarily, due to the fact that the number of particles is very large. In classical statistical mechanics the essential reason is the lack of sufficient information. The search for such a subquantum theory is known as the *hidden variables programmme*. The reason for the name is that the said subquantum theory would involve variables not (yet) known and therefore hidden. In my view the ensemble interpretation and hidden variables (HV) are the natural approach to search for a realistic understanding of quantum physics. However, many supporters of the incompleteness of the quantum description do not propose a search for specific HV models. For example, Einstein did not support explicitly the search for hidden variables, rather he advocated for a new theory from which quantum mechanics could be derived.

The question of hidden variables arose soon after the formulation of quantum mechanics, during the years 1925–26. It was explicitly mentioned in the book by von Neumann in 1932 [**2**], where he derived a celebrated 'no hidden variables' theorem. From that time many books and articles have been devoted to the subject. Nevertheless, there is no sharp and widely accepted definition of what a hidden variables (HV) theory is. I propose the following:

Definition 1. *HV is a theory physically equivalent to quantum mechanics (that is, giving the same predictions for all experiments) which has the formal structure of classical statistical mechanics.*

The definition may be illustrated in the following table giving the correspondence of concepts in experiments, in standard quantum theory and in a possible HV theory:

Table I. Correspondence of concepts

EMPIRICAL	QUANTUM THEORY	HV THEORY
physical system	Hilbert space H	phase space Λ
state	vector $\mid \Psi\rangle \in H$	probability density $\rho(\lambda)$
observable A	self-adjoint operator $\widehat{A}$	function $A(\lambda)$
expectation values	$\langle \Psi \mid \widehat{A}^n \mid \Psi\rangle$	$= \int [A(\lambda)]^n \rho(\lambda)\, d\lambda$
correlations	if $\widehat{A}\widehat{B}=\widehat{B}\widehat{A}$, $\langle \Psi \mid \widehat{A}^k \widehat{B}^l \mid \Psi\rangle$	$= \int [A(\lambda)]^k [B(\lambda)]^l \rho(\lambda)\, d\lambda$

For the sake of clarity, in the table we have considered only quantum pure states. The most general states are associated to density operators, $\hat{\rho}$, whence the quantum expectation value and correlation should be written, respectively

$$\mathrm{Tr}\left(\hat{\rho}\widehat{A}^n\right), \mathrm{Tr}\left(\hat{\rho}\hat{A}^k\hat{B}^l\right).$$

The parameter (or parameters) λ is usually called the *hidden variable*. Two observables, A and B, which are associated to commuting operators, $\widehat{A}$ and $\widehat{B}$, are said *compatible*. The correspondence might be extended to more than two compatible observables, associated to commuting operators. The knowledge of all moments of the observables determines the joint probability distribution. In fact, in order to prove the equality of the characteristic functions of the probability distribution in quantum mechanics and in the hidden variables theory it is enough to substitute $\exp\left(i\xi\widehat{A}\right)$ for $\widehat{A}^n$ and $\exp\left[i\xi A\left(\lambda\right)\right]$ for $\left[A\left(\lambda\right)\right]^n$ in the equality. Hence, it is a simple matter to get the probability distribution. Similarly, for several commuting observables the joint probability distribution is given in quantum mechanics by

$$f\left(a,b\right)=\frac{1}{4\pi^2}\int d\zeta\int d\chi\exp\left(-i\zeta a-i\chi b\right)\mathrm{Tr}\left[\exp\left(i\zeta\hat{A}+i\chi\hat{B}\right)\hat{\rho}\right].$$

This may be trivially extended to any finite set of commuting onservables.

However, as is well known, quantum mechanics does not provide joint distributions of observables that are not compatible (the associated operators not commuting). Therefore it is not obvious that hidden variables models might be generalized to incompatible observables. Actually, there is no good proposal of a joint probability distribution for sets of not-all-commuting observables; in particular, for the position and momentum of a particle, as will be studied in chapter 3.

If quantum mechanics admitted hidden variables for all sets of observables, then a realistic interpretation would be trivial. In fact, it would be compatible with the existence of real (ontic) pure states where all relevant observables would possess a definite value. Mixed states, corresponding to quantum states in the ensemble interpretation, would possess a joint probability distribution of all observables. However, this is not possible because hidden variables theories should fulfil some constraints as studied in more detail in chapter 3.

2.4.7. Conclusions. In comparison with the Copenhagen and the von Neumann interpretations, the many worlds (MWI) has the advantage that it does not require the measurement postulates. It follows rigorously from the universal validity of quantum mechanics. However, in order to avoid a Platonic paradigm (see section 2.3.4), strange to natural science, it should be combined with (or replaced by) an ensemble interpretation, thus giving rise to an interpretation which may be realistic and not bizarre.

None of these interpretations leads to a realistic model of the quantum world. The most obvious alternative is a hidden variables theory. However, in this book we go beyond a simple change in the interpretation of the Hilbert space formalism resting on the standard postulates revisited in section 2.2.1. I believe that we should not be committed with the interpretation of the standard formalism; the relevant question is the interpretation of the empirical facts, may be via a completely different formalism. I believe that this was also the opinion of Einstein, who never supported explicitly hidden variable theories (within the current formalism) but an alternative (say subquantum) theory from which the present QM should be derived in a manner similar to the passage from statistical mechanics to thermodynamics. An intermediate possibility is a modification — or rather weakening — of the standard postulates that would allow a realistic interpretation, as suggested in section 2.3.5.

The hidden variables approach is closely related to the questions of completeness of the quantum formalism, locality, entanglement and the Bell inequalities, that require an extended treatment. Therefore we will devote the whole chapter 3 to that subject.

2.5. The logical structure of quantum theory

Quantum mechanics looks radically different from all classical theories of physics. Is that difference fundamental or is it due to our present lack of understanding of quantum mechanics? In this chapter I revisit several approaches that might suggest that the difference is indeed fundamental, although that belief will be questioned in later chapters of the book. Here I also introduce the Bell inequalities as constraints for hidden variables theories, which will play a relevant role in following chapters.

2.5.1. Quantum logic and quantum probability. According to Birkhoff and von Neumann [**38**] the difference between quantum

and classical theories is radical because it appears at the most fundamental level, the logic. The elements of a logic are the propositions which, using the language of physics, are observables having the possible values 1 (the proposition is true) or 0 (false). Some pairs of propositions are related by the *implication (A implies B)* if B is true whenever A is true. This binary relation endows the logic with the mathematical structure of a partially ordered set ('poset'). Another relation associates every proposition with its negation (for each proposition A there exist another one, A', which is true if and only if the first is false). This makes the poset *orthocomplemented.* The internal operations 'meet' and 'join' endow the poset with a richer structure making it an orthocomplemented *lattice.* Finally it is assumed that there exists the sure proposition I, always true, and the absurd proposition Φ, always false, which makes the lattice *complete.* From now on any complete and orthocomplemented lattice will be called a *logic.* Classical logic is a *distributive* lattice and it is called a *Boolean algebra.*

In the view of Birkhoff and von Neumann the structure of quantum logic may be derived from the correspondence between propositions and *projection operators* (which we shall call projectors in the following). Accordingly, these authors postulated that the proposition associated to the projector $\widehat{P}$ is true (or false), for a physical system in a given state, if the state vector $\mid \Psi\rangle$ is an eigenvector of $\widehat{P}$ (or $\widehat{I}-\widehat{P}$). This assumption gives rise to a *trivalent* logic where propositions may be, in addition to true or false, also undefined (which happens if $\mid \Psi\rangle$ is neither an eigenvector of $\widehat{P}$ nor an eigenvector of $\widehat{I}-\widehat{P}$). As projectors are associated to closed subspaces of the Hilbert space, quantum logic has the mathematical structure of the set of closed subspaces.

From these assumptions it is straightforward to define the fundamental relation of order (or *implication*) of propositions. We say that, for two propositions A and B we have $A \leq B$ (or $A \Rightarrow B$) if the subspace associated to B contains that associated to A. Hence the binary operations 'meet', $\curlywedge$, and 'join' , $\curlyvee$, may be defined in a natural form and it follows that the propositions form a *lattice.* The lattice is *orthocomplemented* (the subspaces associated to the proposition A and its negation A' are orthogonal) and *complete* (there exist the sure proposition, I, corresponding to the whole Hilbert space and its negation, Φ, corresponding to the null vector). Up to here everything is similar to what happens in classical logic. But the quantum lattice is not *distributive (Boolean*), unlike the classical one. As a conclusion, the authors claimed that the non-Boolean character of the lattice of

propositions is the essential characteristic of quantum theory. The details may be seen in the original article [**38**].

In the years elapsed since the work of Birkhoff and von Neumann many articles and several books have been devoted to the subject of quantum logic (see e.g. the book of Hooker [**39**]), in many cases starting from different definitions of quantum propositions. Some criticism has also arose in the sense that 'quantum logic' is not a true logic, but just a propositional calculus. Indeed, in an 'actual' logic the relations amongst proposition like $A \Rightarrow B$ or $A \curlywedge B = C$ should be also considered propositions, which is not necessarily the case in a propositional calculus. But the described approach to the logic of quantum mechanics is still widely used.

It is straightforward to define a probability distribution (or 'state') in any logic (orthocomplemented and complete lattice):

DEFINITION 2. *If $\mathcal{L}$ is a logic, a probability distribution is a mapping $p : \mathcal{L} \to [0, 1]$ with the axioms*

1) $p(\Phi) = 0$, $p(I) = 1$, where Φ (I) is the absurd (sure) proposition,

2) If $\{A_i\}$ is a (countable) sequence such that $A_i \leq A'_j$ for all pairs $i \neq j$, A' being the negation of A, then $\sum_i p(A_i) = p(\curlyvee A_i)$,

3) For any sequence $\{A_i\}$, $p(A_i) = 1\ \forall i \Rightarrow p(\curlywedge A_i) = 1$,

Thus from quantum logic, as defined by Birkhoff and von Neumann, we get quantum probability, whilst the classical, Boolean, logic provides the standard probability theory. Indeed, the above axioms are simply a generalization of the axioms of probability as stated by Kolmogorov [**40**].

2.5.2. The Bell inequalities as tests of classical probability. A discrimination between classical and quantum probability is provided by the Bell inequalities, derived as follows [**41**]. For any two propositions $A, B \in \mathcal{L}$ we may define a function, $d(A, B)$, by

$$d(A, B) = p(A \curlyvee B) - p(A \curlywedge B). \tag{27}$$

That function has the properties

$$0 \leq d(A, B) \leq 1, \qquad d(A, A) = 0, \qquad d(A, A') = 1, \tag{28}$$

and provides some measure of the 'distance' between two propositions in a given state (probability distribution). The function is called a *metric (pseudometric)* if the following additional property holds (does

not hold) true

$$d(A,B)=0 \qquad \Rightarrow \qquad A=B, \tag{29}$$

but this property is not very relevant for our purposes. More important are the following *triangle inequalities*, which are (are not in general) *fulfilled if the lattice is* (is not) *Boolean*

$$|d(A,B)-d(A,C)| \leq d(B,C) \leq d(A,B)+d(A,C). \tag{30}$$

As the Boolean character provides the essential difference between classical and quantum theories, according to Birkhoff and von Neumann [**38**], the triangle inequalities (30) give a criterium to distinguish between them. These inequalities are closely related to the Bell inequalities as shown in the following [**42**], although in the mathematical theory of probability the inequalities (30) were known well before Bell's work.

In quantum mechanics, if we consider three compatible propositions, $\{A,B,C\}$ (associated to pairwise commuting projectors) the inequalities (30) hold true because the lattice of commuting projectors is distributive. On the other hand, if two of the propositions, say A and B, are not compatible then their mutual distance is not defined because quantum mechanics does not provide a joint probability of two incompatible observables (and it is assumed that they cannot be measured simultaneously). However, there are quadrilateral inequalities, derived from the triangular ones (30), which may be violated by quantum mechanics and tested empirically. In fact, if we consider four projectors $\{A,B,C,D\}$ it is easy to see that the inequalities (30) lead to

$$d(A,D) \leq d(A,B)+d(B,C)+d(C,D). \tag{31}$$

In contrast with the inequalities (30), now all four distances may be defined in quantum mechanics if every pair involve commuting projectors (that is, if $[A,D]=[A,B]=[B,C]=[C,D]=0$). We see that the inequality (31) and the other three obtained by permutations involving the four projectors are necessary conditions for the existence of a *classical* joint probability distribution defined on the set of projectors. There are cases where quantum mechanics predicts violations of one of the inequalities, which leads to Bell's theorem (see chapter 3).

Inequality (31) is equivalent to the following one

$$p_B+p_C \geq p_{AB}+p_{BC}+p_{CD}-p_{DA}, \tag{32}$$

where p_A (or p_{AB}) is the probability that A (or $A \curlywedge B$) is true. This is called a Bell inequality [**43**] and, in this form, it was derived by Clauser and Horne [**44**]. Instead of projectors, taking the values 0 and 1, we might use observables taking the values -1 or +1. They are trivially related to the projectors by

$$a = 2A - 1, \qquad b = 2B - 1, \qquad etc. \tag{33}$$

and the inequality (32) takes the form of Clauser-Horne-Shimony-Holt (CHSH) [**45**]:

$$|\langle ab\rangle + \langle bc\rangle + \langle cd\rangle - \langle ad\rangle| \leq 2, \tag{34}$$

where $\langle ab\rangle$ means the expectation value of the product of a and b. Therefore these CHSH inequalities and the Clauser-Horne inequalities (32) are equivalent.

If no constraint is put, the left side of eq.(34) might be as large as 4. However, quantum mechanics forbids it to be larger than $2\sqrt{2}$ which is named the Cirel'son bound [**46**]. A proof of the Cirel'son bound is as follows. In quantum mechanics the left side of the CHSH inequality(34) should be the expectation value, in some state ψ, of the operator

$$\hat{M} = \hat{a}\hat{b} + \hat{b}\hat{c} + \hat{c}\hat{d} - \hat{d}\hat{a}$$

where all operators $\hat{a}, \hat{b}, \hat{c}, \hat{d}$ have eigenvalues ± 1, so that

$$\hat{a}^2 = \hat{b}^2 = \hat{c}^2 = \hat{d}^2 = 1. \tag{35}$$

Thus we get

$$\hat{M}^2 = 4 + \hat{a}\hat{b}\hat{c}\hat{d} - \hat{b}\hat{c}\hat{d}\hat{a} + \hat{c}\hat{d}\hat{a}\hat{b} - \hat{d}\hat{a}\hat{b}\hat{c}, \tag{36}$$

where I have taken into account eqs.(35), and that each Alice operator, $\hat{a}$ and $\hat{c}$, commute with every Bob operator, $\hat{b}$ or $\hat{d}$. What remains is to prove that the expectation of the four last terms in eq.(36) is not greater than four. This may be shown using the Schwartz inequality, taking into account that for instance $\langle\psi \mid \hat{a}\hat{b}\hat{c}\hat{d} \mid \psi\rangle$ may be seen as the scalar product of the vector $\langle\psi \mid \hat{a}\hat{b}$ times the vector $\hat{c}\hat{d} \mid \psi\rangle$. Thus we get

$$\left|\left\langle \psi \left| \hat{a}\hat{b}\hat{c}\hat{d} \right| \psi \right\rangle\right|^2 \leq \left\langle \psi \left| \hat{a}\hat{b}\hat{b}\hat{a} \right| \psi \right\rangle \left\langle \psi \left| \hat{c}\hat{d}\hat{d}\hat{c} \right| \psi \right\rangle = 1,$$

where again I have taken into account eqs.(35) . As a result I obtain

$$\left|\left\langle \psi \left| \hat{M} \right| \psi \right\rangle\right|^2 \leq \left|\left\langle \psi \left| \hat{M}^2 \right| \psi \right\rangle\right| \leq 8,$$

which completes the proof.

2.5.3. The measure of the amount of information. The amount of information is quantified with the concept of *entropy*. In classical physics, if we have a continuous random variable, λ, with a probability distribution $\rho(\lambda)$, the entropy, S^C, as defined by Shannon is

$$S^C = -\int \rho(\lambda) \log \rho(\lambda)\, d\lambda. \tag{37}$$

The quantum entropy was defined by von Neumann in terms of the density operator, $\hat{\rho}$, with an expression which looks similar to that one, namely

$$S^Q = -\mathrm{Tr}\left(\hat{\rho} \log \hat{\rho}\right). \tag{38}$$

In both cases $S \geq 0$ and the entropy increases with the lack of information, so that the pure states (maximal information) correspond to $S = 0$.

There are two other properties which hold true for both classical and quantum entropy:

Concavity:

$$\lambda S(\rho_a) + (1-\lambda) S(\rho_b) \leq S\left(\lambda \rho_a + (1-\lambda)\rho_b\right), \qquad 0 \leq \lambda \leq 1,$$

where ρ_a stands for either the classical probability density, $\rho_a(\lambda)$, or the quantum density operator, $\hat{\rho}_a$, and similarly ρ_b for a different probability density or density operator of the same system.

Subadditivity:

$$S(\rho_{12}) \leq S(\rho_1) + S(\rho_1),$$

where ρ_{12} stands for either the classical probability density $\rho_{12}(\lambda_1, \lambda_2)$ or the quantum density operator $\hat{\rho}_{12}$, the subindex 1 (2) refers to the first (second) subsystem of a composite system, and we have

$$\rho_1(\lambda_1) = \int \rho_{12}(\lambda_1, \lambda_2)\, d\lambda_2, \qquad \hat{\rho}_1 = \mathrm{Tr}_2 \hat{\rho}_{12}. \tag{39}$$

There is, however, a property which dramatically distinguishes classical from quantum entropy. In fact, in the case of a system consisting of two subsystems, the classical Shannon entropy fulfils

$$S^C(\rho_{12}) \geq \max\left\{S^C(\rho_1), S^C(\rho_2)\right\}, \tag{40}$$

whilst the quantum entropy fulfils the weaker triangle inequality

$$S^Q(\hat{\rho}_{12}) \geq \left|S^Q(\hat{\rho}_1) - S^Q(\hat{\rho}_2)\right|. \tag{41}$$

In my opinion, the fact that the quantum entropy does not fulfil an inequality similar to (40) is highly paradoxical, I would even say

bizarre. In fact, (41) allows for the possibility that both $S^Q(\hat{\rho}_1)$ and $S^Q(\hat{\rho}_2)$ are positive whilst $S^Q(\hat{\rho}_{12})$ is zero. This should be interpreted as saying that we have complete information about a composite system whilst we have incomplete information about each subsystem. This contrast with the classical, and intuitive, idea that full information about the whole *means* that we have complete information about every part. In my view this is indicative that the concept of 'complete' information in quantum theory is not the same as in classical physics, and the different meanings of completeness have been the source of misunderstandings about the interpretation of quantum theory, e.g. in the debate between Einstein and Bohr.

The violation of an inequality similar to (40) is closely related to the violation of the Bell inequality. The connection may be stated more easily if we introduce the concept of *linear entropy.* Actually, although the definitions of entropy (37) and (38) are standard and in some sense an optimum, it is possible to give alternative definitions of entropy which fulfil the essential properties of concavity and subadditivity. The most simple is the socalled linear entropy

$$S^{CL} = 1 - \int \rho(\lambda)^2 \, d\lambda, \qquad S^{QL} = 1 - \mathrm{Tr}\left(\hat{\rho}^2\right). \tag{42}$$

It is easy to see that $S^{CL} = 0$ implies a pure state, that is, all probability density concentrated in a single value of λ. Similarly $S^{QL} = 0$ implies pure quantum state, that is, $\hat{\rho}^2 = \hat{\rho}$, see comment after eq.(15).

The desired connection between linear entropy and the Bell inequalities has been studied by several authors in the last few years. For instance, Horodecki et al. [**47**] proved that the inequality (40) is a sufficient condition for the Bell inequalities. A slightly stronger result may be stated as follows

THEOREM 3. *The inequality*

$$MNS^{QL}(\hat{\rho}_{12}) + MN - M - N \geq NS^{QL}(\hat{\rho}_1) + MS^{QL}(\hat{\rho}_2), \tag{43}$$

where M and N are the dimensions of the Hilbert spaces of the two subsystems, is a sufficient condition for all Bell inequalities (32) *or* (34) *which may be got using two dichotomic observables of each subsystem.*

Proof: We consider observables {a, c} for the first particle and {b, d} for the second, all of which may take values 1 or -1, and the associated operators, $\hat{a}$, $\hat{b}$, $\hat{c}$ and $\hat{d}$. We define the Bell operator, $\widehat{B}$, by

$$\widehat{B} = \hat{a} \otimes \hat{b} + \hat{c} \otimes \hat{b} + \hat{c} \otimes \hat{d} - \hat{a} \otimes \hat{d}. \tag{44}$$

It is easy to see that

$$\mathrm{Tr}\widehat{B} = 0, \qquad \mathrm{Tr}\left(\widehat{B}^2\right) = 4MN, \tag{45}$$

and that the Bell inequality (34) is violated if

$$|\beta| > 2, \qquad \beta \equiv \mathrm{Tr}\left(\widehat{B}\hat{\rho}_{12}\right), \tag{46}$$

(whilst quantum mechanics predicts just $|\beta| \leq 2\sqrt{2}$). Now the obvious inequality

(47)

$$\mathrm{Tr}\left(\hat{\rho}_{12} - \frac{1}{N}\hat{\rho}_1 \otimes \widehat{I}_2 - \frac{1}{M}\widehat{I}_1 \otimes \hat{\rho}_2 + \frac{1}{MN}\widehat{I}_1 \otimes \widehat{I}_2 + \lambda\widehat{B}\right)^2 \geq 0, \lambda \in R,$$

where $\widehat{I}_1$($\widehat{I}_2$) is the identity operator for the first (second) particle, gives a quadratic expression in the variable λ. We get, after some algebra

$$MN\mathrm{Tr}\left(\hat{\rho}_{12}^2\right) - N\mathrm{Tr}\left(\hat{\rho}_1^2\right) - M\mathrm{Tr}\left(\hat{\rho}_2^2\right) \geq \frac{1}{4}\left(\beta^2 - 4\right). \tag{48}$$

Hence the inequality (43) implies $|\beta| \leq 2$, which proves the theorem.

Actually, inequality (43) is rather strong, and therefore not very useful, if either $M > 2$ or $N > 2$ or both, and it is trivial if either $M = 1$ or $N = 1$. Consequently its main interest is the case $M = N = 2$, where it is a consequence of the inequality (40), characteristic of classical information theory.

Bibliography

[1] P. A. M. Dirac: *The principles of quantum mechanics.* 4th ed. Clarendon Press, Oxford, 1958. (Original 1930).

[2] J. von Neumann: *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, 1955 (German original 1932).

[3] A. Galindo, P. Pascual: *Quantum Mechanics I, II.* Springer Verlag. 1991.

[4] M. Born: Quantum mechanics of collision processes. *Z. Phys.* **38**, 803 (1926); English translation in G. Ludwig, Ed., *Wave Mechanics*, Pergamon Press, Oxford, 1968.

[5] G. Emch: *Algebraic methods in statistical mechanics and quantum field theory.* Wiley-Interscience, 1972

[6] S. Weinberg: *The quantum theory of fields.* Cambridge University Press, 1995.

[7] E. Schrödinger: The present situation of quantum mechanics. *Naturwissenschaften,* **23**: pp. 807-812; 823-828; 844-849. (1935). (English translation reprinted in Wheeler and Zurek (1983)).

[8] J. S. Bell: *Speakable and unspeakable in quantum mechanics.* Cambridge University Press, 2004 (Second edition. First published in 1987*)*. This book contains reprints of many articles of Bell on foundations of quantum physics.

[9] K. R. Popper: *Quantum theory and the schism of physics.* Totowa, New Jersey, W. W. Bartley, 1982.
[10] M. Bunge: *Philosophy of physics.* Dordrecht, Reidel, 1973.
[11] N. Bohr: The quantum postulate and the recent development of atomic theory. *Nature*, *121*, 580-90 (1928). (Reprinted in Wheeler and Zurek (1983)).
[12] J. A. Wheeler, W. H. Zurek: *Quantum theory and measurement.* Princeton University Press, 1983.
[13] D. Wick: *The infamous boundary: Seven decades of controversy in quantum physics.* Birkhaüser, Boston, 1995.
[14] P. Mittelstaedt: The problem of interpretation of modern physics. *Found. Phys.* **41**, 1667-1676 (2011).
[15] P. Mittelstaedt: *The Interpretation of quantum mechanics and the measurement process.* Cambridge University Press, Cambridge UK, 1998.
[16] J. S. Bell. Are there quantum jumps?, in J. S. Bell: *Speakable and unspeakable in quantum mechanics.* Cambridge University Press, 2004. (Second edition. First published in 1987*).*
[17] E. London, E. Bauer, E. La thórie de l'observation en mécanique quantique. In P. Langevin (Ed.) *Actualités scientifiques et industrielles: Exposés de physique générale, No. 775*, 1939 . Hermann, Paris. (English translation reprinted in Wheeler and Zurek (1983)).
[18] E. P. Wigner: Remarks on the mind-body question. In I. J. Good (Ed.), *The scientist speculates*, pp. 284-302. Heinemann, London, 1961. (Reprinted in Wheeler and Zurek (1983)).
[19] G. C. Ghirardi, A. Rimini, T. Weber: Unified dynamics for microscopic and macroscopic systems. *Phys. Rev. D* **34**: 470 (1986).
[20] G. C. Ghirardi, P. Pearle, A. Rimini: Markov processes in Hilbert space and continuous spontaneous localization of systems of identical particles, *Phys. Rev.* A **42**, 78 (1990).
[21] G. C. Ghirardi: Collapse theories, in Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy.* (Winter 2011 Edition), .
[22] H. Everett: 'Relative state' formulation of quantum mechanics. *Rev. Mod. Phys.* **29***,* 454-462 (1957).
[23] B. S. de Witt, N. Graham: *The many-worlds interpretation of quantum mechanics.* Princeton University Press, 1973.
[24] L. Marchildon: Can Everett be interpreted without extravaganza? *Found. Phys.* **14**, 357 (2011).
[25] R. Omnes: *Understanding Quantum Mechanics.* Princeton University Press, 1999.
[26] E. Joos: *Decoherence and the Appearance of a Classical World in Quantum Theory* (2nd ed.). Springer, Berlin, 2003.
[27] R. J. Griffiths: *Consistent quantum theory.* Cambridge U. P. (2002).
[28] M. Tegmark: The interpretation of quantum mechanics: many worlds or many words? *Forsch. Phys.* **46**, 855-862 (1998).
[29] A. Einstein, B. Podolsky, N. Rosen: Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47**, 777-780 (1935) (Reprinted in Wheeler and Zurek (1983)).

[30] A. Einstein: Remarks concering the essays brought toghether in this cooperative volume. In P. A. Schilpp (Ed.), *Albert Einstein: Philosopher-Scientist* (pp. 665-688). La Salle Illinois, Open Court, 1949.
[31] L. E. Ballentine: *Quantum mechanics, a modern development.* World Scientific, Singapore, 1998.
[32] N. Bohr: Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **48**, 696 (1935).
[33] C. A. Fuchs, A. Peres: Quantum theory needs no 'interpretation'. *Phys. Today* **53**, 70 (2000).
[34] B. G. Englert: On quantum theory, *Eur. Phys. J D* **67**, 238 (2013).
[35] P. G, Lewis, D. Jennings, J. Barrett, T. Rudolph: Distinct quantum states can be compatible with a single state of reality, *Phys. Rev. Lett.* 109, 150404 (2012).
[36] M. F. Pusey, J. Barrett, T. Rudolph: On the reality of the quantum state. *Nature Physics* **8**, 475-478 (2012).
[37] N. Harrigan, R.W. Spekkens: Einstein, incompleteness, and the epistemic view of quantum states, *Found. Phys.* **40**, 125-157 (2010).
[38] G. Birkhoff and J. von Neumann: The Logic of Quantum Mechanics. *Ann. Math.* **37**, 823 (1936).
[39] C. A. Hooker (ed.), *The Logico-algebraic Approach to Quantum Mechanics*, Vols. I and II, Reidel, Dordrecht, 1975, 1979.
[40] E. Santos: Quantum logic, probability and information. The relation with the Bell inequalities. *Internat. J. Phys.* **42**, 2545-2555 (2003).
[41] E. Santos: The Bell inequalities as tests of classical logic. *Phys. Lett. A* **115**, 363-365 (1986).
[42] E. Santos: Mathematical and physical meaning of the Bell inequalities, *Eur. J. Phys.* **37**, 055402 (2016).
[43] J. S. Bell: On the Einstein, Podolsky and Rosen paradox. *Physics.* **1**, 195 (1964).
[44] J. F. Clauser and M. A. Horne: Experimental consequences of objective local theories. *Phys. Rev. D* **10**, 526 (1974).
[45] J. F. Clauser, M. A. Horne, A. Shimony and R. A. Holt: Proposed experiment to test local hidden variables theories. *Phys. Rev. Lett.* **23**, 880 (1969).
[46] B. S. Cirel'son: Quantum generalizations of Bell's inequality. *Lett. Math. Phys.* **4**, 93 (1980).
[47] R. Horodecki, P. Horodecki and M. Horodecki. Quantum α-entropy inequalities: independent condition for local realism? *Phys. Lett. A* **210**, 377 (1996).

CHAPTER 3

Realism, locality and Bell inequalities

3.1. Introduction

During the past 50 years the Bell inequalities have had an important role in all discussions about the interpretation of quantum mechanics. In this chapter we will briefly discuss the origin of the inequalities, which goes back to the early attempts at explaining the statistical character of quantum mechanics by means of hidden variables. There was a vivid discussion as to whether quantum mechanics is a complete theory, as has been discussed in chapter 2. The completeness problem was related by Einstein with relativistic locality, a relation that was the starting point of Bell's work. In section 3.2 we discuss the consequences of realism and the fact that quantum mechanics should be interpreted taking context into account. The conflict between quantum mechanics and local realism will be treated in section 3.3. The empirical tests of the Bell inequalities, proposed or performed, will be the matter of section 3.4 (optical test) and 3.5 (other tests). Finally I will include a section about the origin of irreversibility in physics, that may help to better understand the difference between causality and locality.

3.1.1. Theorems against hidden variables theories. Soon after the proposal of quantum mechanics in the years 1925-26 the hypothesis was put forward that the probabilistic character of the theory might be due to the fact that the description offered by the wave function is not complete. If this is the case additional 'hidden variables' might be included in order to complete the description, thus producing a subquantum theory whence quantum mechanics could be derived by appropriate averaging. Hidden variables have been discussed in chapter 2 section 2.4.6. The mainstream of the scientific community was positioned against hidden variables (HV) and several theorems have been proved with the purpose of either excluding HV theories or restricting the class of theories that are possible. In the following we will comment on the three most relevant, put forward by von Neumann, Kochen-Specker and Bell. We comment on the former now and on the other two in sections 3.2 and 3.3, respectively.

John von Neumann included in his celebrated 1932 book (see chapter 2) a theorem allegedly proving that any hidden variables model would contradict the predictions of quantum mechanics. The theorem was an obstacle for the research on HV during more than three decades (for details see chapter 2 sections 2.4.5 and 2.4.6). Our

dicussion of this theorem serves also as an introduction to the concept of contextuality.

The theorem rests on the hypothesis that HV theories should preserve linear relations. For instance, if three quantum observables represented by the operators $\widehat{A}$, $\widehat{B}$, $\widehat{C}$ fulfil the relation

$$\widehat{C} = \widehat{A} + \widehat{B}, \tag{49}$$

then a HV model should allow ascribing dispersion-free values to all three observables fullfiling a relation similar to eq.(49). It is then proved that this is not possible in general, that is, without a disagreement between quantum and HV predictions. An illustrative example will be given in chapter 5 section 5.2.6.1. At this moment it is enough to mention the criticism of Bell [**1**], namely that the linearity constraint to hidden variables models is actually unphysical. In fact, when the observables $\widehat{A}$ and $\widehat{B}$ commute a joint probability for them exists that provides an explicit HV model, as defined in chapter 2, section 2.4.6. However, if they do not commute then the observables are incompatible and the relation eq.(49) cannot be tested empirically. Nevertheless, HV theories are possible if we just demand that they reproduce the predictions of quantum mechanics for *actual experiments*. This fits in our definition for realistic interpretation made in chapter 1, section 1.2.9. The point is further clarified in section 3.2 below.

3.1.2. The two roles of the Bell inequalities. In 1964 Bell showed [**2**] that an important class of HV models are incompatible with quantum mechanics, namely local HV. This produced a deep impact in the scientific community, specially on quantum physicists interested in foundations. In particular, it seemed to support Bohr's vs. Einstein's position about the completeness of quantum mechanics. In fact, as discussed in chapter 2, section 2.4.5, Einstein had claimed that quantum mechanics is either incomplete or nonlocal, the second alternative being inconceivable to him because it suggests a contradiction with relativity theory. This is the first and most relevant role of the Bell inequalities: To show a possible conflict between quantum mechanics and relativistic causality, with the advantage that it may be put to empirical test. This is the aspect that will be mainly discussed in this chapter, starting in section 3.3.

The other role of the Bell inequalities is to provide a criterion to characterize phenomena that are specifically quantum, that is, phenomena that may not be interpreted within classical theories. This

role of the inequalities is extensively used in quantum information theory because an essential resource in quantum information is quantum entanglement and it is closely connected with the Bell inequalities as will be discussed in section 3.2.5 below. The relevance of this role of the Bell inequalities derives from the exponential growth of work in quantum information theory during the last few years. This growth derives from the hope that quantum laws may allow new technological advances in computation and other areas as well as from the fact that the size of computer's components is reaching the atomic level. In this book I will not deal with this subject which, although extremely important from the technological point of view, is not too relevant for the interpretation of the theory.

3.2. Realism and contextuality

In this section we will start defining contextual HV models. After that we will discuss the following topics: 1) to which extent contextual models contradict, or not, realism, 2) how the standard quantum formalism seems to demand noncontextuality, 3) the Kochen-Specker theorem, 4) the relation between the Bell inequalities and entanglement.

3.2.1. A classification of experiments. Our knowledge of the natural world derives from observations and experiments. In some cases, in particular astronomy, we cannot prepare the systems at will, we may just observe them and measure some quantities. For instance, we may measure the radius, surface temperature and spectrum of a star. In contrast, in laboratory experiments there are usually three steps: first we should prepare the state of a system following an appropriate protocol, then the state evolves in time, and finally we measure the relevant properties of the system in the final state. The preparation defines the initial state (represented by a density matrix in quantum mechanics) and the measurement the values of the relevant observables (self-adjoint operators in the quantum formalism). It is possible to define more general measurements, named positive operator valued measurements (POVM) but for our aim (finding a realistic interpretation) this generalization is not necessary.

It is useful to classify experiments in two kinds: simple and composite.

Definition 4. *A simple experiment consists of the preparation of a state of a physical system, followed by the evolution of the system*

and finishing with the measurement of one or a finite set of compatible observables.

DEFINITION 5. *A composite experiment consists of several simple experiments with the same preparation, as defined by a protocol, and the subsequent evolution, but measuring different sets of compatible observables in each simple experiment. Typically the observables measured in a simple experiment are incompatible with those measured in another simple experiment.*

With these definitions we may state the following consequence for quantum mechanics: Any simple experiment admits a hidden variables model. This follows from the fact that all observables that can be measured in a single experiment are compatible, which implies that the corresponding operators commute and hence there exists a joint probability distribution, as discussed in section 2.4.6 of chapter 2.

3.2.2. The implications of realism. Realistic interpretations of a physical theory rest on the concept of 'ontic states', that is, real physical states of systems, objective and independent of any observer. Ontic states are assumed to possess objective properties that might be measured, maybe indirectly. I will define as realistic any theory aimed to interpret the empirical facts in terms of ontic states, and not just concerned with providing a set of rules for the calculation of the results of experiments or observations. A necessary condition for realism is as follows. Let us assume that in some experiment we want to measure the observable A, which may possess several possible values. The value obtained in the measurement, say a, will depend on the state, say λ, of the system and the observable which we measure, A. We may write

$$a = a\left(\lambda, A, context\right). \tag{50}$$

where we may interpret λ as the set of values of the variables which faithfully determine the ontic state of the system. Actually, in the standard interpretation of eq.(50) it is implicit the 'arrow of time' (for a study of the arrow of time see section 3.6 below) because λ refers to the state of the system *before* the measurement, but it is assumed that the result does not depend on anything that happens *after* it. Thus we might say that eq.(50) is a condition for 'time-ordered realism' which combined with locality is equivalent to realism plus relativistic causality. Actually, any theory is related to *human* knowledge of nature and humans are unavoidably constrained by the

arrow of time. The dependence on *context* takes into account that the result of the measurement may depend on the full experimental equipment used for the measurement of the observable A. That is, the possibility is allowed that two different measurements of the same observable performed on two identical systems in the same state give different results, due to a different environment.

In classical physics the influence of the environment is considered both unavoidable and inconvenient but (most of the times) irrelevant. In fact, all (or most) measurements have errors, which are attributed to an incomplete control of the experimental setup (e.g., due to perturbations caused by the environment). It is inconvenient because an ideal measurement should have no errors. It is not too relevant because we assume that the errors may decrease indefinitely with increasingly better equipment. In quantum physics the situation is different. Indeed, in some cases the errors are not only unavoidable but as big as the main result. This happens for instance if we attempt to measure at the same time the position and the velocity of a particle, as is shown by the Heisenberg uncertainty inequalities. Thus in order to formalize the possible influence of the environment we should include the context in eq.(50) .

For some people eq.(50) should be called a condition for determinism, rather than realism, because the states of the system and the context completely *determine* the result of the measurement. That is, eq.(50) excludes the possibility that natural laws are not strictly causal. In order to circumvent this objection I may replace for eq.(50) the more general one

$$Prob(a) = P_a\left(\lambda, A, context\right), \tag{51}$$

with the meaning that the states of system and context only determine the probability of getting the value a. It is compatible both with the assumption that natural laws are strictly causal and with its denial. In fact when P_a may take only the values 0 or 1 then eq.(51) reduces to eq.(50). We see that realism (the necessary condition for which is eq.(51)) is more general than determinism (or deterministic realism, which needs eq.(50)).

In eq.(51) the state of the system is represented by a continuous variable λ with probability density $\rho\left(\lambda\right)$. Introducing a density we include the possibility that the state of the system is incompletely known and we attach a probability distribution to our ignorance. This

is named a mixed state. Actually, many people do not believe that probability should be always attributed to incomplete information (which is usually named *ignorance interpretation of the probability*), but to the fact that the laws of nature are intrinsically probabilistic. In any case, the introduction of a mixed state via $\rho(\lambda)$ may derive from the incomplete control of the initial state of the system, due to either a practical imperfection of the set-up or an essential indeterminacy. (At this moment I point out that the interpretation of probability is a controversial subject, to which many articles and books have been devoted. Here I will not discuss this question).

The dependence on *context* takes into account that the result of the measurement might depend on the experimental environment. Therefore a composite experiment consisting of two simple ones, with the same preparation but different measurements, corresponds to unequal contexts at least for the environment. If this is accepted then contextual realistic models (or hidden variables models) are always possible [**1**].

3.2.3. Noncontextual models.

DEFINITION 6. *A realistic model for a composite experiment is noncontextual if there exists a joint probability distribution for all observables of the system, even if some pairs are not compatible.*

Models for simple experiments are always noncontextual because these models may predict the joint probability distribution of the observables, all of them obviously compatible. In composite experiments a necessary, but not sufficient, condition for noncontextuality is the following. The marginal for the variable A in the joint distribution of the compatible observables A and B is the same as the marginal for A in the joint distribution of the compatible observables A and C. If this was not the case we could say that the observable A in the first simple experiment is different from A in the second one. With this convention the condition holds true both in quantum mechanics and classical physics. For instance, for a given preparation we may measure the spin of a particle together with its position or measure spin together with its momentum, and quantum mechanics predicts that the result for the spin should be the same.

The existence of a joint distribution is a stronger constraint. What is required is the existence of some function of all the observables, $p(A, B, C \dots)$, fulfilling the mathematical properties of a joint probability distribution and such that the marginals for every subset of

compatible observables are the ones predicted by quantum mechanics. The said joint distribution is just a mathematical object (it cannot be measured if some of the observables are not compatible) but their mere existence puts constraints which may be tested empirically as discussed in the following.

It is not difficult to see that the existence of a joint distribution for the observables $A, B, C, \ldots$, is equivalent to the existence of a positive normalized function, $\rho(\lambda)$ of the variable or set of variables, λ, plus functions like $A(\lambda), B(\lambda), C(\lambda) \ldots$ Then the joint distribution $f(A, B, C, \ldots)$ may be obtained from the moments that may be calculated as follows

$$\left\langle A^{\alpha} B^{\beta} C^{\gamma} \ldots \right\rangle = \int \rho(\lambda)\, d\lambda\; A(\lambda)^{\alpha}\, B(\lambda)^{\beta}\, C(\lambda)^{\gamma} \ldots$$

A joint probability distribution $f(A, B, C, \ldots)$ cannot be measured if not all observables are compatible. In quantum mechanics it is assumed that a necessary condition for compatibility is that the associated operators $\widehat{A}$, $\widehat{B}$, $\widehat{C}$,... commute pairwise. The situation with respect to realistic (ontological) models is the following. We may obtain several ontological (or HV) models, one for each simple experiment. For instance, let us consider a composite experiment consisting of two simple ones. In the first, where we measure A and B, a realistic model should provide the functions $\rho_1(\lambda)$, $A_1(\lambda), B_1(\lambda)$. In the second, where we measure A and C, another model would give $\rho_2(\lambda)$, $A_2(\lambda), C_2(\lambda)$. The two models together might be called a model for the composite experiment. It would be noncontextual if $\rho_1(\lambda) = \rho_2(\lambda)$ and $A_1(\lambda) = A_2(\lambda)$; if this does not happen the model is *contextual*.

For the sake of clarity I put a more involved example. Let us consider four dichotomic observables, A, B, C and D, each of which may take the values 0 or 1. We assume that A and C are incompatible, and B and D are also incompatible, the remaining pairs being compatible. The corresponding operators will be projectors, i.e. fulfilling $\widehat{A}^2 = \widehat{A}$, etc., all pairs of projectors commuting except

$$\left[\widehat{A}, \widehat{C}\right] \neq 0, \qquad \left[\widehat{B}, \widehat{D}\right] \neq 0. \tag{52}$$

Let us label p_A the probability of $A = 1, p_{AB}$ the probability that $A = B = 1$, etc. The existence of a joint distribution means that

there are 15 positive quantities

$$p_A,\ p_B,\ p_C,\ p_D,\ p_{AB},\ p_{AC},\ p_{AD},\ p_{BC},\ p_{BD},$$
$$p_{CD},\ p_{ABC},\ p_{ABD},\ p_{ACD},\ p_{BCD},\ p_{ABCD}, \tag{53}$$

which should fulfil the relations

$$0 \leq p_{ABCD} \leq p_{ABC} \leq p_{AB} \leq p_A \leq 1, \tag{54}$$

and those obtained by all permutations of the labels. Only 8 of these quantities may be measured (and they are predicted by quantum mechanics), namely

$$p_A, p_B, p_C, p_D, p_{AB}, p_{AD}, p_{BC}, p_{CD}. \tag{55}$$

The remaining 7 quantities cannot be measured, the corresponding observables not being compatible, and quantum mechanics predicts no value for them. The question is whether there exist 7 (not measurable) quantities fulfilling all constraints of the type eq.(54) which together with the 8 measurable ones provide the desired joint probability distribution eq.(53).

3.2.4. Kochen-Specker theorem. The impossibility of noncontextual models compatible with the quantum predictions in all (composite) experiments is established by the following theorem.

THEOREM 7. *Noncontextual HV models do not exist in general, i.e. in agreement with quantum predictions for all possible (composite) experiments.*

This is usually called Kochen-Specker theorem [**4**] after the authors who proved it in 1967. However, the theorem had been actually shown one year earlier by Bell [**1**]. Furthermore, the theorem may be proved also via the Bell inequality (BI) in two steps. Firstly we realize that noncontextuality implies BI, which was shown in chapter 2, section 2.5.2. Then we should find examples of composite experiments where the quantum predictions violate BI. Many instances will be exhibited in sections 3.4 and 3.5 below. At this moment it is appropriate to stress that BI holds true for any noncontextual model of an experiment (simple or composite), that is, a model including a joint probability distribution for all observables. As this happens always in classical physics (except some contrived irrelevant examples) then the BI are fulfilled in classical theories.

The correlations that cannot be explained via noncontextual models may appear in two different scenarios: 1) Correlations between

properties of a system localized in a small region of space, 2) Correlations between distant systems. Actually, the difference between the two scenarios is not sharp, but there is an important case which belongs clearly to the latter, namely EPR type experiments to be studied after introducing the constraint of locality. Actually, the Kochen-Specker theorem poses no real problem for a realistic interpretation of quantum mechanics; it is enough to assume that the context is very relevant in the quantum domain. However, a dramatic exception with great relevance is noncontextuality in EPR type experiments as studied in section 3.3 below.

3.2.5. Entanglement and contextuality. I discussed the quantum concept of entanglement in chapter 2, section 2.3.2. Two subsystems of a physical system are said entangled if the quantum (Hilbert space vector) representative of the whole cannot be written as a product of the representatives of the subsystems. That is, if ψ_a is the wave-function of the first subsystem and ψ_b of the second, both normalized, and the wave function of the whole is

$$\psi = \psi_a \psi_b, \tag{56}$$

then the subsystems are *not entangled.* The typical wave-vector of a system consisting of entangled subsystems is (modulo a global phase factor)

$$\mid \psi\rangle = c \mid H_a\rangle \mid V_b\rangle + \sqrt{1 - |c|^2} \mid V_a\rangle \mid H_b\rangle, \qquad |c| < 1. \tag{57}$$

where $\mid H_a\rangle$ and $\mid V_a\rangle$ are possible normalized states of the first subsystem (say measured by Alice) and $\mid H_b\rangle$ and $\mid V_b\rangle$ of the second (Bob), c being a complex number. The labels H and V are written by analogy with the state of two photons entangled in polarization (horizontal or vertical), but our treatment is general. Actually, we might also consider states of systems where $\mid \psi\rangle$ is a sum of more than two terms in the form of products, the essential condition for entanglement being that the state vector of the whole system $\mid \psi\rangle$ cannot be written in the form of eq.(56) via a choice of the subsystem states.

The relevance of entanglement was pointed out in 1935 in two articles by Einstein et al. (EPR) and by Schrödinger, respectively, both critical with the Copenhagen interpretation of quantum mechanics. The former (EPR) has been discussed in chapter 2 section 2.4.5. It exhibited an explicit example of entanglement. In the second one Schrödinger wrote that entanglement is *the* characteristic trait

of quantum mechanics and emphasized the strange consequences of entanglement with his celebrated example of the 'cat', discussed in chapter 2 section 2.4.

Entanglement provides the examples needed to prove Bell's theorem, that is, the (alleged) incompatibility of quantum mechanics with local realism. We will prove in section 3.3 below that local realism implies the Bell inequalities. Now we will show that there are cases of quantum states violating the inequalities. In fact, entanglement is both a necessary and a sufficient condition for the violation of Bell inequalities, in the precise sense stated as follows.

THEOREM 8. *For any system in a state without entanglement (whose wave-function may be written as in eq.*(56)*), and for any set of observables* $\{\widehat{A}_1,\ \widehat{A}_2,\ \widehat{B}_1,\ \widehat{B}_2\}$ *the first (last) two acting on the first (second) subsystem, all with eigenvalues in the range* $[-1,1]$*, the CHSH (Bell type) inequality holds true.*

The CHSH inequality was derived in chapter 2 section 2.5.2 from noncontextuality, that is, the existence of a joint probability distribution. It will be studied again in section 3.3.3 below.

Proof: For any product $\widehat{A}_j\ \widehat{B}_k$, $j,k=1,2$ the expectation factorizes, that is,

$$\left\langle \psi \left| \widehat{A}_j \widehat{B}_k \right| \psi \right\rangle = \left\langle \psi_a \left| \widehat{A}_j \right| \psi_a \right\rangle \left\langle \psi_b \left| \widehat{B}_k \right| \psi_b \right\rangle \equiv a_j b_k; \quad a_j, b_k \in [-1,1]\,.$$

Then the proof reduces to show that any four real numbers all in the range $[-1,1]$ fulfil the following inequalities

$$-2 \leq a_1 b_1 + a_1 b_2 + a_2 b_1 - a_2 b_2 \leq 2,$$

which easily leads to the CHSH eq.(61) below.

THEOREM 9. *For any system in a state with entanglement (whose wave-function may be written as in eq.*(57)*), it is possible to find four observables* $\{\widehat{A}_1,\ \widehat{A}_2,\ \widehat{B}_1,\ \widehat{B}_2\}$ *the first (last) two acting on the first (second) subsystem, all with eigenvalues in the range* $[-1,1]$*, such that the CHSH inequality is violated.*

This is usually named Gisin theorem [**5**] and it has been generalized for all entangled pure states [**6**]. In the following I sketch the proof for the entangled state eq.(57) . In analogy with polarization we associate the operators $\{\widehat{A}_1,\ \widehat{A}_2,\ \widehat{B}_1,\ \widehat{B}_2\}$ with angles in the form

$$\widehat{A}_j = 2 \mid \theta_j\rangle\langle\theta_j \mid -1, \qquad \widehat{B}_j = 2 \mid \phi_j\rangle\langle\phi_j \mid -1,$$

where $\mid \theta_j\rangle$ is a state with polarization at an angle θ_j with the horizontal, so that $\mid \phi_j\rangle\langle\phi_j \mid$ is a projector onto that state. Hence we get

$$\begin{aligned}\langle\psi \mid \widehat{A}_j\widehat{B}_k \mid \psi\rangle = |c|^2 \left\langle H_a \left| \widehat{A}_j \right| H_a \right\rangle \left\langle V_b \left| \widehat{B}_k \right| V_b \right\rangle \\ + (1-|c|^2) \left\langle V_a \left| \widehat{A}_j \right| V_a \right\rangle \left\langle H_b \left| \widehat{B}_k \right| H_b \right\rangle \\ + c\sqrt{1-|c|^2} \left\langle H_a \left| \widehat{A}_j \right| V_a \right\rangle \left\langle V_b \left| \widehat{B}_k \right| H_b \right\rangle \\ + c\sqrt{1-|c|^2} \left\langle V_a \left| \widehat{A}_j \right| H_a \right\rangle \left\langle H_b \left| \widehat{B}_k \right| V_b \right\rangle,\end{aligned}$$

where

$$\begin{aligned}\left\langle H_a \left| \widehat{A}_j \right| H_a \right\rangle &= 2\langle H_a \mid \theta_j\rangle\langle\theta_j \mid H_a\rangle - 1 = 2\cos^2\theta_j - 1, \\ \left\langle V_a \left| \widehat{A}_j \right| V_a \right\rangle &= 2\sin^2\theta_j - 1, \\ \left\langle H_a \left| \widehat{A}_j \right| V_a \right\rangle &= 2\cos\theta_j \sin\theta_j - 1,\end{aligned}$$

and similarly for the other cases. Then after some algebra we have

$$S \equiv \langle\psi \mid \widehat{A}_1\widehat{B}_1 + \widehat{A}_1\widehat{B}_2 + \widehat{A}_{2j}\widehat{B}_1 - \widehat{A}_2\widehat{B}_2 \mid \psi\rangle = S(\theta_1, \theta_2, \phi_1, \phi_2),$$

which is a function of the four angles. What remains is the mathematical proof that there are choices of the angles $\{\theta_1, \theta_2, \phi_1, \phi_2\}$ such that either $S > 2$ or $S < -2$, thus violating the CHSH inequality (61). For $|c| = -\sqrt{1/2}$ the proof is trivial, see section 3.4.1.

These theorems show that quantum predictions for entangled states violate Bell inequalities, thus proving that quantum mechanics is not compatible with noncontextual hidden variables models. The alleged incompatibility with local realism (named Bell's theorem) is studied in the following.

3.3. Local realism

3.3.1. Correlations derived from a common origin. Correlations between events are quite common and usually easy to discover, but it is not always trivial to know whether a simple correlation derives from a causal connection. For instance, in some country a specific food may be common and a specific disease be highly prevalent, which might suggest that the latter is a consequence of the former. However, it may be that the climate in the country favours both the abundance of the said food and illness, without any causal connection between

the two. In any case every correlation that does not correspond to a causal connection should be attributed to a common cause, as the climate in our example. Indeed, both in classical physics and in other sciences the correlations between distant systems are assumed to derive from either a causal connection or a common cause in the past. For instance, the similarity between twins (possibly living in different cities) is a correlation between distant bodies. It is an obvious consequence of the common origin, the twins having a similar genetic code.

The relevant fact for us is that correlations cannot be interpreted as due to a common cause whenever a Bell inequality is violated. Hence a fundamental consequence of Bell's work is to (allegedly) show that *correlations between distant bodies not deriving from a common past may exist in nature*, this being the case in the correlations violating a Bell inequality, as predicted by quantum mechanics. I include the word 'allegedly' because I am convinced that Bell's formulation of local realism is not general enough. However, most of this chapter will be written fitting in the common opinion, questioning it just at the end of this section 3.3 and in chapter 6.

The Bell inequalities are properties of correlations that derive from a common past, as discussed above. Therefore, they may be derived from the laws of (standard) probability (see chapter 2, section 2.5.2), a derivation independent of the existence of quantum mechanics. Quantum mechanics enters because it predicts strange correlations able to violate the inequalities in some cases. The counterintuitive character of refuting the cherished assumption that correlations are either causally connected or due to common causes in the past is strongly reinforced if the measurements are performed in space-like separated systems in the sense or relativity theory. Actually, Bell himself reinterpreted the correlation between two measurements, one by Alice the other one by Bob, within relativity theory [**3**]. To do that he considered the set of variables λ to be the union of two sets, λ_a and λ_b, consisting each of all events in the past light cone of the measurement performed by Alice and Bob, respectively. Therefore, the violation of a Bell inequality with space-like separated measurements, in the sense of relativity theory, would imply that the correlation does not derive from events in the intersection of the said past light cones.

The said incompatibility between quantum mechanics and relativistic causality is called Bell's theorem. The contradiction looks dramatic, but most authors think that there is no real contradiction with

relativity theory because quantum mechanics does not allow sending superluminal signals from Alice to Bob or from Bob to Alice (see below). In order to clarify the subject it is necessary to distinguish bewteen no-signalling and nonlocal correlations, as will be made in section 3.3.4 below. Before that we will make a digression on the meaning of locality and derive from it the BI.

3.3.2. Einstein and Bell locality conditions. It is common opinion that the most celebrated supporter of local realism was Albert Einstein, whence recalling his views may add clarity. His opinions about realism have been commented on in chapter 1 section 1.1.3 and in chapter 2 section 2.4.5, see also [**7**], and will be discussed again now. He supported the assumption that a picture of the world fitting in the tradition of (classical) physics is possible. That is, a world view where physical systems have properties independently of any observation (*the moon is there when nobody looks*), probabilities appear due to incomplete information, maybe unavoidable, rather than by an essential indeterminacy (*God does not play dice*) and actions propagate in spacetime at a speed not greater than that of light (without *spooky actions at a distance*), where I have emphasized three well-known sentences by Einstein.

The third sentence is specifically related to *local* theories, which Einstein supported emphatically in his autobiographical notes [**8**]. He wrote "On one supposition we should, in my opinion, absolutely hold fast: the real factual situation of the system S2 is independent of what is done with the system S1, which is spatially separated from the former". This quotation is usually interpreted as Einstein's support for 'relativistic causality', this believed to be synonymous of 'Einstein locality', a belief actually not correct. The pioneer paper of John Bell [**2**], where he introduced his celebrated inequalities, starts quoting that sentence of Einstein as a justification of his (Bell's) formula defining local realism, eqs.(50) below. However, Bell locality includes two ingredients: Einstein locality, which forbids (relativistic) spacelike influences, plus an asymmetry of time sometimes named 'the arrow of time'. This implies the assumption that the present may influence the future, but not the past, which in (special) relativity would mean that an event may be influenced only by events in its past light cone, that is, influenced neither by spacelike separated events nor by events in the common future light cone. Therefore, Bell's locality corresponds more properly to what is usually named (relativistic) causality.

Einstein sentence did not exclude influences by events in the future light cone. Indeed, he was well aware that the laws of physics do not distinguish future from past, as in the often quoted passage from his letter of condolences upon the death of his friend Michele Besso: "Michele has left this strange world just before me. This is of no importance. For us convinced physicists *the distinction between past, present and future is an illusion*, although a *persistent* one" [**9**] (my emphasis).

I have stressed the difference between Einstein and Bell locality because it is philosophically interesting, but in practice the additional ingredient of the latter with respect to the former is unavoidable. We stress that it involves an anthropic element, namely the fact that we, human beings, are able to prepare a system at a time and study the evolution towards the future, but are unable to prepare a system and study the evolution towards the past. The relevance of the anthropic element for the subject that we are discussing makes it worth to have a closer look at the 'arrow of time'. We will do that in section 3.6.

3.3.3. The Bell inequalities. In his pioneer 1964 paper Bell [**2**] introduced local theories, intermediate between contextual and non-contextual ones. The title of the paper, "On the Einstein-Podolsky-Rosen paradox", acknowledges inspiration in the celebrated 1935 EPR article (see chapter 2 section 2.4.5), where the authors proposed as evident a locality condition, as discussed above. In practice, the experiments devised by Bell consist of two signals produced in a source that travel towards two parties, Alice and Bob. Alice measures the observable A and Bob the observable B, in such a way that the measurements are performed in spatially separated regions in the sense of relativity theory. In this case no signal travelling with a speed smaller than light may inform Bob (Alice) about what Alice (Bob) is measuring. Therefore in the correlation

(58)

$$\langle AB\rangle = \sum_{ab} ab \int \rho\left(\lambda\right) d\lambda P_a\left(\lambda, A, \text{context}(A)\right) P_b\left(\lambda, B, \text{context}(B)\right),$$

P_a might depend on the context of the measurement of the observable A, but it cannot depend on the context of the measurement of B, and similarly the measurement of B cannot depend on the context of A. However, non-contextual models may be local, as it happens for instance in static experiments, where information might go from Alice to Bob or from Bob to Alice at any velocity. On the other hand

there are more general (fully contextual) models where the locality condition is not fulfilled. Thus we may write the following hierarchy

$$\text{non-contextual} \subseteq \text{local} \subseteq \text{general (contextual)}. \tag{59}$$

Within relativity theory eq.(58) might be interpreted assuming that λ consists of all events in the intersection of the past light cones of the measurements of A and B, whilst $context(A)$ and $context(B)$ would consist of the events in the regions of the respective past light cones that are not common. The strange fact is that eq.(58) seems to not always hold true according to quantum mechanics, as we discuss in the next subsection.

The proof that the local realistic eq.(58) is violated by quantum mechanics was achieved by Bell in 1964 [**2**]. The proof is more clear deriving a Bell inequality (BI) from local realism and then showing that the BI contradicts QM in some cases. An appropriate BI testable in experiments was proposed in 1969 by Clauser, Horne, Shimony and Holt (CHSH) [**10**]. As pointed out in section 3.2.4, the BI are necessary conditions for noncontextual models, where there is a joint probability distribution for all measurable quantities in different simple experiments with similar preparations. In chapter 2, section 2.5.2 we derived the CHSH inequality from that condition. In the following we shall derive it proving that local realism implies the existence of a (mathematical) function with the properties of a joint probability distribution. However, the function is not a joint probability distribution in the physical sense because it cannot be measured.

We may consider a composite experiment consisting of four simple ones with the same preparation and four different measurements by two parties. Alice may measure either the observable A_1 or A_2, but only one of these at a time, Bob may measure either B_1 or B_2, one at a time. We assume that the four observables are dichotomic, having values $\{-1, 1\}$ and that Alice and Bob perform four joint measurements. That is, they measure A_1B_1 and get as a result either $\{a_1' = 1, b_1' = 1\}$ or $\{a_1' = 1, b_1' = -1\}$ or $\{a_1' = -1, b_1' = 1\}$ or $\{a_1' = -1, b_1' = -1\}$. Then in another run of the experiment they measure A_1B_2 with the result $a_1''b_2''$. Similarly they get $a_2'''b_1'''$ in the measurement of A_2B_1 and finally $a_2''''b_2''''$ in the measurement of A_2B_2. It is trivial that the following inequalities hold true for any set of results

$$-4 \leq S \equiv a_1'b_1' + a_1''b_2'' + a_2'''b_1''' - a_2''''b_2'''' \leq 4. \tag{60}$$

Stronger inequalities can be obtained adding physical constraints. In fact, if we introduce restrictions derived from local realism then the CHSH inequality is obtained

$$-2 \leq \langle a_1 b_1 \rangle + \langle a_1 b_2 \rangle + \langle a_2 b_1 \rangle - \langle a_2 b_2 \rangle \leq 2, \tag{61}$$

where $\langle a_j b_k \rangle$ means the expectation value of the product $a_j b_k$ in a joint measurement performed of the observables a_j b_k by Alice and Bob, respectively.

In four experiments using identically prepared states of the same system, local realism implies that Alice's and Bob's measurements are independent, see eq.(58) . This implies that the probability P_a that Alice gets the value a cannot depend on whether Bob measures B_1 or B_2, and similar for the Bob probability P_b. Bell formalized these facts writing the single and coincidence probabilities for the Alice and Bob measurements as follows (see eq.(58))

$$\begin{aligned} P_{A_j} &= \langle A_j \rangle = \langle M(A_j) \rangle \equiv \int \rho(\lambda)\, d\lambda M(\lambda, A_j), \\ P_{B_k} &= \langle B_k \rangle = \langle M(B_k) \rangle \equiv \int \rho(\lambda)\, d\lambda M(\lambda, B_k), \\ P_{A_j B_k} &= \langle A_j B_k \rangle = \langle M(A_j) M(B_k) \rangle \\ &\equiv \int \rho(\lambda)\, d\lambda M(\lambda, A_j) M(\lambda, B_k), \end{aligned} \tag{62}$$

where $j = 1, 2, k = 1, 2$, λ is a set of random ('hidden') variables with a probability density $\rho(\lambda)$, and $\langle A_j \rangle, \langle B_k \rangle$ and $\langle A_j B_k \rangle$ are the expectation values of the results of measuring the observables A_j, B_k or their product $A_j B_k$, respectively. The expectations agree with the probabilities if the variables A_j, B_k have values $\{0, 1\}$. It is easy to see that eqs.(62) and their generalizations provide a method to obtain a 'joint probability distribution' for the observables. In fact, if we know the function $M(\lambda, A_j)$ for a set of observables $\{A_j\}$ (with values $\{0, 1\}$) then the joint probability $P(A_1, A_2, A_3, \dots)$ for all these observables may be got from the integral

$$P(A_1, A_2, A_3, \dots) = \int \rho(\lambda)\, d\lambda \prod_j M(\lambda, A_j).$$

This completes the proof that eqs.(62) imply the CHSH inequality (61).

I shall point out that in papers dealing with the Bell inequalities it is frequent to call 'correlation' the expectation value of a product of observables like $\langle AB\rangle$. The name does not agree with the standard one in mathematical statistics, where the correlation between two random variables, A and B, is usually defined by the dimensionless quantity

$$Corr(A,B) \equiv \frac{\langle AB\rangle - \langle A\rangle \langle B\rangle}{\sqrt{\langle A^2\rangle - \langle A\rangle^2}\sqrt{\langle B^2\rangle - \langle B\rangle^2}}. \tag{63}$$

It has been shown that inequalities similar to CHSH with the correlation eq.(63) substituted for the expectation of the product, $\langle AB\rangle$, may be violated by classical (local realistic) models. This fact has lead some authors to misunderstand, and criticize, Bell's work.

3.3.4. No-signalling and nonlocal correlations. No-signalling means that it is not possible to send faster-than-light information. That is, an observation cannot be affected by a disturbance located at a point of spacetime such that the distance divided by the time difference is greater than the speed of light (space-like separation). It is one of the rules of relativity adopted by quantum mechanics [**11**], [**12**]. Indeed, it is a consequence of microcausality, a postulate of quantum field theory [**13**]. Microcausality means that two field operators at spacelike separation commute. Nonlocal correlation is a weaker assumption than no signaling, so that the former may be true but not the latter. In order to better understand the difference it is convenient to introduce the following terminology.

In an experiment where Alice and Bob measure some observables at spacelike separation we may speak about *parameter dependence and outcome dependence* [**14**], [**15**]. The former means that the result of a measurement made by Bob could depend on the specific measurements performed by Alice in a distant place. The latter means that Bob's result depends on (is correlated with) the result obtained by Alice, and it might be named just correlation. It may be realized that a *parameter dependence* could allow Alice to send signals to Bob because Bob might know what Alice is measuring. This would be strange if both measurements are made in distant places, even more strange if they are space-like separated. On the other hand faster-than-light signalling cannot be achieved if locality is violated only by the occurrence of outcome dependence.

For the sake of clarity let us consider the following example with reference to the CHSH inequality (61). A hypothetical system where

Alice may measure either the observable a or c and Bob may measure either b or d has the following expected results, depending on which variables are measured (I will label P the probability):

- $a = 1$, $b = 1$ with $P = \frac{1}{2}$; $\quad a = -1$, $b = -1$ with $P = \frac{1}{2}$.
- $c = 1$, $b = 1$ with $P = \frac{1}{2}$; $\quad c = -1$, $b = -1$ with $P = \frac{1}{2}$.
- $c = 1$, $d = 1$ with $P = 1$.
- $a = 1$, $d = -1$ with $P = 1$.

In this example it is easy to see that the CHSH inequality (61) is violated because the left side has the value 4. There is no 'parameter independence' in the sense that the result of the measurement of d by Bob does depend on what Alice is measuring. In fact, if Alice is measuring a, Bob will get $d = -1$ with certainty, whilst if she is measuring c, he will obtain $d = 1$. It is intuitively obvious that this requires a transmission of information from Alice to Bob that, if the measurements are space-like, should be faster-than-light. This would allow sending superluminal signals from Alice to Bob. If she chooses to measure $a(b)$, he will know Alice's choice because he will get $-1(1)$ in the measurement of d. On the other hand a situation like the one of the example is forbidden by quantum mechanics. In fact, the quantum formalism implies that the expectation of the observable d should be independent on whether d is measured simultaneously with a or simultaneously with c, and given by $\left\langle \psi \left| \hat{d} \right| \psi \right\rangle$, where $\hat{d}$ is the operator associated to the observable d, and ψ represents the state of the system. In summary, in the example: 1) a Bell inequality is violated, 2) there is no parameter independence, 3) it is possible to send superluminal signals, and 4) quantum predictions are violated.

An example with parameter independence is the following:

- $a = 1$, $b = 1$ with $P = \frac{1}{2}$; $\quad a = -1$, $b = -1$ with $P = \frac{1}{2}$.
- $c = 1$, $b = 1$ with $P = \frac{1}{2}$; $\quad c = -1$, $b = -1$ with $P = \frac{1}{2}$.
- $c = 1$, $d = 1$ with $P = \frac{1}{2}$; $\quad c = -1$, $d = -1$ with $P = \frac{1}{2}$.
- $a = 1$, $d = -1$ with $P = \frac{1}{2}$; $\quad a = -1$, $d = 1$ with $P = \frac{1}{2}$.

Here there is parameter independence, because the probability distribution for both Alice observables, a and c, is $p = 1/2$ for every possible value, 1 or -1, independently of what Bob measures and similarly for the Bob observables b and d. However, there is no outcome independence; that is, the result obtained by Bob is correlated with (depends on) the result obtained by Alice. It is easy to see that the example again violates the Bell inequality (61), but there is no possibility to send information from Alice to Bob by means of the said

measurements. On the other hand this example is not compatible with quantum mechanics which would require that the left side of the inequality (61) should not surpass the value $2\sqrt{2}$ (Cirel'son bound, see chapter 2 section 2.5). But there are many examples of quantum mechanical predictions which violate outcome independence.

3.3.5. Attitudes in the presence of Bell inequality violations. After 50 years of experimental and theoretical effort, a loophole free violation of the Bell inequality has been achieved, aside from the small possibility that some as yet unknown loophole still exists. Once the empirical BI violation is confirmed, the possible attitudes may be summarized as follows:

The real world is quantum. The violation of a BI is a new confirmation of quantum mechanics, which we must accept as the correct theory of physics. Consequently, classical physics and our intuition are wrong. Then labelling quantum mechanics as nonlocal is flawed because locality should not be defined using classical reasoning, but within the quantum formalism. Many authors support this view, see for instance Griffiths [**16**].

Superluminal influences. Violation of a BI implies superluminal (faster-than-light) influences between Alice and Bob measurements.

Influence of the future on the present. For instance, assuming that the state (of entangled particles) is determined to some extent by the (future) possible detection by Alice or Bob. This fits in Einstein's view that faster-than-light influences are not possible and the distinction between future and past is an illusion (i.e. the laws of physics do not lead to an arrow of time).

Complete determinism. If the future determines the past, as well as the past determines the future, then an absolute determinism seems unavoidable. Most people dislike this possibility (is there free will in this case?) but again it seemed not too unpleasant to Einstein who famously stated that "God does not play dice" (which may be understood as a support to determinism).

Bell's definition of local realism is too restrictive. The current wisdom is that Bell's definition of local realism, eq.(62), is obvious. However, models in agreement with experiments exist not fitting in eq.(62) but which could be labeled as realistic and local. This is the view supported in this book. Indeed, a local model violating a Bell inequality is exhibited in chapter 6 section 6.6.

3.4. Tests of Bell inequalities with optical photons

The violation of Bell inequalities by the quantum predictions seems to show that either local realism or quantum mechanics is wrong (but see section 3.3.5 above, questioning this common opinion). Therefore it is considered relevant to know whether actual experiments may violate a Bell inequality and many such tests have been performed in the last 50 years. There are many reviews, e.g. the article by Brunner et al. [**17**]. In most performed experiments there are loopholes for the refutation of local realism. Only a few recent ones are reported as loophole-free [**18**], [**19**]. Thus older experiments have only a historical interest, but I will comment on some representative ones in order to stress aspects that are relevant for a realistic interpretation of quantum mechanics. I will study in particular the loopholes in the refutation of local realism, because they are not just practical imperfections of the experimental setup, as usually assumed; they derive from important physical causes as I will point out.

As said above, the first BI suitable for an experimental test was derived by CHSH [**10**] in 1969 and the authors proposed to test it using *entangled photon pairs produced in atomic cascades*, that is, photon pairs produced in the decay of an atom from a stationary state to another one via an intermediate short lived state. Typically the final atomic state has spin 0 and the intermediate state spin 1, whilst the initial state possesses either spin 0 or spin 1. In the 50 years elapsed since the CHSH proposal many photon experiments have been performed, most of them in agreement with the quantum predictions although not suitable to refute local realism.

The loopholes have been attributed to imperfections of the experimental set-up that might be corrected with improved experiments. This state of opinion is well summarized in the following sentence of John Bell [**3**] after the early atomic cascade experiments: "It is hard for me to believe that quantum mechanics works so nicely for inefficient practical set-ups and is yet going to fail badly when sufficient refinements are made." As a consequence the physical origin of the loopholes has been scarcely studied, something that I have done, in collaboration with Trevor W. Marshall from Manchester and other colleagues, in several papers that will be summarized in the following.

3.4.1. Atomic cascade experiments. The first experimental test of a Bell inequality involving optical photons was performed by Freedman and Clauser (FC) in 1972 [**20**]. The experiment will be

discussed here in some detail because it is a prototype of tests involving entangled optical photon pairs. (Optical means that they have frequencies either in the visible or in near ultraviolet or infrared).

FC used photon pairs produced in the decay of excited calcium atoms via a 0-1-0 cascade. That is, the initial and final atomic states had 0 total angular momentum and the intermediate state a unit (i.e. $\hbar$) angular momentum. Hence the two emitted photons were entangled in polarization, that is, the two-photon state might be represented, ignoring other variables like momenta, by

$$| \psi\rangle = \frac{1}{\sqrt{2}} \left(| V(a)\rangle \, | H(b)\rangle - | H(a)\rangle \, | V(b)\rangle \right), \tag{64}$$

where $| V\rangle \, (| H\rangle)$ corresponds to a photon state with vertical (respectively horizontal) polarization along a given direction (say Z) and a, b label the different frequencies and/or momenta of the two photons.

The most relevant quantity to be measured is the expectation of the operator $\hat{\phi}_1(a)\,\hat{\phi}_2(b)$, where $\hat{\phi}_1(a)$ corresponds to the observable taking the value 1(0) if the photon a is found (is not found) by Alice with polarization along a plane at an angle ϕ_1 with the vertical. And similarly for $\hat{\phi}_2(b)$. It may be shown that for the state eq.(64) the expectation value of the operator depends only on the difference of angles, so that it may be written in the form

$$\begin{aligned} \left\langle \psi \left| \widehat{V}(a)\,\hat{\phi}(b) \right| \psi \right\rangle &= \frac{1}{2} \left\langle V(a) \left| \widehat{V}(a) \right| V(a) \right\rangle \left\langle H(b) \left| \hat{\phi}(b) \right| H(b) \right\rangle \\ &= \frac{1}{2} \sin^2 \phi = \frac{1}{4} \left[1 - \cos(2\phi) \right], \end{aligned} \tag{65}$$

where $\phi \equiv \phi_1 - \phi_2$ and I have taken into account that

$$\left\langle V(a) \left| \widehat{V}(a) \right| V(a) \right\rangle = 1, \qquad \left\langle H(b) \left| \hat{\phi}(b) \right| H(b) \right\rangle = \sin^2 \phi,$$

which corresponds to Malus law of classical optics. It is a simple matter to get the corresponding expectations when there is only one polarizer in place or none, and also the expectation for a single photon. With the appropriate corrections for imperfect apparatuses the quantum predictions are summarized below, eq.(68).

In the FC experiment [**20**] a beam of (vaporized) calcium atoms, say moving in the direction Z, and a laser in a perpendicular direction, say Y, cross each other in a small region which becomes the source of the photon pairs. The laser excites the calcium atoms from the $4s^2\ {}^1S_0$ ground state to a $4p^2\ {}^1S_0$ state (via a short lived $3d4p\ {}^1P_1$ state) whence it decays in a short time to a $4s4p\ {}^1P_1$ which in about 5ns

decays to the ground sate, thus emitting two entangled photons. If the two photons of a pair happen to fly in opposite directions (say along the X axis), they are collected by appropriate lens systems and each photon eventually crosses a polarizer and a colour filter, finally arriving at a detector. Each polarizer may be rotated and its position is determined by an angle which I will label ϕ_j. The colour filters make that Alice may detect only 'green' photons and Bob only 'violet' photons (these colours correspond to the actual frequencies of the photons in the experiment).

The dichotomic observables measured were detection or non-detection of every photon, so that the appropriate Bell inequality would be eq.(32), which for variables with values $\{0,1\}$ may be rewritten in terms of probabilities as follows

$$p\left(a_1\right)+p\left(b_2\right)\geq p(a_1,b_1)+p\left(a_2,b_1\right)+p\left(a_1,b_2\right)-p\left(a_2,b_2\right), \tag{66}$$

where $a(b)$ stands for Alice (Bob) and the subindices 1 and 2 label two possible positions of each polarizer. The left side of eq.(66) corresponds to single counts by either Alice or Bob, respectively, and the right side to coincidence counts got by both simultaneously (more properly within a short time window). In practice the probability, p, is approximated as the ratio between the measured rate, R and the production rate R_0; that is, $p=R/R_0$. Actually, the knowledge of R_0 is not needed for the test of eq.(66) because all terms are proportional to R_0. The rate of single counts suffers from a substantial uncertainty due to the possibility of spurious counts. As a consequence only coincidence counts may be accurately measured and this poses a difficulty for the test of the Bell inequality eq.(66).

The predictions of quantum mechanics for the experiment may be summarized as follows, with some simplifications for the sake of clarity. The measurable quantities in the experiment are the single rates, R_a and R_b, and the coincidence rates $R_{ab}(\phi_{ai},\phi_{bj})$, $R_{ab}(\phi_{ai},\infty)$, $R_{ab}(\infty,\phi_{bj})$, $R_{ab}(\infty,\infty)$, where ϕ_{ai} and ϕ_{bj} are the angles, say with respect to the vertical, of the polarizer's planes of Alice and Bob, respectively, and ∞ denotes that the polarizer has been removed. The quantum predictions may be written in terms of the production rate, R_0, the solid angle covered by the apertures, $\Omega=2\pi\left(1-\cos\theta\right)$, and the quantities, $\zeta,\varepsilon_M,\varepsilon_m$, which may be measured in auxiliary experiments involving macroscopic light. I label ζ the quantum efficiency of the detector (i.e. the probability that a photon arriving at it is detected) and $\varepsilon_M\left(\varepsilon_m\right)$ is the transmission of the polarizer for light

polarized parallel (perpendicular) to the polarizer axis. The polarization analyzers used by FC were of the piles of plates type, with the following values of their relevant parameters: $\varepsilon_M \simeq 0.96, \varepsilon_m \simeq 0.04$. Actually, the values of ζ, ε_M and ε_m are different for Alice and Bob, but the small difference is irrelevant for our discussion and I will take them as equal. In order to simplify the expressions that follow it is convenient to introduce the new parameters

$$\eta = \frac{\Omega}{4\pi}\zeta, \qquad V = \left(\frac{\varepsilon_M - \varepsilon_m}{\varepsilon_M + \varepsilon_m}\right)^2 \tag{67}$$

where $\eta \ll 1$ is the overall detection efficiency of a photon, which includes collection efficiency and quantum efficiency of the detectors (for simplicity we put the same efficiency η for the green and the violet photons, which is approximately true in practice).

The small value of the collection efficiency derives from the fact that only a fraction ($\Omega/4\pi \simeq 0.07$ in the discussed experiment) of the emitted photons fly in directions appropriate to enter the apertures placed in front of the lens systems, whence a small fraction of the emissions ($(\Omega/4\pi)^2 \simeq 0.005$) are such that both photons of the pair enter. I have taken into account that in the discussed experiment the photon pair is in a state such that once the linear momentum of one photon is fixed all momenta are equally probable for the partner photon, with good enough approximation [**20**]. The parameter $V < 1$, usually called visibility or contrast, takes into account the departure from ideality of the polarizers (ideal polarizers should have $\varepsilon_M = 1$, $\varepsilon_m = 0 \Rightarrow V = 1$).

The quantum predictions are

$$\begin{aligned} R_a(\phi_i) = R_b(\phi_j) &= \frac{1}{2}R_0\eta\left(\varepsilon_M + \varepsilon_m\right), \\ R_{ab}(\infty, \phi_j) = R_{ab}(\phi_i, \infty) &= \frac{1}{2}R_0\eta^2\left(\varepsilon_M + \varepsilon_m\right), \\ R_{ab}(\infty, \infty) &= R_0\eta^2, \\ R_{ab}(\phi_{ai}, \phi_{bj}) &= \frac{1}{4}R_0\eta^2\left(\varepsilon_M + \varepsilon_m\right)^2\left(1 - F(\theta)V\cos\left(2\phi_{ij}\right)\right). \end{aligned} \tag{68}$$

where $\phi_{ij} = \phi_{ai} - \phi_{bj}$. The quantity $F(\theta)$ takes into account the decrease in the polarization correlation when the photons do not travel with opposite momenta. It may be realized in eq.(68) that the quantum prediction for the coincidence rate with both polarizers in place depends only on the difference of angles, but rates with a single or no

polarizer in place do not depend on the polarizer's position. Actually, the parameters $F(\theta)$ and $(\varepsilon_M + \varepsilon_m)$ may be approximated by 1 in the discussed experiment.

All quantum predictions eq.(68) were confirmed in the experiment, which was certainly a triumph of quantum mechanics in a case where some doubt might have existed about its validity in view of the possible contradiction with local realism. However, the experiment was unable to show the violation of a genuine Bell inequality, as explained in the following. Identifying the probabilities in eq.(66) with the ratios between detection rates and production rate, the quantum predictions eq.(68) inserted into the Bell inequality (66) give

$$\begin{aligned} 1 \quad \geq \quad & r \equiv \frac{1}{2} + \frac{1}{4}\eta F(\varepsilon_M + \varepsilon_m)V \\ & \times \left[\cos\left(2\phi_{11}\right) + \cos\left(2\phi_{21}\right) + \cos\left(2\phi_{12}\right) - \cos\left(2\phi_{22}\right)\right]. \end{aligned}$$

Taking into account that $\phi_{ij} = \phi_{ai} - \phi_{bj}$, it is a simple mathematical exercise to prove that the maximum of the right side corresponds to a choice of angles such that

$$|\phi_{11}| = |\phi_{12}| = |\phi_{21}| = \pi/8, |\phi_{22}| = 3\pi/8. \tag{69}$$

For instance, Alice might choose $\phi_1 = 0$, $\phi_2 = \pi/4$, Bob choosing $\phi_1 = \pi/8$, $\phi_2 = -\pi/8$. Thus the ratio, r, between the right side and the left side of the Bell inequality (66) becomes

$$r \leq \frac{1+\sqrt{2}}{2}\eta F(\varepsilon_M + \varepsilon_m)V \approx 10^{-4}, \tag{70}$$

so that the inequality $r \leq 1$ is very well fulfilled. The conclusion is that the experiment was not a reliable test of quantum mechanics vs. local realism. Indeed, the experiment is compatible with the former, but also with the latter.

In order to provide a quantitative measure of the violation of local realism, FC [**20**] tested a (pseudo)-Bell inequality derived as follows. They introduced a hypothesis named *no-enhancement* (actually proposed by CHSH [**10**]), assumed plausible, which may be stated in several more or less equivalent forms. For instance, the following: "all photons incident in a detector have a probability of detection that is independent on whether or not the photon has passed through a polarizer", used for the derivation of the inequality tested in the first atomic-cascade experiment [**20**]. Or, "if a pair of photons emerges

from the polarizers the probability of their joint detection is independent of the angles ϕ_1 and ϕ_2", used in the first experimental proposal [**10**]. The assumption may be stated in the form of an inequality, namely

$$P_a(\lambda, \phi_{ai}) \leq P_a(\lambda, \infty), P_b(\lambda, \phi_{bj}) \leq P_b(\lambda, \infty), \tag{71}$$

meaning that the detection probability of a photon cannot increase by the fact that it crosses a polarizer. Of course, the inequalities (71) are fulfilled in the average (that is, summing over all values of λ), but the FC assumption was that the inequality holds true for *every photon*, that is, for every value of the hidden variable λ, something which cannot be tested. From (71) the following (pseudo-)Bell inequality may be derived

$$\begin{aligned} p(\infty, \infty) &\geq p(a_1, \infty) + p(\infty, b_2) \\ &\geq p(a_1, b_1) + p(a_2, b_1) + p(a_1, b_2) - p(a_2, b_2), \end{aligned} \tag{72}$$

where again the simbol ∞ means that the corresponding polarizer has been removed. The maximal violation of this inequality by the quantum predictions eqs.(68) occurs also for the angles eq.(69), and for these angles eq.(72) implies

$$\delta \equiv \frac{|R(\pi/8) - R(3\pi/8)|}{R(\infty, \infty)} \leq \frac{1}{4}, \tag{73}$$

which is the so-called Freedman inequality. Here $R(\pi/8)$ and $R(3\pi/8)$ mean $R_{ab}(\phi_{ai}, \phi_{bj})$ with $|\phi_{ai} - \phi_{bj}| = \pi/8$ and $|\phi_{ai} - \phi_{bj}| = 3\pi/8$, respectively. In the discussed experiment the measured value of the Freedman parameter was $\delta = 0.300 \pm 0.008$, in agreement with the quantum prediction $\delta = 0.301 \pm 0.007$ but violating the Freedman (pseudo)-Bell inequality (73) by more than 6 standard deviations.

Several experiments similar to the Freedman-Clauser one were performed in the decade of 1970, that agreed with quantum predictions, with one exception [**21**], but really they did not test local realism. I will comment only on the experiments performed at Orsay (near Paris) by Aspect et al. in 1980-82, that are considered the best in the class. These authors made three atomic-cascade experiments, in the third one achieving for the first time measurements, by Alice and Bob, at space-like separation in the sense of relativity theory.

The first experiment [**22**] was similar to the FC one [**20**] with improved source and polarizers which allowed a much better statistics. However, the price paid for the high intensity of the source was

a very high number of accidental coincidences. The rate of accidental coincidences was subtracted. That is, the values of $R_{ab}(\phi_{ai}, \phi_{bj})$ and $R_{ab}(\infty, \infty)$ used in the calculation of the parameter δ were not the row empirical data, but the difference between those and the accidental coincidences estimated from the single rates. That is, $\mathrm{R}_{ab}\,(\text{accidental}) = R_a R_b$. The procedure has been criticized as inappropriate for a test of the whole family of LHV theories, because it excludes LHV models where a part of the alleged accidental coincidence are not accidental, but might be determined by the hidden variables. The rate of accidental coincidences had not been so relevant in previous experiments, in particular FC [**20**].

The second experiment presented a novelty, namely the use of two-channel polarizers, that has become standard in later experiments. That is, a photon arriving at a polarizer was either transmitted or reflected and in both cases the photon was sent to a detector. Thus four coincidence rates, R_{++}, R_{+-}, R_{-+} and R_{--}, could be detected, where the subindex + (-) means detection in the transmission (reflection) outgoing channel of the polarizer, the first (second) subindex corresponding to Alice (Bob). In the discussed experiment a photon produced in the source may be either detected in the transmission channel, or detected in the reflection channel or not detected at all. The latter may happen because the photon either does not enter the corresponding aperture or is not detected in spite of arriving at a detector. The relevant point is that now the observable A is not dichotomic, it has three possible values rather than two. Consequently the CHSH inequality does not apply. The authors introduced the *fair sampling assumption* that the detected photons are a sample representative of the whole set of photons, that is, including those not detected. This assumption is criticized in the next subsection.

The third experiment [**23**] was similar to the first one, but with improvements that achieved for the first time measurements at spacelike separation. There were two polarizers for Alice, with different orientations, and another two for Bob, also with different orientations. Every photon entering one of the lens system was sent, at random, to one of the polarizers by means of an optical switch. The arrangement was equivalent to one in which a single polarizer on each side is switched quickly between two orientations. The separation between the switches of Alice and Bob was $L = 12m$. Switching between the two channels occurred about once every 10 ns and, since this time interval, as well as the lifetime of the intermediate state of the cascade

(5 ns), was small compared with L/c (40 ns), the detection event by Alice and the change of orientation of Bob's polarizer were separated by a space-like interval.

The experimental results agreed with the quantum predictions and a (pseudo)-Bell inequality was violated by about 5 standard deviations. (The specific inequality appropriate for the experiment was somewhat involved and will not be reported here). This experiment is currently considered to have closed the locality loophole. Hence, modulo the fair sampling assumption, the third Aspect experiment has been quoted as a definite disproof of local realism, until the more conclusive loophole-free experiments to be discussed below.

3.4.2. The loopholes. In atomic-cascade experiments there are loopholes for the violation of Bell inequalities that usually were dismissed as irrelevant non-idealities of the experimental set-up. However, in my view they were relevant supports for a wave, rather than particle, nature of photons. Indeed, although the most popular interpretations of quantum mechanics renounce any intuitive picture of the microworld as mentioned in chapter 2 section 2.3, it is the case that many quantum physicists have in their minds a picture of the photon correlation experiments as follows. Photons are small (maybe pointlike) particles that travel from the source to the polarization analyzers. Of course, photons behave in some cases like waves, this being one of the big mysteries of quantum mechanics. If a polarization analyzer is inserted in the way of a photon then there is some probability, P_1, that the photon crosses the analyzer. There is also a probability, P_2, that a photon arriving at the detector is detected, this probability being the same for all photons (it is assumed, as in quantum statistical mechanics, that photons with the same frequency are identical). In addition there is some probability per unit time that spurious counts may be recorded due to thermal fluctuations in the detectors (this being named dark rate). If this picture is accepted then the following assumptions are plausible in the interpretation of a Bell experiment: 1) *no enhancement*, that is, the detection probability cannot increase when a polarizer is inserted because $P_1 P_2 \leq P_2$; 2) *fair sampling*, namely we may extrapolate the measured detection efficiency to unity because the fact that $P_2 < 1$ may be attributed to technical defects of the photon counters; and 3) it is appropriate *to subtract the background* of spurious counts via an estimate of the dark rate, as in [**22**], discussed above. In my view this picture, not explicit,

has been the reason why these three assumptions have been believed as plausible.

The interpretation supported in this book is quite different. We assume that the electromagnetic radiation is purely wavelike and there is a vacuum (zero-point) radiation filling the whole space. Although a more deep discussion will be provided in chapter 6, we may anticipate that 'photons' produced in atomic decays are long narrow wavepackets (needles of radiation in the words of Einstein) superposed to a background random radiation corresponding to the vacuum fluctuations of quantum electrodynamics (the zero-point field, ZPF). In this picture local realism was not violated in atomic-cascade experiments but the alleged plausible assumptions are violated instead. Actually, that violation is plausible as we discuss in the following.

No enhancement. An essential hypothesis in our interpretation is that vacuum fields are real. Then in a polarizer with two incoming channels the signal enters one of the channels and radiation from the vacuum field enters the other channel. The amplitudes of both signal and vacuum field may interfere constructively, thus making it possible that the relevant intensity in one of the outgoing channels is greater than the intensity of the incoming signal, and plausibly enhancing the detection probability. This point will be discussed in more detail in chapter 6, devoted to the interpretation of quantum optics. Our conclusion is that what the FC experiment refuted was the no-enhancement hypothesis, rather than local realism. Therefore the experimental results reinforce our realistic interpretation of quantum mechanics.

Low detection efficiency. For macroscopic light the most frequent detection method is to measure the energy transferred to the detector. This is the case for instance for the radiation arriving in Earth from stars. However, the experiments aimed at testing Bell inequalities require photon counting. Counters are devices providing a yes-no response and, as is typical in that kind of test, there is the possibility of false positive or false negative answers. Furthermore, if we attempt to minimize the false negative we should increase the sensitivity, but this would increase the probability of false positives. Thus in photon counting there is a trade-off between high efficiency and dark rate, that may be associated to quantum fluctuations. Indeed, the detector may record a count either due to the arrival of a signal or the arrival of an eventual intense vacuum fluctuation. Of course, if the temperature is not low the fluctuations are higher. As a consequence either the

detector is very efficient and detects a high spurious rate or it has low efficiency with a low dark rate. This provides the physical reason for the difficulty to manufacture photon counters with high efficiency and very low dark rate. In contrast, radiation detectors not intended to *count* photons do not have such problem, they may be 100% efficient. Indeed, the average of the zero-point fluctuations would go to zero in a large time interval, so that the time average may effectively remove the fluctuations in photodetectors not attempting to register individual counts.

Up to here we have considered optical photons. High energy photons, say gamma rays, have a wavelength smaller than the atomic size and an energy far greater than typical atomic energies. As a result the interaction of one photon with bulk matter takes place via a single atom, which gives rise to energetic electrons (and maybe positrons) which eventually transfer the energy to many other atoms. The consequence is that high energy photons may be detected with good efficiency. However, the measurement of their polarization should be made indirectly and involves a large uncertainty. Thus no experiment with pairs of gamma rays (e.g. produced in the positronium decay) might show a true violation of a Bell inequality. In contrast, optical photons have a size much bigger than atoms and an energy of order the atomic transitions, so that the detection may not be too efficient, but the polarization can be measured with macroscopic devices (polarization analyzers).

The fair sampling assumption. Fair sampling is the assumption that we may extrapolate the empirical results to what we believe that would be obtained if all experimental devices were ideal. For instance, if our detectors have efficiency $\eta < 1$ we should multiply all single rates by $1/\eta$ and coincidence rates times $1/\eta^2$. The interpretation of an experiment using the fair sampling assumption may test quantum mechanics but not local realism. In fact if the empirical results are extrapolated via the fair sampling assumption and they agree with quantum predictions (as has been the case with few exceptions), this gives a new confirmation of quantum mechanics. If they do not agree, then quantum mechanics would be refuted. However, the fair sampling assumption may not be valid for some local models, whence those models may be compatible with the results of the experiment. For instance, the wave picture implies that ideal photon-counters might not be possible (i.e. with 100% efficiency and nil dark rate) as discussed above.

Decrease of entanglement with lower angular correlation. Atomic cascade experiments are unreliable as tests of local realism, as is proved by the existence of local (hidden variables) models of the experiments [**24, 25**]. Indeed, the predictions of quantum mechanics for atomic-cascade experiments are always compatible with local realism even if the experiment is made with ideal set-up, in particular 100% efficiency detectors, as is shown in the following.

The atomic cascade decay, giving rise to a photon pair, is a three-body problem with the consequence that the angle, χ, between the directions of emission of the two photons is almost uniformly distributed over the sphere. The polarization correlation of the photons in the entangled pair decreases with the angle χ and hence the parameter $F(\theta)$ (see (68)) decreases with the angle θ, which determines de solid angle Ω covered by the apertures. In particular, in the experiments involving 0-1-0 cascades, e.g. those of Aspect et al., we have

$$\frac{\Omega}{4\pi} = \frac{1}{2}(1-\cos\theta), \qquad F(\theta) = 1 - \frac{2}{3}(1-\cos\theta)^2. \tag{74}$$

A necessary condition for the violation of a (genuine) Bell inequality by the quantum prediction is the inequality (see eq.(70))

$$\frac{1+\sqrt{2}}{2}\eta\frac{\Omega}{4\pi}F(\theta)V = \frac{1+\sqrt{2}}{4}\eta(1-\cos\theta)\left[1-\frac{2}{3}(1-\cos\theta)^2\right]V > 1,$$

where eq.(74) has been taken into account. But this inequality is never true, the maximun value of the left side being $(2+\sqrt{2})/12 \simeq 0.28$, where I have taken into account that both V and η are smaller than unity, reaching unity in the ideal case. In summary, taking into account the low angular correlation of the photon pairs produced in atomic cascades, *these experiments cannot discriminate between local realism and quantum mechanics.* The problem might be solved if the recoil atom were detected [**26**], [**27**] but such experiments have not been attempted.

3.4.3. Parametric down converted photons.

The advantage of photons not maximally entangled. A source of entangled optical photon pairs more convenient than atomic cascades is the process of spontaneous parametric down-conversion (SPDC) and it has been used in optical photon experiments since about 1984. Unlike atomic-cascade experiments, here the momenta of the two photons of the entangled pair are fairly well defined. Thus the predictions

of quantum mechanics are similar to either eqs.(68) or (66) with unity substituted for both Ω/π and $F(\theta)$, whence a Bell inequality might be violated if V is close to 1 (which is achievable in actual experiments) provided that $\eta > 2\left(\sqrt{2}-1\right) \simeq 0.82$. Such a high value of the detection efficiency is difficult to obtain and the low efficiency of detectors has been a persistent loophole for the disproof of local realism during more than 40 years.

The problem of the low efficiency of photon counters has been alleviated by an interesting observation made by Eberhard [**28**]. In 1993 he realized that an efficiency $\eta > 2/3$ is enough to test Bell inequalities in some two-photon states not maximally entangled. A state of this type is

$$(75) \qquad |\,\psi\rangle = \frac{1}{\sqrt{1+r^2}}(|\,V\rangle_A \,|\, H\rangle_B + r \,|\, H\rangle_A \,|\, V\rangle_B), \qquad r \neq \pm 1,$$

to be compared with the maximally entangled state eq.(64) . The state eq.(75), with $r = -2.9$, has been used in the experiment by Giustina et al. [**19**], who tested the Eberhard inequality

$$(76) \qquad p_{++}(a_1 b_1) - p_{+0}(a_1 b_2) - p_{0+}(a_2 b_1) - p_{++}(a_2 b_2) \leq 0.$$

Here a (b) refers to Alice (Bob), 1 and 2 to two angles of the polarizers and the subindex + (0) means detection (not detection) in the corresponding photon counter. This inequality is a trivial consequence of the CH inequality, eq.(66), taking into account that

$$p_{+0}(a_1 b_2) = p_{+}(a_1) - p_{++}(a_1 b_2).$$

In the experiment by Giustina et al. detectors were used with efficiency about 77%. With appropriate choice of polarizer's angles the reported measured value of the left side of eq.(76) was 7.27×10^{-6} whilst under local realism the probability of observing that value did not exceed a p-value of 3.74×10^{-31}.

Memory loopholes. Need of random choice for Alice and Bob settings. The memory loophole derives from the fact that tests of the Bell inequalities require performing many similar experiments on different entangled pairs of particles in order to estimate the probabilities from the frequencies actually measured. In practice there is a series of trials with a given experimental set-up, the trials being separated from each other by short time intervals. It is usually assumed that any hidden variables associated with the nth particle pair would be independent of measurement choices and outcomes for the first $(n-1)$ pairs. Models which violate this assumption exploit the possible memory effects.

The strongest type of violation uses a 2-sided memory loophole, in which the hidden variables for pair n may depend on the previous measurement choices and outcomes in both wings of the experiment.

The memory loophole has been studied by Barrett et al. [**29**]. The authors concluded that, although in principle the memory loophole might imply a flaw in existing analyses of Bell experiments, the data still may refute local realistic models. Here I shall follow the current analysis of the experiments, but this analysis will be questioned in chapter 6. In actual experiments the photon pairs are produced via SPDC in a nonlinear crystal. Hence, when neither Alice nor Bob detect a photon in a trial, they cannot know whether two photons arrived to the parties but none was detected or no photon was produced in the source in that trial. In fact, the second case is by far more probable because only a small fraction of the laser pulses incident on the nonlinear crystal produce photon pairs. Thus in the photon experiments we cannot determine true probabilities but only relative probabilities. The question arises whether the memory loophole is also irrelevant in these conditions. In the following I give a simple proof that the answer is affirmative [**29**].

The division of the laser beam in pulses makes an experiment consist of a series of trials. During each trial Alice and Bob randomly choose between one of two measurement settings, denoted a and a' for Alice and b and b' for Bob, and record either a '+' if they observe a detection event or a '0' otherwise. The details of that type of Bell test may be seen, for instance, in the experiments by Shalm et al. [**18**] or by Giustina et al. [**19**]. Alice may get two possible results in each one of the two possible measurements of her photon, and the same for Bob. Therefore there are 16 coincidence probabilities that might be determined. In practice the Bell test requires just 4 that may be combined in the CH/Eberhard inequality [**28**], namely

$$0 \leq B \equiv P\left(0+ \mid ab'\right) + P\left(+0 \mid a'b\right) + P\left(++ \mid a'b'\right) - P\left(++ \mid ab\right). \tag{77}$$

The terms $P(++ \mid ab)$ and $P(++ \mid a'b')$ correspond to the probability that both Alice and Bob record detection events $(++)$ when they choose the measurement settings ab or $a'b'$, respectively. Similarly, the terms $P(+0 \mid ab')$ and $P(0+ \mid a'b)$ are the probabilities that only Alice or Bob record an event for settings ab' and $a'b$, respectively.

The joint probability $P(x, y; a, b)$ that the photons yield the outcomes x and y when subjected to the measurement with the settings

a and b respectively is given, according to Bell [**2**], by

$$P(x,y;a,b)=\int d\lambda\rho(\lambda)P(x;a,\lambda)P(y;b,\lambda), \tag{78}$$

where, in our notation, x, y might be either $+$ or 0. Eq.(78) may be taken as the definition of local hidden variables models. Now let us assume that there are memory effects such that the hidden variables, λ, and the action of the measuring devices in the trial n depend on all previous trials, $j=1,2,\ldots,n-1$. This means that we must substitute the following

$$P_n(x,y;a,b)=\int d\lambda_n\rho_n(\lambda_n)P_n(x;a,\lambda_n)P_n(y;b,\lambda_n), \tag{79}$$

for eq.(78) . That is, all possible memory effects may just change the probability distribution of the hidden variables, λ, and the probability of outcome in the measurement for given settings, a and b. If the probabilities of the 4 settings $ab, ab', a'b, a'b'$ are $1/4$ each, as insured by the random choice, then in eq.(77) we get the CH-Eberhard inequality

$$\begin{aligned} 0\leq B_n \equiv & P_n\left(0+\mid ab'\right)+P_n\left(+0\mid a'b\right) \\ & +P_n\left(++\mid a'b'\right)-P_n\left(++\mid ab\right). \end{aligned} \tag{80}$$

This happens if an entangled pair of photons is produced in the source in the n-th trial. But if no photons are produced then eq.(77) gives

$$B_n=0.$$

In any case, for a set of trials we shall have

$$\sum_j B_j\geq 0,$$

where $\{j\}$ represents the trials chosen for the test of the CH-Eberhard inequality. In practice it is common to choose, within some time interval, all trials such that at least one photon is detected (either by Alice or by Bob).

I emphasize that *the absence of bias in the random choice of the settings is essential for the proof.* For instance, the local model

$$P(+;a,\lambda)=P(+;b,\lambda)=P(+;a',\lambda)=1, P(+;b',\lambda)=0,$$

gives $B=0$, taking eqs.(78) and (77) into account, for unbiased random choices. However, if the choices a and b have probabilities

$(1+\varepsilon)/2$ each and the choices a' and b' probabilities $(1-\varepsilon)/2$, then we get $B=-\varepsilon$ in apparent violation of eq.(77) .

We conclude that for photon tests involving the CH-Eberhard inequality the possible memory loophole is irrelevant provided that the test consists of a large enough number of trials. For a small number there may be fluctuations that could give a wrong answer. The study of fluctuations will not be made here, but it would be similar to the study made by Barrett et al. [**29**] in relation with the CHSH inequality. I point out that the analysis made refers to the loopholes in the violation of a Bell inequality. However, in chapter 6 we will question that the inequalities are necessary conditions for local realism.

3.5. Tests not involving photon pairs

Many tests of Bell's inequalities have been performed not using photon pairs. In the following I will comment only on experiments involving atoms, electrons and hadrons.

3.5.1. Tests with entangled atomic states. In experiments with atoms detection may be very efficient and there is a property corresponding to the polarization of photons, namely a linear combination of different atomic states. A typical Bell experiment with atoms has been performed by Rowe et al. [**30**]. In it the Bell inequality (66) has been violated. As a consequence *the experiment has refuted non-contextual hidden variables theories* (see section 3.3 for the proof that the violation of a Bell inequality refutes non-contextual HV theories). However, the measurements have not been made insuring spacelike separation, and therefore *local* HV theories have not been tested in it.

Another experiment [**31**] has measured the correlation between the quantum states of two Yb^+ ions separated by a distance of about 1 meter. The results of the experiment violate a CHSH [**10**] (Bell) inequality, modulo the locality loophole. The authors claim that the experiment is relevant because it closes the detection loophole. The results of the discussed experiment [**31**] do refute LHV theories if locality is understood in the sense that the measurements (by Alice and Bob) are made in *separated (distant) parts of space*. However, they did not refute LHV theories with the separation understood in the sense of relativity theory, which requires that the measurements are performed *in spacetime regions with space-like separation*.

3.5.2. Experiments with spin-1/2 particles. The spin-0 state of a pair of spin-1/2 particles was introduced by Bohm [**32**] as a substitute for the position-momentum entanglement in the EPR paper (see chapter 2 section 2.4.5). It has become the standard example of entanglement in textbooks. However, tests of Bell inequalities using spin entangled particles have proved to be difficult [**33**].

An experiment measuring correlations of electron spins has been performed in Delft [**34**]. Electrons of nitrogen vacancies (NV) of diamond placed at a long distance (more than 1 km) were measured. The setup comprised three separate laboratories, A, B and C. The boxes at location A and B each contain a single NV centre electron spin. A quantum random number generator was used to provide the input to the box. The spin is read out in a basis that depends on the input bit and the resulting signal provides the output of the box. A third box at location C records the arrival of single photons that were previously emitted by, and are entangled with, the spins at A and B. The detection of two such photons constitutes an event-ready signal. Successful preparation of the spins is signalled by a specific coincidence detection pattern. Independent of the event-ready signal, the setups at location A and B choose a random basis, rotate the spin accordingly, and start the optical spin-readout.

The experiment was claimed to be loophole-free but the raw data indicate violation of the no-signalling principle [**35**]. Violation of this principle would be a groundbreaking revolution in physics (see section 3.3.4 for comments about this principle). Therefore a careful reproduction and analysis of the experiment should be worthwhile.

3.5.3. Entanglement and Bell inequalities in kaon systems. The Bell experiments discussed till now involve electromagnetic interactions, and it is interesting to study entanglement and possible tests of Bell inequalities with systems governed by strong and weak interactions. Several papers have been devoted to study entanglement and possible tests with $K^o - \overline{K}^o$ or $B^o - \overline{B}^o$ systems [**36**]. In this section we shall be concerned with the $K^o - \overline{K}^o$ system produced in the decay of a ϕ-meson resonance. Most of the following may be applied also to other neutral meson-antimeson systems like $B^o - \overline{B}^o$ and ${B_s}^o - \overline{B}_s{}^o$.

Quantum-mechanical predictions for the $K^o - \overline{K}^o$ system. With some simplification (i.e. neglecting a small CP violation) we may write the entangled state of two kaons in terms of four possible quantum

states:

$$|\, K_j\rangle \in \left\{|\, K^o\rangle, |\, \overline{K}^o\rangle, |\, K_S\rangle, |\, K_L\rangle\right\}, \tag{81}$$

$K^o\rangle, |\, \overline{K}^o\rangle$ being eigenstates of strangeness and $|\, K_S\rangle, |\, K_L\rangle$ eigenstates of CP. These states are related by

$$|\, K_S\rangle = \frac{1}{\sqrt{2}}\left(|\, K^o\rangle + |\, \overline{K}^o\rangle\right), \qquad |\, K_L\rangle = \frac{1}{\sqrt{2}}\left(|\, K^o\rangle - |\, \overline{K}^o\rangle\right).$$

If a kaon pair is produced by the decay of a ϕ-meson resonance, the state after the decay is a maximally entangled state, that is,

$$\begin{aligned} |\, \phi\,(0)\rangle &= \frac{1}{\sqrt{2}}\left(|\, K^o\rangle|\, \overline{K}^o\rangle - |\, \overline{K}^o\rangle\, |\, K^o\rangle\right) \\ &= \frac{1}{\sqrt{2}}\left(|\, K_L\rangle\, |\, K_S\rangle - |\, K_S\rangle\, |\, K_L\rangle\right), \end{aligned}$$

with an obvious notation.

In the rest frame of the ϕ-meson the two neutral K-mesons travel in opposite directions, say one of them going to Alice and the other one to Bob. If the kaons move in the vacuum then all quantum predictions reduce to 16 coincidence probabilities that Alice's kaon is in one of the 4 states (81) at time t_a and Bob's kaon in another of these 4 states at time t_b. As an example I give just the following

$$\begin{aligned} Q\left(K^o, K^o\right) &= Q\left(\overline{K}^o, \overline{K}^o\right) \\ &= \frac{1}{8}\left(E_L\left(t_a\right) E_S\left(t_b\right) + E_S\left(t_a\right) E_L\left(t_b\right)\right)\left[1 - \frac{\cos\left(\Delta m \tau\right)}{\cosh(\Gamma\tau)}\right], \\ Q(K^o, \overline{K}^o) &= Q\left(\overline{K}^o, K^o\right) \\ &= \frac{1}{8}\left(E_L\left(t_a\right) E_S\left(t_b\right) + E_S\left(t_a\right) E_L\left(t_b\right)\right)\left[1 + \frac{\cos\left(\Delta m \tau\right)}{\cosh(\Gamma\tau)}\right]. \end{aligned}$$

where $\tau \equiv t_a - t_b$, and the decay rates are $E_L\left(t\right) \equiv \exp\left(-\Gamma_L t\right)$, $E_S\left(t\right) \equiv \exp\left(-\Gamma_S t\right)$, $\Gamma \equiv \left(\Gamma_S - \Gamma_L\right)/2$. In the following we shall use units such that $\hbar = c = \Delta m = 1$, Δm being the mass difference $K_L - K_S$. In these units $\Gamma_S = 2.105$ and $\Gamma_L \simeq 8.21 \times 10^{-4}$.

These probabilities correspond to travel in free space but if there are slabs of nucleonic material in the way of one or both kaons, then after the kaons have crossed the slabs the probabilities are different due to absorption and rotation in isospin space. The quantum predictions may be calculated at any time for any number of slabs of arbitrary width, but the results will not be given here.

For a fixed set-up it is always possible to construct an explicit LHV model. The existence of the model proves that no experiment involving kaon pairs with a fixed set-up may violate a Bell inequality. Therefore any test of LHV theories should involve the variation of the set-up in the four measurements (the variation may be made by just changing the positions of the slabs of nucleonic material). In fact, the model is specific for a fixed set-up and therefore does not guarantee the fulfillment of the Bell inequality in a complex experiment involving measurements with different set-ups, made in different meson pairs.

Actually, Bell inequality cannot be violated by the quantum predictions in the $K^o - \overline{K}^o$ system due to the relatively high value of Γ_S in comparison with Δm, which makes the strangeness oscillations to be rapidly damped, but tests without inequalities have been proposed [**37**]. A Bell inequality test has been performed using the decay products of B^o mesons [**38**]. Again there is damping making impossible the violation of the Bell inequality, and only a normalization of the correlation function to the undecayed pair of B^o leads to a violation. In any case, the experiments are involved and its relevance for the test of local realism is scarce in comparison with the optical photon experiments mentioned above.

3.6. The arrow of time

The name 'arrow of time' was introduced by Arthur Eddington in 1927. He wrote "I shall use the phrase time's arrow to express this one-way property of time which has no analogue in space" [**39**]. Thus the arrow of time refers to the distinction between past and future that we observe in nature. At present it is used more specifically with reference to the problem of explaining the irreversibility that we experience, which is not trivial taking into account that the laws of nature are invariant under time reversal. There are many books and articles devoted to (or discussing) the arrow of time and a review is out of our scope here [**40**]. I will only discuss a few points that sometimes have been a source of confusion. A discussion about the origin of irreversibility is relevant in order to clarify the difference between relativistic locality and causality. As we pointed out in section 3.3.2 Einstein's assumption of no faster-than-light signals is weaker than Bell's 'locality', whence empirical refutation of the latter does not refute the former.

3.6.1. Thermodynamic and statistical Boltzmann explanation. The existence of an arrow of time was formalized by Clausius with the concept of entropy and its postulated increase for any spontaneous evolution of an isolated system. Actually, entropy was introduced in physics as a kind of measure of the 'quality' of energy. For instance, mechanical and gravitational energy have high quality because they may be transformed completely into other forms, but this is not the case for heat because only a part of it can be transformed into work (mechanical energy). In the particular case of energy transfer taking place exclusively in the form of heat, a simple quantitative calculation of the entropy change, ΔS, of a system is possible, namely

$$\Delta S = \int \frac{dQ}{T}, \tag{82}$$

Q being the heat entering the system and T the absolute temperature. For other cases the calculation is more involved. Clausius realized that in the processes that are possible in the laboratory the total entropy never decreases. This led to postulate that entropy never decreases in closed systems, which was the first scientific statement on the existence of an arrow of time. For instance, if we put a hot body in contact with a cold one heat goes spontaneously from the former to the latter until they have equal temperature. This gives an increase of entropy as is easily derived from eq.(82) leading to

$$\Delta S = \int \frac{dQ}{T_{cold}} - \frac{dQ}{T_{hot}},$$

which is positive taking into account that $dQ > 0$ ($dQ < 0$) is defined as energy that enters (leaves) the body and obviously $T_{hot} > T_{cold}$.

The fundamental step towards the solution of the apparent contradiction between the *irreversibility of spontaneous (macroscopic) evolution vs. reversibility of the fundamental (microscopic) laws of nature* was made by Boltzmann, who gave a microscopic interpretation of entropy. Boltzmann realized that irreversibility is always associated to macroscopic systems and he proposed that it is due to the tendency towards more probable states in the spontaneous evolution. Then Boltzmann introduced a relation between the entropy, S, of a composite system and the number N of microscopic states of the system that correspond to a given macroscopic state, that is,

$$S = k_B \log N, \tag{83}$$

where k_B is today named Boltzmann constant. A standard example is a box divided in two equal parts by a wall with a small hole on it, filled with an amount of gas consisting of n molecules. If we define a microscopic state by specifying which gas molecules are present in each part of the box, there is only one state with all molecules in the left (or in the right). In this state $N = 1$ and eq.(83) gives $S = 0$. If at time $t = 0$ the box starts in this state, after some time $t = T$ there will be several, say j, molecules on the left and $n - j$ on the right. Hence the number of microstates equals the number of ways to choose j molecules amongst n, that is,

$$N = \frac{n!}{j!(n-j)!} > 1 \Rightarrow S > 0.$$

The most probable state will correspond to $j = n/2$, whence

$$S_{\max} = k_B \log N_{\max} \simeq k_B n \log 2.$$

Boltzmann's work was one of the great achievements in the history of physics, but it did not solve the problem of the arrow of time as was soon pointed out by several authors, in particular Loschmidt and Poincaré. I think that in order to clarify the subject it is important to distinguish between the evolution of systems in experiments made in the laboratory and what happens on Earth.

In addition to the thermodynamic arrow of time people speak about an electromagnetic one. A moving charge produces outgoing (retarded) radiation in the form of spherical waves, but we never see ingoing (advanced) spherical waves. However, both are solutions of the Maxwell equations. It seems that most of the outgoing radiation goes to infinity and never returns, thus producing an asymmetry between the past and the future. A solution to the problem is the assumption of a real (zero-point) radiation filling the whole space, as we will discuss in detail in chapter 5.

3.6.2. Irreversibility on Earth and laboratory experiments. I will speak about LAB experiments in a wide sense, including processes induced by human beings like those of chemical industry. In any case, I will refer only to evolution of isolated systems because it is obvious that evolution subject to external influences may present irreversibility induced by them. In the example of the box, discussed in the previous section, the irreversibility is related to

$$S(T) > S(0).$$

The Loschmidt argument applied to this example is as follows. If the system was isolated since well before $t = 0$ then at time $t = -T$ the gas would be filling both parts of the box. In fact, the evolution backwards in time between $t = 0$ and $t = -T$ would be identical to the evolution forward in time between $t = 0$ and $t = T$ with all velocities reversed at time $t = 0$. Therefore, in terms of entropy we may write

$$S(-T) = S(T) > S(0).$$

The reversal of velocities is appropriate for classical mechanical systems consisting of particles. In quantum physics complex conjugation of the wave-function is substituted for the reversal of velocities.

Of course, nobody has ever seen an isolated box with a quantity of gas initially having a homogeneous density (say at time $t = -T$) to evolve spontaneously towards a state with all the gas concentrated in a part of the box (at time $t = 0$). This is true, but the point is that we, human beings, are able to prepare a box having gas in only one part of the box and then observe the evolution towards the future, $t = T$, but we are unable to observe towards the past, $t = -T$, the evolution of an isolated system prepared at time $t = 0$. In fact, our preparation implies that the system was not isolated for $t < 0$. We are also unable to prepare a system in such a way that after isolation evolves with decreasing entropy, as for instance a gas with equal (macroscopic) density in the two parts of the box but with velocities of molecules such that their motion carries all of them to one of the parts (making this would require a so-called 'Maxwell demon', a device that we cannot manufacture for macroscopic systems). Summing up, *the irreversibility in the LAB is not a feature of the material systems themselves, but it derives from our fundamental irreversibility as living beings.*

The conclusion is that closed (isolated) systems are actually reversible, this being a straightforward consequence of the reversibility of the fundamental laws of physics. In particular, if a system is isolated between times $-T$ and T and at time $t = 0$ it is out of equilibrium, then it would be more close to equilibrium both at time T and at time $-T$. Of course, this does not apply to the Earth as a whole or to the living beings, including humans, because they are not isolated.

Explaining the irreversibility of living beings, including humans, is rather trivial once we know that the universe is expanding. The universe may be assumed an isolated system, governed by reversible laws, but its initial state was very special. In that state it was far from

equilibrium and consequently its evolution has been irreversible. The expansion, combined with the attractive nature of gravity, caused that the initial almost homogeneous plasma evolved giving rise to galaxies and stars. The stars frequently have associated planets giving rise to solar systems. Every planet receives energy from its star, this causing irreversible evolution. Incidentally, in a stationary universe the existence of (irreversible) living beings would be impossible.

Our solar system was formed about 5 billion years ago. After some period the Earth, initially very hot, became cold arriving at an approximate stationary state with a separation of the solid crust, the sea and the atmosphere. In that cold Earth, life emerged and evolved until the appearance of human beings. The evolution in that period has been clearly irreversible and the reason is obvious. The (almost stationary) Earth is not an isolated system. Asides from minor perturbations, the main cause of irreversibility is the fact that Earth is receiving energy at high temperature ($T_{in} \simeq 5800K$) from the Sun and sending away a similar power as radiation at lower temperature ($T_{out} \simeq 300K$). This produces a net increase in entropy of the universe at a rate

$$\frac{dS}{dt} = \frac{W}{T_{out}} - \frac{W}{T_{in}} > 0,$$

where W is the average power received from the Sun or emitted by the Earth to outer space. The irreversibility of Earth is responsible for the irreversibility of the living beings, including us. That is, life in Earth is an irreversible process because living beings are interacting with the environment and entropy increases in the process [**41**].

Bibliography

[1] J. S. Bell: On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.* **38**, 447 (1966).

[2] J. S. Bell: On the Einstein, Podolsky and Rosen paradox. *Physics,* **1**, 195 (1964).

[3] J. S. Bell, *Speakable and unspeakable in quantum mechanics,* Cambridge University, 1987. This book reproduces the most relevant articles on foundations of John S. Bell.

[4] S. Kochen, E. P. Specker: The problem of hidden variables in quantum mechanics. *J. Math. Mech.* 17, 59 (1967).

[5] N. Gisin: Bell inequality holds for all non-product states, *Phys. Lett. A* **154**, 201 (1991).

[6] S. Yu, Q. Chen, C.. Zhang, C. H. Lai, C. H. Oh: All entangled pure states violate a single Bell inequality, *Phys. Rev. Lett.* **109**,120402 (2012).

[7] N. Harrigan and R. W. Spekkens: Einstein, incompleteness, and the epistemic view of quantum states. *Found. Phys.* **40**, 125-157 (2010).

[8] A. Einstein, in *Albert Einstein: Philosopher-scientist,* P. A. Schilpp, ed. Open Court, Evanston, 1949.

[9] A. Einstein in *Correspondence with Michele Besso 1903-1955*. Hermann, Paris, 1979.

[10] J. F. Clauser, M. Horne, A. Shimony and R. A. Holt: Proposed experiment to test local hidden-variables theories. *Phys. Rev. Lett.* **23**, 880-883 (1969).

[11] G. C. Ghirardi, A. Rimini, T. Weber: A general argument against superluminal transmission through the quantum mechanical measurement process, *Lett. Nuovo Cimento*, **27**, 293 (1980).

[12] G. C. Ghirardi, R. Grassi, A. Rimini, T. Weber: Experiments of the Einstein-Podolsky-Rosen type involving CP-violation do not allow faster-than-light communication between distant observers. *Europhys. Lett.* **6**, 95 (1988).

[13] R. F. Streater, A. S. Wightman: *PCT, spin and statistics, and all that*. Benjamin, New York, 1964.

[14] J. P. Jarrett: On the physical significance of the locality conditions in the Bell arguments, *Nous*, **18:** 569 (1984).

[15] A. Shimony: Controllable and uncontrollable non-locality, in *Proceedings of the International Symposium on the Foundations of Quantum Mechanics*, S. Kamefuchi et al. (eds), Physical Society of Japan, Tokyo. 1983.

[16] R. B. Griffiths: Nonlocality claims are inconsistent with Hilbert space quantum mechanics. *Phys. Rev. A* **101**, 022117 (2020).

[17] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani and S. Wehner: Bell nonlocality. *Rev. Mod. Phys.* **86**, 419478 (2014).

[18] L. K. Shalm et al.: A strong loophole-free test of local realism. *Phys. Rev. Lett.* **115**, 250402 (2015).

[19] M. Giustina et al.: A significant loophole-free test of Bell's theorem with entangled photons, *Phys. Rev. Lett.* **115**, 250401 (2015).

[20] S. J. Freedman, J. F. Clauser: Experimental test of local hidden-variable theories. *Phys. Rev. Lett.* **28**, 938 (1972).

[21] T. W. Marshall, E. Santos: Stochastic optics: a local realistic analysis of optical tests of Bell inequalities. *Phys. Rev. A* **39**, 6271-6283 (1989).

[22] A. Aspect, J. Dalibard, G. Roger: Experimental test of Bell's inequalities using time-varying analyzers. *Phys. Rev. Lett.* **49**, 1804-1807 (1982).

[23] A. Aspect, P: Grangier and G. Roger: Experimental realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A new violation of Bell's inequalities. *Phys. Rev. Lett.* **49**, 91-94 (1982).

[24] E. Santos: Does quantum mechanics violate the Bell inequalities? *Phys. Rev. Lett.* **66**, 1388-1390 , **68**, 2702-2703 (1991).

[25] E. Santos: Critical analysis of the empirical tests of local hidden-variables theories. *Phys. Rev. A* **46**, 3646 (1992).

[26] S. F. Huelga, M. Ferrero, E. Santos: Atomic-cascade experiment with detection of the recoil atom. *Europhys. Lett.* **27**, 181-186 (1994).

[27] S. F. Huelga, M. Ferrero. E. Santos: Loophole-free test of the Bell inequality. *Phys. Rev. A* **51**, 5008-5011 (1995).

[28] P. H. Eberhard: Background level and quantum efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment. *Phys. Rev. A*, **47**, R747 (1993).

[29] J. Barrett, D. Collins, L. Hardy, A. Kent, and S. Popescu. Quantum nonlocality, Bell inequalities, and the memory loophole *Phys. Rev. A*, **66**, 042111 (2002).
[30] M.A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe and D. J. Wineland. Experimental violation of a Bell's inequality with efficient detection. *Nature* **409**, 791-794 (2001).
[31] D. N. Matsukevich, P. Maunz, D. L. Moehring, S. Olmschenk, and C. Monroe: Bell inequality violation with two remote atomic qubits. *Phys. Rev. Lett.* **100**, 150404 (2008).
[32] D. Bohm: *Quantum Theory.* Prentice Hall, Englewood Cliffs, NJ.(1951).
[33] E. S. Fry, T. Walther, S. Li: Proposal for a loophole-free test of the Bell inequalities. *Phys. Rev. A* **52**, 4381 (1995).
[34] B. Hensen et al.: Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometers, *Nature* **526**, 682-686 (2015).
[35] A. Bednorz: Signaling loophole in experimental Bell tests, *Phys. Rev. A* **95**, 042118 (2017).
[36] B. C. Hiesmayr: Nonlocality and entanglement in a strange system. *Eur. Phys. J.* **C 50**, 73-79 (2007).
[37] A. Bramón, G. Garbarino: Tests of local realism with entangled kaon pairs without inequalities. *Phys. Rev. Lett.* **89,** 160401 (2002).
[38] A. Go: Observation of Bell inequality violation in B mesons. *Conference proceedings for Garda Lake workshop*, 2003.
[39] A. S. Eddington: *The Nature of the Physical World,* 1928. Reprinted University of Michigan, 1981.
[40] H. D. Zeh: *The physical basis of the direction of time.* The Frontier Collection. Springer, 2007.
[41] D. Kondepudi: I. Prigogine, *Modern thermodynamics. From heat engines to dissipative structures.* John Wiley and Sons, 1998.

CHAPTER 4

Alternative formulations: de Broglie-Bohm, stochastic mechanics, Weyl-Wigner and Feynman

4.1. Introduction

As said in chapter 1 section 1.1.3, quoting the famous EPR article, we "must take into account the distinction between *the objective reality, which is independent of any theory*, and the physical concepts with which the theory operates... by means of which we picture this reality to ourselves" (my emphasis). Therefore, we shall not bind the quantum behaviour to the standard Hilbert space formalism. There may be other theories that predict the same (correct) results for experiments, at least in restricted domains, and nevertheless might provide a better picture of reality. For this reason it is useful to explore different formulations of the theory.

It is true that the canonical formalism of quantum mechanics, in terms of Hilbert spaces (or C*-algebras in field theory), has proved to be extremely useful and there is no alternative that may compete with it; but some interesting formalisms are worth study as is done in the following. The mathematical theories behind these formalisms are however quite different from each other and, what is more relevant for us, they suggest different physical pictures of nature. I shall consider only 'elementary quantum mechanics', that is, the quantum theory of a finite number of particles in the nonrelativistic approximation, excluding both spin and the required symmetry or antisymmetry of the many-particle wave-function with respect to the exchange of any two particles (e.g. the Pauli principle for electrons).

Asides from the historical interest of the formalisms studied in this chapter, they give hints for possible solutions to some counterintuitive features of the standard Hilbert space formalism, for instance the wave-particle duality. Also, some of them may be useful calculational tools for specific problems. This is the case for the Wigner function and the Feynman path integrals. However, none of the formalisms, or theories, analyzed here provide a clear interpretation of quantum mechanics, although the path integrals studied in section 4.6 suggest a modification that is close to achieve that goal.

A theory that supplies a clear picture of the quantum world is stochastic electrodynamics, although its domain of validity is limited. Due to its relevance for a realistic interpretation of quantum mechanics, we devote the whole chapter 5 to that theory.

4.2. Theory of de Broglie-Bohm

4.2.1. Early interpretations of wave mechanics. The origin of the de Broglie-Bohm theory may be traced back to the proposal by Louis de Broglie [**1**] of associating a wave to every particle. For a free motion the wavelength, λ, is related to the particle's linear momentum, p, via

$$\lambda = \frac{h}{p}.$$

As is well known this hypothesis led Schrödinger to the development of wave mechanics. The cornerstone of the theory is Schrödinger celebrated equation which, for a single particle in a potential, $V(\mathbf{r})$, may be written

$$i\hbar\frac{\partial}{\partial t}\psi(\mathbf{r},t) = -\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r},t) + V(\mathbf{r})\psi(\mathbf{r},t), \tag{84}$$

where ∇^2 is the Laplacian operator.

Initially Schrödinger interpreted this equation as representing a field, the apparent particles (e.g. electrons) actually being wave-packets of that field. That interpretation was soon abandoned due to several strong difficulties to maintain it, pointed out mainly by Bohr and Heisenberg. Here I shall mention two of them. First, a localized wave-packet would spread out indefinitely, according to eq.(84), whilst a particle is always localized at a point when it is detected. Second, when the Schrödinger equation is generalized to N particles, it is necessary to define the wave-function ψ in the $3N$ dimensional configuration space. In particular, the electrostatic interaction between two electrons should appear in the Schrödinger equation in the form e^2/r_{12}, where r_{12} is the distance between the (pointlike) electrons. This fact was confirmed by the detailed calculation, by Hylleraas in 1927, of the ionization energy of the helium atom. Both these facts suggest that electrons are particles.

At about the same time L. de Broglie proposed his 'pilot wave interpretation' assuming that every (pointlike or very small) particle is accompanied by a wave (extended over a much bigger region) that guides the motion of the particle [**2**]. He abandoned it around 1930 and accepted the dominant Copenhagen interpretation. However, de Broglie returned to his early 'particle plus wave' assumption after the work of David Bohm that we discuss below. In fact, de Broglie modified his original interpretation assuming that the particle was actually a singularity in the wave, what he named 'theory of the double

solution' [**3**]. This interpretation never became very popular, whilst Bohm's approach, conceptually very similar to the de Broglie 'pilot wave interpretation', has survived.

The most relevant step for the interpretation of wave mechanics in the early period was the 1926 Born hypothesis [**4**] that the squared modulus of the wave-function, $\rho(\mathbf{r},t) \equiv |\psi(\mathbf{r},t)|^2$, is the probability density for the particle being in the spacetime point $(\mathbf{r},t)$. This interpretation is compatible with eq.(84) because ρ fulfils a continuity equation, that is,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = \mathbf{0}, \mathbf{j} \equiv -\frac{i\hbar}{2m}\left(\psi^* \nabla \psi - \psi \nabla \psi^*\right), \tag{85}$$

where ψ^* is the complex conjugate function of ψ. Born's interpretation (or, rather, postulate) has become a cornerstone of quantum mechanics until today. Any viable interpretation of quantum mechanics must include it either as a postulate or as a consequence of the postulates. In particular, it is a fundamental assumption in the Copenhagen interpretation discussed in chapter 2.

4.2.2. Bohmian mechanics. The continuity equation (85) suggests writing the wave-function in polar form as follows

$$\psi(\mathbf{r},t) = \sqrt{\rho(\mathbf{r},t)} \exp\left[iS(\mathbf{r},t)/\hbar\right], \tag{86}$$

where the introduction of $\hbar$ in the exponential makes S to have dimensions of action. If this is inserted in the Schrödinger eq.(84) and we separate the real and imaginary parts, we get two equations involving only real quantities:

$$\frac{\partial \rho}{\partial t} - \frac{1}{m}\nabla\left(\rho \nabla S\right) = 0, \tag{87}$$

$$\frac{\partial S}{\partial t} = \frac{1}{2m}\left(\nabla S\right)^2 + V(\mathbf{r}) - \frac{\hbar^2}{2m}\left(\frac{\nabla^2 \sqrt{\rho}}{\sqrt{\rho}}\right). \tag{88}$$

The passage from eq.(84) to eqs.(87) and (88) was proposed by Madelung in 1926 [**5**], who gave an hydrodynamical interpretation not agreeing with the mentioned one by de Broglie in terms of particles accompanied by waves (see below section 4.3.1).

Eq.(87) is just the same continuity eq.(85), although writing the current velocity j/ρ as the gradient of a scalar S resembles the equation of an inviscid fluid. The interpretation of eq.(88) is still less obvious, but its classical limit is clear. In fact, when $\hbar \to 0$ it becomes the Hamilton-Jacobi equation of classical mechanics. This led

David Bohm in 1951 [**6**] to interpret eq.(88) as governing the motion of a particle. That is, the particle moves under the action of two forces, the first one derives from the potential $V(r)$ and the second one from the additional 'quantum potential'

$$U(\mathbf{r},t) \equiv -\frac{\hbar^2}{2m}\left(\frac{\nabla^2\sqrt{\rho(\mathbf{r},t)}}{\sqrt{\rho(\mathbf{r},t)}}\right). \tag{89}$$

In this form Bohm elaborated a 'quantum theory of motion' [**7**], sometimes known as Bohmian mechanics.

Essentially, Bohm's theory assumes that particles have well defined trajectories that may be obtained assuming that

$$\mathbf{v} \equiv \frac{d\mathbf{r}}{dt} = -\frac{1}{m}\nabla S(\mathbf{r},t), \tag{90}$$

is the actual velocity of the particle when it is placed at $(\mathbf{r},t)$. If we know the function $S(\mathbf{r},t)$ we may obtain the path of the particle by solving this (vector first order) ordinary differential eq.(90). As in classical mechanics we must previously solve the partial differential 'Hamilton-Jacobi' eq.(88). However, unlike in classical mechanics, here the 'Hamilton-Jacobi equation' contains an unknown function $\rho(\mathbf{r},t)$, so that eqs.(88) and (87) are coupled and it is necessary to solve them simultaneously. These two non-linear first order partial differential equations are equivalent to the complex, but linear, Schrödinger eq.(84), which in practice is easier to solve.

In order to better understand the calculational procedure involved in Bohmian mechanics, I will consider the simple example of a one-dimensional harmonic oscillator. The Schrödinger equation may be written

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x_2} + \frac{1}{2}kx^2\psi.$$

I shall study only two particular solutions, corresponding to the ground state and a coherent state, respectively. These solutions are [**8**]

$$\psi = \sqrt{\frac{m\omega}{\pi\hbar}}\exp\left(-\frac{1}{2}\xi^2\right), \qquad \omega \equiv \sqrt{\frac{k}{m}}, \qquad \xi \equiv \sqrt{\frac{m\omega}{\hbar}}x, \tag{91}$$

$$\begin{aligned}\psi &= \sqrt{\frac{m\omega}{\pi\hbar}}\exp\left[-\frac{1}{2}(\xi-\xi_0\cos(\omega t))^2\right] \\ &\times\exp\left[-i\left(\frac{1}{2}\omega t + \xi\xi_0\sin(\omega t) - \frac{1}{4}\xi_0^2\sin(2\omega t)\right)\right],\end{aligned} \tag{92}$$

where ξ_0 is a constant measuring the amplitude of the oscillation (see below). The study of the ground state eq.(91) is straightforward because there is no phase depending on position. Thus $\mathbf{v}$, eq.(90), is zero and the picture provided by Bohmian mechanics is that the particle is at rest but its position is unknown, the probability density of the possible positions being determined by the squared modulus of the wave-function ψ, eq.(91) . The result is general, the particle is at rest in any stationary state where the ψ-function may be taken as real. The wave is stationary and possesses an amplitude given by eq.(91) . As an aid to the intuition we might replace the actual particle by a statistical ensemble of particles whose density would agree with the "probability density of finding the particle at $\mathbf{r}$ if a position measurement is performed", using the standard (Copenhagen) language of quantum mechanics.

I shall now study the coherent state in terms of the dimensionless position variable ξ. The velocity of the particle placed at (ξ, t) is given by eq.(95) applied to the phase of the complex function eq.(92) . Introducing a 'dimensionless velocity' $\nu \equiv \sqrt{m/(\hbar\omega)}v$ we get

$$\nu(\xi, t) = \frac{\partial}{\partial \xi}\left(\frac{1}{2}\omega t + \xi\xi_0 \sin(\omega t) - \frac{1}{4}\xi_0^2 \sin(2\omega t)\right) = \xi_0 \sin(\omega t).$$

Thus the motion of a particle belonging to the statistical ensemble depends on the position only via the time dependence of ξ. In fact, it fulfils

$$(93)\quad \nu(\xi, t) = \frac{1}{\omega}\frac{d\xi}{dt} = \xi_0 \sin(\omega t) \quad \Rightarrow \quad \xi(t) = -\xi_0 \cos(\omega t) + const.$$

The probability density is given by the squared modulus of eq.(92), that is,

$$(94)\qquad \rho(\xi, t)\, d\xi = \frac{1}{\sqrt{\pi}} \exp\left[-(\xi - \xi_0 \cos(\omega t))^2\right] d\xi,$$

which is consistent with the velocity field derived in eq.(93), as it should because both are related by eqs.(88) and eq.(87) . The picture that emerges is a 'cloud' of particles that oscillates of frequency ω without deformation. In addition to the actual particle there is a wave whose amplitude, ψ, evolves according to Schrödinger eq.(92) .

4.2.3. The ontology of Bohm's theory. The interpretation of eq.(88) as the quantum counterpart of the Hamilton-Jacobi equation poses a difficult problem, at least for a realistic interpretation of Bohm's theory. In fact, in classical mechanics the potential $V(\mathbf{r})$, or

at least its gradient, corresponds to a *real field* present at all points of some region, e.g. an electric field. In sharp contrast a probability density like $\rho(\mathbf{r}, t)$, and consequently the quantum potential eq.(89), does not correspond to any real field, it is just a mathematical function containing the available information about the presence of the particle. Thus it is difficult to understand how the quantum potential, devoid of physical reality, may have any effect on an actual particle. The solution to the difficulty is to assume that $\rho(\mathbf{r}, t)$ and therefore U and S, which are related to ρ via eqs.(88) and (90), are real fields. This leads to an interpretation of Bohm's law of motion very similar to the pilot wave theory of L. de Broglie. That is, for every actual particle, e.g. an electron, there is an associated wave. The wave is governed by Schrödinger eq.(84) (or, what is equivalent, eqs.(88) and (90)), and the wave guides the motion of the particle.

For us the relevant question is: Does Bohmian mechanics provide the correct ontology for a realistic understanding of quantum theory? On the positive side I must concede that it is a viable hidden variables theory, it offers a realistic ontology without changing at all the (successful) predictions of quantum mechanics (at least in the non-relativistic domain here studied). Indeed, Bohmian mechanics has become the paradigm of hidden-variable theories. For these reasons a group of physicists, relatively small, has supported Bohmian mechanics as the solution of all interpretational problems of quantum mechanics [**7**]. In contrast, the mainstream of the scientific community is not very enthusiastic about the theory, precisely because they do not like hidden variables theories. Indeed, from a pragmatic point of view Bohmian mechanics does not present any calculational advantage over standard wave mechanics because it is necessary to solve the Schrödinger equation in the first place. However, it has been pointed out that the theory may allow calculating some quantities which cannot be found using standard quantum mechanics, such as first passage times. It also provides interesting pictures of some phenomena like interference in the two-slit experiment or the evolution of molecules during a collision.

There is however an important difficulty for Bohmian mechanics. It is the problem for the interpretation of systems of N particles. In fact, in this case the Schrödinger equation is an equation in the configuration space of $3N$ dimension. That is, with reference to eq.(84)

the wave-function and the Laplacian operator become

$$\psi = \psi\left(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N, t\right), \nabla^2 = \sum_{j=1}^{N} \left(\frac{\partial^2}{\partial x_j^2} + \frac{\partial^2}{\partial y_j^2} + \frac{\partial^2}{\partial z_j^2}\right).$$

It is very difficult to understand the nature of a 'physical wave living in a non-physical 3N-dimensional space'. Actually, a similar problem appears in the early Schrödinger interpretation, as discussed above. In Bohmian mechanics it is possible to assume that there is a wave associated to every particle, but the set of waves gives rise to a force on every particle which happens to be calculable via the generalization to $3N$ dimensions of the 3 dimensional eqs.(88) and (90). This assumption however seems contrived and not too satisfactory. Also if the hypothesis is not accompanied by the evolution equations of the different waves then it is not too informative. Indeed, it does not explain how the complexity of multiple waves leads to the relatively simple Schrödinger equation in 3N dimension. The use of the configuration space might be avoided formulating the theory for relativistic quantum fields, rather than nonrelativistic quantum mechanics. Indeed, quantum field theory is elaborated in ordinary 3D space (plus time). Attempts in these directions have been made with partial success, but at the price of losing simplicity. In particular, the intuitive picture of particles moving along deterministic paths, which is one of the virtues of Bohmian mechanics, is lost.

Finally there is also another shortcoming of the theory, namely the lack of any stochastic element. This implies that the probability needs to be introduced so to speak 'by hand'. For instance, in any stationary state there is a probability density for the position of the particle, but the particle is at rest and there is no apparent dynamical mechanism able to produce the randomness. For all these reasons my personal attitude towards Bohmian mechanics is negative, although I admit that this opinion is to some extent a matter of taste, and therefore it is difficult to provide more clear arguments. In any case, if I am compelled to choose amongst the existing ontologies aimed at a realistic interpretation of wave mechanics, I prefer stochastic theories, like stochastic mechanics, the theory to be studied in the following; or, even better, the stochastic electrodynamics studied in chapter 5.

4.3. Stochastic mechanics

4.3.1. Madelung hydrodynamical interpretation. As said above, Madelung [**5**] proposed in 1926 to write the Schrödinger wavefunction in terms of modulus and phase, as in eq.(86), but interpreting eqs.(87) and (88) as corresponding to a fluid with density, ρ, and velocity, $\mathbf{v}$, given by

$$\mathbf{v}(\mathbf{r},t) \equiv -\nabla S/m. \tag{95}$$

Thus the continuity eq.(87) becomes

$$\frac{\partial \rho}{\partial t} + \nabla(\rho \mathbf{v}) = 0, \tag{96}$$

and eq.(88) leads to an equation which may be written in terms of ρ and $\mathbf{v}$. In fact, if we apply the gradient operator, ∇, to eq.(88) we get, taking eq.(95) into account,

$$\frac{d\mathbf{v}}{dt} = \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}.\nabla)\,\mathbf{v} = -\frac{1}{m}\nabla\left[V(\mathbf{r}) + U(\mathbf{r})\right], \quad U(\mathbf{r}) = -\frac{\hbar^2}{2m^2}\frac{\nabla^2\sqrt{\rho}}{\sqrt{\rho}}. \tag{97}$$

This allows a hydrodynamical interpretation of Schrödinger equation. Indeed, for $\hbar = 0$ eq.(97) becomes the Euler equation of motion of an inviscid fluid under the action of an external force. The modification introduced by quantum mechanics derives from the term $U(\mathbf{r},t)$, the 'quantum potential' in the de Broglie-Bohm theory.

If we accept Born's interpretation of $\rho(\mathbf{r},t)$ as a probability density, it is natural to look at the Madelung fluid as a fictitious fluid consisting of a statistical ensemble of particles that represent our information about the unique particle of our system. Thus the term $U(\mathbf{r},t)$ has the resemblance of a diffusion term produced by some random field, which would provide an explanation for the quantum behaviour. Indeed, many people have attempted to make a derivation of Schrödinger equation or to provide an interpretation of wave mechanics, via explaining the origin of the term eq.(97) (e.g. [**9**]). The most elaborated approach along these lines is stochastic mechanics, proposed in 1966 by E. Nelson [**10**], [**11**].

4.3.2. Dynamics of Markov stochastic processes. I now start presenting a theory which rests upon the assumption that the motion of a (quantum) particle may be treated as a stochastic process, that is, as a probability distribution on the set of possible paths. I shall consider the particle in one dimension for simplicity, and later I

shall comment on the modifications needed for a 3N-dimensional theory. The presentation here follows L. de la Peña [**12**]. We assume that the particle's path may be represented by a continuous, but nowhere differentiable, function of time, $x(t)$. Lack of differentiability means that it is not possible to get the velocity as the time derivative of the position. However, if we consider a Markov stochastic process, it is possible to define two stochastic time derivatives, forward, D_+ and backward, D_-, as follows

$$\begin{aligned} v_+(t) &\equiv D_+x(t) \equiv \lim_{\Delta t\to 0}\frac{\langle x\left(t+\Delta t\right)-x\left(t\right)\rangle}{\Delta t}, \\ v_-(t) &\equiv D_-x(t) \equiv \lim_{\Delta t\to 0}\frac{\langle x\left(t\right)-x\left(t-\Delta t\right)\rangle}{\Delta t}, \end{aligned} \tag{98}$$

where $\langle\rangle$ means ensemble average, and $v_+(t)$ and $v_-(t)$ are called forward and backward velocities, respectively. A stochastic process is a Markov process in case the past and future are conditionally independent given the present. The notion is time symmetric. We also assume that the following limits exist and they are equal and independent of time:

$$\begin{aligned} 2D &= \lim_{\Delta t\to 0}\frac{\left\langle\left[x\left(t+\Delta t\right)-x\left(t\right)\right]^2\right\rangle}{\Delta t} \\ &= \lim_{\Delta t\to 0}\frac{\left\langle\left[x\left(t\right)-x\left(t-\Delta t\right)\right]^2\right\rangle}{\Delta t}. \end{aligned} \tag{99}$$

On the other hand we assume that

$$\lim_{\Delta t\to 0}\frac{\langle\left[x\left(t+\Delta t\right)-x\left(t\right)\right]^n\rangle}{\Delta t} = \lim_{\Delta t\to 0}\frac{\langle\left[x\left(t\right)-x\left(t-\Delta t\right)\right]^n\rangle}{\Delta t} = 0,$$

for $n > 2$. As we shall see below these assumptions lead to a correct dynamics in the case of Brownian motion, what makes them plausible.

Now we may consider an analytic function of x and t, $f(x,t)$, and apply to it the operators of time derivative. We get

$$\begin{aligned} D_+f(x,t) &= \frac{\partial f}{\partial t}+v_+(x,t)\frac{\partial f}{\partial x}+D\frac{\partial^2 f}{\partial x^2}, \\ D_-f(x,t) &= \frac{\partial f}{\partial t}+v_-(x,t)\frac{\partial f}{\partial x}-D\frac{\partial^2 f}{\partial x^2}. \end{aligned}$$

The particle's position $x(t)$ being a stochastic process, a convenient procedure to study the evolution is to introduce a statistical ensemble of particles with density $\rho\left(x,t\right)$, so that ρdx is the probability that the actual particle is placed between x and $x+dx$ at time t. It is

interesting to obtain equations for the evolution of the density. To do that we consider a function $f(x,t)$ that represents a conserved quantity; that is, we assume

$$D_{\pm} f = 0.$$

If we multiply these equations times the density, ρ, and perform an integration with respect to x we obtain

$$0 = \int_{-\infty}^{\infty} dx \rho D_{\pm} f = \int_{-\infty}^{\infty} dx f \left[-\frac{\partial \rho}{\partial t} - \frac{\partial}{\partial x} \left(v_{\pm} \rho \right) \pm D \frac{\partial^2 \rho}{\partial x^2} \right],$$

where the last equality derives from an integration by parts (assuming $\rho(x,t) \to 0$ for $x \to \pm\infty$ and for $t \to \pm\infty$). This equality being true for any (conserved) quantity f, the quantity inside the squared bracket should vanish, which gives rise to two equations. The sum and the difference of them lead to

$$\begin{array}{ll} \dfrac{\partial \rho}{\partial t} + \dfrac{\partial}{\partial x} \left(\rho v \right) = 0, & \dfrac{\partial}{\partial x} \left(u \rho \right) = D \dfrac{\partial^2 \rho}{\partial x^2}, \\ v \equiv (v_+ + v_-)/2, & u \equiv (v_+ - v_-)/2 \end{array} \tag{100}$$

where I have introduced the current velocity, v, and the osmotic velocity, u. The first eq.(100) is the continuity equation, and the latter may be integrated to give

$$u = D \rho^{-1} \frac{\partial \rho}{\partial x} = D \frac{\partial}{\partial x} \left(\log \rho \right). \tag{101}$$

Actually, the integration is not so simple in three dimensions, because the solution contains an additional term of the form $\rho^{-1} \nabla \times \mathbf{w}(\mathbf{r}, t)$, where $\mathbf{w}$ is an arbitrary vector field. Here I shall ignore that possibility (which however could be interesting because the term might be associated to the spin angular momentum).

In order to have a complete description of the particle's motion we need, in addition to the continuity equation, another equation for the systematic velocity. Depending on our choice of that equation we may obtain different theories, and I study briefly the most relevant: Brownian motion and stochastic mechanics.

The theory of Brownian motion was studied by Einstein in 1905 for the case where no external force acts on the particle, and it was generalized to the case involving external forces by Smoluchowski. In the former case there are no privileged directions in space which implies that the (ensemble) mean position at time $t + \Delta t$ equals the

position at time t. This implies that the forward velocity v_+, eq.(98), is zero, which leads to

$$0 = v_+ = v + u = v + D\frac{\partial}{\partial x}\left(\log \rho\right).$$

This is the second equation we were searching for. Combined with the continuity equation we obtain the diffusion equation

$$\frac{\partial \rho}{\partial t} = D\frac{\partial^2 \rho}{\partial x^2}. \tag{102}$$

If there is an external force, $f(x,t)$, a plausible assumption is that the forward velocity is proportional to the force, which is appropriate for a dissipative phenomenon. Thus we get

$$f = Cv_+ \qquad \Rightarrow \qquad \frac{\partial \rho}{\partial t} = -\frac{1}{C}\frac{\partial\left(\rho f\right)}{\partial x} + D\frac{\partial^2 \rho}{\partial x^2}, \tag{103}$$

which is Smoluchowski equation for Brownian motion in the presence of an external force field (e.g. gravity.) Clearly eqs.(102) and (103) may be used only for the evolution towards the future. If we know the position of the particle at a time and we ask for the previous positions we should use the conditions $v_- = 0$ or $f = Cv_-$ leading to similar equations with $-D$ substituted for D. The Einstein-Smoluchowsky theory of Brownian motion involves particle's paths where the instantaneous velocity is not defined ('it is infinite'). A more physical (although more involved) theory was later developed by Ornstein and Uhlenbeck, but we will not comment on it here.

4.3.3. Stochastic mechanics. Now I will study a non-dissipative theory where the force is proportional to the acceleration. Actually, there are four accelerations, obtained by application of the two operators $D_\pm$ to the two velocities $v_\pm$. It is more convenient to define two new operators of time derivative as follows

$$\begin{aligned} D_c &\equiv \frac{1}{2}\left(D_+ + D_-\right) = \frac{\partial}{\partial t} + v\frac{\partial}{\partial x}, \\ D_s &= \frac{1}{2}\left(D_+ - D_-\right) = u\frac{\partial}{\partial x} + D\frac{\partial^2}{\partial x^2}, \end{aligned} \tag{104}$$

and to write the accelerations in terms of these operators and the velocities v and u defined in eq.(100), that is,

$$D_c v, \quad D_s u, \quad D_c u, \quad D_s v. \tag{105}$$

In order to postulate the law of dynamics we shall choose amongst the possible linear combinations of these four accelerations. Firstly it

is necessary to choose a time-reversal invariant acceleration in order to obtain a non-dissipative theory. Only the two first accelerations in eq.(105) fulfil this condition, the last two changing sign under the reversal of time. Thus the required acceleration must be of the form $aD_cv + bD_su$. Now D_cv should enter with coefficient $a = 1$ in order to insure the correct deterministic limit (see below). Thus the only free parameter will be the coefficient, b. However, from eqs.(101) and (104) we obtain

$$D_su = D^2\frac{\partial}{\partial x}\left[\frac{1}{2}\rho^{-2}\left(\frac{\partial\rho}{\partial x}\right)^2 + \rho^{-1}\frac{\partial}{\partial x}\left(\frac{\partial\rho}{\partial x}\right)\right] = 2D^2\frac{\partial}{\partial x}\left(\frac{1}{\sqrt{\rho}}\frac{\partial^2\sqrt{\rho}}{\partial x^2}\right), \tag{106}$$

so that the parameter b may be absorbed in D^2 except for its sign. Consequently our freedom reduces to choosing a sign, and I postulate $b = -1$. Hence

$$f = m\left(D_cv - D_su\right), \tag{107}$$

which is the dynamical law generalizing Newton's second law. Eq.(107) may be rewritten, taking eqs.(100) and (106) into account

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} - 2D^2\frac{\partial}{\partial x}\left(\frac{1}{\sqrt{\rho}}\frac{\partial^2\sqrt{\rho}}{\partial x^2}\right) = \frac{f}{m} = -\frac{1}{m}\frac{\partial V(x)}{\partial x}, \tag{108}$$

where m is the mass of the particle. The last equation corresponds to the case where the force f is conservative and it agrees with the Madelung eq.(97) provided we identify $2D = \hbar/m$. This is the basis for the interpretation of the Schrödinger equation as the equation for evolution of the probability density of a particle moving under the action of a given conservative force plus a non-dissipative random force of some kind characterized by the assumptions made above.

4.3.4. Does stochastic mechanics provide a realistic interpretation of quantum mechanics? Firstly I shall comment on a subject related to stochastic mechanics named 'stochastic quantization' [**13**]. It consists of using a fictitious time, τ, related to the real time, t, by $\tau = -it$, where i is the imaginary unit. In this form the time-dependent Schrödinger eq.(84) looks like a diffusion equation and the expansion in terms of eigenfunctions of the Hamiltonian

$$\Psi(x,t) = \sum_j \psi_j \exp\left(iE_jt\right) = \sum_j \psi_j \exp\left(-\tau E_j\right),$$

looks like a partition function of statistical mechanics, provided we identify τ with the inverse of the temperture. The method may be extended to quantum field theory. Thus the change to imaginary time allows dealing with quantum mechanics with the methods of statistical mechanics or stochastic processes. Stochastic quantization adds no hint for a realistic interpretation of quantum mechanics, it is just a formal convenient method for some calculations; therefore, I will not study it further in this book. In contrast, stochastic mechanics apparently provides an ontology, that is, a physical model where particles have actual trajectories. Thus it is similar to Bohmian mechanics with the advantage that the paths present stochasticity, which fits in the probabilistic character of quantum mechanics. In this respect stochastic mechanics has had successes but also failures which will be summarized in what follows [**14**].

The main success is of course the derivation of Schrödinger equation with the correct relation between the wave-function and the probability density of the particle positions (i.e. Born's rule). It is also possible to explain why identical particles satisfy either Bose-Einstein or Fermi-Dirac statistics and why particles have integer or half-odd spin [**15**]. Nelson ([**14**], Section 17) has pointed out that it is also possible to get a stochastic picture of the two-slit experiment, explaining how particles have trajectories going through just one slit or the other, but nevertheless produce a probability density as for interfering waves. All these developments are rather formal and it is difficult to get an intuitive picture from them. The stochastic processes associated to these phenomena are rather strange, to say the least. In particular, the existence of nodal surfaces in the stationary states of the Schrödinger equation is counterintuitive. We might believe that there are stochastic processes which, for unexplained reasons, give rise to barriers which cannot be crossed ([**14**]). However, those processes look bizarre. Actually, it is not necessary to attribute a real existence to the stationary states of the Schrödinger equation, as discussed in chapter 2 section 2.3.1.

More difficult is to understand many-body (nonrelativistic) quantum mechanics. It is necessary to assume that every particle has an influence on all other particles via the quantum potential, that is, the third term of eq.(108) . In fact, in the many-body equation that term contains derivatives with respect to all the particle's positions. The result is that, if there are two uncoupled systems, an alteration of the

second one affects the first one. This makes unrealistic to regard the trajectories as physically real.

In summary stochastic mechanics does not provide a clear intuitive model of the behaviour of quantum systems. However, some modification of stochastic mechanics could provide the ontology we are searching for. The modification might come from a different more physical approach, resting also upon stochastic processes, namely *stochastic electrodynamics*, a theory studied in chapter 5.

4.4. The Weyl-Wigner formalism

4.4.1. Introduction. Historical note. The Weyl-Wigner formalism may be seen as an attempt to formulate nonrelativistic quantum mechanics in the phase space of coordinates and momenta. A relevant question is whether it is possible to get a realistic interpretation of quantum mechanics from the formalism. The answer is negative as will be discussed in section 4.4.4.

Hermann Weyl [**16**], in a context related to representation theory in mathematics, proposed in 1927 to map classical phase space functions into operators. The procedure gives a method of phase space quantization. However, it is now understood that Weyl quantization is not always well defined and sometimes gives unphysical answers. More interesting is the inverse Weyl transform that leads from operators in Hilbert space to phase space functions. For states this representation is known as Wigner function, reviewed in the following. An application to optics will be studied in chapter 6.

The Wigner function (or Wigner quasiprobability distribution) was introduced by Eugene Wigner in 1932 [**17**] in order to study quantum corrections to classical statistical mechanics. It provides an invertible mapping between functions in the quantum phase space formulation and Hilbert space operators in the Schrödinger picture. Within quantum mechanics the mapping is well defined and, in contrast with Weyl original intention in seeking a consistent quantization scheme, this map amounts to a mere change of representation.

In the 1940s, H. J. Groenewold [**18**] and José E. Moyal [**19**] elucidated how the Wigner function provides the integration measure (analogous to a probability density function) in phase space, which allows to formulate quantum mechanics in phase space. For reviews see [**20**] and [**21**].

In 1948 Jean Ville [**22**] rederived the Wigner function, out of the context of quantum mechanics, as a quadratic representation of the

local time-frequency energy of a signal. For instance, in music it may be used to study the frequencies of sound as a function of time. In this case eq.(122) should be interpreted so that ψ corresponds to the sound wave amplitude, $\mathbf{r}$ and t fix a spacetime point and $\mathbf{p}$ is the wave-vector of one plane-wave component of the sound. The use of the Wigner function in (classical) signal theory will not be studied here, and the interested reader may look at the books dealing with the subject [**23**].

4.4.2. Weyl transform. A naive method to associate phase space functions to the Hilbert space operators in the quantum mechanics of particles is as follows. If we consider a single particle in one dimension, a method to get the image of the operator $\widehat{M}$ may be

$$F_{\widehat{M}}(x,p) = \mathrm{Tr}\left\{\widehat{M}\delta\left(\hat{x}-x\right)\delta\left(\hat{p}-p\right)\right\},$$

where $\delta\left(\right)$ is Dirac's delta. However, the mathematical problem of defining properly the delta distribution of an operator is not trivial. We might define $F_{\widehat{M}}(x,p)$ in the form

(109)

$$F_{\widehat{M}}(x,p) = \\ = \frac{1}{4\pi^2\hbar^2}\int_{-\infty}^{\infty}d\alpha\int_{-\infty}^{\infty}d\beta\mathrm{Tr}\left\{\widehat{M}\exp\left[i\alpha\left(\hat{x}-x\right)/\hbar\right]\exp\left[i\beta\left(\hat{p}-p\right)/\hbar\right]\right\},$$

where $\mathrm{Tr}\widehat{A}$ means the trace of the operator $\widehat{A}$ and the Planck constant $\hbar$ is introduced here to agree with common usage. As the operators $\hat{x}$ and $\hat{p}$ do not commute, a *different* definition would be

(110)

$$F'_{\widehat{M}}(x,p) = \\ = \frac{1}{4\pi^2\hbar^2}\int_{-\infty}^{\infty}d\beta\int_{-\infty}^{\infty}d\alpha\mathrm{Tr}\left\{\widehat{M}\exp\left[i\beta\left(\hat{p}-p\right)/\hbar\right]\exp\left[i\alpha\left(\hat{x}-x\right)/\hbar\right]\right\}.$$

The Weyl transform is intermediate amongst the two possibilities:

$$W_{\widehat{M}} = \frac{1}{4\pi^2\hbar^2}\int_{-\infty}^{\infty}d\alpha\int_{-\infty}^{\infty}d\beta\exp\left(-i\alpha x/\hbar - i\beta p/\hbar\right) \\ \times\mathrm{Tr}\left\{\widehat{M}\exp\left[\left(i/\hbar\right)\left(\alpha\hat{x}+\beta\hat{p}\right)\right]\right\}, \tag{111}$$

where we have taken into account that the numbers x and p may be put outside of the trace operation. I point out that the 3 transforms,

eqs.(109) , (110) and (111) might be, and have been, used but here I will study only the last one.

The order of the operators $\hat{x}$ and $\hat{p}$ within an exponential is irrelevant, as may be shown expanding the exponential in powers. But, as the operators $\hat{x}$ and $\hat{p}$ do not commute, the product of the exponentials is not equal to the exponential of the sum. In fact, the Campbell-Haussdorf formula

$$\exp\left(\hat{A}+\hat{B}\right)=\exp\left(\hat{A}\right)\exp\left(\hat{B}\right)\exp\left(-\frac{1}{2}\left[\hat{A},\hat{B}\right]\right), \tag{112}$$

that holds true if $\left[\hat{A},\hat{B}\right]$ commutes with $\hat{A}$ and with $\hat{B}$, applied to eq.(111) gives

$$\begin{aligned}\exp\left[(i/\hbar)\left(\alpha\hat{x}+\beta\hat{p}\right)\right] &= \exp\left[(i/\hbar)\alpha\hat{x}\right]\exp\left[(i/\hbar)\beta\hat{p}\right]\exp\left((i/2\hbar)\alpha\beta\right)\\ &= \exp\left[(i/\hbar)\beta\hat{p}\right]\exp\left[(i/\hbar)\alpha\hat{x}\right]\exp\left(-(i/2\hbar)\alpha\beta\right),\end{aligned} \tag{113}$$

where we have taken into account the commutation rules, eqs.(114) below. This exhibits the relation of the Weyl transform of the operator $\widehat{M}$, eq.(111), with the choices eqs.(109) and (110) .

The generalization to the Hilbert space associated to a system with n degrees of freedom is straightforward. The coordinates and momenta operators $\{\hat{x}_j,\hat{p}_j\}\,, j=1,2,\ldots,N$ fufill the standard commutation relations

$$[\hat{x}_j,\hat{x}_k]=[\hat{p}_j,\hat{p}_k]=0,[\hat{x}_j,\hat{p}_k]=i\hbar\delta_{jk}. \tag{114}$$

To any trace-class operator $\hat{\rho}$ in that space we may associate a function, W, in a 2n-dimensional space spanned by the complex numbers $\{x_j,p_j\}$, $j=1,2,\ldots,N$, through the following Weyl-Wigner transform

$$\begin{aligned}W_{\hat{\rho}}\left(\{x_j,p_j\}\right) &= (2\pi\hbar)^{-n}\int \mathrm{Tr}\Big\{\hat{\rho}\exp\Big[(i/\hbar)\sum_j\left(\alpha_j\hat{x}_j+\beta_j\hat{p}_j\right)\Big]\Big\}\\ &\times\exp\Big[(-i/\hbar)\sum_j\left(\alpha_j x_j+\beta_j p_j\right)\Big]\prod_j d\alpha_j d\beta_j.\end{aligned} \tag{115}$$

The transform is invertible in the sense that given the function $W(\{x_j,p_j\})$ we get the operator $\hat{\rho}$ through

$$\hat{\rho} = (2\pi\hbar)^{-2n} \int \prod_k \exp\left[(-i/\hbar)\sum_k (\alpha_k \hat{x}_k + \beta_k \hat{p}_k)\right] d\alpha_k d\beta_k$$

$$\times \int W_{\hat{\rho}}(\{x_j,p_j\}) \prod_j \exp\left[(i/\hbar)(\alpha_j x_j + \beta_j p_j)\right] dx_j dp_j. \tag{116}$$

Here the operator $\hat{\rho}$ is obtained as a function of the coordinate and momentum operators $\{\hat{x}_j, \hat{p}_j\}$.

Several relevant properties of the Weyl-Wigner transform are:

1) **Linearity**. If $\hat{\rho}_1$ and $\hat{\rho}_2$ are two trace-class operators (of the same Hilbert space) with transforms $W_{\hat{\rho}_1}(\{x_j,p_j\})$, $W_{\hat{\rho}_2}(\{x_j,p_j\})$, the transform of the sum is the sum of the transforms, that is,

$$W_{\hat{\rho}_1 \hat{+} \rho_2}(\{x_j,p_j\}) = W_{\hat{\rho}_1}(\{x_j,p_j\}) + W_{\hat{\rho}_2}(\{x_j,p_j\}). \tag{117}$$

The proof is straightforward.

2) **Trace**. The trace of an operator in terms of the Weyl-Wigner transform is

$$\mathrm{Tr}\hat{\rho} = \int W_{\hat{\rho}}(\{x_j,p_j\}) \prod_j dx_j dp_j, \tag{118}$$

as may be easily proved.

3) If the operator $\hat{\rho}$ is a function of the position (momentum) operators alone, then the transform is obtained through the simple change $\hat{\mathbf{r}}_j \to \mathbf{r}_j$ ($\hat{\mathbf{p}}_j \to \mathbf{p}_j$, respectively). The proof is not difficult but it is omitted.

As a consequence, taking linearity into account, the Hamiltonian operator is transformed as follows

$$\widehat{H} = \sum_j \left(\frac{\hat{\mathbf{p}}_j^2}{2m_j} + V(\hat{\mathbf{r}}_j)\right) \quad \to \quad W_{\widehat{H}} = \sum_j \left(\frac{\mathbf{p}_j^2}{2m_j} + V(\mathbf{r}_j)\right). \tag{119}$$

4.4.3. Wigner function. The most interesting Weyl transform happens when $\hat{\rho}$ is a density operator representing the state (generally mixed) of the physical system. In this case $\hat{\rho}$ fulfils the conditions of being Hermitian, normalized and positive, that is,

$$\hat{\rho}^\dagger = \hat{\rho}; \quad \mathrm{Tr}\hat{\rho} = 1; \quad \forall \psi \in H,\ \langle \psi \mid \hat{\rho} \mid \psi \rangle \geq 0. \tag{120}$$

Then $W_{\hat{\rho}}(\{x_j,p_j\})$ becomes the Wigner function of the quantum state. In particular, for a pure state $\hat{\rho} = |\psi\rangle\langle\psi|$ and eq.(115) becomes

$$W(\mathbf{r},\mathbf{p}) = (2\pi\hbar)^{-1} \int \exp(-i\alpha\cdot\mathbf{r}/\hbar)\exp(-i\beta\cdot\mathbf{p}/\hbar)$$
$$\times \langle\psi \mid \exp\left[i\left(\alpha\cdot\hat{\mathbf{r}}+\beta\cdot\hat{\mathbf{p}}\right)/\hbar\right] \mid \psi\rangle d\alpha d\beta. \tag{121}$$

(From now on we consider systems of one particle, although the generalization to many particles is straightforward). I stress that writing the exponential of a sum as a product of exponentials is possible for numbers, but not for operators. Eq.(121) leads to

$$W(\mathbf{r},\mathbf{p},t) = \frac{1}{(2\pi\hbar)^3}\int \psi\left(\mathbf{r}+\frac{1}{2}\mathbf{u},t\right)\psi^*\left(\mathbf{r}-\frac{1}{2}\mathbf{u},t\right)\exp(-i\mathbf{p}\cdot\mathbf{u}/\hbar)\,d\mathbf{u}, \tag{122}$$

for any wave-function $\psi(\mathbf{r},t)$ in the Schrödinger representation. It may be generalized to density matrices and many-particle systems.

The Wigner function encodes all information about the quantum system and therefore it gives rise to an alternative formulation of quantum mechanics. Specifically it yields expectation values from a phase-space scalar function $g(\mathbf{r},\mathbf{p})$ uniquely associated to the operator $\widehat{G}$ through an appropriate (Weyl) transform. Thus the expectation of the operator looks like 'a phase-space average' for the Weyl transform of the operator, that is,

$$\left\langle \widehat{G} \right\rangle = \int d\mathbf{r}\, d\mathbf{p}\, W(\mathbf{r},\mathbf{p}) g(\mathbf{r},\mathbf{p}). \tag{123}$$

The three properties eqs.(120) of the density matrices give rise to the following properties of the Wigner function:

1) It is a real function, that is, $W(\mathbf{r},\mathbf{p}) = W^*(\mathbf{r},\mathbf{p})$,

2) It is normalized, that is, $\int W(\mathbf{r},\mathbf{p})\, d\mathbf{r}\, d\mathbf{p} = 1$,

3) It is 'positive' in the following sense

$$\int W(\mathbf{r},\mathbf{p})\, W_{\widehat{\Theta}}(\mathbf{r},\mathbf{p})\, d\mathbf{r}d\mathbf{p} \geq 0,$$

for any pure state $\widehat{\Theta}$.

However, this kind of positivity property does not guarantee the usual positivity, that is, $W(\mathbf{r},\mathbf{p}) \geq 0$. Indeed, it has been shown [**24**] that for pure quantum states a necessary and sufficient condition for nonnegativity in one dimension is that the wave-function is Gaussian, which for instance corresponds to either the ground state or a coherent state of the one-dimensional oscillator. This theorem has been

generalized to many dimensions [**25**]. For mixed states there is no simple criterion. The fact that the Wigner function is not always positive prevents its interpretation as a probability distribution in phase space (for this reason it is called a pseudo-probability distribution). In section 4.4.4 we will return to this problem.

Other relevant properties of the Wigner function are the following

1) **Marginals**. The quantum **r** and **p** distributions are given by the marginals, i.e.

$$\rho(\mathbf{r}) = \int W(\mathbf{r},\mathbf{p},t)\, d\mathbf{p}, \qquad f(\mathbf{p}) = \int W(\mathbf{r},\mathbf{p},t)\, d\mathbf{r},$$

which easily follows from the definition in eq.(122) . It is interesting that in points where $\rho(\mathbf{r}) = 0$ we may have $W(\mathbf{r},\mathbf{p},t) \neq 0$. This would not be possible if $W(\mathbf{r},\mathbf{p},t) \geq 0$ everywhere because the integral of a nonnegative quantity is zero only if the quantity is zero (except for a set of zero measure.)

2) **Overlap**. State overlap is calculated as

$$|\langle\psi \mid \chi\rangle|^2 = \int W_\psi(\{x_j,p_j\})\, W_\chi(\{x_j,p_j\}) \prod_j dx_j dp_j. \tag{124}$$

The proof in one dimension is as follows. Writing the right side of eq.(124) taking eq.(122) into account we get

$$\begin{aligned} I &= \frac{1}{(2\pi\hbar)^2} \int dx\, dp\, du\, dv\, \psi\left(x+\frac{u}{2}\right) \psi^*\left(x-\frac{u}{2}\right) \\ &\quad \times\chi\left(x+\frac{v}{2}\right) \chi^*\left(x-\frac{v}{2}\right) \exp\left[-ip\,(u+v)/\hbar\right] \\ &= \frac{1}{2\pi\hbar} \int dx\, du\, \psi\left(x+\frac{u}{2}\right) \psi^*\left(x-\frac{u}{2}\right) \chi\left(x+\frac{u}{2}\right) \chi^*\left(x-\frac{u}{2}\right), \end{aligned}$$

where we have performed the p integral and then the v integral. Hence the change of variables $s = x + \frac{u}{2}, w = x - \frac{u}{2}$ leads to

$$I = \frac{1}{2\pi\hbar} \int ds\, dw\, \psi(s)\, \chi^*(s)\, \psi^*(w)\, \chi(w) = |\langle\psi \mid \chi\rangle|^2 .$$

3) **Boundedness**. By virtue of the Cauchy-Schwartz inequality, for a pure state in one dimension W is constrained by

$$-\frac{2}{\hbar} \leq W(x,p) \leq \frac{2}{\hbar}$$

Property eq.(124) leads to the possibility of obtaining the expectation value of an 'observable' operator $\widehat{A}$ in the quantum state with

density operator $\hat{\rho}(t)$ through

$$\text{Tr}\left(\hat{\rho}\widehat{A}\right) = \int W_{\widehat{A}}(\mathbf{r},\mathbf{p})\, W_{\hat{\rho}}(\mathbf{r},\mathbf{p},t)\, d\mathbf{r}\, d\mathbf{p}. \tag{125}$$

This is crucial in order to allow the formulation of quantum mechanics using the Wigner formalism. The proof of eq.(125) is easy taking into account that any observable $\widehat{A}$ may be written in terms of its eigenvalues, a_j, and eigenvectors, $\mid \psi_j\rangle$, as follows

$$\widehat{A} = \sum_j a_j \,|\psi_j\rangle\langle\psi_j|\,.$$

Hence the linearity eq.(117) and the overlap property eq.(124) prove eq.(125) .

Time dependence may be studied with the evolution equation of the Wigner function, which may be obtained from the von Neumann evolution equation for the density operators

$$\frac{d\hat{\rho}}{dt} = \frac{1}{\hbar}\left(\widehat{H}\hat{\rho} - \hat{\rho}\widehat{H}\right), \qquad \widehat{H} = \frac{1}{2m}\hat{\mathbf{p}}^2 + V(\hat{\mathbf{r}})\,. \tag{126}$$

The Weyl-Wigner transform of this equation gives,

$$\frac{\partial W}{\partial t} = -\frac{1}{m}\mathbf{p}\cdot\nabla W - \frac{1}{\hbar}\int \frac{d\mathbf{p}'}{(2\pi\hbar)^3}\widetilde{V}(\mathbf{r},\mathbf{p}')\, W(\mathbf{r},\mathbf{p}+\mathbf{p}',t) \tag{127}$$

$$\widetilde{V}(\mathbf{r},\mathbf{p}') \equiv \int d\mathbf{u} \sin\left(\mathbf{p}'\cdot\mathbf{u}/\hbar^3\right)\left[V(\mathbf{r}+\mathbf{u}/2) - V(\mathbf{r}-\mathbf{u}/2)\right].$$

The proof involves the Weyl transform eq.(111) of the different terms and it is straightforward although lengthy. As an illustration we present the transform of one of the terms in one dimension assuming that $\hat{\rho}$ corresponds to a pure state, i.e. $\hat{\rho} = |\psi_0\rangle\langle\psi_0|$. I will calculate the transform of $\widehat{V}\hat{\rho}$, that is,

$$\begin{aligned} W_{\widehat{V}\hat{\rho}} &= (2\pi\hbar)^{-1}\int \exp\left[(-i/\hbar)(\alpha x+\beta p)\right] d\alpha\, d\beta \\ &\quad\times \text{Tr}\left\{V(\hat{x})\,\hat{\rho}\exp\left[(i/\hbar)(\alpha\hat{x}+\beta\hat{p})\right]\right\} \\ &= (2\pi\hbar)^{-1}\int \exp\left[(-i/\hbar)(\alpha x+\beta p+\alpha\beta/2)\right] T d\alpha\, d\beta, \\ T &\equiv \text{Tr}\left\{V(\hat{x})\,\hat{\rho}\exp(i\alpha\hat{x}/\hbar)\exp(i\beta\hat{p}/\hbar)\right\}. \end{aligned} \tag{128}$$

Now we use a basis of vectors, $\{\mid \psi_j\rangle\}$, and consider that $\hat{\rho} = \mid \psi\rangle\langle\psi \mid$, that is, it represents a pure state. Thus we get

$$\begin{aligned} T &= \sum_j \langle\psi_j \mid V(\hat{x}) \mid \psi\rangle\langle\psi \mid \exp(i\alpha\hat{x}/\hbar)\exp(i\beta\hat{p}/\hbar) \mid \psi_j\rangle \\ &= \langle\psi \mid \exp(i\alpha\hat{x}/\hbar)\exp(i\beta\hat{p}/\hbar)V(\hat{x}) \mid \psi\rangle, \end{aligned}$$

where the projective decomposition of the identity, $\sum_j \mid \psi_j\rangle\langle\psi_j \mid = 1$, has been used. In the coordinates representation this leads to

$$T = \int dx'\psi^*(x')\exp(i\alpha x'/\hbar)V(x'+\beta)\,\psi(x'+\beta)\,dx'.$$

Inserting this into eq.(128) leads, after integrating over α and x',

$$W_{\widehat{V}\hat{\rho}} = \int \exp(-i\beta p/\hbar)d\beta\psi^*(x+\beta/2)\,V\left(x-\frac{1}{2}\beta\right)\psi\left(x-\frac{1}{2}\beta\right),$$

which may be written introducing a Dirac's delta as

$$\begin{aligned} V\left(x-\frac{1}{2}\beta\right) &= \int \delta(\beta-u)\,V\left(x-\frac{1}{2}u\right)du \\ &= \frac{1}{2\pi\hbar}\int dp'\int du\exp\left[i(u-\beta)p'/\hbar\right]V\left(x-\frac{1}{2}u\right). \end{aligned}$$

Hence it follows

$$W_{\widehat{V}\hat{\rho}} = \int dp' W_{\hat{\rho}}(x,p+p')\int du\exp(iup'/\hbar)\,V\left(x-\frac{1}{2}u\right).$$

$W_{\widehat{V}\hat{\rho}}$ gives an analogous result with $x+\frac{1}{2}u$ substituted for $x-\frac{1}{2}u$. The calculation of $W_{\widehat{V}\hat{p}^2}$ and $W_{\hat{p}^2\widehat{V}}$ is more involved but similar. Combining these results leads to the one-dimensional analog of eq.(127) .

It is interesting to write this equation explicitly in powers of Planck's constant, $\hbar$. I will do that in one dimension. Firstly we change the variable $p' = \hbar v$, which leads, after expansion in powers of $\hbar$, to

$$\begin{aligned} \frac{\partial W}{\partial t} &= -\frac{1}{m}p\frac{\partial W}{\partial x} - \frac{1}{2\pi\hbar}\int_\infty^\infty dv\sum_{k=0}^\infty \frac{(\hbar v)^k}{k!}\frac{\partial^k W(r,p,t)}{\partial p^k} \\ &\quad\times\int_\infty^\infty du\sin(vu)\left[V(x+u/2)-V(x-u/2)\right]. \end{aligned} \tag{129}$$

Performing the v integral first, we get zero if k is even, while for $k = 2l+1$ we obtain

$$I_{2l+1}(u) \equiv \int_\infty^\infty v^{2l+1}\sin(\hbar vu)\,dv = -2\pi(-1)^l\frac{d^{2l+1}}{du^{2l+1}}\delta(u),$$

whence, after integrating by parts $2l+1$ times, we obtain

$$\int_{\infty}^{\infty} du\left[V\left(x+u/2\right)-V\left(x-u/2\right)\right]I_{2l-1}\left(u\right)=-\pi\left(-1\right)^{l}\frac{d^{2l+1}}{du^{2l+1}}V(x).$$

If this is put in eq.(129) we get

$$\frac{\partial W}{\partial t}=-\frac{1}{m}p\frac{\partial W}{\partial x}+\sum_{l=0}^{\infty}\frac{\left(-1\right)^{l}\hbar^{2l}}{(2l+1)!}\frac{\partial^{2l+1}V(x)}{\partial x^{2l+1}}\frac{\partial^{2l+1}W(x,p,t)}{\partial p^{2l+1}}. \tag{130}$$

This is called the Moyal equation [**19**], although it is usually written for the more general case where the Hamiltonian's dependence on p is not so simple as eq.(126). In this case,

$$\frac{\partial W}{\partial t}=\left\{\frac{2}{\hbar}\sin\left[\frac{\hbar}{2}\left(\frac{\partial}{\partial x'}\frac{\partial}{\partial p''}-\frac{\partial}{\partial x''}\frac{\partial}{\partial p'}\right)\right]W(x',p',t)H(x'',p'')\right\}_{x,p}, \tag{131}$$

where $\{\}_{x,p}$ means making $x'=x''=x, p'=p''=p$ after performing the derivatives. The right hand side is named Moyal bracket and becomes the classical Poisson bracket in the limit $\hbar\to 0$. In this limit eq.(131) reproduces the classical Liouville equation.

The Wigner function may be extended to field theory, although its meaning is quite different there. In particular, the application to the electromagnetic field (quantum optics) will be studied in detail in chapter 6. There are also other distribution functions in phase space, in particular the Husimi function that is positive semidefinite. However, the Wigner function presents clear advantages over the alternatives, that will not be presented here [**20**], [**21**].

4.4.4. Does the Wigner function allow a realistic interpretation? In classical mechanics, given a collection of particles, the probability of finding a particle at a certain position in phase space is specified by a probability distribution, the Liouville density. This strict interpretation fails in quantum mechanics. Instead, the above quasiprobability Wigner distribution plays an analogous role, but it does not satisfy all the properties of a conventional probability distribution; and, conversely, it satisfies boundedness properties not always true in classical distributions.

For instance, the Wigner distribution can and normally does go negative for states which have no classical model. Smoothing the Wigner distribution through a filter of size larger than h results in a positive-semidefinite function, i.e. it may be thought to have been coarsened to a semi-classical one. Regions of such negative value are

provable (by convolving them with a small Gaussian) to be 'small': they cannot extend to compact regions larger than a few times h, and hence they disappear in the classical limit. They are shielded by the uncertainty principle, which does not allow precise location within phase-space regions smaller than h, and thus renders such 'negative probabilities' less paradoxical. Nevertheless, it is obvious that we cannot interpret W as a probability distribution of the particles in phase space.

We may try to get some hint about the possible paths of particles if we integrate both sides of eq.(127) with respect to $\mathbf{p}$. It is easy to see that the integral of the second term vanishes and we obtain the continuity eq.(85) with the density ρ and current density $\mathbf{j}$ given by

$$\rho = \int W(\mathbf{r},\mathbf{p},t)\mathbf{dp}, \qquad \mathbf{j} = \frac{1}{m}\int \mathbf{p}W(\mathbf{r},\mathbf{p},t)\,d\mathbf{p}. \tag{132}$$

Thus the moments of orders 0 and 1 of the momentum p play the same role as if W was a true probability distribution. However, this is no longer true for higher moments.

As a conclusion the Wigner function does not provide a realistic interpretation of nonrelativistic quantum mechanics. However, we shall see in chapter 6 that the Weyl-Wigner formalism may be applied to the electromagnetic field and provides an alternative to at least some domain within quantum optics that allows a realistic interpretation.

4.5. Path integrals

4.5.1. The Feynman formulation of quantum mechanics. The path integral formulation of quantum mechanics was introduced by Feynman in 1948 [**26**]. In a book on the subject [**27**] the formalism is shown to be a straightforward consequence of the superposition principle with the following argument. Let us assume that there is a source of particles at point $x = x_0, y = 0$, a screen with two slits at $x = \pm a, y = b$, and the particles may be detected at any point in the plane $y = c$. (For simplicity I ignore here the z coordinate). Assuming that a particle leaves the source at time $t = 0$, crosses the screen with the slits at t_1 and arrives at the detector at t_2, the probability, $P(x)$, of reaching a point with coordinate x in the plane $y = c$ is proportional to the squared modulus of a 'probability amplitude', the latter being

the sum of two amplitudes, that is,

$$P(x) \propto |A(x_0, 0 \mid x, t_2)|^2 , \tag{133}$$
$$A(x_0, 0 \mid x, t_2) = A(x_0, 0 \mid a, t_1)A(a, t_1 \mid x, t_2) + A(x_0, 0 \mid -a, t_1)A(-a, t_1 \mid x, t_2).$$

Now we consider the case of many screens at positions $y_1, y_2, \ldots$, where the particle may arrive at times $t_1, t_2, \ldots$, respectively, and that every screen possesses many slits. In this case the amplitude eq.(133) should be written

$$A(x_0, 0 \mid x, t) = \tag{134}$$
$$\sum_j \sum_k \cdots \sum_q A(x_0, 0 \mid x_j, t_1)A(x_j, t_1 \mid x_k, t_2) \ldots A(x_q, t_n \mid x, t).$$

Here the set of positions $\{x_0, x_j, x_k, \ldots, x_q, x\}$ may be called a path, so that the amplitude $A(x_0, 0 \mid x, t)$ is a sum of amplitudes, every one corresponding to one possible path. If the number of slits in every screen is increased indefinitely, at the end there will be no screen at all. Then the discrete values x_l, x_k,... become continuous and the sums become integrals, that is,

$$A(x_0, 0 \mid x, t) = \tag{135}$$
$$\int dx_1 \ldots \int dx_{n-1} A(x_0, 0 \mid x_1, t_1) \ldots A(x_{n-1}, t_{n-1} \mid x, t).$$

The time intervals may be chosen identical, that is, $t_{j+1} - t_j = \varepsilon$, with ε as small as desired. In the limit $\varepsilon \to 0$, $A(x_0, 0 \mid x, t)$ becomes an integral of paths.

The given derivation shows that eq.(135) is not specific of Feynman's formulation of quantum mechanics, but it may be valid for any field fulfilling a superposition property, e.g. classical optics. In this case, however, the $P(x)$ of eq.(133) would not be a probability but a light field intensity, as we will illustrate below with Huygen's principle. Also, an expression formally similar to eq.(134) may be used in the study of random motion when the probability that a particle goes from one position to another one in some time interval is the sum of probabilities corresponding to different paths, a method pioneered by Norbert Wiener in the study of Brownian motion. This will be illustrated below with a derivation of the diffusion law via a path integral. In this case the probability density is given directly, rather than via the squared modulus of an amplitude.

What is specific of Feynman's formulation of Schrödinger equation is the choice of A to be the exponential of $i/\hbar$ times the (classical) Lagrangian, L, of the particle's motion, an idea taken from Dirac [**28**]. For instance, in the case of one-dimensional motion in a potential $V(x)$ we have $L = \frac{1}{2}m\dot{x}^2 - V(x)$ whence

$$A\left(x_{j-1}, t_{j-1} \mid x_j, t_j\right) = \sqrt{\frac{m}{2\pi i\hbar\varepsilon}} \exp(\frac{i\varepsilon}{\hbar}L), \tag{136}$$

$$L \equiv \left[\frac{1}{2}m\left(\frac{x_j - x_{j-1}}{\varepsilon}\right)^2 - \frac{1}{2}V\left(x_{j-1}\right) - \frac{1}{2}V\left(x_j\right)\right]$$

where m is the mass of the particle. (This expression differs from the original one of Feynman [**27**] because I have substituted $\frac{1}{2}V\left(x_{j-1}\right) + \frac{1}{2}V\left(x_j\right)$ for $V\left[\left(x_{j-1} + x_j\right)/2\right]$ for later convenience. Both formulations agree in the limit $\varepsilon \to 0$). It is common to take the continuous limit and write eq.(134) in the form

$$\begin{aligned} A(x_0, 0 \mid x, t) &= \int \mathcal{D}(paths) \exp\left[\frac{i}{\hbar}\int_0^t dt L(x, \dot{x})\right] \\ &= \int \mathcal{D}(paths) \exp\left[\frac{i}{\hbar}\int_0^t dt \left(\frac{1}{2}m\dot{x}^2 - V(x)\right)\right]. \end{aligned} \tag{137}$$

But we should take into account that this symbolic equation actually means the limit $\varepsilon \to 0$ of eq.(135) with A given by eq.(136) .

The amplitude eq.(137) may be calculated explicitly for simple potentials. For illustrative purposes I will reproduce the case of a free particle, i.e. $V(x) = 0$. From eqs.(135) and (136) we obtain

$$A(x_0, 0 \mid x, t) = \tag{138}$$

$$\lim_{\varepsilon\to 0}\left(\frac{m}{2\pi i\hbar\varepsilon}\right)^{n/2} \int dx_1 \cdots \int dx_{n-1} \prod_{j=1}^{n} \exp\left[\frac{im}{2\hbar\varepsilon}\left(x_j - x_{j-1}\right)^2\right],$$

where $x_n = x$. It is straightforward to calculate $A(x_0, 0 \mid x, t)$ by iteration, starting with the integral in x_1, then the integral in x_2, and so on [**27**]. Here I will use an alternative procedure. I change to new variables $u_j = x_j - x_{j-1}$ and integrate over u_1, u_2, ..., u_{n-1}. However, the variables u_j are not independent, but fulfill

$$x - x_0 = u_1 + \cdots + u_n.$$

This condition may be introduced in eq.(138) via a Dirac's delta, that is,

$$\delta\left(u_1+\cdots+u_n-x+x_0\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left[i\zeta\left(u_1+\cdots+u_n-x+x_0\right)\right]d\zeta,$$

where I have used a well known integral representation of the delta distribution. After that, changing the order of the integrals it is possible to perform them as if all u_j were independent, leading to

$$\begin{aligned} A(x_0,0\mid x,t) &= \lim_{\varepsilon\to 0}\left(\frac{m}{2\pi i\hbar\varepsilon}\right)^{n/2}\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left[-i\zeta(x-x_0)\right]d\zeta \\ &\times\left[\int_{-\infty}^{\infty}du\exp\left(\frac{im}{2\hbar\varepsilon}u^2+i\zeta u\right)\right]^n \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left[-i\zeta(x-x_0)-\frac{i\hbar n\varepsilon\zeta^2}{2m}\right]d\zeta \\ &= \sqrt{\frac{m}{2\pi i\hbar t}}\exp\left[\frac{im(x-x_0)^2}{2\hbar t}\right]. \end{aligned} \tag{139}$$

The calculation of the path integrals eqs.(138) or (139) either by iteration [**27**] or via the method presented above lacks mathematical rigour, because the integrals involved are not convergent. The problem may be easily surmounted via including either the factor $\exp\left[-\gamma\sum x_j^2\right]$ or $\exp\left[-\gamma\left(\sum u_j^2-\zeta^2\right)\right]$, respectively, and taking the limit $\gamma\to 0$ after performing the integrals. In the derivation of eq.(139) I have used implicitly this regularization procedure. More relevant is the difficulty to get a physical interpretation of the path integral, which will be discussed below.

In the general case, when $V(x)\neq 0$, the amplitude $A(x_0,0\mid x,t)$ is named the 'propagator'. The interesting property is that the propagator allows getting the wave-function at time t from the wave-function at time 0, that is,

$$\psi(x,t)=\int dx_0\psi(x_0,0)\,A(x_0,0\mid x,t), \tag{140}$$

I omit the general proof but for one dimension it is not difficult. It is enough to check that eq.(140) with the amplitude A given by eq.(139) fulfils Schrödinger equation and that the limit $t\to 0$ of the amplitude gives a Dirac's delta, that is

$$\lim_{t\to 0}A(x_0,0\mid x,t)=\delta(x-x_0).$$

Thus the propagator is the Green's function of the Schrödinger equation. The conclusion is that Feynman's formalism resting on the propagator eq.(137) combined with eq.(140) is equivalent to Schrödinger's wave mechanics.

The path integral formulation can be generalized to $3N$ dimensions, that is, N particles in 3D, and also to relativistic field theory. It has an extremely important role in modern theoretical physics, both because it is well adapted in order to derive general properties, e.g. symmetries, and due to its relevance for actual calculations, as shown by the use of Feynman graphs in covariant perturbation theory [**29**].

The use of Feynman's formalism in order to make perturbative calculations may be understood with an example in one dimension. Let us assume that we must calculate the probability amplitude that a particle placed in the point x_0 at time $t = 0$ arrives to point x at time t. If we consider discrete time, the amplitude is given by eqs.(135) and (136) . Now we assume that the potential consists of two terms, that is,

$$V(x) = V_0(x) + \lambda V_1(x), \tag{141}$$

where the calculation involving $V_0(x)$ is assumed easy and we want to get the general solution as an expansion in powers of the parameter λ. It may be checked that the term of order λ is obtained putting λV_1 in one of the factors in eq.(135) and $V_0(x)$ in all other factors. Thus we get

$$A_1(x_0, 0 \mid x, t) = \frac{i\varepsilon}{\hbar}\lambda \sum_{k=1}^{n-1} \int dx_k A(x_0, 0 \mid x_k) V_1(x_k) A(x_k \mid x, t).$$

Similarly the term of order λ^2 will be

$$A_2(x_0, 0 \mid x, t) = -\frac{\varepsilon^2}{\hbar^2}\lambda^2 \sum_{l=1}^{n-2} \sum_{k=l+1}^{n-1} \int\int dx_l dx_k A(x_0, 0 \mid x_l) V_1(x_l) \times A(x_l \mid x_k) V_1(x_k) A(x_k \mid x, t),$$

and so on. Going to the continuous limit we get an easy method to get the terms of the perturbation expansion. In quantum field theory the procedure is much more involved because there are usually several interacting fields. In this case an efficient method to take account of the different terms is the use of the Feynman graphs, each graph corresponding to a term in the perturbation calculation.

Dealing with calculational aspects lies out of the scope of this book, which is devoted to the physical interpretation of the quantum formalism. For the path integrals the question is studied in section 4.5.4 below.

The use of path integrals is not exclusive of Feynman's formalism. From the mathematical point of view path integrals might be used for any evolution governed by a partial differential equation of first order in time. In physics such equations appear in two types of evolution: 1) The motion of particles that may be represented by a stochastic process, and 2) the evolution of fields. In the former case the condition that the differential equation is first order in time means that the evolution after a time $t+\Delta t$ depends only on the state at time t, with independence of the states at earlier times. This is called Markov property and gives rise to an equation formally similar to eq.(135) . The typical example is Brownian motion or, more generally, diffusion theory, see section 4.5.3 below.

The evolution of fields (e.g. classical optics or acoustics) is usually governed by second order equations whence the above criterion does not apply. However, it is the case that the second order equation involving real numbers may be rewritten as a first order equation involving complex numbers. The use of complex numbers to change second order to first order equations is common in optics, as discussed in the following. Also it may be that this is also the reason for complex functions in Schrödinger's wave mechanics as suggested in the Madelung interpretation studied in section 4.3.1 above.

In the following I will revisit the use of a formula similar to eq.(135) in classical optics and in diffusion theory.

4.5.2. Huygens principle. Historically the antecedent of a formulation in terms of 'path integrals' goes back to Christian Huygens, more than three centuries ago. In fact, Huygens proposed that light are waves (in contrast with the corpuscular theory of his contemporary Isaac Newton) and he was able to explain the straight line propagation and other properties of light from his celebrated principle. Huygens principle states that light propagation may be understood as if every point where light arrives becomes the source of a spherical wave, and the waves coming from different points are able to interfere. In practice this implies that from each point of a given wavefront at time t spherical wavelets originate. At time $t+\Delta t$ each wavelet will have a radius $c\Delta t$ and the envelopes of these wavelets will form a

new wavefront. Huygens principle may be formalized stating that the light arriving at time t at a point $\mathbf{r}$ may be calculated from the three-dimensional generalization of eq.(135) with an appropriate transition amplitudes $A(\mathbf{r}_j, t_j \mid \mathbf{r}_{j+1}, t_{j+1})$. In contrast with the cases of Feynman's formalism, eq.(137) or diffusion (see below eq.(147)), where the velocities of the particles are not fixed, light travels in vacuum with the velocity, c, which puts the constraint $|\mathbf{r}_{j+1} - \mathbf{r}_j| = c(t_{j+1} - t_j)$. Hence the transition amplitude should be of the form

(142)
$$A(\mathbf{r}_j, t_j \mid \mathbf{r}_{j+1}, t_{j+1}) = f(\mathbf{r}_j \mid \mathbf{r}_{j+1})\, \delta(|\mathbf{r}_{j+1} - \mathbf{r}_j| - c(t_{j+1} - t_j)).$$

Light consists of transverse waves (that is, the electric and magnetic fields are perpendicular to the direction of propagation) which implies that the function f should depend also on the angle between the electric field and the $\mathbf{r}_j - \mathbf{r}_{j+1}$ vectors. I shall avoid this complication, irrelevant for our purposes, considering in the following longitudinal waves, e.g. sound propagation in air instead of light in vacuum.

For monocromatic sound waves in air the function f of eq.(142) is specially simple, namely

$$f(\mathbf{r}_j \mid \mathbf{r}_{j+1}) = \frac{\exp(ik\,|\mathbf{r}_{j+1} - \mathbf{r}_j|)}{2\pi i\,|\mathbf{r}_{j+1} - \mathbf{r}_j|}. \tag{143}$$

where the denominator takes into account that the intensity from a point source decreases as the inverse of the distance squared. I stress that here the use of complex amplitudes, i.e. the introduction of the imaginary unit number i, has no deep meaning, it is just a convenient mathematical procedure to simplify the calculations. Actually, the wave amplitudes might be always represented by real numbers (indeed the light amplitude may be taken to be the electric field of the radiation and the amplitude of sound vaves in air to be the excess pressure).

Due to the constraint of constant velocity, see eq.(142), it is not useful to study the case of a source emitting a signal at time $t = 0$ from the point $\mathbf{r} = 0$, as in the Feynman's formalism (or diffusion, see below). We will consider instead that at time $t = 0$ there is an amplitude, $\psi_0(\mathbf{r})$, extended in space and we want the amplitude at time t, $\psi(\mathbf{r},t)$. This may be calculated inserting eqs.(142) and (143) in the path integral formula eq.(135). For instance, we may consider

n equal time intervals $t_{m+1} - t_m = \varepsilon$, thus getting

$$\psi(\mathbf{r},t) = \tag{144}$$
$$= \int ds_1 \cdots \int ds_{n-1} \psi_0(\mathbf{r}_1) \frac{\exp(ik|\mathbf{r}_1 - \mathbf{r}_2|)}{2\pi i |\mathbf{r}_1 - \mathbf{r}_2|} \cdots \frac{\exp(ik|\mathbf{r}_{n-1} - \mathbf{r}|)}{2\pi i |\mathbf{r}_{n-1} - \mathbf{r}|},$$

where ds_m means integration over a spherical surface of radius

$$|\mathbf{r}_{m-1} - \mathbf{r}_m| = c\varepsilon$$

centered at the point $\mathbf{r}_{m-1}$, where c is the sound velocity.

Eq.(144) is a natural formulation of Huygen's principle by means of path integrals. In practice, however, it is most frequent to use very simple paths consisting of just three points, namely $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}\}$, with a straight line propagation between points. For instance, a typical problem solved via eq.(144) is the diffraction by a small hole. In this case we might consider an initial signal represented by

$$\psi_0(\mathbf{r}_1) = F(\mathbf{r}_1) \exp(ikz_1),$$

if the wave is monochromatic. Here $F(\mathbf{r}_1)$ is a slowly varying function of $\mathbf{r}_1$ with support in a domain having a size small in comparison with the distance from the source to the hole, but large in comparison with the wavelength, k^{-1}. The values of $\mathbf{r}_2 \equiv (x_2, y_2, z_2)$ are constrained by the hole. For instance, in a simple model we might write

$$z_2 = d, \qquad x_2^2 + y_2^2 \leq R^2,$$

R being the radius of the hole. Finding the diffraction pattern in a screen after the hole is straightforward although lengthy and will not be given here (see any book of electromagnetic theory, e.g. [**30**]).

4.5.3. Diffusion. We may also derive the diffusion probability via a path integral approach. This is possible provided that the associated stochastic process is Markovian whence the transition probabilities fulfil a Chapman-Kolmogorov equation, that is,

$$P(x_0, 0 \mid x, t) = \int dx P(x_0, 0 \mid x_1, t_1) P(x_1, t_1 \mid x, t). \tag{145}$$

which by iteration leads to an equation similar to eq.(135) but with probabilities P substituted for the amplitudes A. In the simple case of Brownian motion, or the mathematically equivalent problem of diffusion with neither drift nor external forces, we assume that in every step of duration T the particle may travel any distance, Δx, with a

probability proportional to $\exp\left(-\frac{\Delta x^2}{2DT}\right)$. Then the probability of a path going from 0 to x will be

$$P \propto \int dx_1 \int dx_2 \cdots \int dx_{n-1} \exp\left[-\frac{1}{4D}\sum_{k=0}^{n-1}\left(\frac{x_{k+1}-x_k}{t_{m+1}-t_m}\right)^2\right],$$

(146) $\quad x_n = x.$

In the limit of large n, but given $t = nT$, this may be written substituting integrals for the sums, that is,

$$P \propto \int \mathcal{D}(paths) \exp\left(\frac{1}{4D}\int_0^t dt\dot{x}^2\right), \tag{147}$$

which is formally similar to the Feynman path integral eq.(137). There are however two fundamental differences. Firstly in the diffusion case eq.(147) gives directly the probability, rather than the amplitude, and secondly the quantity inside the exponential is real positive, so that the sum involves positive probabilities, rather than complex numbers. The integrals in eq.(146) are rather simple and the result is the Green's function of the diffusion equation

$$\frac{\partial}{\partial t}\rho(0,0 \mid x,t) = D\frac{\partial^2}{\partial x^2}\rho(0,0 \mid x,t).$$

A technique similar to the one illustrated here for diffusion may be applied to any process where the probability of going from an initial state to a final one is the sum of the probabilities of the different paths.

4.5.4. Difficulties for a physical interpretation of Feynman paths. It is sometimes stated that Feynman's path integral formalism provides an intuitive picture of the quantum mechanical evolution, but this is not true. In fact, unlike Huygens' formalism for diffusion via path integrals, Feynman's formalism is counterintuitive for two reasons. Firstly because in an intuitive picture the probability of travel of a particle between two points should be a sum of probabilities rather than amplitudes involving complex numbers. Secondly because Feynman 'paths' are just sets of disconnected points, rather than (continuous) paths in the usual sense of the word. In the following I comment on the last shortcoming that might be less obvious.

The alleged path represented by the successive positions $\{x_0, x_1, x_2, \ldots, x_n, x\}$ is actually a set of points, each one separated from the previous one by a long distance. Indeed, the position x_{j-1} may have

values in the interval of integration $(-\infty, \infty)$, all values having the same weight according to the Riemann measure of the integral (see eq.(135)). Similarly the position x_j may have values in the interval $(-\infty, \infty)$ all values having the same weight. Thus the step $u_j = x_j - x_{j-1}$ between two positions may be *arbitrarily large.* Indeed, its mean square value, $\langle u_j^2 \rangle$, diverges. The counterintuitive character of this is enhanced by the fact that the (indefinitely long) step u_j *takes place in an infinitesimal time interval* ε. Thus Feynman's eq.(135) should be seen as a purely mathematical construction, a useful calculational tool where the physical meaning appears only in the final result of the calculation.

In sharp contrast, Huygens' path integrals, eqs.(142), are continuous. In fact, the quantities $|\mathbf{r}_{j+1} - \mathbf{r}_j|$ are never too large due to the denominator in eq.(143) and, above all, they decrease to zero when the time interval, $t_{j+1} - t_j$, goes to zero. Thus the sum or integral involved is over *continuous paths.* A similar thing happens in the diffusion problem defined via eq.(146) where the probability that the mean velocity in a step, $(x_{k+1} - x_k)/(t_{m+1} - t_m)$, surpasses a value K goes exponentially to zero when $K \to \infty$.

From a physical rather than formal point of view, Feynman's path integral is more similar to Huygen's principle of classical optics than to the diffusion problem. Indeed, the use of an amplitude suggests a wave picture, although the fact that the squared modulus of the amplitude is a probability, rather than an intensity, gives a particle appearance. However, the particle picture is misleading. In fact, in spite of the frequent use of the expression 'probability amplitude' in quantum theory, it is not possible to give a real meaning to these two words in order to get a realistic interpretation of quantum mechanics. In field theory the counterintuitive character of Feynman's path integrals is less obvious if we assume that an amplitude squared represents field intensity, i.e. energy flux per unit time. Then the apparent particle behaviour at detection might be interpreted assuming that the detection probability is proportional to the intensity arriving at the detector during some activation time interval. See chapter 6 for an extended discussion for the case of quantum radiation field.

4.6. Motion of particles in terms of path probabilities

In the following I propose an elaboration of Feynman's formalism that leads to a transition probability $P(a \to b)$ as a sum of path's weights. The idea is to substitute a single path of the new formalism

for an equivalence class of pairs of Feynman paths, in such a way that the sum (or integral) of the probabilities of the former paths gives directly the transition probability, rather than the transition amplitude as in the original Feynman's proposal.

This formalism may allow getting predictions for the results of some experiments in agreement with Feynman and Schrödinger formalisms. However, the advantage of the new formalism is not calculational, its virtue is that it offers a realistic interpretation as a stochastic picture of the quantum motion. Also it might allow predictions going beyond standard quantum mechanics, that might be tested empirically. Some of them will be mentioned in section 4.6.2 below.

4.6.1. Transition probability in terms of particle paths.

General formulation. In the Feynman path integral formulation the transition probability is the squared modulus of the transition amplitude. That is, from eq.(137) we have

$$\begin{aligned} P(\mathbf{r}_a, t_a \;\; \rightarrow \;\; \mathbf{r}_b, t_b) &= |A(\mathbf{r}_a, t_a \rightarrow \mathbf{r}_b, t_b)|^2 \\ &= \int \mathcal{D}(\text{path pairs}) \exp\left[i \int_{t_a}^{t_b} dt \left(\frac{1}{2}\dot{\mathbf{y}}^2 - V(\mathbf{y}) \right) \right] \\ &\quad \times \exp\left[-i \int_{t_a}^{t_b} dt \left(\frac{1}{2}\dot{\mathbf{x}}^2 - V(\mathbf{x}) \right) \right], \end{aligned} \tag{148}$$

where $\mathbf{x}_0 = \mathbf{y}_0 = \mathbf{r}_a$, $\mathbf{x}_n = \mathbf{y}_n = \mathbf{r}_b$, and $\int \mathcal{D}$ represents a functional integral. Taking eq.(136) into account the symbolic eq.(148) may be written explicitly as

$$P = \lim_{\varepsilon \to 0} \left(\frac{m}{2\pi\hbar\varepsilon} \right)^{3n} \int d\mathbf{x}_{n-1} \int d\mathbf{y}_{n-1} \cdots \int d\mathbf{x}_1 \int d\mathbf{y}_1 \prod_{j=1}^{n} \exp L_j \tag{149}$$

$$L_j \equiv \frac{im}{2\varepsilon\hbar} \left[(\mathbf{y}_j - \mathbf{y}_{j-1})^2 - (\mathbf{x}_j - \mathbf{x}_{j-1})^2 \right] - \frac{i\varepsilon}{\hbar} \left[V(\mathbf{x}_j) - V(\mathbf{y}_j) \right],$$

where $\int d\mathbf{x}_j$, $\int d\mathbf{y}_j$ are (triple) integrals over the whole space and I have included the parameters m and $\hbar$ following Feynman [**27**]. With this choice the quantity $P(a \rightarrow b)$ has dimensions of probability per square volume and it should be interpreted as a relative probability. I shall point out that we use the words 'transition probability' in a wide

sense, but in a restricted sense they are used only when (Chapman-Kolmogorov) equations of the form of eq.(146) hold, which is not the case here. See section 4.5.3 above.

Our aim is to obtain, from eqs.(149), a formulation for the motion of a particle in terms of probabilities (rather than amplitudes!) of paths; that is, to get the transition probability as a sum of probabilities of paths, that we may write as

$$P(\mathbf{r}_a, t_a \to \mathbf{r}_b, t_b) = \sum_k W_k(\mathbf{r}_a, t_a \to \mathbf{r}_b, t_b), \tag{150}$$

where every value of the index k corresponds to a possible path of the particle with end points $(\mathbf{r}_a, t_a)$ and $(\mathbf{r}_b, t_b)$. Our problem is to find positive 'weights' W_k in order to be able to interpret them as probabilities, which would provide an intuitive picture of the quantum evolution as a random motion of particles.

In this section I will study a simple method to get 'weights' W_k such that eq.(150) is fulfilled, ignoring for the moment the positivity requirement $W_k \geq 0$. I start from eq.(149) and make a change of variables, that is,

$$\mathbf{r}_j = \frac{1}{2}\left(\mathbf{x}_j + \mathbf{y}_j\right), \quad \mathbf{u}_j = (\mathbf{x}_j - \mathbf{y}_j)/\hbar, \quad 0 \leq j \leq n-1. \tag{151}$$

Hence eq.(149) becomes, reordering the terms,

$$\begin{aligned} P(\mathbf{r}_a, t_a \to \mathbf{r}_b, t_b) &= \lim_{\varepsilon \to 0} \left(\frac{m}{2\pi\hbar\varepsilon}\right)^3 \int d\mathbf{r}_{n-1} \cdots \int d\mathbf{r}_1 \prod_{j=1}^{n-1} Q_j, \\ Q_j &= \left(\frac{m}{2\pi\hbar\varepsilon}\right)^3 \int d\mathbf{u}_j \exp\left[-\frac{im}{\varepsilon}\mathbf{u}_j \cdot (\mathbf{r}_{j-1} - 2\mathbf{r}_j + \mathbf{r}_{j+1})\right] \\ &\quad \times \exp\left\{\frac{i\varepsilon}{\hbar}\left[V(\mathbf{r}_j - \frac{\hbar}{2}\mathbf{u}_j) - V(\mathbf{r}_j + \frac{\hbar}{2}\mathbf{u}_j)\right]\right\}, \end{aligned} \tag{152}$$

where $\mathbf{r}_0 = \mathbf{r}_a$, $\mathbf{r}_n = \mathbf{r}_b$ and $\mathbf{u}_0 = \mathbf{u}_n = 0$.

A sufficient, although not necessary, condition for the positivity of W_k is that all quantities Q_j are positive, a question that is studied below.

Classical limit, free particle and quadratic potentials. Before proceeding with the general treatment we exhibit a few simple examples. For the quantum free particle the potential is a constant, whence eq.(152) with $V = 0$ may be easily integrated via a method similar to eq.(139) . However, we shall use a different more convenient procedure here, although lacking mathematical rigour, that involves Dirac

deltas. After performing the $\mathbf{u}_j$ integrals we get

$$P(\mathbf{a}\to\mathbf{b}) = \tag{153}$$

$$\lim_{\varepsilon\to 0}\left(\frac{m}{2\pi\hbar\varepsilon}\right)^3\int d\mathbf{r}_{n-1}\ldots\int d\mathbf{r}_1\prod_{j=1}^{n-1}\delta^3\left(\frac{\mathbf{r}_{j-1}-2\mathbf{r}_j+\mathbf{r}_{j+1}}{\varepsilon^2}\right).$$

The $\mathbf{r}_j$ integrals are now easy and we obtain

$$P(\mathbf{a}\to\mathbf{b}) = \lim_{\varepsilon\to 0}\left(\frac{m}{2\pi\hbar\varepsilon}\right)^3 n^{-3} = \left[\frac{m}{2\pi\hbar\,(t_b-t_a)}\right]^3.$$

The final result says nothing interesting although it shows the need of the global factor ε^{-3} in eq.(152), in order to get a finite result when $\varepsilon\to 0$. Actually, the interesting information of the formalism appears before integrating the coordinates $\{\mathbf{r}_j\}$, namely to provide the probabilities of different paths. Indeed, eq.(153) predicts that the particle travels with certainty in straight line with constant velocity.

When the potential is quadratic we also get a single path with probability one, namely the classical path. In fact, let us assume the potential

$$V = \mathbf{A}\cdot\mathbf{r} + B\mathbf{r}^2, \tag{154}$$

where $\mathbf{A}$ and B are constant, vector and scalar respectively. Then the last exponential in eq.(152) becomes

$$\exp\left\{\frac{i\varepsilon}{\hbar}\left[V(\mathbf{r}_j-\frac{\hbar}{2}\mathbf{u}_j)-V(\mathbf{r}_j+\frac{\hbar}{2}\mathbf{u}_j)\right]\right\}$$
$$=\exp\left\{-i\varepsilon\left(\mathbf{A}\cdot\mathbf{u}_j+2B\mathbf{r}_j\cdot\mathbf{u}_j\right)\right\},$$

whence performing the $\mathbf{u}_j$ integrals gives the classical equation of motion. This might provide an explanation as to why the quantum harmonic oscillator has a behaviour sometimes resembling the classical motion (e.g. in the coherent states). See also chapter 5 sections 5.2 and 5.3.

The classical limit $\hbar\to 0$ in eq.(152) gives, after performing the $\mathbf{u}_j$ integrals,

$$W \propto \prod_{j=1}^{n-1}\delta^3\left(\frac{\mathbf{r}_{j-1}-2\mathbf{r}_j+\mathbf{r}_{j+1}}{\varepsilon^2}+\frac{1}{m}\nabla\cdot V\left(\mathbf{r}_j\right)\right), \tag{155}$$

where the argument of the three-dimensional Dirac's delta leads again to Newton's law in the limit $\varepsilon\to 0$. There is a single path corresponding to the classical motion, all other possible paths having zero

probability. This is a form of showing the passage from quantum to classical mechanics in the limit $\hbar \to 0$.

Path weights in terms of the Fourier transform of the potential. From now on I shall use units such that $m = \hbar = 1$ for simplicity of writing, but these parameters will be restored sometimes for clarity. For later developments it is convenient to introduce velocity variables

$$\begin{aligned} \mathbf{v}_j =& \frac{\mathbf{r}_{j+1} - \mathbf{r}_j}{\varepsilon}, \\ \mathbf{s}_j =& \mathbf{v}_j - \mathbf{v}_{j-1} = (\mathbf{r}_{j+1} - 2\mathbf{r}_j + \mathbf{r}_{j-1})/\varepsilon, \quad j = 1, \ldots, n-1. \end{aligned} \tag{156}$$

The quantity $\mathbf{s}_j$ has the physical meaning of the velocity change, so that $\mathbf{s}_j/\varepsilon$ might be interpreted as the mean acceleration in the time interval ε. However, the limit of that quantity for $\varepsilon \to 0$ does not exist in general, that is, the instantaneous acceleration may not be well defined. Indeed, the ratio $\mathbf{s}_j/\varepsilon$ may diverge in that limit. It may be realized that in the definition of a path there are $n+1$ coordinates $\{\mathbf{r}_j\}$, n velocities $\{\mathbf{v}_j\}$ and $n-1$ velocity changes $\{\mathbf{s}_j\}$ In order to define a path it is necessary to choose $n-1$ of either, in addition to the fixed end points. The initial and final velocities of the path correspond to $\mathbf{v}_a = \mathbf{v}_0, \mathbf{v}_b = \mathbf{v}_{n-1}$ with the definitions of eqs.(156).

With these new variables we may write the Q_j of eq.(152) (with $m = \hbar = 1$) in the form

$$\begin{aligned} Q_j(\mathbf{r}_j, \mathbf{s}_j) \quad \equiv \quad & \frac{1}{8\pi^3} \int d\mathbf{u}_j \exp(-i\mathbf{u}_j \cdot \mathbf{s}_j) \\ & \times \exp\left\{ i\varepsilon \left[V(\mathbf{r}_j - \frac{1}{2}\mathbf{u}_j) - V(\mathbf{r}_j + \frac{1}{2}\mathbf{u}_j) \right] \right\}. \end{aligned} \tag{157}$$

The advantage of eq.(157) in comparison with eq.(152) is that it will allow studying the positivity of the quantities Q_j^k as functions of two variables, $\mathbf{r}_j$ and $\mathbf{s}_j$, ignoring the fact that actually the variables $\mathbf{r}_j$ and $\mathbf{s}_j$ are not independent but related by eqs.(156) .

A more convenient description of the path weights is possible, leading however to the same total transition probability. In fact, I will show that a form equivalent to eq.(152), modulo terms of order $O(\varepsilon)$, is obtained substituting the following for eq.(157)

$$\begin{aligned} Q_j(\mathbf{r}_j, \mathbf{s}_j) \quad = \quad & \frac{1}{8\pi^3} \int d\mathbf{u}_j \exp[-i\mathbf{u}_j \cdot \mathbf{s}_j] \\ & \left\{ 1 + i\varepsilon \left[V(\mathbf{r}_j - \frac{1}{2}\mathbf{u}_j) - V(\mathbf{r}_j + \frac{1}{2}\mathbf{u}_j) \right] \right\}. \end{aligned} \tag{158}$$

The equivalence is proved as follows, labelling for short

$$B_j \equiv V(\mathbf{r}_j - \frac{1}{2}\mathbf{u}_j) - V(\mathbf{r}_j + \frac{1}{2}\mathbf{u}_j).$$

Expanding in powers of ε the last exponential in eq.(157) we have

$$\begin{aligned} I &\equiv \prod_{j=1}^{n-1} \exp\{i\varepsilon B_j\} = \prod_{j=1}^{n-1} \left\{ \frac{1}{k!} \sum_{k=0}^{\infty} (i\varepsilon B_j)^k \right\} \\ &= \left(1 + i\varepsilon B_1 + \frac{(i\varepsilon)^2}{2} B_1^2 + \frac{(i\varepsilon)^3}{6} B_1^3 + \dots \right) \\ &\quad \times \left(1 + i\varepsilon B_2 + \frac{(i\varepsilon)^2}{2} B_2^2 + \dots \right) \times \dots \end{aligned} \tag{159}$$

Grouping terms by increasing powers of ε we get

$$\begin{aligned} I = 1 &+ i\varepsilon \sum_{j=1}^{n-1} B_j - \varepsilon^2 \left(\sum_{j=1}^{n-1} \sum_{l=1}^{j-1} B_j B_l + \frac{1}{2} \sum_{j=1}^{n-1} B_j^2 \right) \\ &- i\varepsilon^3 \left(\sum_{j=1}^{n-1} \sum_{l=1}^{j-1} \sum_{i=1}^{l-1} B_j B_l B_i + \frac{1}{2} \sum_{j=1}^{n-1} \sum_{l=1}^{j-1} (B_j B_l^2 + B_j^2 B_l) + \frac{1}{6} \sum_{j=1}^{n-1} B_j^3 \right) \\ &+ \dots, \end{aligned}$$

where $l \neq j$, $i \neq j$, $i \neq l$. It may be realized that, asides from unity, there are two different kinds of terms: Terms where every B_j appears to zeroth or first power, and terms containing some B_j to a power greater than one. The relevant result is that only terms of the former class may contribute in the limit $\varepsilon \to 0$, provided that all terms B_j are finite (or zero). For instance, the sum $\sum_j B_j$ consists of $n-1$ terms whence the quantity $\varepsilon \sum_j B_j$ may have a finite (or zero) limit when $\varepsilon \to 0$ because we assume $n\varepsilon = t_b - t_a$, finite. Similarly there are $(n-1)(n-2)/2$ terms in the double sum $\sum_l \sum_{j>l} B_l B_j$ so that its product times ε^2 may have also a finite limit. However, in a sum like $\sum_j (i\varepsilon)^k B_j^k$ with $k > 1$ there are $(n-1)$ terms times $(i\varepsilon)^k$, a product that goes to zero in the limit $\varepsilon \to 0$. The argument may be extended to all terms, which proves the equivalence between eqs.(158) and (157) .

The u_j integral involved in $Q_j(\mathbf{r}_j, \mathbf{s}_j)$, eq.(157), is not convergent. Indeed, the integrand oscillates rapidly when $u_j \to \infty$ so that the upper limit of the integral is not well defined. The difficulty may be solved including a regularization factor, depending on a parameter γ,

that goes to unity in the limit $\gamma \to 0$, as discussed in the following. A convenient factor is $\exp(-\gamma |\mathbf{u}_j|), \gamma > 0$ that resembles the regularization introduced before eq.(139) in order to make the integrals that appear in Feynman's calculation of the transition amplitude convergent. Thus we shall perform all the calculations with a finite $\gamma > 0$ and take the limit $\gamma \to 0$ at the end. Therefore, we substitute the following for eq.(152)

$$P(\mathbf{r}_a, t_a \to \mathbf{r}_b, t_b) = \lim_{\varepsilon \to 0} \frac{\varepsilon^{-3}}{8\pi^3} \int d\mathbf{r}_{n-1} \cdots \int d\mathbf{r}_1 \prod_{j=1}^{n-1} R_j (\mathbf{r}_j, \mathbf{s}_j), \tag{160}$$

whence, after performing the $\mathbf{u}_j$ integral taking the factor $\exp(-\gamma |\mathbf{u}_j|)$ into account, we get

$$R_j (\mathbf{r}_j, \mathbf{s}_j; \gamma) \equiv D_j (\mathbf{r}_j, \mathbf{s}_j; \gamma) + F_j (\mathbf{r}_j, \mathbf{s}_j; \gamma), \tag{161}$$

(162)

$$D_j = \frac{\pi^{-2}\gamma}{\left(\gamma^2 + |\mathbf{s}_j|^2\right)^2},$$

$$F_j \equiv \frac{i\varepsilon}{8\pi^3} \int d\mathbf{u} \exp(-i\mathbf{u} \cdot \mathbf{s}_j - \gamma |\mathbf{u}_j|) \times \left[V(\mathbf{r}_j - \frac{1}{2}\mathbf{u}) - V(\mathbf{r}_j + \frac{1}{2}\mathbf{u}) \right].$$

As a conclusion we shall substitute the following for eq.(160)

$$P(a \to b) = \lim_{\gamma \to 0} \lim_{\varepsilon \to 0} \frac{\varepsilon^{-3}}{8\pi^3} \int d\mathbf{r}_{n-1} \cdots \int d\mathbf{r}_1 \prod_{j=1}^{n-1} R_j, \tag{163}$$

where R_j is given by eqs.(161) and (162) with $\mathbf{s}_j = \mathbf{v}_j - \mathbf{v}_{j-1}$. This will be the starting point for the applications of our formalism.

Actually, the integrals involved in F_j are convergent for $\gamma = 0$ provided that the potential $V(\mathbf{r}_j)$ goes to zero for $\mathbf{r}_j \to 0$ at least as rapidly as $|\mathbf{r}_j|^{-3}$, which follows from the constraint eq.(166), see below. Therefore, in practice we may substitute the following for F_j in eq.(162)

$$F_j \equiv F(\mathbf{r}_j, \mathbf{s}_j; 0) = \frac{2}{\pi^3} \mathrm{Im} \left[\tilde{V}(2\mathbf{s}_j) \exp(-2im\mathbf{r}_j \cdot \mathbf{s}_j / \hbar) \right], \tag{164}$$

where $\mathbf{r}_j$and $\mathbf{s}_j$ are related via eqs.(156) and $\tilde{V}(2\mathbf{s})$ is the Fourier transform of the potential, here defined by

$$\tilde{V}(2\mathbf{s}) \equiv \int V(\mathbf{x}) \exp(2mi\mathbf{x} \cdot \mathbf{s}/\hbar)\, d\mathbf{x}. \tag{165}$$

Positivity of the path weights. The transition probability $P(a \to b)$, eq.(163), is equivalent to the squared modulus of the Feynman amplitude eq.(137) for $\gamma \to 0$. However, the functions R_j are not positive definite in general. Therefore, it is not obvious that the weight of a path is positive, which would be necessary in order to allow an interpretation of the weight as the relative probability of the path consisting of the positions $\{\mathbf{r}_j\}$ (or the velocities $\{\mathbf{v}_j\}$ related to the positions via eqs.(156)). The difficulty is solved for physically realistic potentials that I define as follows:

1. The space integral of the absolute value of the potential $|V(\mathbf{r})|$ is bounded, i.e. there is a positive constant K such that

$$\int |V(\mathbf{r})|\, d\mathbf{r} \leq K. \tag{166}$$

2. The gradient of the potential ∇V is also bounded, that is,

$$|\nabla V(\mathbf{r})| \leq L. \tag{167}$$

Then positivity is guaranteed by the following result.

THEOREM 10. *For every fixed value of $\gamma > 0$ there is a parameter $\varepsilon_0 > 0$ such that the quantity*

$$R_j(\mathbf{r}_j, \mathbf{s}_j; \gamma) \equiv \frac{\pi^{-2}\gamma}{\gamma^2 + |\mathbf{s}_j|^2} + F(\mathbf{r}_j, \mathbf{s}_j; \gamma), \tag{168}$$

with F given by eq.(162) is not negative whenever $\varepsilon < \varepsilon_0$.

Proof. Eq.(166) implies a bound on the function $F(\mathbf{r}_j, \mathbf{s}_j; \gamma)$, eq.(162), namely

$$\begin{aligned} |F(\mathbf{r}_j, \mathbf{s}_j; \gamma)| &= \frac{\varepsilon}{4\pi^3} \left| \int d\mathbf{u} \sin(\mathbf{u}\cdot\mathbf{s}_j) \exp(-\gamma|\mathbf{u}|)\, V(\mathbf{r}_j - \frac{1}{2}\mathbf{u}) \right| \\ &\leq \frac{\varepsilon}{4\pi^3} \int d\mathbf{u} \left|\sin(\mathbf{u}\cdot\mathbf{s}_j)\right| \exp(-\gamma|\mathbf{u}|) \left| V(\mathbf{r}_j - \frac{1}{2}\mathbf{u}) \right| \\ &\leq \frac{\varepsilon}{4\pi^3} \int d\mathbf{u} \left| V(\mathbf{r}_j - \frac{1}{2}\mathbf{u}) \right| = \frac{\varepsilon}{\pi^3} \int d\mathbf{z} \, |V(\mathbf{z})| \leq \frac{\varepsilon}{\pi^3} K, \end{aligned}$$

where the equality takes into account that $V(\mathbf{r}_j - \frac{1}{2}\mathbf{u}) - V(\mathbf{r}_j + \frac{1}{2}\mathbf{u})$ is an odd function of $\mathbf{u}$ so that only the odd part of $\exp(-i\mathbf{u}\cdot\mathbf{s}_j - \gamma|\mathbf{u}|)$ contributes. The first inequality takes into account that the integral of a function consisting of a product of terms is not greater than the integral of the product of the moduli of these terms. The last

inequality follows after removing from the integral a term not greater than unity. Now the quantity R_j, eq.(168), is positive if

$$\frac{\pi^{-2}\gamma}{\left(\gamma^2+|\mathbf{s}_j|^2\right)^2} \geq \frac{\varepsilon}{\pi^3}K \geq |F(\mathbf{r}_j,\mathbf{s}_j;\gamma)|, \tag{169}$$

which is fulfilled whenever

$$\varepsilon \leq \varepsilon_0 = \frac{\pi\gamma}{K\left(\gamma^2+|\mathbf{s}_j|^2\right)^2}. \tag{170}$$

The proof of positivity for $R_j(\mathbf{r}_j,\mathbf{s}_j;\gamma)$ requires that $\mathbf{s}_j$ remains finite in the limit $\varepsilon \to 0$ so that the left side of eq.(169) remains finite (not zero). However, if $\mathbf{s}_j$ diverges in that limit the first inequality eq.(169) may not hold true. In this case we shall use the equality

$$V(\mathbf{r}_j-\frac{1}{2}\mathbf{u})-V(\mathbf{r}_j+\frac{1}{2}\mathbf{u}) = \mathbf{u}\cdot\nabla V(\mathbf{r}_j)+O\left(|u|^2\right)$$

whence eq.(162) may be written as follows

$$\begin{aligned} F(\mathbf{r}_j,\mathbf{s}_j;\gamma) &\equiv \\ &\equiv -i\frac{\varepsilon}{8\pi^3}\int d\mathbf{u}\exp\left(-i\mathbf{u}\cdot\mathbf{s}_j-\gamma|\mathbf{u}_j|\right)\left[\mathbf{u}\cdot\nabla V(\mathbf{r}_j)+O\left(|u|^2\right)\right] \\ &= \frac{3\varepsilon\pi^{-2}|\mathbf{s}_j|^4}{\left(\gamma^2+|\mathbf{s}_j|^2\right)^4}\mathbf{n}\cdot\nabla V(\mathbf{r}_j)+O\left(|\mathbf{s}_j|^{-5}\right), \end{aligned}$$

where $\mathbf{n}$ is a unit vector in the direction of $\mathbf{s}_j$. Therefore, taking eq.(167) into account, we have

$$\begin{aligned} \frac{\pi^{-2}\gamma}{\left(\gamma^2+|\mathbf{s}_j|^2\right)^2} &\simeq \pi^{-2}\gamma|\mathbf{s}_j|^{-4} \geq |F(\mathbf{r}_j,\mathbf{s}_j;\gamma)| \\ &\simeq 3\varepsilon\pi^{-2}|\mathbf{s}_j|^{-4}|\mathbf{n}\cdot\nabla V(\mathbf{r}_j)|+O\left(|\mathbf{s}_j|^{-5}\right) \end{aligned}$$

for small ε and large $|\mathbf{s}_j|$ so that again $R_j \geq 0$. This completes the proof.

Actually, the velocity change $|\mathbf{s}_j|$ cannot be larger than twice the velocity of light, c, but it would be inconsistent to include this relativistic condition in our nonrelativistic treatment. Therefore, in the proof of the theorem we have taken that possibility into account studying the case where $\mathbf{s}_j$ diverges for $\varepsilon \to 0$. Nevertheless, the probability

of paths where $|\mathbf{s}_j| > 2c$ should be small in order for the nonrelativistic approximation to be reliable. I assume that this is the case for realistic examples, but the issue will not be investigated further here.

This result implies that in the limit $\varepsilon \to 0$ we may write, instead of eq.(163) ,

$$P_\gamma(a \to b) = \sum_k W_k^\gamma(\mathbf{r}_a, t_a \to \mathbf{r}_b, t_b), \tag{171}$$

$$W_k \propto \lim_{\gamma \to 0} \prod_{j=1}^{n-1} R_j(\mathbf{r}_j, \mathbf{s}_j; \gamma),$$

where $k \to \{\mathbf{r}_0, t_0; \ldots \mathbf{r}_j, t_j; \ldots \mathbf{r}_n, t_n\}$. As $W_k^\gamma \geq 0$ for any $\gamma > 0$ such that $\varepsilon \leq \varepsilon_0(\gamma)$, we may take the limit $\gamma \to 0$ provided that this is made *after* the limit $\varepsilon \to 0$, whence all weights W_k are positive or zero.

I point out that the bounds eqs.(166) and (167) are not fulfilled for strong 'classical forces' (i.e. large gradients of the potential) or if $V(\mathbf{r})$ does not go to zero rapidly enough for $|\mathbf{r}| \to \infty$. This is the case for instance, for quadratic potentials like eq.(154), which nevertheless gives nonnegative weight for all paths. This does not contradict Theorem 10 which provides sufficient, but not necessary, conditions of positivity. In any case, the bounds eqs.(166) and (167) may be considered plausible physical requirements for a nonrelativistic treatment.

4.6.2. A picture of quantum motion. The aim of the formalism here developed is to provide a realistic description of (nonrelativistic) quantum motion, which is represented by the possible paths with some probability every one, given by eqs.(163) and (161) . Systems where the spin or the associated statistical effects are relevant will be excluded, for instance the motion of electrons in an atom. Also the validity is restricted to a domain of quantum mechanics, namely when the forces are conservative (i.e. the Hamiltonian is the sum of the kinetic energy plus a potential energy $V(\mathbf{x})$) and the potential is bounded. The description might be extended to systems of many particles, but in the following I will restrict attention to a single particle. This formalism suggests a picture of nonrelativistic quantum particles without any explicit reference to waves.

In classical mechanics if the initial position $\mathbf{r}_a$ of the particle at time t_a and the final position $\mathbf{r}_b$ at time t_b are given, then Newton equation allows the determination of the path $\mathbf{r}(t)$ for the intermediate

times. In the formalism here studied the particle may follow many different paths and the aim is to determine the relative probability of each path. As we have shown above the probability of every path is positive or zero. Thus we may ascribe to the motion a probability measure in the space of all paths, but in practical calculations we may consider paths defined by a finite set of spacetime points each and calculate the relative probability of the different paths so defined.

The formalism may provide information not given by the standard quantum formalism, for instance the probability that the particle's position is in some space region at a given intermediate time t_c. To get this information it is enough to perform the integral of the position $\mathbf{r}_c$, eq.(163), over the desired region rather than in the whole space. In particular, this might provide knowledge about the probabilities of passage of a particle for each slit in an interference experiment.

In the following I will study another relevant information that may be obtained from the formalism, namely the probability $P(\mathbf{r}_a, \mathbf{v}_a; \mathbf{v}_b)$ that the final velocity is $\mathbf{v}_b$, at time t_b, for a particle that started in position, $\mathbf{r}_a$, with velocity $\mathbf{v}_a$ at time t_a. This may be obtained from eq.(163) as follows. Firstly we avoid performing the integrals with respect to $\mathbf{r}_1$ and $\mathbf{r}_{n-1}$ but fix these positions at the values $\mathbf{r}_1 = \mathbf{r}_a + \varepsilon \mathbf{v}_a$ and $\mathbf{r}_{n-1} = \mathbf{r}_b - \varepsilon \mathbf{v}_b$, respectively. Then we change variables taking eqs.(156) into account so that an integral with respect to $\mathbf{v}_{j-1}$ is substituted for the integral with respect to $\mathbf{r}_j$ for $j = 2, 3, \ldots, n-2$. Finally we substitute a parameter C for $(2\pi\varepsilon)^{-3}$ in front of the integrals. This parameter should be obtained from the condition that $P(\mathbf{r}_a, \mathbf{v}_a; \mathbf{v}_b)$ becomes a normalized probability distribution of the variable $\mathbf{v}_b$. Thus we get

$$(172)\quad P(\mathbf{r}_a, \mathbf{v}_a; \mathbf{v}_b) = C \int d\mathbf{v}_{n-3} \int d\mathbf{v}_{n-4} \cdots \int d\mathbf{v}_1 \prod_{j=1}^{n-1} R_j \{\mathbf{r}_j, \mathbf{s}_j\},$$

with R_j given by eqs.(160) to (165) . This may be the starting point for some applications, for instance the derivation of Born's approximation for scattering, reported in the following.

Born approximation for scattering. The Born approximation for scattering is one of the most useful applications of nonrelativistic quantum mechanics. In the following I will rederive Born's formula from the formalism proposed. The purpose is not to make a new

derivation but to illustrate our approach providing an intuitive picture of the phenomenon, which is not offered by the standard quantum formalism.

We may consider a scattering experiment with the following description. There are three regions namely the source, the target and the detector, separated by macroscopic distances from each other. A particle from the source may move freely (see eq.(155)) to the target region where the potential, $V(\mathbf{r})$, is different from zero. In that region the velocity will change and then the particle will travel until the detector again with constant velocity. We fix the origin of the coordinate system at a point on the target and assume that the particle leaves the source from the position $(0, 0, -z_0)$, with a velocity $\mathbf{v}_a$ in the direction of the Z axis, that is, $\mathbf{v}_a(0, 0, v_a)$. A more convenient model for the calculation is to consider a probability distribution of the particle with the said initial velocity $\mathbf{v}_a$, but distributed in space according to a number density $\rho(\mathbf{r}_a)$ that we will assume uniform within a narrow slab along the plane XY, that is,

$$\rho = \rho(z_a) \tag{173}$$

independent of the initial coordinates x_a, y_a within the slab, zero outside. In practice, the probability distribution of the particle position might be seen as an appropriate ensemble of particles.

Once in the target region the particle suffers accelerations by the action of the potential, leaving the region and eventually arriving at the detector with velocity $\mathbf{v}_b$ at time t_b. A particle may follow different paths, and we are interested in the probability density, $P(\mathbf{v}_b)$, eq.(172), for the final velocity.

The quantity to be derived is the differential cross section, $\sigma(\theta, \phi)$, defined by

$$\sigma(\theta, \phi) = \frac{R(\theta, \phi)}{\Phi_{in}} = \frac{R(\theta, \phi)}{\rho v_a}, \tag{174}$$

where Φ_{in} is the incoming probability flux, $\Phi_{in} = \rho v_a$ in our model. $R(\theta, \phi)$ is the probability per unit time that a particle leaves the target with a velocity vector pointing in a given solid angle. It may be got from $P(\mathbf{v}_b)$ as follows

$$R(\theta, \phi) = \int_0^\infty v_b^2 dv_b \int_{-\infty}^\infty dx_a \int_{-\infty}^\infty dy_a \rho v_a P(\mathbf{r}_a, \mathbf{v}_a; \mathbf{v}_b). \tag{175}$$

The integral of the flux times $P(\mathbf{r}_a, \mathbf{v}_a; \mathbf{v}_b)$ gives the total probability per unit time that one of the particles in the ensemble leaves the target

region with velocity $\mathbf{v}_b$. The integral in dv_b takes into account that only the direction, but not the modulus, of the final velocity matters in our calculation.

I will perform the calculation using particle paths consisting of a discrete set of spacetime points each, going to the continuous at some stage. The main physical assumption in Born approximation is that the potential V is weak in comparison with the kinetic energy of the particles. (The validity of Born approximation in the QM calculation is discussed in most textbooks). Therefore, we will approximate $P\left(\mathbf{r}_a, \mathbf{v}_a; \mathbf{v}_b\right)$ eq.(172), taking eqs.(163) and (161) into account, as follows

(176)

$$P \propto \prod_{j=1}^{n-1} R_j = \prod_{j=1}^{n-1} D_j + \sum_{k=1}^{n-1} \left(\prod_{j=1}^{k-1} D_j \times F_k \times \prod_{l=k+1}^{n-1} D_l \right) + \sum_{k=1}^{n-1} \sum_{s=k+1}^{n-1} \left(\prod_{j=1}^{k-1} D_j \times F_k \times \prod_{l=k+1}^{s-1} D_l \times F_s \times \prod_{i=s+1}^{n-1} D_i \right) + O\left(F^3\right),$$

where $O\left(F^3\right)$ means products involving more than 2 terms of type F. In the limit $\gamma \to 0$ (see eqs.(162) to (165)) the quantities D become Dirac deltas, meaning that the particle motion is uniform. Then the potential acts every time that a term F appears, where the velocity changes, F giving the probabilities of the possible changes. Thus eq.(176) provides the following picture: In the zeroth order approximation the particle travels with constant velocity all the time. First order gives the probability of paths where the particle travels with velocity $\mathbf{v}_a$ until the discrete time k, it changes there its velocity to $\mathbf{v}_b$ and travels with that velocity until the final time n. The second order consists of paths where the particle travels with velocity $\mathbf{v}_a$ until time k, then changes to velocity $\mathbf{v}$ until time t_s and from that time on with velocity $\mathbf{v}_b$, to the end. I point out that this 'picture' is just introduced in order to visualize the mathematical approximations, but we should refrain from taking it as a physical interpretation. Only the total probability $P\left(\mathbf{r}_a, \mathbf{v}_a; \mathbf{v}_b\right)$ could be interpreted as physical, but not each term of eq.(176).

The (relative) probability associated to the first order path $P_0\left(\mathbf{v}_b\right)$ is

(177) $$P_0\left(\mathbf{v}_b\right) = \delta(\mathbf{v}_b - \mathbf{v}_a),$$

where $\delta(\ )$ is the three-dimensional Dirac delta. In order to get the normalized probability distribution eq.(177) we have chosen $C = 1$ in eq.(172). Actually, the parameter C should be chosen in order that the total probability $P = \sum P_l$ is normalized, but the choice $C = 1$ is a fairly good approximation, as will be discussed below. Eq.(177) shows that the path of zeroth order does not contribute to scattering towards a direction different from the initial one, which is our interest here. (In the standard quantum treatment people speak about a cross section in the forward direction, but this should require some elaboration that will not be made here).

The relative probability distribution of the final velocity $\mathbf{v}_b$ due to the first order paths is, taking eq.(172) into account,

$$P_1(\mathbf{v}_a \to \mathbf{v}_b) = 2\pi^{-3} \sum_{k=1}^{n-1} \mathrm{Im}\left\{\tilde{V}(2\mathbf{v}_b - 2\mathbf{v}_a) \exp\left[-2i\mathbf{r}_k \cdot (\mathbf{v}_b - \mathbf{v}_a)\right]\right\},$$

again with $C = 1$. The integer number k is defined so that εk is the time travel from the source to some point in the target. The motion until time εk is uniform with velocity $\mathbf{v}_a$, whence the position vector $\mathbf{r}_k$ is

$$\mathbf{r}_k = \mathbf{r}_a + \mathbf{v}_a \varepsilon k \equiv \mathbf{r}_a + \mathbf{v}_a t$$

The sum in k may be calculated in the continuous time limit that corresponds to $\varepsilon \to 0$. Thus we shall write $\varepsilon \to dt$, and substitute an integral in t for the sum in k. We get

$$P_1(\mathbf{v}_a \to \mathbf{v}_b) = 2\pi^{-3} \int_{t_a}^{t_b} dt \mathrm{Im}\left\{\tilde{V}(2\mathbf{v}_b - 2\mathbf{v}_a) \exp\left[-2i(\mathbf{r}_a + \mathbf{v}_a t) \cdot (\mathbf{v}_b - \mathbf{v}_a)\right]\right\}.$$

The travel times from both the source to the target and the target to the detector are macroscopic, and may be taken as infinite in comparison with typical atomic times. Therefore, we may substitute $-\infty$ for t_a and similarly $+\infty$ for t_b. Hence there is a contribution only when $\mathbf{v}_a \cdot (\mathbf{v}_b - \mathbf{v}_a) = 0$. Now the integral in x_a, y_a, see eqs.(175) and (173), gives the result that $\mathbf{v}_b - \mathbf{v}_a$ has a finite component only in the Z-direction, whence $\mathbf{v}_b = \mathbf{v}_a$. Thus paths of first order do not contribute to scattering either.

The term of order F^2 in eq.(176) may be written taking eq.(164) into account,

$$P_2 = \tag{178}$$

$$= \frac{4}{\pi^6}\sum_{k=1}^{n-1}\sum_{j=k+1}^{n-1}\mathrm{Im}\Big[\tilde{V}(2\mathbf{s}_k)\exp\left(-2i\mathbf{r}_k\cdot\mathbf{s}_k\right)\Big]\mathrm{Im}\Big[\tilde{V}(2\mathbf{s}_j)\exp\left(-2i\mathbf{r}_j\cdot\mathbf{s}_j\right)\Big]$$

$$= \frac{4}{\pi^6}\sum_{k=1}^{n-1}\sum_{j=k+1}^{n-1}\mathrm{Im}\left[\tilde{V}(2\mathbf{v}-2\mathbf{v}_a)\exp\left(-2i\left(\mathbf{r}_a+\mathbf{v}_a t'\right)\cdot\left(\mathbf{v}-\mathbf{v}_a\right)\right)\right]$$

$$\times\mathrm{Im}\left[\tilde{V}(2\mathbf{v}_b-2\mathbf{v})\exp\left(-2i\left(\mathbf{r}_a+\mathbf{v}t+\mathbf{v}_a t'\right)\cdot\left(\mathbf{v}_b-\mathbf{v}\right)\right)\right].$$

where the velocity has changed firstly from $\mathbf{v}_a$ to $\mathbf{v}$ and then from $\mathbf{v}$ to $\mathbf{v}_b$, whence we have made the substitutions

$$\mathbf{r}_j = \mathbf{r}_k + \mathbf{v}t, \mathbf{r}_k = \mathbf{r}_a + \mathbf{v}_a t', \mathbf{s}_k = \mathbf{v}-\mathbf{v}_a, \mathbf{s}_j = \mathbf{v}_b - \mathbf{v}. \tag{179}$$

The product of two 'imaginary parts' may be transformed taking into account the following equality (with a_j complex and b_j real)

$$\mathrm{Im}\left[a_1\exp\left(ib_1\right)\right]\mathrm{Im}\left[a_2\exp\left(ib_2\right)\right] \tag{180}$$

$$= \frac{1}{2}\mathrm{Re}\left\{a_1 a_2^*\exp\left[i\left(b_1-b_2\right)\right]\right\} - \frac{1}{2}\mathrm{Re}\left\{a_1 a_2\exp\left[i\left(b_1+b_2\right)\right]\right\}.$$

We get two terms for P_2 that I will label P_{21} and P_{22} respectively. The former may be written, in the continuous time limit,

$$P_{21} = 2\pi^{-6}\int_{t_a}^{t_b} dt\int_{t_a}^{t} dt'\int d\mathbf{v}\mathrm{Re}I, \tag{181}$$

$$I \equiv \tilde{V}^*\left(2\mathbf{v}_b-2\mathbf{v}\right)\tilde{V}\left(2\mathbf{v}-2\mathbf{v}_a\right)\exp\left\{2i\left[\left(\mathbf{r}_a+\mathbf{v}_a t'\right)\cdot\left(\mathbf{v}-\mathbf{v}_a\right)\right]\right\}$$

$$\times\exp\left\{-2i\left[\left(\mathbf{r}_a+\mathbf{v}_a t'+\mathbf{v}(t-t')\right)\cdot\left(\mathbf{v}_b-\mathbf{v}\right)\right]\right\}.$$

Putting this in eq.(175) we get

$$R_{21}\left(\theta,\phi\right) = 2\pi^{-6}v_a$$

$$\times\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}\rho dy\int v_b^2 dv_b\int_{t_a}^{t_b}dt\int_{t_a}^{t}dt'\int d\mathbf{v}\mathrm{Re}I. \tag{182}$$

The times t_a and t_b may be assumed to correspond to the particle being near the source and near the detector, respectively; therefore we may take $t_a\to-\infty$, $t_b\to\infty$, whence the integrals with respect to

dt' and dt lead to

$$\begin{aligned} & \mathrm{Re}\int_{-\infty}^{t} dt' \exp\left\{2it'\left[\mathbf{v}_a\cdot(\mathbf{v}-\mathbf{v}_a)-(\mathbf{v}_a-\mathbf{v})\cdot(\mathbf{v}_b-\mathbf{v})\right]\right\} \\ = & \frac{\pi}{2}\delta\left[(\mathbf{v}-\mathbf{v}_a)\cdot(\mathbf{v}_a+\mathbf{v}_b-\mathbf{v})\right] \\ = & \frac{\pi}{2}\delta\left[\mathbf{v}\cdot(\mathbf{v}_b-\mathbf{v})+\mathbf{v}_a\cdot(2\mathbf{v}-\mathbf{v}_a-\mathbf{v}_b)\right], \end{aligned}$$

$$\int_{-\infty}^{\infty} dt \exp\left[2it\mathbf{v}\cdot(\mathbf{v}_b-\mathbf{v})\right] = \pi\delta\left[\mathbf{v}\cdot(\mathbf{v}_b-\mathbf{v})\right].$$

The product of the results gives, after some algebra,

$$\frac{\pi^2}{2} v_a^{-1}\delta\left[(2\mathbf{v}-\mathbf{v}_a-\mathbf{v}_b)_z\right] 4\delta\left(v_b^2-v_a^2\right), \tag{183}$$

where we have taken into account that the vector $\mathbf{v}_a$ points in the direction Z, and $()_z$ means Z component. The last Dirac delta insures energy conservation.

The integrals in dx and dy may be extended to infinity taking into account that we assume the density ρ uniform in the XY-plane over a region in that plane that is large in comparison with the size of the target. Thus we get

$$\begin{aligned} & \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \exp\left\{2i\mathbf{r}_a\cdot(2\mathbf{v}-\mathbf{v}_a-\mathbf{v}_b)\right\} \\ = & \pi^2\delta\left[(2\mathbf{v}-\mathbf{v}_a-\mathbf{v}_b)_x\right]\delta\left[(2\mathbf{v}-\mathbf{v}_a-\mathbf{v}_b)_y\right], \end{aligned}$$

where the first delta of eq.(183) has been taken into account. In summary, after integrating with respect to $d\mathbf{v}$ we have

$$\begin{aligned} R_{21}(\theta,\phi) &= \frac{1}{2\pi^2}\rho\int v_b^2 dv_b \delta\left(v_b^2-v_a^2\right)\left|\tilde{V}(\mathbf{v}_b-\mathbf{v}_a)\right|^2 \\ &= \frac{1}{4\pi^2}\rho v_a\left|\tilde{V}(\mathbf{v}_b-\mathbf{v}_a)\right|^2, \end{aligned} \tag{184}$$

whence, taking eqs.(174) and (165) into account, we get

$$\sigma(\theta,\phi) = \frac{1}{4\pi^2}\left|\int d\mathbf{x}\exp\left[-i\mathbf{x}\cdot(\mathbf{v}_b-\mathbf{v}_a)\right]\right|^2,$$

which is Born's formula. I stress that in our calculation the conservation of energy, that is, $|\mathbf{v}_b| = |\mathbf{v}_a|$, appears explicitly in eq.(183).

The calculation of the term $P_{22}(\mathbf{v}_b)$ is similar to the one of $P_{21}(\mathbf{v}_b)$ eq.(181). I skip the details, but the main difference is that $\delta^3(\mathbf{v}_b-\mathbf{v}_a)$ is substituted for $\delta^3(2\mathbf{v}-\mathbf{v}_a-\mathbf{v}_b)$. This implies that the velocity

does not change in the scattering process and therefore the term does not contribute to the desired cross section.

Two comments are in order. Firstly, we have chosen the parameter in front of eq.(172) to be $C = 1$ although this choice normalizes the (partial) probability P_0 rather than the sum $P_0 + P_1 + P_2$ which is our approximation for the full probability P, see eq.(176) . The justification for the choice is that with the assumption that F is small in comparison with D, basic in the Born approximation, we have $P_0 \gg P_1 \gg P_2$ and normalizing just P_0 is fair. Secondly, I comment on an apparently paradoxical fact, namely the coordinates $\mathbf{r}_j$ and $\mathbf{r}_k$ in eq.(179) should correspond to the target, where the potential V may act on the particle, but in the calculation it seems that the velocity change may happen everywhere between the source and the detector. However, this is not the case. In fact, in eq.(178) we have, taking eq.(165) into account,

$$\begin{aligned} &\mathrm{Im}\left[\tilde{V}\left(2\mathbf{s}_k\right)\exp\left(-2i\mathbf{r}_k{\cdot}\mathbf{s}_k\right)\right] \\ &\quad = \mathrm{Im}\int V\left(\mathbf{x}\right)\exp\left(2i\mathbf{x}\cdot\mathbf{s}_k\right)d\mathbf{x}\exp\left(-2i\mathbf{r}_k{\cdot}\mathbf{s}_k\right), \end{aligned}$$

meaning that a velocity change $\mathbf{s}_k \neq 0$ is produced only if $\mathbf{r}_k \approx \mathbf{x}$ belongs to the region where $V \neq 0$.

As a conclusion, the result of our calculation is the same as the standard one of quantum mechanics, as expected. But *in our calculation we deal only with particles* rather than waves as the QM derivation of the Born formula suggests. The picture that emerges from the calculation is as follows. Every single particle, that travels from source to the detector crossing the target region, may have many different paths. The weight of every path should be positive in order to be interpreted as a probability, which is guaranteed by the discussion in section 4.6.1. However, we should not pretend to have a physical picture of the intermediate *mathematical* approximations in the calculation.

'Interference' experiment. Born's approximation, eq.(184) may allow a simple calculation of the result of some interference experiments. For instance, let us consider the potential

(185)
$$V\left(\mathbf{r}\right) = C\left\{\exp\left[-\lambda\left(\mathbf{r}+\mathbf{a}\right)^2\right] + \exp\left[-\lambda\left(\mathbf{r}-\mathbf{a}\right)^2\right]\right\}, \quad \mathbf{a} \equiv (a,0,0)\,,$$

and a particle with initial velocity $\mathbf{v}_0 = (0, 0, v_0)$, moving in the Z direction. Obtaining the cross section via Born's approximation is a simple exercize. We get

$$\sigma = C^2 \frac{\pi}{\lambda^3} \exp\left[-\frac{\mathbf{v}^2 + \mathbf{v}_0^2 - 2v_0 v_z}{2\lambda}\right] \cos^2(av_x),$$

where $\mathbf{v} = (v_x, v_y, v_z)$ is the final velocity. Hence the cross section becomes, taking the conservation of energy eq.(183) into account,

$$\sigma = C^2 \frac{\pi}{\lambda^3} \exp\left[-\frac{2v_0^2 \sin^2(\theta/2)}{\lambda}\right] \cos^2(av_0 \sin\theta \cos\phi). \tag{186}$$

Assuming that the particles are detected by the spots produced in a screen parallel to the XY plane we would observe typical interference fringes with a decreasing intensity in both directions X and Y with a maximum at $x = y = 0$.

The point of this calculation is that no waves are involved, only particles. Therefore, in our formalism *the interference picture is an effect of the nonlocal action of the potential*. Of course, we might assume that the action is mediated by some 'hidden' waves that do interfere. In agreement with the general interpretation of quantum mechanics supported in this book, I propose that those waves are associated to the quantum vacuum fields modified by the presence of either macroscopic devices or the particles and by additional fields involved in the experiments.

4.6.3. Conclusions. I have shown that in nonrelativistic quantum mechanics (without spin) it is possible to picture the transition probability in terms of particle paths, every path having a (positive) probability. A path may be defined by the positions

$$\{\mathbf{r}_a \equiv \mathbf{r}_0, \ldots \mathbf{r}_j, \ldots, \mathbf{r}_b \equiv \mathbf{r}_n\}$$

at times $t_a, \ldots, t_j \equiv t_a + j\varepsilon, \ldots, t_b$ or, what is equivalent, the end positions plus the velocity changes $\{\mathbf{s}_j\}$ at times t_j. Eventually we should consider the limit $n \to \infty$ with $n\varepsilon = t_b - t_a$. This provides a stochastic picture where particles travel along continuous paths.

The quantity P, eq.(152), has been defined for the calculation of the probabilities of the possible paths that start at the spacetime point $(\mathbf{r}_a, t_a)$ and finish at $(\mathbf{r}_b, t_b)$. However, our formalism may be used for other relevant problems, for instance for the calculation of the probability of the different paths starting at a given point with a given velocity, $(\mathbf{r}_a, \mathbf{v}_a, t_a)$. Hence we might study a kind of 'evolution

in phase space', something that quantum mechanics does not provide. Whether some of the possible additional predictions may be tested empirically will not be discussed here. In any case, no contradiction with quantum mechanics is expected.

Is summary our stochastic approach gives an intuitive picture of nonrelativistic quantum mechanics (without spin) in terms of probabilities of the possible paths of *particles*. The particle's motion is represented by a stochastic process such that there is a random change of velocity at every time t_j, with a probability depending on *the potential over a large region* around the position of the particle (indeed it derives from the Fourier transform of the potential, see eq.(165)). The wave behaviour, e.g. in experiments of atom interference, may be interpreted assuming that the motion of the particles is governed by a law (different from Newton's) where the 'acceleration' depends on the potential on a large spatial region, in contrast with the local action of classical dynamics. With this interpretation the interference experiments with particles, e.g. atoms, may be explained without assuming that those particles possess any wave nature or that they may cross two distant slits at the same time. But I stress again that we remain at the level of nonrelativistic quantum mechanics. I do not claim that a similar interpretation may be extended to relativistic quantum field theory, e.g. photons, electrons when spin plays a role, or even atoms or molecules when (Bose or Fermi) statistics is relevant.

Bibliography

[1] L. de Broglie: *Recherches sur la théorie des quanta*. Thesis, Paris, 1924; *Ann. Phys. (Paris)* **3**, 22 (1925).

[2] L. de Broglie: Remarques sur la nouvelle mécanique ondulatoire. *Comptes Rendus Acad. Sci. Paris,* **183**, 447 (1926); Sur le rôle des ondes continues Ψ en Mécanique ondulatoire. *Comptes Rendus Acad. Sci. Paris,* **184**, 380 (1927).

[3] L. de Broglie: *Une tentative d'interpretation causale et non-linéaire de la mécanique ondulatoire: La theorie de la double solution.* Gauthier-Villars, Paris, 1961. (English translation: *Non-linear wave mechanics: A causal interpretation*, Elsevier, Amsterdam, 1963).

[4] M. Born: Quantum mechanics of collision processes. *Z. Phys.* **38**, 803 (1926); English translation in G. Ludwig, Ed., *Wave Mechanics*, Pergamon Press, Oxford, 1968.

[5] E. Madelung: Quantentheorie in hydrodynamischer Form. *Z. Phys.* **40**, 322 (1926).

[6] D. Bohm: A suggested interpretation of the quantum theory in terms of 'hidden' variables, I and II. *Phys. Rev.* **85**, 166-193 (1952).

[7] P. R. Holland: *The Quantum Theory of Motion*. Cambridge U. P, Cambridge, U. K., 1993.
[8] L. I. Schiff: *Quantum Mechanics*. McGraw-Hill, New York, 1955.
[9] E. Santos: A lagrangian formulation of the theory of random motion. *Nuovo Cimento* **59B**, *65-81* (1969).
[10] E. Nelson: Derivation of the Schrodinger Equation from Newtonian Mechanics. *Phys. Rev.* **150**, 1079 (1966).
[11] E. Nelson: *Dynamical theories of Brownian motion*. Princeton U. P., 1967.
[12] L. de la Peña: New formulation of stochastic theory and quantum mechanics. *J. Math. Phys.* **10**, 1620 (1969).
[13] M. Masujima: *Path integral quantization and stochastic quantization*. Springer, Berlin, 2009.
[14] E. Nelson: *Quantum fluctuations*. Princeton U. P., 1985.
[15] T. Wallstrom: *On the derivation of Schrödinger equation from stochastic mechanics*. Princeton University Press, Princeton, 1988.
[16] H. Weyl: *The theory of groups and quantum mechanics*. Dover, New York, 1931. (German original, 1928).
[17] E. Wigner: On the quantum correction for thermodynamic equilibrium. *Phys. Rev.* **40**, 749 (1932).
[18] H. J. Groenewold: On the principles of elementary quantum mechanics. *Physica* **12**, 405-446 (1946).
[19] J. E. Moyal: Quantum mechanics as a statistical theory. *Proc. Cambridge Phil. Soc.* **45**, 99-124 (1949).
[20] M. Hillery, R. F. O'Connell, M. O. Scully, E. P. Wigner: Distribution functions in physics: Fundamentals. *Phys. Rep.* **106**, 121-168 (1984).
[21] C. K. Zachos, D. B. Fairlie, T. L. Curtright: *Quantum mechanics in phase space*. World Scientific, Singapore, 2005.
[22] J. Ville: Théorie et applications de la notion de signal analytique. *Cables et Transmission*, **2**, 6174 (1948).
[23] L. Cohen: *Time-frequency analysis*. Prentice-Hall, New York, 1995.
[24] R. L. Hudson: When is the Wigner quasi-probability density non-negative. *Rep. Math. Phys.* (Torun) **6**, 249 (1974).
[25] F. Soto, P. Claverie: When is the Wigner function of multi-dimensional systems non-negative. *J. Math. Phys.* **23**, 753 (1982).
[26] R. P. Feynman: Space-time approach to nonrelativistic quantum mechanics. *Rev. Mod. Phys.* **20**, 367 (1948).
[27] R. P. Feynman, A. R. Hibbs: *Quantum mechanics and path integrals*. Mc.Graw-Hill, New York, 1965.
[28] P. A. M. Dirac: The Lagrangian in quantum mechanics. *Physikalische Zeitschrift der Sowjetunion* **3**, 64-72 (1933).
[29] J. Zinn Justin: *Path integrals in quantum mechanics*. Oxford University Press, 2004.
[30] J. D. Jackson: *Classical electrodynamics*. John Wiley and Sons, New York, 1975.

CHAPTER 5

Stochastic electrodynamics, a clue for the interpretation of quantum mechanics

5.1. Introduction

In this chapter it is shown that many phenomena that are assumed purely quantum-mechanical may be interpreted with a classical theory except for the assumption that there is a random background radiation filling the whole space. The statistical properties of the radiation are Lorentz invariant whence they can be characterized by a single parameter that fixes the scale. Identifying the parameter with the Planck constant gives rise to the theory called stochastic electrodynamics (SED in the following). Actually, the postulated random radiation closely parallels the vacuum electromagnetic radiation of quantum electrodynamics, assumed a real stochastic field. The relevant fact is that SED offers an intuitive picture of microphysics that provides hints for a realistic interpretation of many quantum phenomena. The hints offered by SED are emphasized in italics throughout this chapter.

5.1.1. Electrodynamics in the presence of a random radiation field. As stated in chapter 1 the basic assumption in this book is that the quantum vacuum fields are real stochastic fields. For the sake of clarity let us consider the best known vacuum field, the electromagnetic radiation. The spectrum, that here I define as the energy per unit volume and unit frequency interval, is given by eq.(187) below. The interesting question is whether the reality of the vacuum electromagnetic field, combined with classical physics, allows explaining some phenomena believed as typically quantal, thus providing a

hint for the realistic interpretation of quantum theory. We restrict our study to a domain defined by:

1) Only one of the interactions of nature, namely electromagnetic.

2) Nonrelativistic energies. Thus we shall study charged particles in given electromagnetic fields and/or interacting with other charges.

3) The Planck constant $\hbar$ appears exclusively in the vacuum electromagnetic radiation. Consequently $\hbar$ does not appear in the evolution equations, whence we shall use the dynamical laws of classical electrodynamics throughout.

SED, also known as random electrodynamics, is the theory so defined. It has been developed by a small number of people during the last sixty years. A review of the work made until 1995 is the book by L. de la Peña and A. M. Cetto [**1**] and new results are included in more recent reviews [**2**], [**3**]. The application of similar ideas to optics will be reviewed in chapter 6.

The origin of SED may be traced back to Walter Nernst, who extended to the electromagnetic field the zero-point fluctuations of oscillators assumed by Planck in his second radiation theory of 1912. Nernst also suggested that the zero-point fluctuations might explain some empirical facts, like the stability of atoms and the chemical bond. The proposal was soon forgotten due to the success of Bohr's model in 1913 and the subsequent development of the (old) quantum theory. Many years later the idea has been put forward again several times (e.g. by Braffort et al. in 1954 [**5**] and by Marshall in 1963 [**6**]).

SED studies the motion of charged particles immersed in ZPF, but the back actions of the charged particles on the ZPF are neglected so that the random field of free space is used. Assuming that the field is Lorentz invariant (at not too high frequencies) determines the spectrum, that is, the energy per unit volume and unit frequency interval [**4**], [**1**]. It is given by

$$\rho_{ZPF}(\omega) = \frac{1}{2\pi^2 c^3}\hbar\omega^3, \tag{187}$$

which corresponds to an average energy $\frac{1}{2}\hbar\omega$ per normal mode. The Planck constant $\hbar$ enters the theory via fixing the scale of the assumed universal random radiation. Of course, the spectrum eq.(187) implies a divergent energy density and any cutoff would break Lorentz invariance. However, we may assume that it is valid for low enough frequencies, the behaviour at high frequencies requiring the inclusion of other vacuum fields and general relativity theory. The spectrum eq.(187) is appropriate for systems at zero Kelvin, but SED may also

be studied at a finite temperature, where we should add to eq. (187) the thermal Planck spectrum. In addition, SED may provide an interpretation of phenomena where the free spectrum is modified by boundary conditions derived from macroscopic bodies, but the average energy $\frac{1}{2}\hbar\omega$ per normal mode still holds true. An example is the Casimir effect, see section 5.6 below.

The study of some simple systems provides an intuitive picture of several phenomena considered as purely quantum, like *the stability of the classical (Rutherford) atom, Heisenberg uncertainty relations, entanglement, specific heats of solids, behaviour of atoms in cavities, etc.* For this reason I propose that SED may be considered as a clue in the search for a realistic interpretation of quantum theory.

5.1.2. Scope of stochastic electrodynamics. Not all predictions that have been claimed to follow from SED derive from the theory as defined above. In some cases additional assumptions are introduced in order to agree with the quantum predictions. With these extra assumptions most of nonrelativistic quantum mechanics might be derived from SED [**2**]. However, the physical bases of the theory and the physical picture of the quantum world become less clear.

In this chapter we will study SED, in the strict sense defined in section 5.1.1. With that definition there are many examples where SED predicts results in contradiction with quantum mechanics and with experiments, as discussed in section 5.6 below. In particular, SED deals only with charged particles whilst QM laws are valid for both charged and neutral particles. It has been claimed that this restriction may be avoided taking into account that neutral particles may contain charged parts (e.g. neutrons possess a magnetic moment). I think this is flawed, the application to those neutral particles might be valid in order to explain the stationary equilibrium state, which is effectively defined over an infinite time and it results independent on the total charge, as seen for instance in eqs.(205) and (206) below. However, this is not the case for time dependent properties like eq.(238) where the value of the charge is relevant.

A different approach is to consider that SED is "the closest classical approximation to quantum theory" [**3**]. This suggests that there are two different theories, namely classical and quantum, but it is assumed that the validity of classical theory may be extended if we include the hypothesis of a radiation field with a Lorentz invariant spectrum in free space. I believe that this approach does not solve

the problem of the 'infamous boundary' between macro (classical) and micro (quantum) physics. Indeed, in the standard view the boundary is defined by the relative value of the Planck constant as compared with the typical values of the action variables in the system. That is, classical theories are approximations of quantum theories when the Planck constant may be neglected.

In my view *SED is a semiclassical theory*, that is, an approximation to quantum theory better than purely classical theories, in some limited domain. The domain is constrained as follows:

(1) SED ignores metric fluctuations (the gravitational vacuum field) that I believe are essential to understand quantum theory, as discussed in chapter 1 section 1.3, and in chapter 7 section 7.2.3;
(2) it also ignores all fields in the Standard Model of fundamental particles, except the electromagnetic (EM) radiation;
(3) SED studies motion of (charged) particles under the action of the vacuum EM field but it neglects the back action of the charges on the field.

The last restriction gives rise to the fact that SED usually agrees with quantum mechanics for linear systems but not for nonlinear ones, as shown below in sections 5.2, 5.3, 5.4 and 5.7. In the latter, this point will be discussed in more detail.

5.1.3. Plan of the chapter. In the following some results of SED are reviewed and the analogies and differences with nonrelativistic quantum mechanics (QM in the following) will be pointed out. Most of the results have been reviewed in more detail elsewhere [**1**]. The novelty here is a more careful comparison of SED with QM and the emphasis on those quantum phenomena that might be better understood via the analogy with the picture provided by SED.

In sections 5.2 and 5.3 the harmonic oscillator is revisited, with an application to oscillators in several dimensions in section 5.4. In sections 5.5 and 5.6 SED is applied to other linear systems, namely the free particle and the particle in a homogeneous magnetic field. Section 5.7 is devoted to the application of SED to some nonlinear systems, showing that a disagreement with QM and with the experiments usually appears in this case. Section 5.8 presents conclusions. This chapter includes many calculations and, in order for the reader not to lose the essential points, I will write in italics the relevant aspects for the comparison between SED and QM.

5.2. The oscillator: SED equilibrium with radiation vs. QM ground state

5.2.1. The stability of matter in classical physics, SED and QM. It is well known that a system of electric charges cannot reach a state of equilibrium in classical electrodynamics. This is the case in particular for the 'miniature solar system' (Rutherford) model of the atom, as discussed in chapter 1. In contrast, both QM and SED predict states of equilibrium, but those states look quite different in both theories. In QM the ground state, the lowest energy solution of the time-independent Schrödinger equation, seems static, nothing changing with time. However, in SED the equilibrium state is dynamical: the charged particles are continuously emitting radiation but also absorbing it from the background random radiation (ZPF) and the stationary state is reached when emission and absorption are balanced on the average. I propose that the SED prediction is the correct realistic interpretation of the phenomenon, whilst the QM formalism hides that interpretation. This is the first lesson in our attempt at a realistic interpretation of quantum mechanics: *we should not try to interpret the quantum formalism, but only the QM predictions for actual or possible observations or experiments.*

5.2.2. Equation of motion of the harmonic oscillator in SED. The harmonic oscillator in one dimension is the most simple system to be treated within SED (the free particle requires a more careful study in order to avoid divergences). It is not strange that it was the first system studied. In this and the following sections we revisit a well known treatment of the oscillator in SED [**7**], [**1**], but the study of the aspects that may provide a clue for the interpretation of quantum mechanics is original.

I shall write the differential equation for the one-dimensional motion of the particle. Radiation acts on the particle via the Lorentz force, that is

$$F_x = e\left[\mathbf{E} + \frac{\dot{\mathbf{r}}}{c} \times \mathbf{B}\right]_x , \tag{188}$$

E and **B** being the electric and magnetic fields of the ZPF. The passage to more dimensions is straightforward. Here we will neglect magnetic effects of the ZPF and the dependence of the field on the position coordinate, which corresponds to the electric dipole approximation, plausible in a nonrelativistic treatment. Thus the differential equation

of motion of the particle in a harmonic oscillator potential is

$$m\ddot{x} = -m\omega_0^2 x + m\tau\dddot{x} + eE(t), \tag{189}$$

where $m(e)$ is the particle mass (charge) and $E(t)$ is the x component of the electric field of the radiation (the zero-point field, ZPF). The equation of the mechanical (classical) oscillator is modified by the two last terms. The second term on the right side of eq.(189), is the damping force due to emission of radiation. It should appear also in the classical electrodynamical treatment. Only the third term is specific to SED because it involves the Planck constant (it is of order $\mathrm{O}\left(\hbar^{1/2}\right)$). The parameter τ is given by

$$\tau = \frac{2e^2}{3mc^3} \quad \Rightarrow \quad \tau\omega_0 = \frac{2}{3}\frac{e^2}{\hbar c}\frac{\hbar\omega_0}{mc^2} \ll 1., \tag{190}$$

so that the dimensionless quantity $\tau\omega_0$ is very small, it being the product of two small numbers: the fine structure constant, $\alpha \equiv e^2/\hbar c \sim 1/137$, and the nonrelativistic ratio $\hbar\omega_0/mc^2 \simeq v^2/c^2 \ll 1$. Thus the two last terms of eq.(189) may be considered small, which allows some useful approximations.

Eq.(189) is a stochastic differential equation of Langevin's type with coloured (non-white) noise. It has been named Braffort-Marshall equation by the early workers on SED [**5**], [**6**]. Solving an equation of this kind usually means finding the evolution of the probability distribution of the relevant quantities as a function of time, starting from given initial conditions. When time goes to infinity the probability distributions become independent of the initial conditions, giving rise to the stationary or equilibrium distribution.

5.2.3. Average values of the potential and kinetic energies. Several solutions of eq.(189) have been published [**7**], [**1**]. The most simple one is the stationary solution, which may be found by Fourier transform of eq.(189) as follows. Firstly, we define the Fourier transform of the stationary process $E(t)$ in a finite time interval by

$$\widetilde{E}(\omega, T) \equiv \frac{1}{\sqrt{4\pi T}} \int_{-T}^{T} E(t) \exp(-i\omega t)\, dt. \tag{191}$$

Hence it may be shown that $\left|\widetilde{E}(\omega, T)\right|^2 /8\pi$ is the mean energy density per unit frequency interval associated to one electric field component in the time interval $(-T, T)$. Thus the total energy density per unit frequency interval, $\rho(\omega)$ eq.(187), should be 6 times that quantity (6 because in the ZPF there are 3 components of the electric field

and another 3 of the magnetic field, all contributing equally on the average). Consequently we define the spectral density, $S_E(\omega)$, of the field $E(t)$ as follows

$$S_E(\omega) \equiv \lim_{T\to\infty} \left|\widetilde{E}(\omega, T)\right|^2 = \frac{4\pi}{3}\rho(\omega) = \frac{2}{3\pi c^3}\hbar\omega^3, \tag{192}$$

the equality giving the relation between the spectral density and the energy density of the ZPF, eq.(187) . For short, the spectral density will be called spectrum in the following.

A Fourier transform of all terms of eq.(189) similar to eq.(191) provides a relation between the spectrum of the field component and the spectrum of the coordinate, $x(t)$; namely

$$m(\omega_0^2 - \omega^2 + i\tau\omega^3)\widetilde{x}(\omega) = e\widetilde{E}(\omega), \tag{193}$$

where $\widetilde{x}(\omega)$ and $\widetilde{E}(\omega)$ are the Fourier transforms of $x(t)$ and $E(t)$ respectively. Hence the spectrum of $x(t)$ is easily got in terms of the spectrum of $E(t)$ that is

$$S_x(\omega) = \frac{3c^3\tau}{2m\left[\left(\omega_0^2 - \omega^2\right)^2 + \tau^2\omega^6\right]} S_E(\omega), \tag{194}$$

whence we obtain, taking eq.(192) into account,

$$S_x(\omega) = \frac{\hbar\tau\omega^3}{\pi m\left[\left(\omega_0^2 - \omega^2\right)^2 + \tau^2\omega^6\right]}. \tag{195}$$

From the spectrum it is easy to get the quadratic mean of the relevant variables, namely

$$\left\langle x^2\right\rangle = \int_0^\infty S_x(\omega)\, d\omega, \left\langle v^2\right\rangle = \int_0^\infty \omega^2 S_x(\omega)\, d\omega, \tag{196}$$

where $\langle\rangle$ means time average, and the quantities in eq.(196) are the coordinate of the oscillator and its velocity, respectively. The spectrum of the velocity is ω^2 times the spectrum of the coordinate because the time derivative leads to multiplication of the Fourier transform times $i\omega$. In our treatment of stationary states in SED an ergodic hypothesis is made; ensemble averages are assumed equal to time averages for the stationary stochastic processes involved.

Calculating the integral of $S_x(\omega)$ is lengthy but it is easy in the limit $\tau \to 0$ where the integrand becomes a Dirac's delta. In fact, if τ

is small then the main contribution to the integral comes from values of ω close to ω_0. In the limit $\tau \to 0$ the integral becomes trivial:

$$\begin{aligned}\left\langle x^2 \right\rangle &= \int_0^\infty S_x(\omega)\, d\omega \\ &\simeq \lim_{\tau\to 0} \int_0^\infty \frac{\hbar\tau\omega_0^3}{\pi m\left[4\omega_0^2(\omega_0-\omega)^2 + \tau^2\omega_0^6\right]} d\omega \\ &= \frac{\hbar}{2m\omega_0}\int_{-\infty}^{\infty} \delta(\omega-\omega_0)\, d\omega = \frac{\hbar}{2m\omega_0}, \end{aligned} \tag{197}$$

whence the mean potential energy is

$$\langle V\rangle = \frac{1}{2}m\omega_0^2\left\langle x^2\right\rangle = \frac{1}{4}\hbar\omega_0.$$

The quantity $\left\langle x^2\right\rangle$ as a function of $\tau\omega_0$ is not analytical and we cannot calculate it as an expansion in powers of $\tau\omega_0$. The correction, $\left\langle x^2\right\rangle_{hf}$, to the result eq.(197) is mainly a contribution of the high frequencies $\omega \gg \omega_0$. It may be estimated by the integral of the spectrum eq.(195) with zero substituted for ω_0 and $2\omega_0$ put as lower limit in order to exclude the contribution of the low frequency part, that was included in the result eq.(197) . That is,

$$\begin{aligned}\left\langle x^2\right\rangle_{hf} &\simeq \int_{2\omega_0}^\infty \frac{\hbar\tau\omega^3}{\pi m\left[\omega^4+\tau^2\omega^6\right]} d\omega \\ &\simeq -\frac{\hbar\tau}{\pi m}\log(\tau\omega_0), \end{aligned} \tag{198}$$

which is positive (see eq.(190)). We see that the result depends, but only slightly, on the lower limit of the integral, provided it is of order $2\omega_0$. A more rigorous calculation will be made in section 5.3.

A similar procedure might be used for the quadratic mean velocity, by performing the integral of the velocity spectrum. However, that integral is divergent and we shall assume that there is some frequency cut-off, ω_c. The result of the integral is the sum of two terms. One of them comes from frequencies near ω_0 and is independent of the cut-off in the limit $\tau \to 0$, giving

$$\left\langle v^2\right\rangle = \int_0^{\omega_c} \omega^2 S_x(\omega)\, d\omega \simeq \frac{\hbar\omega_0}{2m} \Rightarrow \frac{1}{2}m\left\langle v^2\right\rangle = \frac{1}{4}\hbar\omega_0. \tag{199}$$

The other term comes from the high frequency region and is divergent when the cut-off goes to infinity. It may be estimated as in the case

of $\langle x^2 \rangle$, although here we may put zero as lower limit of the integral, that is,

$$\langle v^2 \rangle_{hf} \simeq \int_0^{\omega_c} \frac{\hbar \tau \omega^5}{\pi m \left[\omega^4 + \tau^2 \omega^6\right]} d\omega = \frac{\hbar}{2\pi m \tau} \log\left(1 + \tau^2 \omega_c^2\right). \tag{200}$$

Actually, calculating that term is a rather academic exercise because for those frequencies the nonrelativistic approximation breaks down (see the discussion of the velocity dispersion in the free particle case below). Adding eqs.(197) and (199) gives the total mean energy in the limit $\tau\omega_0 \to 0$, namely

$$\langle U \rangle = \left\langle \frac{1}{2} m\omega_0^2 x^2 + \frac{1}{2} m v^2 \right\rangle = \frac{1}{2} \hbar \omega_0. \tag{201}$$

An alternative definition of the energy is possible in terms of the canonical momentum, p, which avoids problems of divergence. The momentum is defined by

$$p \equiv mv - \frac{e}{c} A, U \equiv \frac{p^2}{2m} + \frac{1}{2} m\omega_0^2 x^2. \tag{202}$$

Now we take into account that the potential vector, whose x component we label A, contains two parts; one coming from the ZPF and the other one from the particle self-field, the latter producing the radiation reaction. These two terms give rise to the last two terms of eq.(189) . Taking this relation into account it is straightforward to get the spectrum of the canonical momentum, that is,

$$\begin{aligned} \frac{d}{dt} p &= -m\omega_0^2 x \qquad \Rightarrow \\ S_p(\omega) &= \frac{m^2 \omega_0^4}{\omega^2} S_x(\omega) = \frac{\hbar m \tau \omega_0^4 \omega}{\pi \left[\left(\omega_0^2 - \omega^2\right)^2 + \tau^2 \omega^6\right]}. \end{aligned} \tag{203}$$

Hence we get

$$\langle p^2 \rangle = m^2 \omega_0^4 \int_0^{\infty} \omega^{-2} S_x(\omega)\, d\omega = \frac{m\hbar\omega_0}{2} \Rightarrow \frac{\langle p^2 \rangle}{2m} = \frac{1}{4} \hbar \omega_0, \tag{204}$$

in the limit $\tau \to 0$. We see that the energy defined from the velocity is divergent (a cut-off was needed), whilst the one derived from the canonical momentum is finite. Thus the use of the canonical momentum in the definition of the energy seems more convenient. We might expect that in a more correct relativistic treatment the former would be also convergent and not too different from the latter.

5.2.4. Radiative corrections. Calculating the corrections due to the finite value of the parameter τ in eqs.(197) to (201) is lengthy and uncertain [**7**], [**1**]. The corrections for the variances of x^2 and v^2 have been estimated roughly in eqs.(198) and (200) . As said above these corrections are not analytical in τ (or in the fine structure constant α, proportional to τ, see eq.(190)), but the leading term agrees with the radiative corrections of quantum electrodynamics (Lamb shift). In any case, they depend on the high frequency region of integrals like eq.(195), where the nonrelativistic approximation breaks down.

An interesting aspect in the comparison of SED with QM is that in the latter the relevant quantities, like variances of coordinate or velocity or the mean energy, may be calculated simply from the formalism (as expectations of the appropriate observables in the ground state) without any reference to the vacuum radiation. After that we may calculate radiative corrections adding the coupling with the quantized radiation field. In contrast, in SED the coupling with the zero-point field is essential in order to get the stationary state, as is shown in eq.(189) . In other words, in SED the main values and the radiative corrections appear together, e.g. in eqs.(197) and (199), the separation of both contributions being a mathematical recourse for easy calculation. As a consequence the QM calculation of the main contribution is easier than in SED, but at the price of hiding the physics, namely the fact that the stability is essentially due to the ZPF.

5.2.5. Probability distributions of position, momentum and energy. In order to fully define the stationary state of the oscillator immersed in ZPF it is necessary to get the probability distributions, not just the mean values. Before doing that we need to clarify the meaning of the probability distributions involved. Up to now we have considered averages over infinite time intervals, see eq.(192) . However, we assume that the time dependent quantities are stochastic processes, that is, probability distributions over functions of time. Thus we should write $x(t, \lambda)$ (as is standard in the mathematical theory of stochastic processes) rather than just $x(t)$, where $\lambda \in \Lambda$ and there is a probability distribution on the set Λ. For a fixed value of t this provides a probability distribution of the random variable $x(t)$. We assume that the probability distribution of each component, $E(t, \lambda)$, of the ZPF (in free space) is Gaussian with zero mean and also that it is a stationary ergodic process. That is, any time average

(over an infinite time interval) equals the ensemble average over the probability distribution of Λ at any single time. The properties of the ZPF will be studied in more detail in chapter 6.

Eq.(189) is linear, whence the Gaussian character of $E(t,\lambda)$ gives rise to Gaussian distributions (with zero mean) for both positions and velocities. Thus eq.(197) fixes completely the normalized probability distribution of the positions to be

$$W(x)\,dx = \sqrt{\frac{m\omega_0}{\pi\hbar}}\exp\left[-\frac{m\omega_0 x^2}{2\hbar}\right]dx. \tag{205}$$

Similarly eq.(204) fixes the distribution of momenta, that is

$$W(p)\,dp = \sqrt{\frac{m}{\pi\hbar\omega_0}}\exp\left[-\frac{p^2}{2\hbar m\omega_0}\right]dp, \tag{206}$$

which is also normalized. The distribution of velocities is similar to that one, with mv substituted for p (modulo ignoring the part due to high frequencies).

In order to get the distribution of energy, U, to lowest order in the Planck constant $\hbar$ we take into account that, as eqs.(205) and (206) already contain $\hbar$, the relation between x, v and U should be written to zeroth order in $\hbar$, that is, using the classical relation. Then we get the following exponential distribution of energies, U,

$$W(U)\,dU = \int W(x)\,dx \int W(p)\,dp\,\delta\left(U - \frac{1}{2}m\omega_0^2x^2 - \frac{1}{2m}p^2\right)dU$$

$$= \frac{2}{\hbar\omega_0}\exp\left(-\frac{2U}{\hbar\omega_0}\right)dU, U \geq 0. \tag{207}$$

where $\delta\,()$ is Dirac's delta. Hence the fluctuation of the energy is

$$\sqrt{\langle U^2\rangle - \langle U\rangle^2} = \langle U\rangle = \frac{1}{2}\hbar\omega_0. \tag{208}$$

The distributions of positions and momenta, eqs.(205) and (206), agree with the QM predictions. But the same is not true for the energy, because QM predicts a sharp energy in the ground state, in disagreement with the SED stationary state eq.(207) . Below we shall study this discrepancy, relevant for our realistic interpretation of quantum theory.

Eqs.(205) and (206) show that the Heisenberg uncertainty relations,

$$\Delta x \Delta p \geq \hbar/2, \tag{209}$$

appear in a natural way in SED. Indeed, the probability distributions eqs.(205) and (206) correspond to a 'minimum uncertainty wave-packet', meaning by this the quantum state where the Heisenberg inequality, eq.(209), saturates becoming an equality.

5.2.6. Comparison between the stationary state in SED and the ground state in QM. The conclusion of the study of the stationary state of the oscillator in SED is that it is rather similar to the ground state of the oscillator en QM. Indeed, the probability distribution of positions and momenta in the stationary state of SED agree with the predictions of QM for the ground state, in the limit $\tau \to 0$, eqs.(205) and (206), whilst the corrections for finite τ, that depend on the small quantity $\tau\omega_0$, correspond to the radiative corrections of quantum electrodynamics. However, the probability distribution of the energy does not agree with QM. In the following I study more carefully this discrepancy.

John von Neumann's theorem against hidden variables. Firstly I should mention that the conflict between the QM prediction and the SED eq.(207) is an example of the general argument used by von Neumann [**8**] in his celebrated theorem of 1932 proving that hidden variable theories are incompatible with QM. That theorem prevented research in hidden variables theories until Bell's rebuttal in 1966 [**9**], as discussed in chapter 3 section 3.1.1. J. von Neumann starts with the assumption that any linear relation between quantum observables should correspond to a similar linear relation between the possible (dispersion free) values in a hypothetical hidden variables theory.

In our case SED may be seen as a hidden variables theory. The quantum oscillator fulfills

$$\hat{H} = \frac{1}{2m}\hat{p}^2 + \frac{1}{2}m\omega_0^2\hat{x}^2.$$

Thus as the energy predicted by quantum mechanics, $U = \hbar\omega_0/2$, is sharp, any pair of values of v^2 and x^2 in the hidden variables theory should fulfil, according to von Neumann's hypothesis,

$$U = \frac{1}{2}m(v^2 + \omega_0^2 x^2) = \frac{1}{2}\hbar\omega_0, \qquad p = mv, \tag{210}$$

which is not compatible with the distributions eqs.(205) and (206) (for instance, the possible value $v^2 = 2\hbar\omega_0/m$ is incompatible with eq.(210) because it would imply $x^2 < 0$). Bell rebutted von Neumann pointing out that the contradiction only arises when two of the quantum observables involved do not commute and in this case the

measurement of the three observables should be made in, at least, two different experiments. Thus a contextual hidden variables theory is possible, that is, a theory where it is assumed that the value obtained in the measurement depends on both the state of the observed system and the full experimental context. The matter was studied in more detail in chapter 3 section 3.2.

In our hidden variables theory (i.e. SED) a sharp energy U would be obtained in a measurement different from those leading to eqs. (205) and (206), as explained in the following.

The apparent contradiction between QM and SED. In our case the apparent contradiction between SED eq.(207) and the QM prediction of a sharp energy disappears if we take into account how the energy of a state is defined *operationally* (i.e. how it may be measured). In SED the stationary state corresponds to a dynamical equilibrium between the oscillator and the ZPF. Checking empirically whether a dynamical equilibrium exists requires a long time, ideally infinite time. If we define the energy of the oscillator in equilibrium as the average over an infinite time, it would be obviously sharp. In fact, the probability distribution of the 'mean energies over time intervals of size Δt' has a smaller dispersion for greater Δt, and will be dispersion free in the limit $\Delta t \to \infty$. Thus it is natural to assume that the ground state energy as defined by QM actually corresponds to measurements made over infinitely long times. This fits fairly well with the quantum energy-time uncertainty relation

$$\Delta U \Delta t \geq \hbar/2, \tag{211}$$

which predicts that the measured energy does possess a dispersion ΔU if the measurement involves a finite time Δt. Thus no contradiction exists between the *observable predictions* by SED and QM for the energy in the ground state.

It is remarkable that QM and SED lead to the same result via rather different paths. In fact, in QM the state vector of the ground state of a system is an eigenstate of the Hamiltonian, which implies a nil dispersion of the state energy, but the uncertainty relation gives rise to some uncertainty for any actual measurement. The instantaneous energy is a badly defined concept. Indeed, the SED distribution eq.(207) derives from the (classical) definition of total energy in terms of positions and momenta, but it does not possesses any operational (measurable) meaning.

5.2.7. Frequencies of the light emitted or absorbed by the SED oscillator. There is another trivial agreement between the SED and QM predictions for the oscillator, namely the spectrum of emitted or absorbed light. In fact, the usual quantum method to derive the spectrum of a system starts solving the stationary Schrödinger equation and then calculating the frequencies using the rule

$$\omega_{jk} = \frac{E_j - E_k}{\hbar} = (j-k)\omega_0,$$

where the eigenvalues of the oscillator Hamiltonian, $E_n = n\hbar\omega_0$, have been taken into account. However, in the oscillator there is a selection rule that, within the electric dipole approximation, forbids transitions except if $j-k = \pm 1$, whence the spectrum has a single frequency that agrees with the classical one. Actually, the spectrum contains also the frequencies $n\omega_0$, which correspond to electric multipole transitions, although these transitions have low probability. The multipoles of the fundamental frequency may be found also in SED calculations if the electric dipole approximation is not used, that is, if the Lorentz force eq.(188) is substituted for the last term of eq.(189). In order to be consistent, the other terms of eq.(189) should be also changed to become relativistic. Then the differential equation of motion becomes nonlinear and it is far more difficult to solve, but this may be achieved numerically and good agreement with quantum predictions is obtained for the spectrum of emitted or absorbed light [**10**].

5.2.8. Lessons for a realistic interpretation of quantum theory. *Our study of a particle in a potential well shows that inclusion of the (vacuum) random electromagnetic field leads to predictions resembling those of quantum electrodynamics:*

The QM ground state corresponds to a SED stationary state.

The spectrum of the field determines the spectrum for the motion of the particle, whence the SED variances of the coordinate and the momentum agree with QM.

No contradiction arises between the exponential distribution of energy in SED and the sharp energy in QM because they are different operational definitions. This is a good illustration of the flaw in the celebrated von Neumann theorem against hidden variables.

Radiative corrections (e.g. Lamb shift) appear naturally in SED with a transparent interpretation.

5.3. Evolution. Commutators of stochastic processes

5.3.1. Particle motion as a stochastic process. In Newtonian mechanics the study of the evolution consists of finding the position of bodies as a function of time for given initial conditions, for instance the initial positions and velocities of particles. Thus the evolution describes a curve in phase space parametrized by time. If there are forces not fully known, which we represent as noise, the evolution is a stochastic process.

From the mathematical point of view a stochastic process $x(t;\lambda)$ is a function of two variables $t \in (0,T)$, $\lambda \in \Lambda$ with a probability distribution on the set Λ. The parameter t usually represents time in applications to physics and Λ is the set of possible realizations of the function $x(t)$. It is possible to characterize a stochastic process by the probability distributions of values at one time $t, P_1[x(t)]$, at two times $P_2[x(t), x(t')]$, at three times, etc. In some cases, e.g. if the probability distributions are Gaussian, the required information consists of just the mean value of x at one time, $\langle x(t)\rangle$, and at two times $\langle x(t)\, x(t')\rangle$; the latter is named self-correlation function.

In the study of the oscillator in SED made below we will be interested in processes that are stationary and ergodic, that is, $\langle x(t)\, x(t')\rangle$ depends only on the difference of times, $t-t'$, and ensemble averages agree with time averages, that is,

$$\langle x(0)\, x(\tau)\rangle = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} x(t)\, x(t+\tau)\, dt.$$

In stationary processes that are ergodic the self-correlation is the cosine Fourier transform of the spectrum, a result known as Wiener-Khinchine theorem. That is,

$$\langle x(t)\, x(t+\tau)\rangle = \int_0^\infty S_x(\omega)\cos(\omega\tau)\, d\omega. \tag{212}$$

Particular cases of this relation are eqs.(196). Also the connection between spectrum and Fourier transform used in section 2.2 is closely related to eq.(212).

A standard method to study time dependent properties is via the self-correlations or cross-correlations of the relevant stochastic processes. We shall use this method for the free particle, but for the oscillator we will use commutators of stochastic processes, a concept that we define in the following.

5.3.2. The free particle. A free particle might be seen as the limiting case of an oscillator whose characteristic frequency decreases to zero. This fact may suggest that the differential equation of motion is like the oscillator's eq.(189) with $\omega_0 = 0$, that is,

$$m\ddot{x} = m\tau\dddot{x} + eE(t). \tag{213}$$

However, this equation is not appropriate because there is a qualitative difference between the two systems. In fact, the motion in the oscillator is always bounded, which is not the case for the free particle. Eq.(213) is a third order equation and therefore it has three independent solutions, but one of them is *runaway*, that is, it predicts that the energy increases without limit, which is physically nonsense. The reason is that the radiation reaction term, that is, the first term on the right side of eq.(213), is a linearized approximation not valid for a free particle (in the oscillator the runaway solution is effectively cut-off by the potential and an approximation like eq.(189) is good enough). Thus we shall substitute the following integro-differential equation for eq.(213)

$$\ddot{x} = -\frac{e}{m\tau}\exp\left(\frac{t}{\tau}\right)\int_t^\infty E(t')\exp\left(-\frac{t'}{\tau}\right)dt'.$$

It has the same solutions as eq.(213) except the runaway ones. Hence, it is easy to get the following equations of evolution for the velocity and the coordinate, respectively:

$$v(t) = v_0 - \frac{e}{m\tau}\int_0^t \exp\left(\frac{s}{\tau}\right)ds\int_s^\infty E(u)\exp\left(-\frac{u}{\tau}\right)du, \tag{214}$$

$$x(t) = x_0 + v_0 t - \frac{e}{m\tau}\int_0^t ds\int_0^s \exp\left(\frac{u}{\tau}\right)du\int_u^\infty E(w)\exp\left(-\frac{w}{\tau}\right)dw,$$

where x_0 is the initial position and v_0 the initial velocity at time $t = 0$. Taking into account that the ensemble average of $E(t)$ is zero, it is easy to get the mean position of the particle, that is

$$\langle x(t)\rangle = x_0 + v_0 t. \tag{215}$$

The most interesting quantities are the dispersions of velocity and position with time. The velocity dispersion may be got from the first

eq.(214) putting $v_0 = 0$. We obtain, taking eq.(190) into account,

$$\left\langle v\left(t\right)^2\right\rangle = \frac{3c^3}{2m\tau}\int_0^t \exp\left(\frac{s}{\tau}\right) ds \int_s^\infty \exp\left(-\frac{u}{\tau}\right) du$$
$$\times \int_0^t \exp\left(\frac{s'}{\tau}\right) ds' \int_s^\infty \exp\left(-\frac{u'}{\tau}\right) du' \left\langle E\left(u\right) E\left(u'\right)\right\rangle . \tag{216}$$

The $E(t)$ self-correlation is the Fourier transform of the spectrum eq.(192)); that is,

$$\left\langle E\left(u\right) E\left(u'\right)\right\rangle = \int_0^\infty S_E\left(\omega\right) \cos\left[\omega\left(u-u'\right)\right] d\omega$$
$$= \frac{1}{2}\int_{-\infty}^\infty \left|S_E\left(\omega\right)\right| \exp\left[i\omega\left(u-u'\right)\right] d\omega. \tag{217}$$

I point out that this relation is correct because $E(t)$ is a stationary process, but it is not possible to get eq.(216) from the spectrum of $v\left(t\right)$ because in the free particle case we cannot get the spectrum of $v(t)$ from that of $E(t)$ (as we made in the derivation of eq.(195) for the equilibrium state of oscillator, where both $x(t)$ and $v(t)$ are stationary processes). Inserting eq.(217) in eq.(216) we get, after changing the order of integrations,

$$\Delta v^2 \equiv \left\langle v\left(t\right)^2\right\rangle = \tag{218}$$
$$= \frac{3c^2}{2m\tau}\int_{-\infty}^\infty d\omega \left|S_E\left(\omega\right)\right| \left|\int_0^t \exp\left(\frac{s}{\tau}\right) ds \int_s^\infty \exp\left(-\frac{u}{\tau}+i\omega u\right) du\right|^2$$
$$= \frac{\hbar}{4\pi m\tau}\int_{-\infty}^\infty \frac{\left|\omega\right| d\omega}{\omega^2+\tau^{-2}} \left|1-\exp(i\omega t)\right|^2$$
$$= \frac{\hbar\tau}{\pi m}\int_0^\infty \frac{\omega d\omega}{1+\tau^2\omega^2}\left[1-\cos(\omega t)\right].$$

Thus the velocity dispersion consists of two terms. The former is an ultraviolet divergent integral that may be regularized introducing a cut-off frequency ω_c. We get

$$\Delta v_1^2 = \frac{\hbar\tau}{\pi m}\int_0^{\omega_c} \frac{\omega d\omega}{1+\tau^2\omega^2} = \frac{\hbar}{2\pi m\tau}\log(1+\omega_c^2\tau^2) \simeq \frac{\hbar\omega_c^2\tau}{2\pi m}. \tag{219}$$

The last term is convergent for $t > 0$ and it has an analytical solution that is

$$\begin{aligned}\Delta v_2^2 &= \frac{2\hbar}{\pi m\tau}\int_0^\infty \frac{z dz}{1+z^2}\cos(zt/\tau) \\ &= \cos(t/\tau)Ci\,(t/\tau) + \sin(t/\tau)\left[Si\,(t/\tau) + \frac{\pi}{2}\right] \\ &\simeq \frac{2\hbar}{\pi m\tau}\left[C + \ln\,(t/\tau)\right], \qquad t \lesssim \tau, \end{aligned} \tag{220}$$

where $C = 0.577\ldots$ is the Euler constant. In order that the dispersion Δv is not greater than the velocity of light the frequency cutoff should fulfil

$$\omega_c < \sqrt{\frac{3\pi}{\alpha}}\frac{mc^2}{\hbar} \approx 0.2\frac{c}{\lambda_C} \ll \frac{1}{\tau}.$$

A correct calculation would require a relativistic theory, which will not be attempted here. Nevertheless, the results obtained show that the particle performs a random motion with relativistic speed although the mean velocity remains a constant (see eq.(215)). The results also suggest that in a relativistic calculation the most relevant wavelengths would be those not far from the Compton one. The increase of the velocity of a free charged particle by the action of the ZPF has been proposed as the possible origen of the observed ultrahigh-energy X rays coming to Earth from outside the Solar System [**15**].

It is interesting to compare the velocity dispersion of the free particle in SED, eq.(219), with the particle in Rayleigh-Jeans (classical) radiation. Taking into account that the ZPF and the Rayleigh-Jeans radiation correspond to $\frac{1}{2}\hbar\omega$ and kT per normal mode, respectively, the replacement $\frac{1}{2}\hbar\omega \to kT$ in eq.(219) leads to

$$\begin{aligned}\Delta v^2 &= \frac{\tau kT}{\pi m}\int_0^\infty \frac{d\omega}{1+\tau^2\omega^2}\left[1 - \cos(\omega t)\right] \\ &\sim \frac{kT}{m} \text{ for } t \gg \tau. \end{aligned}$$

We see that the velocity dispersion of the charged free particle does not increase indefinitely but becomes, after a long enough time, a constant corresponding to the kinetic energy $kT/2$ (which is the equipartition of the energy of classical statistical mechanics).

The dispersion of position of the free particle according to SED may be obtained by a similar method, that is, inserting the last

eq.(217) in eq.(214) . We obtain

$$\begin{aligned}\Delta x^2 &= \frac{\hbar\tau}{2\pi m}\int_{-\infty}^{\infty}\frac{|\omega|\,d\omega}{1+\tau^2\omega^2}\left|\int_0^t\left[1-\exp(i\omega s)\right]ds\right|^2 \\ &= \frac{\hbar\tau}{\pi m}\int_0^{\infty}\frac{\omega d\omega}{1+\tau^2\omega^2}\left[t^2-\frac{2t\sin(\omega t)}{\omega}+\frac{2-2\cos(\omega t)}{\omega^2}\right] \\ &\simeq \frac{2\hbar\tau\omega_c^2}{\pi m}t^2+\frac{2\hbar\tau}{\pi m}\left[\log\left(\frac{t}{\tau}\right)-C-1\right], \qquad (221)\end{aligned}$$

The first term, which dominates at long times, is a consequence of the velocity dispersion as may be seen by a comparison with eq.(219).

For the (canonical) momentum the first eq.(203) remains valid when we put $\omega_0 = 0$, showing that the momentum is constant in time without dispersion. This agrees with the quantum prediction that the momentum of a free particle is a constant. For a particle with zero canonical momentum, the typical distance from the original position increases with about one tenth the velocity of light so that a relativistic treatment would give a quite different picture.

The picture that emerges, and gives hints for the interpretation of the free particle in QM, is as follows. The free particle possesses a conserved canonical momentum with an associated inertial motion but, superposed to this, it has a random motion with a velocity close to that of light. This produces an apparently contradictory behaviour that derives from the form of the spectrum, $S_E(\omega) \propto \omega^3$, of the zero-point field: At short times the motion is governed by the high frequencies where $S_E(\omega)$ is large, thus inducing a rapid erratic motion; at long times it is governed by the low frequencies where $S_E(\omega)$ is small, whence the memory of the initial velocity is lost very slowly. This fact contrasts with what happens in Brownian motion and what our intuition may suggest, namely that memory of the initial conditions should be quickly lost under the action of a noise. The special behaviour of diffusion in SED is a consequence of the fact that the spectrum $S_E(\omega)$ is very different from the popular (Brownian) white noise, $S_{white}(\omega) \simeq constant$.

5.3.3. Commutator vs. self-correlation. Stochastic electrodynamics provides a clue for the interpretation of commutation rules, which are essential ingredients in the standard formulation of quantum mechanics. In fact, we may define the commutator at two times, $\{t, t'\}$, of a stationary stochastic process, $x(t)$, via the Fourier *sine* transform of the spectrum. Then I shall show that the stochastic

commutator applied to the SED oscillator closely resembles the quantum commutator in the Heisenberg picture of QM.

The introduction of the stochastic commutator is suggested by the standard Fourier transform of the spectrum, $S_x(\omega)$. It consists of two terms, namely,

$$\int_0^\infty S_x(\omega)\exp\left[i\omega\left(t'-t\right)\right]d\omega=\left\langle x(t)\,x(t')\right\rangle+\frac{1}{2}\left[x(t),x(t')\right],$$

$$\left\langle x(t)\,x(t')\right\rangle=\int_0^\infty S_x(\omega)\cos\left[\omega\left(t'-t\right)\right]d\omega,$$

$$\left[x(t),x(t')\right]=2i\int_0^\infty S_x(\omega)\sin\left[\omega\left(t'-t\right)\right]d\omega, \tag{222}$$

where the spectrum is defined to be zero for negative frequencies, that is, $S_x(\omega)=0$ if $\omega<0$. The real part, $\langle x(t)\,x(t')\rangle$, is the self-correlation function of the stochastic process as said above. Therefore, it is plausible that the imaginary part, $[x(t),x(t')]$, is also relevant. The factor 2 is chosen in order to stress the similarity with the quantum commutator.

The relation between spectrum and stochastic commutator is also suggested by the fact that in QM there is a similar relation between the spectrum and the two-times commutator of the coordinate operator in the Heisenberg picture. That relation is fulfilled for the ground state of a particle in any potential well. For the proof I consider a one-dimensional (quantum) problem defining the spectrum, $S_x(\omega)$, as follows

$$S_x(\omega)\equiv\sum_n\left|\left\langle\psi_0\left|\hat{x}(0)\right|\psi_n\right\rangle\right|^2\delta\left(\omega-\omega_{0n}\right), \tag{223}$$

where $\hat{x}$ is the quantum position operator of the particle. The coefficients of the Dirac's deltas are proportional to the transition probabilities in QM from the ground state to all possible excited states. (Although I stress that in QED the deltas are approximations of highly peaked functions with a finite width when radiative corrections are taken into account). The analogy with the last eq.(222) is shown as follows. From the Heisenberg equation of motion

$$\hat{x}(t)=\exp(i\hat{H}t/\hbar)\hat{x}(0)\exp(-i\hat{H}t/\hbar)$$

we may obtain the expectation value of the commutator in the ground state,

$$\left\langle\left[\hat{x}(0),\hat{x}(t)\right]\right\rangle=\left\langle\psi_0\left|\hat{x}(0)\,\hat{x}(t)\right|\psi_0\right\rangle-\left\langle\psi_0\left|\hat{x}(t)\,\hat{x}(0)\right|\psi_0\right\rangle.$$

After introducing the resolution of the identity between $\hat{x}(0)$ and $\hat{x}(t)$ and between $\hat{x}(t)$ and $\hat{x}(0)$ in terms of eigenvectors of the Hamiltonian $\hat{H}$, this gives

$$\begin{aligned}\langle[\hat{x}(0),\hat{x}(t)]\rangle &= 2i\sum_n |\langle\psi_0|\hat{x}(0)|\psi_n\rangle|^2 \sin(\omega_{0n}t)\\ &= 2i\int S_x(\omega)\sin(\omega t)\,d\omega,\end{aligned}$$

where in the last equality we have taken eq.(223) into account. This equality, similar to the stochastic term in eq.(222), played an important role in the origin of quantum mechanics. In fact, the derivative with respect to t leads to

$$2im\sum_n \omega_{0n}|\langle\psi_0|\hat{x}(0)|\psi_n\rangle|^2\cos(\omega_{0n}t) = [\hat{x}(0),\hat{p}(t)],$$

which in the limit $t\to 0$ becomes an example of the well known Thomas-Reiche-Kuhn sum rule. The rule is usually applied to atoms where a sum over the three coordinates of the Z electrons is performed, so that it reads

$$2m\sum_n \omega_{0n}\left|\left\langle\psi_0\left|\sum_{j=1}^{Z}\mathbf{r}_j\right|\psi_n\right\rangle\right|^2 = -i\sum_{k=1}^{3Z}[\hat{x}_k(0),\hat{p}_k(0)] = 3Z\hbar.$$

For a stationary process both the self-correlation and the commutator depend only on the time difference $(t-t')$. In this case the last term in eq.(222) may be inverted via a time integral. In fact, we get

$$\int_{-\infty}^{\infty}\sin[\nu(t-t')][x(t),x(t')]\,dt = \tag{224}$$

$$= \int_{-\infty}^{\infty}\sin[\nu(t-t')]dt 2i\int_0^{\infty}S_x(\omega)\sin[\omega(t'-t)]\,d\omega$$

$$= -2i\int_0^{\infty}S_x(\omega)\,d\omega\int_{-\infty}^{\infty}\sin[\nu(t-t')]\sin[\omega(t-t')]\,dt = -i\pi S_x(\nu),$$

where in the last equality we take into account that $S_x(\nu)=0$ for $\nu<0$.

5.3.4. Application to the SED oscillator. All stationary properties of the SED oscillator studied in section 2 may be equally

well obtained either from the spectrum eq.(195), from the self-correlation or from the commutator, the latter being

$$\begin{aligned}\left[x\left(t'\right),x\left(t\right)\right] &= 2i\int_0^\infty \frac{\hbar\tau\omega^3\sin\left[\omega\left(t-t'\right)\right]d\omega}{\pi m\left[\left(\omega_0^2-\omega^2\right)^2+\tau^2\omega^6\right]} \\ &= i\int_{-\infty}^\infty \frac{\hbar\tau\omega^3\exp\left[i\omega\left(t-t'\right)\right]d\omega}{\pi m\left[\left(\omega_0^2-\omega^2\right)^2+\tau^2\omega^6\right]}. \qquad (225)\end{aligned}$$

This equality shows the advantage of the commutator with respect to the self-correlation in the stochastic process associated to the SED oscillator. In fact, the last integral may be performed analytically via the method of residues, whilst getting the self-correlation is more involved. Of course, all properties derived in the following might be obtained using self- or cross-correlations (see e.g. [**1**]), but the use of commutators will exhibit their advantages and shortcomings.

I propose that the success of the use of noncommuting observables in QM, and indirectly the adequacy of the Hilbert space formulation, might be related to the fact that the basic stochastic processes involved have spectra that are odd with respect to the change $\omega \to -\omega$. This contrasts with classical physics where the stochastic processes that appear most times have spectra that are *even* in the frequency, as is the case for white noise in Brownian motion.

Performing the integral eq.(225) is straightforward, extending the variable ω to the complex plane. First we shall find the poles of the integrand, which requires the solution of an algebraic equation of third degree, that is, finding the 6 zeroes of the denominator of the spectrum eq.(195). The solution is involved, but a fairly good approximation is as follows, resting on the fact that $\tau\omega_0^2 \ll 1$. There are poles close to $\omega_0 = \pm\omega$ that are solutions of the equation

$$\left(\omega_0^2-\omega^2\right)^2 = -\tau^2\omega^6 \simeq -\tau^2\omega_0^6 \qquad \Rightarrow \qquad \omega = \pm(\omega_0 \pm \frac{1}{2}i\tau\omega_0^3).$$

Two other poles correspond to high values of $\omega \gg \omega_0$, that is,

$$\omega^6 = -\tau^{-2}\left(\omega_0^2-\omega^2\right)^2 \simeq -\tau^{-2}\omega^4 \qquad \Rightarrow \qquad \omega = \pm i\tau^{-1}. \qquad (226)$$

Finding the zeroes of the denominator allows factorizing it so that we approximate eq.(225) as

$$[x(0), x(t)] \simeq \frac{i\hbar}{\pi m}\int_{-\infty}^{\infty} \frac{\omega^3 \exp(i\omega t)\, d\omega}{\tau\left[(\omega_0+\omega)^2+\tau^2\omega_0^4\right]\left[(\omega_0-\omega)^2+\tau^2\omega_0^4\right](\omega^2+\tau^{-2})},$$

where we have put $t'=0$ without loss of generality because the commutator depends only on the difference of times in a stationary stochastic process. The integral may be evaluated by the method of residues, taking into account that the 3 poles in the upper (lower) half of the complex plane of the variable ω contribute for $t>0$ (for $t<0$). This leads to a change of sign of the commutator for $t\to -t$, as expected. Thus we get

(227)
$$[x(0), x(t)] = \frac{i\hbar}{m\omega_0}\left\{\sin(\omega_0 t)\exp\left(-\tau\omega_0^2|t|\right)+\tau\omega_0\frac{t}{|t|}\exp\left(-\frac{|t|}{\tau}\right)\right\}.$$

For the canonical momentum we obtain, taking eq.(203) into account,

$$[p(0), p(t)] = \frac{i\hbar m}{\omega_0}\sin(\omega_0 t)\exp\left(-\tau\omega_0^2|t|\right), \tag{228}$$

where we neglect terms of order $\tau^2\omega_0^2$. There are no terms proportional to τ here, like the last one of eq.(227) .

I point out that the commutator of the particle coordinate in the SED oscillator may be derived from the commutator of the electric field of the zero-point radiation, taking eqs.(192) and (222) into account. Actually, the relation between the quantum commutator of the vacuum field and the commutator of the Heisenberg (time dependent) position operator of a charged particle may be obtained in the context of quantum mechanics without passing through the spectrum. Indeed, this was made long ago by Schiller [**1**], [**4**].

It is possible to define the derivative of a commutator with respect to time:

$$\frac{d}{dt'}[x(t), x(t')] = \lim_{t''\to t'}\frac{[x(t), x(t'')]-[x(t), x(t')]}{t''-t'} = \lim_{t''\to t'}\frac{[x(t), x(t'')-x(t')]}{t''-t'} = \left[x(t), \frac{dx(t')}{dt'}\right],$$

where the linearity of the commutator has been used. Hence we obtain from eq.(227)

$$[x(0), \dot{x}(t)] = \frac{i\hbar}{m}\cos(\omega_0 t)\exp\left(-\tau\omega_0^2|t|\right) + \frac{2i\hbar\tau}{m}\delta(t) = -[\dot{x}(0), x(t)].$$

Similarly it is easy to get the velocity commutator:

$$[v(0), v(t)] = [\dot{x}(0), \dot{x}(t)] = \frac{2i\hbar\tau\omega_0^2}{m}\delta'(t) + \frac{i\hbar\omega_0}{m}\left\{\sin(\omega_0 t) + 2\tau\omega_0\frac{t}{|t|}\cos(\omega_0 t)\right\}\exp\left(-\tau\omega_0^2|t|\right). \tag{229}$$

A most relevant result is that to zeroth order in τ we obtain

$$[x(0), p(t)] = m[x(0), \dot{x}(t)] = i\hbar\cos(\omega_0 t) \quad \Rightarrow \quad [x(0), p(0)] = i\hbar,$$

the latter being similar to the fundamental commutation rule of quantum mechanics for one particle. The first equality is correct because the canonical momentum p differs from the product $m\dot{x}$ by terms of order τ.

The stochastic commutator provides a hint for a realistic interpretation of the quantum commutation rules as a disguised form of establishing the properties of some peculiar stochastic processes. The main peculiarity is the fact that the spectra of the processes involved are usually odd in the frequency.

However, I shall point out that the analogy of the stochastic commutator with the commutators of quantum mechanics holds only for linear problems, this being also the domain of validity of SED (see section 5.6 below). In fact, the noncommutativity of operators in quantum mechanics is an essential consequence of the Hilbert space formulation and it plays a far more relevant role than the commutators of stochastic processes here studied.

5.3.5. The Hilbert transform. Our definition of commutator may be generalized to two different stationary stochastic processes as follows:

DEFINITION 11. *Given two stationary stochastic processes, $x(t)$ and $y(t)$, I define their (stochastic) commutator, $[x(t), y(t')]$, as $2i$ times the Hilbert transform of the cross-correlation, $\langle x(t)\, y(t')\rangle$.*

The Hilbert transform, $g(u)$, of a function $f(t), t \in (-\infty, \infty)$ is defined by

$$g(u) = \frac{1}{\pi}P\int_{-\infty}^{\infty} f(t)\frac{1}{u-t}dt, \quad f(t) = \frac{1}{\pi}P\int_{-\infty}^{\infty} g(u)\frac{1}{u-t}du, \tag{230}$$

where P means principal part and the second equality corresponds to the inverse transform. However, the inverse transform does not always recover the original. For instance, the Hilbert transform of a constant is zero and the inverse of zero is also zero. The relevant property for us is that the Hilbert transform changes $\sin(\omega t)$ into $\cos(\omega u)$ and $\cos(\omega t)$ into -$\sin(\omega u)$, provided that $\omega \neq 0$.

Using the inverse Hilbert transform, the second eq.(230), we may obtain the self-correlations and cross-correlations of stationary stochastic processes. It is convenient to get them as integrals in the interval $(0,\infty)$ rather than $(-\infty,\infty)$, writing

$$\begin{aligned} f(t) &= \frac{2}{\pi} P \int_0^\infty g(u) \frac{u}{u^2 - t^2} du \text{ if } g(-u) = -g(u), \\ f(t) &= \frac{2}{\pi} P \int_0^\infty g(u) \frac{t}{u^2 - t^2} du \text{ if } g(-u) = g(u). \end{aligned}$$

Hence taking eq.(227) into account we get the self-correlation

(231)
$$\langle x(0) x(t) \rangle = \frac{1}{2i} \times \frac{2}{\pi} P \int_0^\infty g(u) \frac{u}{u^2 - t^2} du, \quad g(u) \equiv [x(0), x(u)],$$

leading to

(232) $$\langle x(0) x(t) \rangle = \frac{\hbar}{2m\omega_0} \cos(\omega_0 t) \gamma(t) + \frac{2\hbar\tau}{\pi m} \left[\log\left(\frac{t}{\tau}\right) - C - 1 \right],$$

where the first term is the product of the result in the limit $\tau \to 0$ times a slowly varying function $\gamma(t)$ that fulfils

$$\gamma(t) \to 1 \text{ for } t \to 0, \qquad \gamma(t) \to 0 \text{ for } t \to \infty.$$

The last term is valid for $t \ll \omega_0^{-1}$ because it derives from the last term of eq.(227) and that term came from the pole eq.(226) that was obtained neglecting ω_0. For that range of values of t the last term of eq.(232) is the same that appears in the free particle eq.(221). Indeed, for these values of t the evolution is dominated by the ZPF and the effect of the oscillator potential is irrelevant. The self-correlation of the canonical momentum may be obtained from the commutator eq.(228) and we get

(233) $$\langle p(0) p(t) \rangle = \frac{\hbar m \omega_0}{2} \cos(\omega_0 t) \gamma(t).$$

The results obtained lead to a picture of the SED oscillator as follows. The motion of the charged particle may be seen as a superposition of two motions, one of them smooth and the other one very

irregular, represented by the two terms of eq.(232) . The latter prevents the possibility of defining an instantaneous velocity, but only the mean velocity over not too small time intervals. This fits qualitatively with the restrictions derived from the Heisenberg uncertainty relations. Nevertheless, the irregular motion does not produce a loss of memory of the initial conditions, as shown by the first term, that corresponds to a motion close to the classical one. Indeed, if we ignored the factor $\gamma(t)$ the self-correlation given by the first term alone would correspond precisely to an ensemble of classical oscillators. However, there is a slow passage from one classical path to another neighbour path.

An alternative way to study the motion of the oscillator in SED is to introduce new variables, $a(t)$ and $b(t)$, as follows

$$\begin{aligned} x(t) &= a(t)\cos(\omega_0 t) + b(t)\sin(\omega_0 t), \\ \dot{x}(t) &= -a(t)\omega_0 \sin(\omega_0 t) + b(t)\omega_0 \cos(\omega_0 t). \end{aligned} \tag{234}$$

It may be realized that these variables correspond to constants of the motion of the mechanical oscillator. That is, they would be time-independent if we ignored the last two terms of the SED eq.(189) . In order to study the variation we may invert eqs.(234) leading to

$$\begin{aligned} a(t) &= x(t)\cos(\omega_0 t) - \omega_0^{-1}\dot{x}(t)\sin(\omega_0 t), \\ b(t) &= x(t)\sin(\omega_0 t) + \omega_0^{-1}\dot{x}(t)\cos(\omega_0 t). \end{aligned} \tag{235}$$

It is straightforward to derive the self-correlations of $a(t)$ and $b(t)$ and their cross-correlation from the correlations of $x(t)$ and its derivatives or, alternatively, to use the commutators whence to get the desired correlations. The results will not be reported here.

It is possible to derive differential equations for the probability densities of a and b. They have the form of Fokker-Planck (or diffusion) equation and the result is [**1**]

$$\frac{\partial \rho(a,t)}{\partial t} = \frac{1}{2}\tau\omega_0^2 \frac{\partial}{\partial a}(a\rho) + \frac{\hbar\tau}{4m}\frac{\partial^2 \rho}{\partial a^2}, \tag{236}$$

and a similar one with b substituted for a. But I point out that the stochastic processes $a(t)$ and $b(t)$ are correlated.

The conclusion of the calculations is that the classical constants of the motion, like the parameters a and b or the energy, U, perform a slow random motion with typical relaxation time $1/\left(\tau\omega_0^2\right)$. In particular, the energy is related to these parameters as follows:

$$U(t) = \frac{1}{2}m\omega_0^2 x(t)^2 + \frac{1}{2}m\dot{x}(t)^2 = \frac{1}{2}m\omega_0^2\left[a(t)^2 + b(t)^2\right], \tag{237}$$

where eqs.(234) have been taken into account. The change of the classical constants of motion of the oscillator in SED is obviously due to the two last terms of eq.(189) . The term involving the vacuum field produces diffusion, characterized by D, and the radiation reaction term gives rise to drift, characterized by A. The diffusion rate is independent of the particle's velocity, whence D is a constant, but the drift increases with the velocity and the result is that A is proportional to a. The effect of the diffusion is reduced to some extent by the drift, with the consequence that the probability densities remain localized. In fact, when time increases indefinitely the densities approach the stationary solution studied in section 5.2. In particular, the stationary solution of eq.(236) (with normalized ρ, which implies that ρ vanishes for $a \to \pm\infty$) is

$$\rho = \sqrt{\frac{\hbar}{\pi m\omega_0^2}}\exp\left(-\frac{\hbar}{m\omega_0^2}a^2\right),$$

and similar for b. The energy tends to the density given by eq.(207) .

5.3.6. States of the oscillator in stochastic electrodynamics and in quantum mechanics. Every nonnegative definite function in phase space may be taken as an initial probability density and thus be considered a state of the SED oscillator. The set of states in SED is quite different from the set of states in QM (given by a density operator each). In particular, the pure states in SED are those whose initial conditions correspond to points in phase space, whilst the pure states in QM correspond to state vectors (or wave functions). The comparison between SED and QED becomes more clear if we define the quantum states by means of functions in phase space, which might be achieved via the Wigner function formalism revisited in chapter 4. Thus only a small fraction of pure quantum states correspond to states in SED. Actually, asides from the ground state there are only two interesting pure quantum states that correspond precisely to SED states, namely coherent states and squeezed states. The latter are relevant in case of radiation (squeezed states of light), but not so much for matter oscillators and they will not be studied here.

Coherent states in SED appear as solutions of the oscillator eq. (189) obtained by combining the stationary solution of the equation with the general solution of the homogenous equation, that is,

$$\ddot{x} + \omega_0^2 x + \tau\omega_0^2\dot{x} = 0 \qquad \Rightarrow \qquad x \simeq A\cos\left(\omega_0 t + \phi\right)\exp\left(-\tau\omega_0 t\right),$$

where I have approximated $\dddot{x} \simeq -\omega_0^2 \dot{x}$ and neglected a small shift, of order τ, in the frequency ω_0. Hence, taking eq.(205) into account, we see that the solution of eq.(189) leads to the following time dependent probability distribution of positions

(238)

$$W(x,t) \simeq \sqrt{\frac{m\omega_0}{\pi\hbar}} \exp\left[-\frac{m\omega_0}{2\hbar}\left[x - A\cos(\omega_0 t + \phi)\exp\left(-\tau\omega_0^2 t\right)\right]^2\right],$$

which contains two integration constants, A and ϕ. It must be stressed that this expression for the probability density derives from eq.(189) and the ZPF spectrum eq.(192) with the approximation of putting $\tau \to 0$ except in the exponential decay. It may be seen that when $\tau = 0$ the evolution of the position probability density eq.(238) fully agrees with the one of the coherent states of quantum mechanics, whilst the expression for finite τ contains the most relevant contribution of the radiative corrections of quantum electrodynamics to these states (a decay towards the stationary state with relaxation time $\left(\tau\omega_0^2\right)^{-1}$) [**7**].

In summary we see that a few states of the oscillator in SED correspond to pure quantum states in a phase-space representation. But no pure state of SED corresponds to a state of QM. Furthermore, most of the pure states of QM do not correspond to states of SED. However, for mixed states the agreement is greater, and actually all (mixed) quantum states possessing a positive Wigner function closely correspond to mixed states of SED with the same phase-space distribution.

In spite of these differences the quantum theory of the harmonic oscillator admits a realistic interpretation via SED, provided that the same predictions may be obtained for actual experiments. In particular, the quantum states, solutions of Schrödinger equation, might be just mathematical auxiliary functions used in the QM formalism but not required in the SED approach for the prediction of empirical results.

5.4. Coupled oscillators in SED

The generalization of the harmonic oscillator in SED to many dimensions is straightforward using the appropriate extension of eq. (189) . In the following I will study two simple examples of coupled oscillators. Firstly a system of two one-dimensional oscillators at a long distance as an example of van der Waals force. The system is interesting because it shows that a phenomenon similar to quantum

entanglement appears also in SED. The second example is an array of coupled three-dimensional oscillators at a finite temperature, that reproduces Debye theory of the specific heat of solids.

5.4.1. A model for quantum entanglement. Entanglement is a quantum property of systems with several degrees of freedom, which appears when the total state vector cannot be written as a product of vectors associated to one degree of freedom each. In formal terms a typical entangled state fulfils

$$| \psi (1,2)\rangle = \sum_{m,n} c_{mn} \mid \psi_m (1)\rangle \mid \psi_n (2)\rangle, \tag{239}$$

where 1 and 2 correspond to two different degrees of freedom, usually belonging to different subsystems. The essential condition is that the state eq.(239) cannot be written as a product, that is, the sum cannot be reduced to just one term via a change of basis in the Hilbert space. See chapter 2, section 2.3.2. Entanglement appears as a specifically quantum form of correlation, which is claimed to be dramatically different from the correlations of classical physics. The latter may be sometimes written in the form

$$\rho (1,2) = \sum_{m,n} w_{mn} \rho_m (1) \rho_n (2) , \tag{240}$$

where the quantities $\rho \gtrsim 0$ are probability densities and the coefficients play the role of weights fulfilling $w_{mn} \gtrsim 0$, in sharp contrast with eq.(239) where $\mid \psi\rangle$ are vectors in a Hilbert space and c_{mn} are complex numbers.

Actually, entanglement is quite common in nonrelativistic quantum mechanics of many-particle systems, e.g. for electrons in atoms or molecules. However, it is most relevant when the state vectors $\mid \psi_m (1)\rangle$ and $\mid \psi_n (2)\rangle$ of eq.(239) belong to different systems placed far from each other. A study of entanglement and its relation with 'local realism' was made in chapter 3 section 3.2.5 and examples of photon entanglement will be provided in chapter 6. Here I will illustrate, with a simple example, that entanglement might be understood as a correlation induced by quantum vacuum fluctuations acting in two different places.

An example: London-van der Waals forces. I shall study the London theory of the van der Waals forces in a simple model of two one-dimensional oscillating electric dipoles. Each dipole consists of a particle at rest and another particle (which we will name electron)

with mass m and charge e. In the model it is assumed that every electron moves in a harmonic oscillator potential and that there is an additional interaction between the electrons. Thus the Hamiltonian is

$$H = \frac{p_1^2}{2m} + \frac{1}{2}m\omega_0^2 x_1^2 + \frac{p_2^2}{2m} + \frac{1}{2}m\omega_0^2 x_2^2 - Kx_1x_2, \tag{241}$$

where $x_1(x_2)$ is the position of the electron of the first (second) dipole with respect to the equilibrium position. The positive parameter $K < m\omega_0^2$ depends on the distance between the dipoles, but the dependence is irrelevant for our purposes. (For a more complete study of this problem within SED see Refs. [**1**], [**2**]). We shall revisit both the QM and the SED calculations.

Quantum theory of the model. An exact quantum calculation is not difficult. We take x_j, p_j and H as operators in the Hilbert space of the full system, fulfilling the standard commutation relations

$$[\hat{x}_j, \hat{x}_l] = [\hat{p}_j, \hat{p}_l] = 0, \qquad [\hat{x}_j, \hat{p}_l] = i\hbar\delta_{jl}. \tag{242}$$

Now we introduce the new operators

$$\begin{aligned} \hat{x}_+(t) &= \frac{1}{\sqrt{2}}\left[\hat{x}_1(t) + \hat{x}_2(t)\right], \quad \hat{x}_-(t) = \frac{1}{\sqrt{2}}\left[\hat{x}_1(t) - \hat{x}_2(t)\right], \\ \hat{p}_+(t) &= \frac{1}{\sqrt{2}}\left[\hat{p}_1(t) + \hat{p}_2(t)\right], \quad \hat{p}_-(t) = \frac{1}{\sqrt{2}}\left[\hat{p}_1(t) - \hat{p}_2(t)\right]. \end{aligned} \tag{243}$$

It is easy to derive the commutation relations of the new operators, that are similar to eqs.(242) with the subindices $+, -$ substituted for $1, 2$. The Hamiltonian eq.(241) in terms of the new operator is

$$\hat{H} = \frac{\hat{p}_+^2}{2m} + \frac{1}{2}\left(m\omega_0^2 + K\right)\hat{x}_+^2 + \frac{\hat{p}_-^2}{2m} + \frac{1}{2}\left(m\omega_0^2 - K\right)\hat{x}_-^2.$$

This is equivalent to two uncoupled harmonic oscillators with the same mass, m, and frequencies

$$\omega_+ = \sqrt{\omega_0^2 + K/m}, \qquad \omega_- = \sqrt{\omega_0^2 - K/m}, \qquad \left(K < m\omega_0^2\right)$$

respectively. Thus the wave-function of the two-electron system is

$$\begin{aligned}(244)\quad \psi &= \psi(x_+)\,\psi(x_-)\\ &= \sqrt{\frac{m}{\pi\hbar\sqrt{\omega_+\omega_-}}}\exp\left[-\frac{m}{2\hbar}\left(\omega_+x_+^2+\omega_-x_-^2\right)\right]\\ &= \sqrt{\frac{m}{\pi\hbar\sqrt{\omega_+\omega_-}}}\exp\Big\{-\frac{m}{4\hbar}\big[(\omega_++\omega_-)\left(x_1^2+x_2^2\right)\\ &\quad +2(\omega_+-\omega_-)\left(x_1x_2\right)\big]\Big\},\end{aligned}$$

and the interaction energy of the system is

$$\Delta E=\frac{\hbar}{2}\left(\sqrt{\omega_0^2-K/m}+\sqrt{\omega_0^2-K/m}-2\omega_0\right)=-\frac{\hbar K^2}{4m^2\omega_0^3}+O\left(K^4\right),$$

which to lowest nontrivial order in the coupling constant K gives

$$(245)\qquad \psi=\sqrt{\frac{m\omega_0}{\pi\hbar}}\left(1+\frac{2Kx_1x_2}{m\omega_0}\right)\exp\left[-\frac{m\omega_0}{2\hbar}\left(x_1^2+x_2^2\right)\right].$$

This may be written in terms of the wave-functions of the ground state, $\psi_0(x)$, and the first excited state, $\psi_1(x)$, of the simple oscillator as follows

$$\psi=\psi_0(x_1)\,\psi_0(x_2)+\frac{K}{m\omega_0^2}\psi_1(x_1)\,\psi_1(x_2).$$

(The function is not normalized because the normalization was lost when we truncated at first order the expansion in powers of K). In quantum language this wave-function ψ may be interpreted saying that the two-system state is a sum of two amplitudes; one of them corresponds to both oscillators being in the ground state and the other one to both being in the first excited state. It is true that eq.(245) is not an irreducible sum of products like eq.(239) . However, it cannot be factorized in terms of wave-functions of individual electrons and therefore it is not a classical correlation that might be represented as eq.(240). Therefore, it may be considered an entangled state involving two distant systems.

Although quantum mechanics usually does not offer intuitive pictures of phenomena, in this case it is difficult to refrain from interpreting the entanglement in this example as a correlation of the (random) motions of the electrons. Indeed, the modulus squared of the wave-function eq.(245) gives the probability density for the positions of the electrons, which is larger when the electrons are far from each other

so that their mutual repulsion energy is smaller. Then the correlation (entanglement) lowers the energy, giving rise to an attractive force between the oscillators. Of course, this explanation departs from the Copenhagen interpretation (see chapter 2), that should not speak about the probability that *one electron is* in the region $x_1 > 0$ and the *other one is* in the region $x_2 > 0$. Instead it compels us to say something like "if we perform a measurement of the simultaneous positions of the electrons *the probability that we get* one of them in the region $x_1 > 0$ and the other one is in the region $x_2 > 0$ is given by the squared modulus of eq.(245)". (Simultaneous measurements are possible because the observables commute). In any case, the origin of the correlation is not clear in quantum mechanics.

The model in stochastic electrodynamics. In sharp contrast with QM the interpretation offered by SED is transparent: the random motion of the electrons is induced by the ZPF, and the correlation is produced by the interaction. The SED calculation is as follows. The differential equations of motion may be obtained from eq.(241) adding the forces due to the random ZPF and the radiation reaction, see eq.(189), that is,

$$\begin{aligned} m\ddot{x}_1 &= -m\omega_0^2 x_1 - Kx_2 + \frac{2e^2}{3c^3}\dddot{x}_1 + eE_1(t), \\ m\ddot{x}_2 &= -m\omega_0^2 x_2 - Kx_1 + \frac{2e^2}{3c^3}\dddot{x}_2 + eE_2(t). \end{aligned} \tag{246}$$

However, neglecting the x dependence of the field, $E(\mathbf{x},t)$, as made in eq.(189), is not good if the dipoles are at a long distance (on the other hand the Hamiltonian eq.(241) is not valid for short distances). We may neglect the x dependence within each dipole; that is we can approximate $E(\mathbf{x}_1,t) \simeq E(\mathbf{a},t)$, $E(\mathbf{x}_2,t) \simeq E(\mathbf{b},t)$, where $\mathbf{a}$ and $\mathbf{b}$ are the positions of the first and second dipole, respectively. Also we will simplify the notation writing $E_1(t)$ for $E(\mathbf{a},t)$ and $E_2(t)$ for $E(\mathbf{a},t)$. Furthermore, as we assume that the distance between dipoles is large, we shall take the stochastic processes $E_1(t)$ and $E_2(t)$ as uncorrelated.

The coupled eqs.(246) may be decoupled via writing new equations which are the sum and the difference of the former, and introducing the new position variables

$$x_+(t) = \frac{1}{\sqrt{2}}\left[x_1(t) + x_2(t)\right], \quad x_-(t) = \frac{1}{\sqrt{2}}\left[x_1(t) - x_2(t)\right], \tag{247}$$

and similarly definitions for $E_+(t)$ and $E_-(t)$. We get

$$m\ddot{x}_+ = -(m\omega_0^2 - K)x_+ + \frac{2e^2}{3c^3}\dddot{x}_+ + eE_+(t),$$

$$m\ddot{x}_- = -(m\omega_0^2 + K)x_- + \frac{2e^2}{3c^3}\dddot{x}_- + eE_-(t), \tag{248}$$

where the stochastic processes $E_+(t)$ and $E_-(t)$ are statistically independent as a consequence of $E_1(t)$ and $E_2(t)$ being uncorrelated. It is not difficult to show that the spectra of both $E_+(t)$ and $E_-(t)$ are again given by eq.(192). With the method used to solve eqs.(197) and (199) we get

$$\left\langle x_\pm^2 \right\rangle = \frac{\hbar}{2m\sqrt{\omega_0^2 \mp K/m}}, \qquad \left\langle v_\pm^2 \right\rangle = \frac{\hbar\sqrt{\omega_0^2 \mp K/m}}{2m}. \tag{249}$$

The Hamiltonian eq.(241) may be written in terms of $x_+(t)$, $x_-(t)$ leading to

$$H = \frac{p_+^2}{2m} + \frac{1}{2}m\omega_0^2 x_+^2 + \frac{p_-^2}{2m} + \frac{1}{2}m\omega_0^2 x_-^2 - \frac{1}{2}K\left(x_+^2 - x_-^2\right).$$

Hence, defining $p_\pm = mv_\pm$, it is easy to get the total energy, $\langle H \rangle$, taking eqs.(249) into account. The result is in agreement with the quantum prediction. The joint probability distribution of positions is Gaussian and factorizes because eqs.(248) are decoupled. That is,

$$\rho(x_+, x_-)\,dx_+dx_- = \rho_+(x_+)\,\rho_-(x_-)\,dx_+dx_-.$$

The densities $\rho_\pm$ should be normalized whence we get

$$\rho_\pm(x) = \sqrt{\frac{2m}{\pi\hbar}}\left(\omega_0^2 \mp K/m\right)^{-1/4} \exp\left[-\frac{m}{2\hbar}\sqrt{\omega_0^2 \mp K/m}\,x_\pm^2\right].$$

Hence it is easy to get the joint probability in terms of the variables x_1 and x_2 taking eqs.(247) into account. The result is in agreement with the quantum prediction, eq.(245).

In the equation of motion eq.(246) I have assumed that the ZPF components, $E_1(t)$ and $E_2(t)$, acting on the two particles are uncorrelated. This is a good approximation if the particles are at a distance which is large in comparison with wavelength, $\lambda \simeq c/\omega_0$, corresponding to the typical frequencies involved. However, if the distance is of that order or smaller, the ZPF components will be correlated, which would cause a much stronger correlation between the particle's motions. We might speculate that correlations induced by the ZPF are

the SED alternative to quantum statistics, that is, behaviour of particles as either bosons or fermions. But this possibility will not be further discussed here.

The conclusion of our study of the two coupled oscillators in SED is the suggestion that quantum entanglement is a correlation between the quantum fluctuations of different systems, mediated by the vacuum fields, these fields not being apparent in the quantum formalism.

5.4.2. Specific heats of solids. An application of SED at a finite temperature is the calculation of the specific heat of solids, which we summarize in the following [**16**]. We shall consider a solid as a set of positive ions immersed in an electron gas. As is well known the electrons contribute only slightly to the specific heat at not too high temperatures. In SED we shall study the motion of the ions under the action of three forces. The first one derives from the interaction with the neighbour ions and the electron gas, that may be modelled by an oscillator potential when the distance between neighbour ions departs from the equilibrium configuration. The second is the random background radiation with Planck spectrum (including the ZPF) and the third one is the radiation reaction. The motion under the action of these forces is governed by a discrete set of coupled third order differential equations that may be decoupled by the introduction of normal mode coordinates. After that, every equation is similar to eq.(189) and may be solved in analogous form. The net result is that the mean (potential plus kinetic) energy in equilibrium is

$$E(\omega) = \frac{1}{2}\hbar\omega \coth\left(\frac{\hbar\omega}{2kT}\right), \tag{250}$$

where ω is the frequency of the mode. With an appropriate distribution, $\rho(\omega)$, of modes this leads to the quantum result derived by Debye [**17**], the specific heat being the derivative of the total energy with respect to the temperature.

There are other interesting results of SED at a finite temperature, in particular about magnetic properties. They may be seen in the books of de la Peña et al. [**1**], [**2**] and references therein.

The SED calculation of the specific heat of solids provides another argument for the continuity (as opposed to discreteness) of the energies of quantum oscillators. It is hard to accept that electromagnetic radiation consists of particles (photons) in a realistic interpretation of quantum physics, but it is still harder to assume that quantized oscillations of the ions in a solid (phonons) are particles. It is more

plausible to assume that the energies of the normal modes of the set of ions have a continuous, although random, distribution of energies such that the average for a mode is given by eq.(250). *It is also plausible that the mean energy of a vibration mode of every ion is the same as the mean energy of the radiation mode of the same frequency, which is the result here obtained.*

5.5. The particle in a homogeneous magnetic field

Another linear problem that has been extensively studied within SED is the motion of a charged particle in a homogeneous magnetic field [**1**]. The most relevant result is the prediction of diamagnetic properties of a free charge (without magnetic moment), which departs from classical physics and agrees with QM. Here I shall revisit the SED calculation of the free charged particle in a homogeneous field of magnitude B.

5.5.1. Classical and quantum theory. The classical motion may be got from Newton's law with the Lorentz force, that is,

$$m\ddot{\mathbf{r}} = -\left(e/c\right)\dot{\mathbf{r}} \times \mathbf{B}. \tag{251}$$

If we choose the Z axis in the direction of the magnetic field $\mathbf{B}$ the motion in that direction is uniform and in the perpendicular plane it is given by

$$\begin{aligned} x &= R\cos\left[2\omega_0\left(t-t_0\right)\right]+x_0, \\ y &= R\sin\left[2\omega_0\left(t-t_0\right)\right]+y_0, \qquad \omega_0 \equiv \frac{eB}{2mc}, \end{aligned} \tag{252}$$

with four integration constants, namely $\{R, t_0, x_0, x_0\}$. The motion is circular with radius R and constant (Larmor) angular frequency ω_0. The total energy E may be identified with the Hamiltonian, that is

$$H = \frac{p_x^2}{2m}+\frac{p_y^2}{2m}-\omega_0\left(xp_y-yp_x\right)+\frac{1}{2}m\omega_0^2\left(x^2+y^2\right). \tag{253}$$

Taking Hamilton equations into account we get

$$E = \frac{1}{2}m\left(\dot{x}^2+\dot{y}^2\right) = 2mR^2\omega_0^2. \tag{254}$$

Actually, in a classical electrodynamical calculation we should include the radiation reaction (similar to the second term of the right side in the oscillator eq.(189)). This term would give rise to a loss of energy by radiation whence the system will eventually arrive at the

state of minimal energy, which is zero. This shows that no diamagnetic effects can be expected to occur in classical physics.

The QM treatment starts from a quantum Hamiltonian operator which may be got from eq.(253) by promoting the classical coordinates and momenta to operators in a Hilbert space (for a detailed study see [**18**]). The Z component of the angular momentum operator and the Hamiltonian commute and we may search for simultaneous eigenvectors having eigenvalues

$$\begin{aligned} L_z &\equiv xp_y - yp_x \to m_l\hbar, \qquad m_l = 0, \pm 1, \pm 2, \ldots, \\ H &\to E_r = (2k+1)\,\hbar\,|\omega_0|\,, \qquad k = 0, 1, 2, \ldots \end{aligned} \tag{255}$$

We see that the quantum ground state, given by $k = 0$ and $E_r = \hbar\,|\omega_0|$, has an infinite degeneracy because this energy is shared by states with all possible values of m_l. For this reason it is common to add to the Hamiltonian a two-dimensional oscillator potential with characteristic frequency $\omega_1 (> 0.)$ Then the energy eigenvalues have an additional term $(2n+1)\,\hbar\omega_1$ with $n = 2k - m_l \geq 0$, which breaks the degeneracy, the ground state now corresponding to $k = n = m_l = 0$. From eq.(255) we may get the most relevant parameter, which is the magnetic moment. In the ground state it is

$$\mathbf{M} = -\nabla_{\mathbf{B}} E = -\hbar\nabla_{\mathbf{B}}\,|\omega_0| = -M_B \frac{\mathbf{B}}{B}, \qquad M_B = \frac{\hbar\,|e|}{2mc} \tag{256}$$

where M_B is the Bohr magneton. This (or the ground state energy, second eq.(255)) is the result that we may expect to be reproduced in SED.

5.5.2. SED treatment. In SED we should add the action of the ZPF (plus the radiation reaction) to the force derived from the homogeneous magnetic field, see eq.(251). We will study only the motion in the XY plane. If $u \equiv \dot{x}$ and $v \equiv \dot{y}$ are the components of the velocity vector then the equations of motion are

$$m\dot{u} = \frac{e}{c} vB + m\tau\ddot{u} + eE_u, \qquad m\dot{v} = -\frac{e}{c} uB + m\tau\ddot{v} + eE_v, \tag{257}$$

where the first term is the component of the Lorentz force, the second is the radiation reaction and the third one the action of the ZPF (in the long wavelength approximation, see eq.(189)). The components of the electric ZPF, $E_u\,(t)$ and $E_v\,(t)$, are assumed statistically independent stochastic processes.

The small value of $\tau \ll 1/\omega_0$ allows an approximation similarly to the one made in the free particle case, section 5.3.4. We may

substitute $evB/(cm\tau)$ for $\ddot{u}$ and similar for $\ddot{v}$, thus obtaining two first order equations from eqs.(257) . Then the solution is straightforward and we get, with steps similar to those involved in the solution of eq.(189),

$$\left\langle u^2 \right\rangle = \int_0^\infty \frac{\hbar\tau\omega^3\left(4\omega_0^2+\omega^2\right)}{\pi m\left[\left(4\omega_0^2-\omega^2\right)^2+4\tau^2\omega^6\right]}d\omega \simeq \frac{\hbar\left|\omega_0\right|}{m}, \tag{258}$$

and the same result for $\left\langle v^2 \right\rangle$. Actually, the integral in eq.(258) is ultraviolet divergent so that a a high frequency cutoff, ω_c, should be included. It may be seen that for small τ, i.e. $\tau\omega_0 \ll 1$, the main contributions to the integral eq.(258) come either from frequencies ω close to $2\omega_0$ or for high frequencies $\omega \gg 2\omega_0$ (a similar case happens in the oscillator, see eqs.(197) and (199)). The first contribution, given by eq.(258), is independent of both the cut-off frequency and the precise value of τ. The last, high frequencies, contribution may be obtained neglecting ω_0 in comparison with ω, and putting $4\omega_0$ as lower limit of the integral in order to exclude the frequency region around $2\omega_0$ calculated in eq.(258) . Thus we get

$$\begin{aligned}\left\langle u^2 \right\rangle_{hf} &\simeq \int_{4\omega_0}^{\omega_c} \frac{\hbar\tau\omega^5 d\omega}{\pi m\left(\omega^4+4\tau^2\omega^6\right)} \\ &= \frac{\hbar}{8\pi m\tau}\log\left(1+4\tau^2\omega_c^2\right) \simeq \frac{\hbar\tau}{2\pi m}\omega_c^2,\end{aligned}$$

where we have assumed $2\tau\omega_c \ll 1$. A comparison with eq.(219) shows that this contribution is the same for a free particle. Indeed, it is independent of the magnetic field, which does appear in eq.(258).

The mean energy in SED is obtained putting eq.(258) in the expression of the energy (see eq.(254)), giving

$$\langle E\rangle = \frac{1}{2}m\left\langle u^2+v^2\right\rangle = \hbar\left|\omega_0\right|, \tag{259}$$

in agreement with the quantum result. Hence there is also agreement for the magnetic moment, eq.(256).

Another interesting result from SED is the mean value of the angular momentum, which is

$$\langle L_z\rangle = \langle xp_y - yp_x\rangle = m\langle xv-yu\rangle = -\frac{e}{|e|}\hbar, \tag{260}$$

independently of the magnitude of the magnetic field and the mass of the particle. I omit the proof, which is straightforward. Thus the angular momentum is parallel to the magnetic field if the charge

is negative and antiparallel if it is positive. We saw that the magnetic moment is always antiparallel to the magnetic field. The results eqs.(259) and (260) correspond to the limit $\tau \to 0$. In both cases there are corrections for finite τ which would requiere a relativistic treatment. If an appropriate cutoff is introduced, say $\omega_c = mc^2/\hbar$, the high frequencies contribution is small.

There is however a disagreement between QM and SED for the angular momentum in the stationary state. SED predicts a finite value given by eq.(260) but in QM there are many possible angular momenta in the ground state as shown in eq.(255) . On the other hand if we include an additional oscillator potential with characteristic frequency ω_1, then the quantum prediction for the ground state angular momentum is zero, a result that also disagrees with the SED result. Thus in QM there are two features whose realistic interpretation is difficult. Firstly, eq.(256) strongly suggests that the angular momentum in the ground state is the same of the SED prediction, eq.(260) rather than the degeneracy eq.(255) . Secondly, that an additional oscillator potential no matter how small breaks the degeneracy, but leading to zero angular momentum, rather than the most intuitive value eq.(260) . These facts show that a realistic interpretation of the angular momentum in some quantum mechanical states is difficult. A possible solution is proposed in section 6.4.

In summary, the SED treatment of the particle in a homogeneous magnetic field reproduces the most relevant results of QM and provides a realistic interpretation for the QM prediction of a diamagnetic behaviour of the charged particle in the presence of a homogeneous magnetic field.

5.6. Stochastic electrodynamics in confined space

The vacuum radiation field, which in this book we assume to be a stochastic real field, may be modified by the presence of macroscopic bodies. Usually the modification consists of changing the normal modes of the field, although maintaining the average energy per normal mode.

5.6.1. The Casimir effect. The most celebrated example for the change of the vacuum fields in confined space is the Casimir effect [**11**]. It consists of the attraction between two parallel perfectly conducting plates. As discussed in chapter 1 section 1.3.1 the reason for the attraction may be understood qualitatively as follows. The

electric field of the zero-point radiation, ZPF, should be zero or normal at any plate surface. Otherwise a current would be produced if the plates are perfectly conducting. This fact constrains the possible normal modes of the radiation, whence the total energy of the ZPF will be different from the energy inside the volume when the plates are removed.

The SED calculation is straightforward using classical electrodynamics but with the assumption that the ZPF is a real stochastic field. We should determine the possible radiation modes and assign a mean energy $\frac{1}{2}\hbar\omega$ to every mode. Then we subtract the energies found in both cases, with the plates in and out for frequencies up to ω_c. After that we take the limit $\omega_c \to \infty$. In practice. more sophisticated methods than just the subtraction may be used. In any case, the picture of the phenomenon is that the radiation pressures in both faces of each plate are different and this is the reason for the force on each plate. The quantum calculation is quite similar except that the mode energies are operators and the quantum vacuum expectation is substituted for the ZPF average. The details of the calculations appear in many places [**1**], [**4**] and will not be reproduced here.

I point out that similar calculations have been made with different geometries, like hollow conducting spheres [**1**] and for other vacuum fields different from the electromagnetic one, like the positron-electron field [**4**].

The Casimir effect is the most relevant support for the reality of the vacuum fields.

5.6.2. Cavity electrodynamics. SED may also offer intuitive pictures for cavity quantum electrodynamics, a well established experimental field of research [**1**]. An atom in a cavity gets its properties modified, in particular its lifetime. In fact, the atom does not decay if the modes having the frequency of the emitted radiation are not possible inside the cavity. In the quantum treatment the intriguing question is how does the atom 'know' in advance that it should not decay in these conditions. In SED the explanation is simple: spontaneous decay is actually stimulated by appropriate modes of the ZPF, and the modes required for the stimulation do not exist inside the cavity. For instance, in an early experiment by Haroche et al. [**12**] the excited atoms propagate between two metallic mirrors separated by 1.1 μm for about 13 natural lifetimes without appreciable decay. The

experiment involved a small applied magnetic field in order to demonstrate the anisotropy of spontaneous emission between mirrors. This experiment has been studied within SED via modelling the atom by a harmonic oscillator whence the empirical results have been reproduced quantitatively, but I will not review that work here [**13**], [**14**].

5.7. SED application to nonlinear systems

Several nonlinear systems have been studied in stochastic electrodynamics that provide some results in semiquantitative agreement with quantum mechanics, but badly fail in other cases. Actually, SED reproduces quantum results, and agrees with experiments, in a limited domain, namely for systems of charged particles that may be treated linearly and with a nonrelativistic approximation. In sharp contrast the treatment of nonlinear systems gives results that usually disagree with the quantum predictions. The explanation of this fact is that SED, as defined in the introduction section, is an approximation to QED to lowest order in Planck constant $\hbar$, but quantum mechanics gives predictions for nonlinear systems that involve $\hbar$ to higher order. Therefore, to be valid for all physical systems, SED should be generalized, likely including all vacuum fields and taking into account the back action of the particles on the fields. There is here a paradox; we might foresee that the final theory should be rather cumbersome due to the large number of fields involved and the nonlinearity of equations, while quantum theory has a relatively simple formalism. This is the magic of quantum theory and one of the reasons for the difficulty of getting a realistic interpretation.

In the following I comment on some calculations for nonlinear systems. The best method for the SED study in these cases is to get the evolution of the classical mechanical 'constants of the motion', one of them being the total energy. These parameters are no longer constant due to the interaction with the ZPF and the radiation reaction, but may be slowly varying. The method was used in section 5.3.5 for the oscillator and it will be illustrated in the following for a nonlinear system, the rigid planar rotor. After that I will comment on the hydrogen atom in SED and the problem of equilibrium between radiation and matter.

5.7.1. The planar rigid rotor. The planar rigid rotor is the most simple nonlinear system studied in SED [**19**]. A model of the rotor is a particle of mass m and charge e constrained to move in the

XY plane always at a distance R of a fixed point. Thus the problem has a single degree of freedom and the SED equation of motion may be written in terms of the polar angle ϕ as follows

$$\begin{aligned} mR\ddot{\phi} &= -m\tau R\dot{\phi}^3 + m\tau\dddot{\phi} + eE, \\ E &= -\cos\phi E_x(t) + \sin\phi E_y(t). \end{aligned}$$

The terms of the right side give the tangential component of the radiation reaction force, $m\tau\ddot{\mathbf{r}}$, and the tangential component of the force due to the ZPF, respectively. In terms of the angular velocity, $\omega = \dot{\phi}$, the equation becomes

$$\dot{\omega} = -\tau\omega^3 + \tau\ddot{\omega} + \frac{e}{mR}E. \tag{261}$$

Eq.(261) may be solved perturbatively in two steps. In the first step we solve the classical equation of motion $\dot{\omega} = 0$, which trivially gives $\omega = \omega_0 =$ constant. That constant becomes slowly varying when we take into account the radiation reaction and the action of the ZPF. In order to get that variation eq.(261) may be solved substituting ω_0 for ω in the perturbation, that is, in all terms of the right side. The solution with initial condition $\omega(0) = \omega_0$ becomes

$$\omega(t) = \omega_0 + \int_0^t dt' \Big[-\tau\omega_0^3 + \frac{e}{mR}\left(-\cos(\omega_0 t')E_x(t) + \sin(\omega_0 t')E_y(t')\right)\Big]$$

The first term within the integral sign represents drift and the last term diffusion. The diffusion constant may be calculated via the limit

$$\begin{aligned} D &= \lim_{t\to\infty}\frac{1}{t}\left\langle\left[\int_0^t dt'\frac{e}{mR}\left(-\cos(\omega_0 t')E_x(t') + \sin(\omega_0 t')E_y(t')\right)\right]^2\right\rangle \\ &= \left(\frac{e}{mR}\right)^2 \lim_{t\to\infty}\frac{1}{t}\int_0^t dt'\int_0^t dt''\cos\left[\omega_0(t'-t'')\right]\langle E_x(t')E_x(t'')\rangle, \end{aligned} \tag{262}$$

where I have taken into account that

$$\langle E_x(t')E_x(t'')\rangle = \langle E_y(t')E_y(t'')\rangle \quad \text{and} \quad \langle E_x(t')E_y(t'')\rangle = 0.$$

The field correlation may be easily obtained from the spectrum, eq. (192) as follows

$$\langle E_x(t')E_x(t'')\rangle = \frac{2\hbar}{3\pi c^3}\int_0^\infty u^3\cos\left[u(t'-t'')\right]du.$$

When this is inserted in eq.(262) the variables t' and t'' may be changed to $w \equiv (t'+t'')/2$ and $t'-t'' \equiv s$. With good approximation

the integration may be performed from 0 to t for the w integral and for the whole real line for the variable s. Then the limit $t \to \infty$ in eq.(262) is trivial and we get, taking the definition of τ, eq.(190), into account,

$$\begin{aligned} D &= \frac{\hbar\tau}{\pi m R^2} \int_0^\infty u^3 du \int_{-\infty}^\infty \cos(us)\cos(\omega_0 s)\, ds \\ &= \frac{\hbar\tau}{\pi m R^2} \int_0^\infty u^3 du \pi \left[\delta(\omega_0+u) + \delta(\omega_0-u)\right] = \frac{\hbar\tau}{mR^2}\omega_0^3. \end{aligned}$$

From the diffusion constant and the damping it is possible to obtain the following Fokker-Planck equation for the probability density, $\rho(\omega_0)$, of frequencies of the rotor

$$\frac{\partial\rho}{\partial t} = \frac{\partial}{\partial\omega_0}\left(\tau\omega_0^3\rho\right) + \frac{1}{2}\frac{\partial^2}{\partial\omega_0^2}\left(\frac{\hbar\tau}{mR^2}\omega_0^3\rho\right),$$

whose (regular) stationary solution is

$$\rho = \frac{2mR^2}{\hbar}\exp\left(-\frac{2mR^2\omega_0}{\hbar}\right).$$

A similar method may be used for the three-dimensional rotor [**19**] and the result is

$$\rho = \left(\frac{2mR^2}{\hbar}\right)^2 \omega_0 \exp\left(-\frac{2mR^2\omega_0}{\hbar}\right). \tag{263}$$

5.7.2. Comparison between SED and QM. The predictions of SED for the rigid rotor disagree with those of QM at least in four aspects, that will be illustrated in the following for the particular case of the three-dimensional rigid rotor:

1. *The distribution of positions or momenta in the minimal energy state.* In quantum mechanics the eigenstates of the angular momentum squared and the Hamiltonian of the rotor are, respectively,

$$\mathbf{L}^2 = \hbar^2 l(l+1), \qquad E_l = \frac{\hbar^2}{2I}l(l+1), \qquad l = 0, 1, 2, \ldots \tag{264}$$

so that the ground state corresponds to $\mathbf{L}^2 = E = 0$. In contrast, the stationary solution in SED is given by eq.(263) where there is a spherical distribution of angular momenta given by

$$W(L)LdL = \frac{4}{\hbar^2}\exp\left(-\frac{2L}{\hbar}\right)LdL. \tag{265}$$

2. *The set of states.* As in the oscillator studied in section 5.3.3 the set of possible states is quite different in QM and SED.

3. *The spectrum.* In QM the spectrum consists of the set of frequencies

$$\omega_{lj} = (\hbar/2I)\,[j(j+1) - l(l+1)] \quad \rightarrow \quad (\hbar/I)\,(l+1)\,, \tag{266}$$

the latter corresponding to the transitions allowed in the atomic dipole approximation. In sharp contrast, SED predicts a continuous spectrum, although the most intense absorption from the stationary state eq.(264) corresponds to the maximum absorption, which may be shown to be $\omega = \hbar/I$ [**1**], in agreement with the QM result for the transition from the ground to the first excited state, see eq.(266) . However, QM predicts a sharp frequency whilst the SED prediction corresponds to a wide band. In experiments the frequency is not sharp, but it is less wide than the SED prediction. The disagreement between the QM prediction and experiments is usually explained because the rigid rotor is not a realistic model of a molecule. For instance, molecules are not completely rigid. The disagreement with experiments is greater in SED and it cannot be explained as easily as in QM.

4. *The specific heat.* There is also a discrepancy as shown in the comparison between quantum and SED treatments [**19**]. This point will not be discussed here.

5.7.3. A difficulty with the angular momentum. The disagreement between the quantum prediction, eq.(264), and the SED prediction, eq.(265), for the rigid rotor is actually general and it poses a problem for any realistic model of rotation in quantum physics. For instance, if we try to get a picture of a rotating molecule. The quantum ground state of the rigid rotor possesses *zero angular momentum* and *spherical symmetry,* but these two properties are contradictory for any realistic interpretation. For the sake of clarity let us consider for instance the molecule of carbon oxide, CO, which may be modelled by a three-dimensional rigid rotor. It consists of an oxygen atom and a carbon atom at a distance which is very well known empirically. The ground state of this molecule possesses zero angular momentum and *therefore* (according to the quantum formalism) spherical symmetry. Discarding explanations which are bizarre for any realistic interpretation, like saying that 'the form of the molecule emerges during the act of measurement', the meaning of spherical symmetry is unclear. The only meaning compatible with a physical picture is that the molecule is rotating randomly in such a way that the probability distribution of the orientations of the axis in space possesses spherical symmetry. However, this is in conflict with the quantum prediction that the total

angular momentum is zero, *dispersion-free*. That is, the squared mean angular momentum is also zero. This situation is quite common, it appears in many molecules, atoms or nuclei. It seems that either the standard quantum prediction is wrong (e.g. the ground state is not physically achievable) or a *realistic* physical model is not possible.

A possible solution to the dilemma is that the quantum formalism actually provides the total angular momentum of the molecule *plus* the vacuum fields that interact with it. If the ground state corresponds to an equilibrium of the system (e.g. the molecule) with the vacuum fields it is plausible to assume that there is a continuous exchange of angular momentum between the system and the vacuum fields so that the total angular momentum (a conserved quantity) remains always zero. This is the case in the SED treatment of the planar rigid rotor of the previous section. In fact, eq.(261) may be interpreted as the equation for the balance of angular momentum. Indeed, the equation may be rewritten as the Z component of the angular momentum vector, that is,

$$\frac{d}{dt}(I\omega) = -\tau I\omega^3 + \tau I\ddot{\omega} + e\left(\mathbf{R}\times\mathbf{E}\right)_z ,$$

where the change of the rotor angular momentum equals the radiated momentum (the first two terms) minus the momentum absorbed from the ZPF. In summary, there is no real contradiction between the fact that SED predicts a distribution of angular momenta of *the rotor alone* and our interpretation of the QM prediction that the angular momentum of *rotor plus field* is stricly zero. I do not believe that SED is the correct reinterpretation of QM, but I think that it illustrates adequately the possible solution to the problem.

A similar solution may be given to the strange, if not paradoxical, quantum prediction that a charged particle in a homogeneous magnetic field has zero component of the angular momentum in the direction of the field, but the energy is precisely the product of the field times the magnetic moment. The SED results are more intuitive, namely the energy of the equilibrium state, eq.(259), agrees with the quantum ground energy, but there is a component of the angular momentum in the direction of the field, see eq.(260) .

5.7.4. The hydrogen atom. The hydrogen atom is the most relevant nonlinear system within elementary quantum mechanics, and therefore a crucial test for the validity of SED. Once the stationary state of the harmonic oscillator had been solved with success, several

authors devoted a big effort during the 1970's to study the hydrogen atom in SED. Several approximation methods were proposed for calculating the stationary state of the atom (modelled as two particles with opposite charge, one of them at rest). The most successful method devised for the study of a charged particle in a potential well rests upon the assumption that the classical constants of the motion change slowly. That is, the motion is close to the classical one, the action of the ZPF and the radiation reaction giving rise to a slow diffusion in the space of classical orbits. As every classical orbit is determined by the initial position and velocity, $\{\mathbf{r}_0, \mathbf{v}_0\}$, the final result of the calculation is a probability distribution in the phase space of positions and velocities, $\{\mathbf{r}, \mathbf{v}\}$, which is the same as the distribution of initial positions and velocities, $\{\mathbf{r}_0, \mathbf{v}_0\}$, if the state is stationary. This is similar to what happens in the planar rigid rotor studied above.

In the case of the hydrogen atom the result of the calculation did not provide a stationary solution. In fact, the prediction was that the atom is not stable but ionizes spontaneously due to the orbits passing close to the nucleus [**20**].

That work has been criticized because such orbits cannot be treated with a nonrelativistic approximation, and a relativistic treatment could produce an important change in the results. Actually, the prediction of spontaneous ionization made by SED analytical calculation is not a too strong argument against the SED prediction. In fact, the result depends crucially on the electron orbits passing close to the nucleus, which would requiere a relativistic treatment. Also, quantum theory predicts that the free atom is unstable against ionization at any finite temperature, no matter how small. This trivially follows from the fact that the quantum partition function is divergent, that is,

$$Z = \sum_{n=1}^{\infty} \sum_{l=0}^{n-1} (2l+1) \exp\left(-\frac{E_0}{n^2}\right) \to \infty.$$

Therefore, it is not too relevant if the approximation method used in SED has an effect (spontaneous ionization) similar to the effect of a thermal radiation in QM.

Furthermore, numerical solutions of the hydrogen atom in SED have been made [**21**] since 2003 that explain the stability of the atom. They led to a stationary distribution fairly close to the quantum prediction for the position distribution in the ground sate. However, more powerful calculations made in 2015 [**22**] predict a ionization of

the atom. Numerical calculations have the advantage that do not require approximations in the differential equations, like neglecting the dependence on position of the ZPF (the electric dipole approximation). However, the numerical methods have uncertainties that may explain the discrepancy, as discussed above for the early analytical treatment. See also [**23**].

5.7.5. Thermal equilibrium between radiation and matter. SED derivation of Planck's Law. Several authors have claimed that Planck's law may be derived from classical postulates, usually within the framework of SED [**2**], [**24**]. A derivation of the thermal radiation should follow from the study of the thermal equilibrium between radiation and matter. In the framework of standard quantum theory it leads to Planck's law, but here we are considering the question whether the law may be obtained from classical electrodynamics. The difficulty is related to the fact that the equilibrium radiation-matter should involve nonlinear systems. In particular, the study of equilibrium requires a balance between absorption of energy from the radiation at a frequency and emission at a different frequency. Only in these conditions it is possible to study the distribution of energy amongst the different frequencies, which is the essential purpose of a radiation law. If we deal only with linear (harmonic) oscillators then both the absorption and emission of radiation take place at the same frequency.

The problem of thermal equilibrium was extensively studied in the first decades of the 20th century and the conclusion was uncontroversial in my opinion: If one assumes classical dynamics then thermal equilibrium is achieved when the particles have the Maxwell-Boltzmann distribution and the radiation the Rayleigh-Jeans spectrum [**25**]. Thus there is a contradiction between the derivation reported by van Vleck and the derivations claiming that the classical equilibrium spectrum is given by Planck's law. It was suggested that early derivations involved Newtonian dynamics and that a study with relativistic dynamics might led to Planck's law. However, it has been shown that thermal equilibrium of relativistic particles also leads to the Rayleigh-Jeans law [**26**], [**27**].

Actually, the contradiction is only apparent. In fact, it may be shown that according to classical electrodynamics a system of particles may reach dynamical equilibrium with radiation if this has a ZPF plus Planck spectrum, that is, an energy per normal mode given

by eq.(250) . But this is so for any values of the parameters T (temperature) and $\hbar$ (Planck constant), in particular in the limit $\hbar \to 0$ when eq.(250) becomes the Rayleigh-Jeans law. The point is that the derivation of Planck's law (with ZPF) requires assuming both equilibrium between matter and radiation and that the limit $T \to 0$ gives the ZPF spectrum rather than zero. With these assumptions Planck spectrum has been derived [**24**], [**28**].

A related result is the classical derivation of the Davies-Unruh effect initially derived from quantum electrodynamics [**29**], [**30**]. It is interpreted in quantum theory as the production of photons with Planck distribution of frequencies when a detector moves in the vacuum with accelerated motion. The result may be got in SED with the interpretation that the spectrum of the ZPF appears as thermal when seen from an accelerated reference frame [**31**], [**4**].

5.8. Conclusions

We have seen that calculations of several linear systems within SED provide a remarkable agreement with the predictions of QM. On the other hand the realistic interpretation of SED is rather obvious. Thus the question arises, does SED offer the realistic interpretation of QM which we are searching for? Unfortunately the answer is in the negative; the difficulties of SED for the interpretation of phenomena associated to nonlinear systems seem insurmountable.

I propose that getting a realistic interpretation of QM would be possible accepting the general ideas of SED but rejecting many of the particular assumptions. The general ideas to be retained are the following: 1) Nuclei, atoms or molecules (but maybe not elementary particles like electrons) are bodies with well defined size and form following definite, but highly irregular, trajectories. (If the bodies are composite, like atoms, they may also suffer deformations). 2) Their motion is strongly influenced by the fluctuations of the vacuum fields.

In summary the most relevant clue provided by SED for a realistic interpretation of (nonrelativistic) quantum mechanics is the following:

The attempt at a realistic interpretation of the quantum mechanics of particles is not possible if the vacuum fields are not included.

Bibliography

[1] L. de la Peña, A. M. Cetto: *The quantum dice. An introduction to stochastic electrodynamics.* Kluwer Academic Publishers, Dordrecht, 1996.

[2] L. de la Peña, A. M. Cetto, A. Valdés Hernández: *The emerging quantum. The physics beyond quantum mechanics.* Springer, Heidelberg. 2015.
[3] T. H. Boyer: Stochastic electrodynamics: The closest classical approximation to quantum theory. *Atoms* **7**, 29 (2019).
[4] P. W. Milonni: *The Quantum Vacuum. An introduction to quantum electrodynamics.* Academic Press, San Diego. 1994.
[5] P. Braffort, C. Tzara: *C. R. Acad. Sci. Paris* **239**, 1775 (1954).
[6] T. W. Marshall: Random electrodynamics. *Proc. Roy. Soc.* **A276**, 475 (1963).
[7] E. Santos: The harmonic oscillator in stochastic electrodynamics. *Nuovo Cimento* **19B**, 57-89 (1974). There is a repeated misprint in the Appendix A of that paper, namely sinh (cosh) should read si (ci) everywhere.
[8] J. von Neumann: *Mathematishe Grundlagen der Quantenmechanik*. Springer-Verlag, Berlin, 1932. English translation Princeton University Press, 1955.
[9] J. S. Bell: On the problem of hidden variables in quantum mechanics. *Rev. Modern Phys.* **38**, 447-52 (1966).
[10] W. C. Huang, H. Batelaan: Discrete excitation spectrum of a classical harmonic oscillator in zero-point radiation. *Found Phys.* **45**, 333-353 (2015).
[11] H. B. G. Casimir: On the attraction between two perfectly conducting plates. *Proc. K. Ned. Acad. Wet.* **51**, 793 (1948).
[12] S. Haroche, J. M. Raimond: *Exploring the Quantum - Atoms, Cavities and Photons*. Oxford University Press, 2006.
[13] H. M. França, T. W. Marshall, E. Santos: Spontaneous emission in confined space according to stochastic electrodynamics. *Phys. Rev. A* **45,** 6436-6442 (1992).
[14] R. Blanco, H. M. França, E. Santos, R. C. Sponchiado: Radiative noise in circuits with inductance. *Phys. Lett. A* **282,** 349-356 (2001).
[15] A. Rueda: The vacuum electromagnetic field and the energetics of the intergalactic medium. *Phys. Rev.* **A30**, 2221 (1984).
[16] R. Blanco, H. M. França, E. Santos: Classical interpretation of the Debye law for the specific heat of solids. *Phys. Rev. A* **43,** 693-699 (1991).
[17] P. Debye: *On the Theory of Specific Heats*, Ann. Phys. (Leipzig) 39, 789 (1912)
[18] R. H. Dicke, J. P. Wittke: *Introduction to Quantum Mechanics*. Addison-Wesley, Reading, Mass., 1960.
[19] T. H. Boyer: Specific heat of classical, plane, rigid, dipole rotator in electromagnetic zero-point radiation. *Phys. Rev.* **A3**, 2035 (1971).
[20] T. W. Marshall, P. Claverie: Stochastic electrodynamics in non-linear systems. I. Particle in a central field of force. *J. Math. Phys.* **32**, 1819 (1980).
[21] D. C. Cole, Yi Zou: Quantum mechanical ground state of the hydrogen atom obtained from classical electrodynamics. *Phys. Lett.* **317**, 14-20 (2003).
[22] T. M. Nieuwenhuizen, M. T. P. Liska: Simulation of hydrogen ground state in stochastic electrodynamics. *Found. Phys.* **45**, 1190-1202 (2015).
[23] D. C.Cole: Subharmonic resonance and critical eccentricity for the classical hydrogen atomic system. *Eur. Phys. J. D* **72**, 200 (2018).
[24] T. H. Boyer: The blackbody radiation spectrum follows from zero-point radiation and the structure of relativistic spacetime in classical physics. *Found. Phys.* **42**, 595-614 (2012), and references therein.

[25] J. H. van Vleck and D. L. Huber: Absorption, emission, and linebreadths: a semihistorical perspective. *Rev. Mod. Phys.* **49**, 939 (1977)
[26] R. Blanco, L. Pesquera, E. Santos: Equilibrium between radiation and matter for classical relativistic multiperiodic systems. Derivation of Maxwell-Boltzmann distribution from Rayleigh-Jeans spectrum. *Phys. Rev. D* **27**, 1254-1287 (1983);
[27] R. Blanco, L. Pesquera, E. Santos: Equilibrium between radiation and matter for classical relativistic multiperiodic systems. Study of radiative equilibrium with Rayleigh-Jeans radiation. *Phys. Rev. D* **29**, 2240-2254 (1984).
[28] T. H. Boyer: Blackbody radiation in classical physics. A historical perspective. *Am. J. Phys.* **86**, 495-509 (2018).
[29] P. C. Davies: Scalar particle production in Schwarzshild and Rindler sapcetime. *J. Phys. A* **6**, 609-616 (1975).
[30] W. G. Unruh: Notes on blackbody evaporation. *Phys. Rev. D* **14**, 870-892 (1976).
[31] T. H. Boyer: Contrasting classical and quantum vacuum states in non-inertial frames. *Found. Phys.* **43**, 923-947 (2013), and references therein.

CHAPTER 6

Interpretation of the quantum radiation field

6.1. Introduction

This chapter is devoted to the realistic interpretation of the quantized electromagnetic field. We shall restrict attention to radiation with frequencies in the visible or near it, except for a short reference

to the Compton effect. Also, the study of electromagnetic interactions with microscopic (quantum) objects like atoms, molecules or nuclei, will be excluded. Therefore, the aim of the chapter is a realistic interpretation of some relevant phenomena in the domain of quantum optics.

The essential difference between the classical and the quantum treatment of the radiation field is the particle behaviour predicted in the latter, in addition to the common wave behaviour. Thus in section 6.2 I will present a qualitative picture of four typical examples of alleged particle behaviour: Localized radiation energy in atomic emissions, the Compton effect, anticorrelation after a beam splitter and discrete detection events (photocounts).

In section 6.3 I will present a short introduction to the Weyl-Wigner formalism, equivalent to the common Hilbert-space formalism but more suited for a realistic interpretation. Then the formalism will be applied to the interpretation of experiments involving entangled photon pairs produced via spontaneous parametric dawn-conversion (SPDC). In section 6.4 we will study the SPDC process, in section 6.5 realistic models of a few relevant experiments, and finally in section 6.6 we will discuss experimental tests of Bell inequalities and show that their violation may be interpreted via local realistic models.

6.2. The particle behaviour of light

6.2.1. Photons as localized objects. The wave-particle behaviour of quantum objects is possibly the main obstacle for an intuitive understanding of quantum theory. This is the case in particular for the electromagnetic radiation field. In fact, after Maxwell's work it seemed clear that radiation consists of electric and magnetic fields propagating in space, that is, waves. Einstein changed this simple picture in 1905 proposing that radiation consists of discrete amounts of energy, later named photons. Furthermore, in his 1916 study of absorption and (spontaneous and stimulated) emission of radiation by atoms, he concluded that photons have also a well defined momentum, the energy and momentum being related to the frequency ν and wavelength λ, respectively, that is,

$$E = h\nu, p = h/\lambda.$$

As a consequence of this discovery two facts troubled Einstein. Firstly, the direction of the emitted photon momentum is random, thus suggesting a breakdown of causality as we discussed in chapter 1, section

1.2.4. Secondly, photons should have a length much greater than typical atomic dimension in order for their wavelengths to be defined. In fact, typical photons of atomic emissions belong to the visible or near ultraviolet, with wavelengths of the order of microns, whilst atomic radii are of the order of nanometers, which is about 1000 times smaller. But the transverse size should not be much larger than atomic dimensions. Consequently, Einstein proposed that photons should be 'needles of radiation'. Both troubles of Einstein are solved with the hypothesis that radiation consists of just waves but the quantum vacuum fields are real stochastic fields. This will be explained in section 6.2.2 below.

The increasing empirical evidence for the particle nature of radiation was reinforced by the Compton experiments in 1924, and from that date it appears that we can renounce neither the wave nor the corpuscle nature of light. However, saying that photons are neither waves nor particles is a paradox rather than a solution. In fact, particle means energy concentrated in a small region of space and wave implies energy extended over a large region. The contradiction may be seen also as follows. In quantum theory it is frequently assumed that all (normalized) vectors of the Hilbert space are possible pure states of the system. This would imply that single-photon states are states of the radiation, in spite of the fact that in free space single-photon states correspond to plane waves extended over the whole space (more properly over a very large normalization volume). Nevertheless, it is also assumed the photons are localized particles. An obvious solution would be to believe that only some localized wave-packets could correspond to physical photon states, but then the frequency of the single photon is not well defined, as discussed above. Without solving the contradiction it is not possible to get a realistic picture of the quantized radiation. The standard 'solution' has been to renounce any picture of photons (and quantum phenomena in general) and to deal only with the (mathematical) formalism, this being the precept of the Copenhagen interpretation, as we mentioned in chapter 2, section 2.4.

The particle aspect of radiation is more manifest than the wave aspect in high energy physics because the photon energies are large but the wavelength is too small to be easily measured. In contrast, the wave aspect is more apparent in the macroscopic domain, the radiation energies being much larger than the energy of one photon with macroscopic wavelength, so that observable phenomena involve a very large number of photons. This dependence on the scale suggests

that the appropriate setting to study the dual behaviour or radiation is quantum optics, where the photon energies and wavelengths (say of visible light) are neither too large nor too short in comparison with atomic energies and sizes, respectively.

6.2.2. Needles of radiation in atomic emissions. In the stochastic approach to the electromagnetic field, supported in this book, the states of the electromagnetic radiation are probability distributions of a stochastic field. In particular, if we deal with radiation in free space (which I will assume except otherwise stated), the amplitudes of plane waves would be random variables. The interplay of the vacuum zero-point field (ZPF) with any additional radiation may give rise to fluctuations with a strong concentration of energy, which we may identify with 'photons'. An example is the 'spontaneous' emission of radiation by an atom. In our approach the emission is not spontaneous but induced by the ZPF. Then let us assume that in a fluctuation a strong plane wave of the ZPF of frequency ν arrives at an atom and it happens that ν is also one of the possible frequencies for emission from the excited atom. The arriving plane wave component of the ZPF may induce the emission of radiation with the same frequency in phase with the incoming wave. Thus the emitted radiation should correspond to the addition of the amplitudes of the incoming (plane) wave plus the emitted (spherical) wave. The frequencies being equal there would be interference and it is not difficult to show that it will be constructive in the forward direction and mainly destructive in all other directions.

More quantitatively, the outgoing energy will be concentrated in a region with a boundary defined by the following relation between the half angle, θ, as seen from the atom and the distance, d,

$$\frac{d}{\cos\theta} - d \sim \frac{\lambda}{2} \quad \Rightarrow \quad \theta \sim \sqrt{\frac{\lambda}{d}},$$

where λ is the wavelength. If we take d to be coherence length of the emitted 'photon', for typical atomic emissions we have $d \sim 1$m and $\lambda \sim 1\mu$, so that $\theta \sim 10^{-3}$. This fits with Einstein's proposal of 'needles of radiation' and, in addition, it explains the random direction of emission. In our interpretation the stochastic character of the ZPF is the cause of the randomness. This provides a picture of the photon as a concentration of radiation energy in the plane perpendicular to the motion and nevertheless with a frequency relatively well defined. Indeed, the frequency range may correspond to the inverse of the

atomic lifetime, roughly the coherence length divided by the velocity of light.

We may apply the photon model to the case of an atomic cascade where two photons are emitted within a short time interval. Then the picture that emerges is the existence of two 'needles of radiation' moving in different directions. In particular, if both the initial and the final state of the atom have zero spin and the photons are emitted in opposite directions, then the angular momenta of the two photons should be opposite by angular momentum conservation, whence they will be strongly correlated in polarization. The quantum formalism predicts that they will be maximally entangled, but I will not provide an interpretation of photon entanglement at this moment, see below section 6.5. The correlation will diminish if the photons are emitted at an angle smaller than 180^o, whence no Bell inequality may be violated in experiments using photon pairs from atomic cascades (see chapter 3 section 3.4.1 for a survey of those experiments).

6.2.3. A qualitative picture of the Compton effect. As another example I present a qualitative model for the Compton effect. As is well known, Compton's was the experiment that the scientific community accepted as the final proof of the particle character of radiation, that is, the existence of photons. The experiment is usually understood as a collision between one photon of X-ray of frequency ω_1 and an electron, giving rise to another photon of smaller frequency ω_2 at an angle θ with the incident one and a recoil electron. Indeed, the (relativistic) kinematics may be explained assuming that the incident and outgoing photons have energies $\hbar\omega_1$ and $\hbar\omega_2$, respectively, and the electron is initially at rest. Quantum electrodynamics gives a quantitative account of the phenomenon, including the cross section of the process, but it does not offer an intuitive picture except the proof that photons *behave* like particles. On the other hand there have been several attempts at a semiclassical explanation that I will not revisit here. A qualitative stochastic interpretation is as follows.

We may assume that the incoming X-ray beam is monochromatic of frequency ω_1 and intensity I_1. An electron placed in the beam will suffer a force in the direction of the Poynting vector of the radiation, which is the field resulting from the addition of the incoming X-ray beam plus the ZPF. The Pointing vector of the ZPF is zero on the average due to its mean rotational invariance, but it is fluctuating. At a given time a large field fluctuation may exist of frequency ω_1 and

phase ϕ in a direction at angle θ_1 with respect to the X-ray beam. The interference of the incident beam field and the fluctuation of ZPF will produce another field with maximum intensity in a direction $\theta_2 < \theta_1$. That intensity I_2 at the maximum may be greater than I_1, and accelerate the electron in the direction θ_2. The accelerated electron will emit radiation that would interfere with the incoming field in such a way that the total energy and momentum are conserved. The emitted radiation will act like a secondary X-ray radiation, which is the observed Compton effect.

The moral of these examples is that if we search for a stochastic interpretation of quantum radiation we should not try to interpret literally the standard quantum (Hilbert space) formalism but just the actual experiments or observations. In the following I will study representative experiments involving radiation, macroscopic bodies and photon counters. For the interpretation of these experiments the Hilbert space formalism is not appropriate. The alternative, but physically equivalent, Weyl-Wigner formalism reviewed below is better.

6.2.4. The action of beam-splitters. In the interaction with macroscopic bodies like lenses or polarizers, light may be usually treated as waves. However, for the interpretation of some experiments, relevant for our purpose of a realistic understanding of quantum theory, it is essential to devote attention to the behaviour of beam-splitters and photon counters, as is done in the following two subsections.

A simple beam-splitter (BS) may consist of just a slab of transparent material. If a light beam impinges on a point of the slab, a part of the beam intensity is transmitted and another part reflected. The relative intensities of the outgoing fields depend on the refractive index of the material and the angle of incidence. In this way we have an elementary beam-splitter with one incoming channel and two outgoing channels. Now we may have another incoming channel via a light beam arriving in the opposite side of the slab, which gives rise to two new outgoing channels. In practice the plate is used so that the transmitted light from one incoming channel is superposed to the reflected light of the other channel. In this way we would have two incoming channels and two outgoing ones. Real beam-splitters may be more sophisticated, e.g. involving piles of plates (used for instance in many of

the tests of Bell inequalities studied in chapter 3). Sometimes the BS polarizes light, thus acting as a polarizer or a polarization analyzer.

In the following I study in more detail *a balanced nonpolarizing BS*. If the field amplitudes of the incoming beams are E_1 and E_2, then the amplitudes in the outgoing channels will be

$$E_{out1} = \frac{1}{\sqrt{2}}\left(E_1 + iE_2\right), \qquad E_{out2} = \frac{1}{\sqrt{2}}\left(E_2 + iE_1\right). \tag{267}$$

The imaginary unit i is appropriate if we treat the electromagnetic fields in the complex representation, as we will do throughout this chapter. From eq.(267) it is obvious that the energy is conserved in the BS. In fact, the sum of intensities in the incoming channels equals the similar sum in the outgoing ones, that is,

$$I_{out1} + I_{out2} = |E_{out1}|^2 + |E_{out2}|^2 = |E_1|^2 + |E_2|^2 = I_{in1} + I_{in2}.$$

In most cases studied in this chapter the field arriving at one of the incoming channels will be a signal and the field at the other one will be a vacuum field. An application of the BS is the following experiment.

6.2.5. Anticorrelation-recombination experiment. A dramatic exhibition of the wave-particle behaviour of light is the anticorrelation-recombination experiment [**1**]. A weak radiation signal consisting of well separated photons is sent to one of the incoming channels of a balanced beam splitter $BS1$, and two photon detectors, A and B, are placed in front of the outgoing channels. No coincidences are observed, which shows the corpuscular behaviour of light: a photon is not divided, but goes to one of the detectors. If the detectors are removed and the two outgoing radiation beams are recombined via the two incoming channels of another beam splitter $BS2$, then the detection in one of the outgoing channels depends on the length difference between the two paths from $BS1$ to $BS2$, this being a typical wave behaviour.

Our stochastic interpretation is as follows [**2**]. If we assumed that the vacuum quantum fields are not real fields then only the signal field E entering $BS1$ should produce outgoing fields in each of the two outgoing channels. However, if the vacuum fields are real, there is another (vacuum) field E_0 entering in $BS1$ via the second incoming channel, whence the outgoing fields will be

$$E_A = \frac{E + iE_0}{\sqrt{2}}, \qquad E_B = \frac{iE + E_0}{\sqrt{2}}. \tag{268}$$

Depending on the relative phases, one of the intensities may be large and the other one small, that is,

$$I_A = |E_A|^2 = \frac{1}{2}(|E|^2 + |E_0|^2) + |E|\,|E_0|\cos\phi, \tag{269}$$
$$I_B = |E_B|^2 = \frac{1}{2}(|E|^2 + |E_0|^2) - |E|\,|E_0|\cos\phi,$$

where ϕ is the relative phase of the fields E and E_0. On the other hand the vacuum intensities would ideally be $I_{A0} = I_{B0} = I_0 = |E_0|^2$.

If we assume that detection is roughly proportional to the part of the arriving intensity that surpasses the ZPF level, then with $I = |E|^2$ and $I_0 = |E_0|^2$ we obtain

$$R_A = R_B = \frac{1}{2}(\langle I\rangle - \langle I_0\rangle), \qquad R_{AB} = \frac{1}{4}(\langle I^2\rangle - \langle I_0\rangle^2) - \frac{1}{2}\langle I\rangle\langle I_0\rangle.$$

This result shows that for weak signals, that is, when I is not much greater than I_0, the coincidence detection rate is inhibited, that is, $R_{AB} \ll R_A R_B$, as observed in the experiment described. (We define the rate as a dimensionless probability of detection per time window). In contrast for macroscopic (classical) light $I \gg I_0$ and the ratio would be

$$r \equiv \frac{R_{AB}}{R_A R_B} \simeq \frac{\langle I^2\rangle}{\langle I\rangle^2}.$$

Hence if the radiation has fixed (nonfluctuating) intensity, like laser light, then

$$\langle I^2\rangle = \langle I\rangle^2 \quad \Rightarrow \quad r = 1$$

meaning that the detections are uncorrelated. On the other hand for chaotic light, where the field fluctuations are Gaussian, we would have

$$\langle I^2\rangle = 2\langle I\rangle^2 \quad \Rightarrow \quad r = 2,$$

meaning that the detections by Alice and Bob are positively correlated. The change from $r = 1$ to $r = 2$, a phenomenon known as 'photon bunching', has been interpreted as a quantum effect attributed to the Bose character of photons. In our stochastic interpretation it is the consequence of the correlated fluctuations derived from the Gaussian character of chaotic light.

In the recombination process the fields of eq.(268) will enter $BS2$ giving rise to the following intensity in one of the outgoing channels

$$I_{rec} = \left|\frac{|E_A|}{\sqrt{2}} + \exp(i\theta)\,\frac{|E_B|}{\sqrt{2}}\right|^2 = \frac{1}{2}(I + I_0) + \frac{1}{2}(I - I_0)\cos\theta, \tag{270}$$

where θ is the relative phase due to the different path lengths. The device used in the experiment [**1**], consisting of two beam splitter and two mirrors in between, is called Mach-Zehnder interferometer. The detection rate will be

$$R_A = \langle I_{rec} \rangle - \langle I_0 \rangle = \frac{1}{2} \langle I - I_0 \rangle (1 + \cos\theta) ,$$

meaning that a 100% visibility may be achieved. Thus we have a wave explanation of one of the most dramatic particle behaviour of light, the anticorrelation after a beam splitter. The anticorrelation is usually named 'photon antibunching' and it is considered a typically quantum phenomenon, that cannot be explained by classical theories. Of course, it can be explained if we do assume that the vacuum fields are real stochastic fields. The evolution of these fields is classical (Maxwellian) but the assumption of real vacuum fields is alien to classical physics. I stress that the Planck constant appears as a scale of the vacuum fields, see eq.(271) .

6.2.6. Photocounters. A fundamental hypothesis in this book is that the electromagnetic vacuum field is a real stochastic field (the zero-point field, ZPF). If expanded in normal modes the ZPF has a (positive) probability distribution of the amplitudes given by

$$W_0 = \prod_j \frac{2}{\pi} \exp\left(-2 |a_j|^2\right) . \tag{271}$$

(See chapter 5. For the connection of this distribution with the Weyl-Wigner formalism see section 6.3.2, eq.(277) below). According to that assumption any photodetector in free space would be immersed in an extremely strong radiation, infinite if no cut-off existed. Thus, how might we explain that detectors are not activated by the vacuum radiation? Firstly, the strong vacuum field is effectively reduced to a weaker level if *we assume that only radiation within some (small) frequency interval is able to activate a photodetector*, that is, the interval of sensitivity (ω_1, ω_2). However, the problem is not yet solved because signals involved in experiments may have intensities of the order of the vacuum radiation in the said frequency interval, whence the detector would be unable to distinguish a signal from the ZPF noise. Our proposal is to assume that *a detector may be activated only when the Poynting vector (i.e. the directional energy flux) of the incoming radiation is different from zero, including both signal and*

vacuum fields. This hypothesis leads us to model the *detector as possessing an active area, the probability of a photocount depending on the integrated energy flux crossing that area from the front side during some specific activation time, T, and the probability being zero if the net flux crosing the area is from rear.*

This assumption allows understanding why the signals, but not the vacuum fields, activate detectors. Indeed, the ZPF arriving at any point (in particular the detector) would be isotropic on the average, whence its associated Poynting vector would vanish on the average, therefore only the signal, which is directional, should produce photocounts. A problem remains because the vacuum fields are fluctuating whence the Poynting vector also fluctuates. Hence we may predict the existence of some dark rate even at zero Kelvin. The problems derived from the fluctuations diminish assuming that *photocounts are not produced by an instantaneous interaction of the fields with the detectors; the activation requires some time interval*, a known fact in experiments. Therefore, the effective energy flux is a time average whose fluctuations are plausibly small.

These arguments are qualitative and a quantitative estimate is worthwhile. However, making such an estimate is rather involved and I will just sketch the steps required. We should start calculating the expectation value of the Poynting vector $\mathbf{P}$ due to the ZPF at a point $\mathbf{r}$ in space at time t. If we expand the ZPF in plane waves as usual, the expectation of the Z component of $\mathbf{P}$ may be written as a sum involving all radiation modes, that is,

$$\langle P_z \rangle = \left[\sum_j \sum_k Prob\left(a_j, a_k\right) \mathbf{E}_j\left(\mathbf{r},t\right) \times \mathbf{B}_k\left(\mathbf{r},t\right) \right] \cdot \mathbf{u}_z, \tag{272}$$

where a_j (a_k) is the amplitude of the mode $j\,(k)$, $\mathbf{E}_j$ $(\mathbf{B}_k)$ its associated electric (magnetic) field and $\mathbf{u}_z$ is a unit vector in the Z direction. The probability density $Prob\left(a_j, a_k\right)$ is related to eq.(271) . We assume that only field components with frequencies in the sensitivity interval (ω_1, ω_2) are effective for detection, whence we should restrict the double sum to modes with frequencies in that interval. Hence we might obtain the probability distribution of the integrated energy flux, Φ, crossing the active area of the detector during the activation time T. We may assume that a detection event takes place whenever Φ surpasses a threshold $\Phi_{th} > 0$ that may be chosen within some limits by the experimental physicist. Low Φ_{th} may increase the detection

efficiency but also increase the number of spurious counts and the opposite happens for high Φ_{th}. In fact, if the ZPF is isotropic the flux crossing the active area of the detector might be positive or negative with equal probabilities but the ratio Φ/T would be smaller for larger T, zero for $T \to \infty$. In actual detectors the choice may be made via adjusting the voltage bias. However, for any finite T there will be a finite probability that $\Phi > \Phi_{th}$, thus producing photocounts even in the absence of signals. These spurious counts give rise to a dark rate r usually attributed to thermal fluctuations. Indeed, the experimental dark rate is known to diminish with decreasing temperature, but we propose that it would remain finite at zero Kelvin. Now we study the case when there is a signal, superposed to the ZPF, arriving at the detector. The signal may be weak with respect to a typical short-time fluctuation of the ZPF, but it is *positive* at all times because signals arrive at the detector in the correct (front) direction. Thus a positive quantity due to the signal should be added to the fluctuating energy flux derived from the ZPF, calculated via eq.(272) . In a particular experiment we should choose the sensitivity interval T and the threshold Φ_{th} such that we have high detection efficiency η and small dark rate r, but there are obvious constraints. For instance, in order to have a small r for given T we need high Φ_{th}, but in this case some signals will become undetected leading to a decrease of η. This is a typical behaviour of tests having yes-no answer, where a high sensitivity gives rise to false positive results and lowering the sensitivity would increase false negative answers. I propose that this trade-off is the physical reason for the difficulties in manufacturing very efficient photon counters.

In the following sections of this chapter we will use repeatedly a detection rule stating that the detection probability, during a time window T, when a field with intensity $I(t)$ arrives at the photocounter, is proportional to the integrated intensity during the time T; that is,

$$P \propto \Phi = \int_0^T I(t)dt. \tag{273}$$

This rule is convenient for calculations, but it is less plausible than the assumption that a detection event is produced when Φ surpasses some threshold Φ_{th} , as proposed above. Furthermore, in some calculations we will have cases where $I(t) < 0$, eq.(273) becoming meaningless. A negative value of I corresponds to field intensity arriving at the detector from rear. I will assume that the threshold rule is more

plausible while eq.(273) is an approximation whose validity should be studied.

With this threshold rule the probability of detection during an activation time (that we may identify with the time window used in the experiments to be discussed in sections 6.5 and 6.6 below) is given by the probability that $\Phi \geq \Phi_{th}$, which we calculate as follows. We shall deal with Gaussian fields, whence the probability distribution of the energy flux Φ would be exponential, that is,

$$\rho(\Phi) = \langle\Phi\rangle^{-1} \exp\left[-\frac{\Phi}{\langle\Phi\rangle}\right], \qquad \Phi \geq 0, \tag{274}$$

where $\langle\Phi\rangle$ is the mean value of Φ. Hence the probability of surpassing the threshold, thus producing a photocount, will be

$$P_{th} = \int_{\Phi_{th}}^{\infty} \rho(\Phi)\, d\Phi = \exp\left[-\frac{\Phi_{th}}{\langle\Phi\rangle}\right], \qquad P_{mean} = K\langle\Phi\rangle, \tag{275}$$

where for comparison we have written also the detection probability P_{mean} predicted from the alternative detection rule, K being a constant. With the P_{th} rule the detection probability is negligible for small $\langle\Phi\rangle$, that is, for weak signals, while for $\langle\Phi\rangle$ large it becomes independent on $\langle\Phi\rangle$. This behaviour is much more plausible than the one derived from the rule P_{mean} which gives the absurd prediction $P > 1$ for big enough signals.

6.3. The Weyl-Wigner formalism in quantum optics

6.3.1. Definition. The WW formalism was developed for non-relativistic quantum mechanics, where the basic observables are positions, $\hat{x}_j$, and momenta, $\hat{p}_j$, of the particles, as reviewed in chapter 4. It may be trivially extended to quantum optics provided we interpret $\hat{x}_j$ and $\hat{p}_j$ to be the sum and the difference of the creation, $\hat{a}_j^\dagger$, and annihilation, $\hat{a}_j$, operators of the j-th normal mode of the radiation. That is,

$$\hat{x}_j \equiv \frac{c}{\sqrt{2}\omega_j}\left(\hat{a}_j + \hat{a}_j^\dagger\right), \qquad \hat{p}_j \equiv \frac{i\hbar\omega_j}{\sqrt{2}c}\left(\hat{a}_j - \hat{a}_j^\dagger\right) \tag{276}$$

$$\Rightarrow \quad \hat{a}_j = \frac{1}{\sqrt{2}}\left(\frac{\omega_j}{c}\hat{x}_j + \frac{ic}{\hbar\omega_j}\hat{p}_j\right), \qquad \hat{a}_j^\dagger = \frac{1}{\sqrt{2}}\left(\frac{\omega_j}{c}\hat{x}_j - \frac{ic}{\hbar\omega_j}\hat{p}_j\right).$$

Here $\hbar$ is the Planck constant, c the velocity of light and ω_j the frequency of the normal mode. In the following I will use units $\hbar = c = 1$. For the sake of clarity I shall represent the operators in a Hilbert space

with a 'hat', e.g. $\hat{a}_j, \hat{a}_j^\dagger$, and the amplitudes in the WW formalism without 'hat', e.g. a_j, a_j^*.

A complex field representation will be used throughout. Thus we will speak about the positive frequency part, E^+, and its complex conjugate field, the negative frequency part $E^- = (E^+)^*$ such that the field intensity will be $I = E^+E^- = E^-E^+$. Note that both fields commute (they are scalars), unlike their Hilbert-space (HS in the following) counterparts, $\widehat{E}^+$ and $\widehat{E}^-$, whose commutator is not zero.

The connection with the HS formalism is made via the Weyl transform as follows, see chapter 4 section 4.4.2. For any trace class operator $\widehat{M}$ of HS we define its Weyl transform to be an integral transform involving the field operators $\left\{\hat{a}_j, \hat{a}_j^\dagger\right\}$, that is,

$$\begin{aligned} W_{\widehat{M}} &= \frac{1}{(2\pi^2)^n}\prod_{j=1}^{n}\int_{-\infty}^{\infty}d\lambda_j\int_{-\infty}^{\infty}d\mu_j \exp\left[-2i\lambda_j \mathrm{Re}a_j - 2i\mu_j \mathrm{Im}a_j\right] \\ &\quad \times \mathrm{Tr}\left\{\widehat{M}\exp\left[i\lambda_j\left(\hat{a}_j + \hat{a}_j^\dagger\right) + \mu_j\left(\hat{a}_j - \hat{a}_j^\dagger\right)\right]\right\}. \end{aligned}$$

The transform is *invertible* as follows:

$$\widehat{M} = \frac{1}{(2\pi^2)^{2n}}\prod_{j=1}^{n}\int_{-\infty}^{\infty}d\lambda_j\int_{-\infty}^{\infty}d\mu_j \exp\left[i\lambda_j\left(\hat{a}_j + \hat{a}_j^\dagger\right) + \mu_j\left(\hat{a}_j - \hat{a}_j^\dagger\right)\right]$$

$$\times \prod_{j=1}^{n}\int_{-\infty}^{\infty}d\mathrm{Re}a_j\int_{-\infty}^{\infty}d\mathrm{Im}a_j W_{\widehat{M}}\left\{a_j, a_j^*\right\}\exp[-2i\lambda_j \mathrm{Re}a_j - 2i\mu_j \mathrm{Im}a_j].$$

The transform is *linear*, that is, if f is the transform of $\hat{f}$ and g the transform of $\hat{g}$, then the transform of $\hat{f} + \hat{g}$ is $f + g$.

It is standard wisdom that the WW formalism is unable to provide any intuitive picture of the quantum phenomena. The reason is that the Wigner function, which may represent a quantum state, is not positive definite in general and therefore cannot be interpreted as a probability distribution (of positions and momenta in quantum mechanics, or field amplitudes in quantum optics). As discussed in chapter 4 this is true in nonrelativistic quantum mechanics. However, we shall see that in quantum optics the formalism in the Heisenberg picture, where the evolution goes in the field amplitudes, allows the interpretation of the experiments using the Wigner function for the vacuum state alone, which is positive definite.

The use of the WW formalism in quantum optics has the following features in comparison with the HS formalism:

1. It is just quantum optics, therefore the predictions for experiments are the same.

2. The calculations using the WW formalism are sometimes easier than those in HS because no problem of noncommutativity arises in the former. However, the commutation rules of the standard formalism frequently produce substantial simplifications, as will be seen in the following.

3. The formalism suggests a physical picture in terms of random variables and stochastic processes. In particular, the counterparts of creation and annihilation operators look like random amplitudes. In this sense the WW formalism is well adapted for a realistic interpretation.

As said above, here we shall use the formalism in the Heisenberg picture, where the evolution appears in the observables. On the other hand the concept of photon, as a particle, does not appear in the WW formalism. However, localized concentrations of energy and momentum may appear in the form of radiation needles, as mentioned in the previous section 6.2.2.

6.3.2. Properties. All properties of the WW transform in particle systems may be translated to quantum optics via eqs.(276) . The transform allows getting a function of (scalar) amplitudes for any trace-class operator (e.g. any function of the creation and annihilation operators of 'photons'). In particular, we may get the (Wigner) function corresponding to any quantum state. For instance, the vacuum state, represented by the density matrix $|0\rangle\langle 0|$, is associated with the following Wigner function (used already in the previous section, see eq.(271))

$$W_0 = \prod_j \frac{2}{\pi} \exp\left(-2\left|a_j\right|^2\right). \tag{277}$$

This function might be interpreted as a (positive) probability distribution. Hence the picture that emerges is that the quantum vacuum of the electromagnetic field (the *zero-point field, ZPF*) consists of stochastic fields that may be defined in terms of radiation modes. The ZPF probability distribution is independent for every mode and it has a Gaussian distribution with mean energy $\frac{1}{2}\hbar\omega$ per mode.

Similarly, in the WW formalism there are functions associated with the observables. For instance, the following Weyl transforms are obtained

$$
\begin{aligned}
(278)\quad & \hat{a}_j \leftrightarrow a_j, \qquad \hat{a}_j^\dagger \leftrightarrow a_j^*, \qquad \frac{1}{2}\left(\hat{a}_j^\dagger \hat{a}_j + \hat{a}_j \hat{a}_j^\dagger\right) \leftrightarrow a_j a_j^* = |a_j|^2 , \\
& \hat{a}_j^\dagger \hat{a}_j = \frac{1}{2}\left(\hat{a}_j^\dagger \hat{a}_j + \hat{a}_j \hat{a}_j^\dagger\right) + \frac{1}{2}\left(\hat{a}_j^\dagger \hat{a}_j - \hat{a}_j \hat{a}_j^\dagger\right) \leftrightarrow |a_j|^2 - \frac{1}{2}, \\
& \left(\hat{a}_j^\dagger + \hat{a}_j\right)^n \leftrightarrow \left(a_j + a_j^*\right)^n , \\
& \left(\hat{a}_j^\dagger - \hat{a}_j\right)^n \leftrightarrow \left(a_j - a_j^*\right)^n , \quad n \text{ an integer.}
\end{aligned}
$$

I stress that the quantities a_j and a_j^* are numbers and therefore they commute with each other. The first eqs.(278) mean that in expressions that are *linear in creation and/or annihilation operator* the Weyl transform just implies '*removing the hats*'. However, this is not the case in nonlinear expressions in general. In fact, from the last two eqs.(278) plus the linearity property it follows that for a product in the WW formalism the HS counterpart is

$$
(279)\qquad a_j^k a_j^{*^l} \leftrightarrow (\hat{a}_j^k \hat{a}_j^{\dagger l})_{sym},
$$

where the subindex *sym* means writing the product with all possible orderings and dividing by the number of terms. Hence the WW field amplitudes corresponding to products of field operators may be easily obtained putting first the operators in symmetrical order via the commutation relations. Particular instances that will be useful later are the following

$$
\begin{aligned}
(280)\qquad & \hat{a}_j^\dagger \hat{a}_j \to |a_j|^2 - \frac{1}{2}, \qquad \hat{a}_j \hat{a}_j^\dagger \to |a_j|^2 + \frac{1}{2}, \\
& {\hat{a}_j}^2 \to a_j^2, \qquad \hat{a}_j^{\dagger 2} \to \hat{a}_j^{*2} \\
& \hat{a}_j^\dagger \hat{a}_j \hat{a}_j^\dagger \hat{a}_j \to |a_j|^4 - |a_j|^2 , \qquad \hat{a}_j \hat{a}_j^\dagger \hat{a}_j \hat{a}_j^\dagger \to |a_j|^4 + |a_j|^2 , \\
& \hat{a}_j^\dagger \hat{a}_j^\dagger \hat{a}_j \hat{a}_j \to |a_j|^4 - 2\,|a_j|^2 + \frac{1}{2}, \qquad \hat{a}_j \hat{a}_j \hat{a}_j^\dagger \hat{a}_j^\dagger \to |a_j|^4 + 2\,|a_j|^2 + \frac{1}{2}.
\end{aligned}
$$

Other properties may be obtained from well known results of the standard Weyl-Wigner formalism in particle quantum mechanics. I will present them omitting the proofs.

Expectation values may be calculated in the WW formalism as follows. In the HS formalism they read $\mathrm{Tr}(\hat{\rho}\widehat{M})$, or in particular $\langle \psi \mid \widehat{M} \mid \psi \rangle$, whence the translation to the WW formalism is obtained

taking into account that the trace of the product of two operators becomes

$$\mathrm{Tr}(\hat{\rho}\hat{M}) = \int W_{\hat{\rho}}\left\{\hat{a}_j, \hat{a}_j^\dagger\right\} W_{\widehat{M}}\left\{\hat{a}_j, \hat{a}_j^\dagger\right\} \prod_j d\mathrm{Re}a_j d\mathrm{Im}a_j.$$

That integral is the WW counterpart of the trace operation in the HS formalism. Particular instances are the following expectations that will be of interest later

$$\begin{aligned}
\left\langle |a_j|^2 \right\rangle &\equiv \int d\Gamma W_0 \, |a_j|^2 = \frac{1}{2}, \qquad \left\langle a_j^n a_j^{*m} \right\rangle = 0 \text{ if } n \neq m. \\
\left\langle 0 \left| \hat{a}_j^\dagger \hat{a}_j \right| 0 \right\rangle &= \int d\Gamma (a_j^* a_j - \frac{1}{2}) W_0 = 0, \\
\left\langle 0 \left| \hat{a}_j \hat{a}_j^\dagger \right| 0 \right\rangle &= \int d\Gamma (|a_j|^2 + \frac{1}{2}) W_0 = 2\left\langle |a_j|^2 \right\rangle = 1, \\
\left\langle |a_j|^4 \right\rangle &= 1/2, \quad \langle |a_j|^n \, |a_k|^m \rangle = \langle |a_j|^n \rangle \langle |a_k|^m \rangle \text{ if } j \neq k.
\end{aligned} \tag{281}$$

where W_0 is the Wigner function of the vacuum, eq.(277). This means that in the WW formalism the field amplitude a_j (coming from the vacuum) behaves like a complex random variable with Gaussian distribution and mean squared modulus $\left\langle |a_j|^2 \right\rangle = 1/2$. I point out that the integral for any mode not entering in the function $M\left(\left\{a_j, a_j^*\right\}\right)$ gives unity in the integration due to the normalization of the Wigner function eq.(277). An important consequence of eq.(281) is that *normal (antinormal) ordering of creation and annihilation operators in the Hilbert space formalism becomes subtraction (addition) of 1/2 in the WW formalism. The normal ordering rule is equivalent to subtracting the vacuum contribution.*

6.3.3. Evolution. In the Heisenberg picture of the HS formalism the density matrix is fixed and any observable, say $\widehat{M}$, evolves according to

$$\frac{d}{dt}\widehat{M} = i\left(\widehat{H}\widehat{M} - \widehat{M}\widehat{H}\right), \qquad \widehat{M} = \widehat{M}(t),$$

where $\widehat{H}$ is the Hamiltonian. Translated to the WW formalism this evolution of the observables is given by the Moyal equation with the sign changed. The standard Moyal equation applies to the evolution of the Wigner function, which represents a quantum state and it is the counterpart of the density matrix in the Schrödinger picture of

the HS formalism. Thus in the WW formalism we have

$$\begin{aligned}\frac{\partial W_{\widehat{M}}}{\partial t} &= 2\Bigg\{ \sin\left[\frac{1}{4}\left(\frac{\partial}{\partial \mathrm{Re}a_j'}\frac{\partial}{\partial \mathrm{Im}a_j''} - \frac{\partial}{\partial \mathrm{Im}a_j'}\frac{\partial}{\partial \mathrm{Re}a_j''}\right)\right] \\ &\times \left[W_{\widehat{M}}\left(a_j', a_j^{*\prime}, t\right) H\left(a_j'', a_j^{*\prime\prime}\right)\right] \Bigg\}_{a_j}, \end{aligned} \tag{282}$$

where $\{\}_{a_j}$ means making $a_j' = a_j'' = a_j$ and $a_j^{*\prime} = a_j^{*\prime\prime} = a_j^*$ after performing the derivatives.

A simple example is the free evolution of the field amplitude of a single mode. The Hamiltonian in the WW formalism is easy to obtain translating the Hamiltonian of the HS formalism, that is,

$$\begin{aligned}\widehat{H}_{free} &= \omega_j \hat{a}_j^\dagger \hat{a}_j = \frac{1}{2}\omega_j(\hat{a}_j^\dagger \hat{a}_j + \hat{a}_j \hat{a}_j^\dagger) - \frac{1}{2}\omega_j \\ \rightarrow H_{free} &= \omega_j(|a_j|^2 - \frac{1}{2}) = \omega_j\left[(\mathrm{Re}a_j)^2 + (\mathrm{Im}a_j)^2 - \frac{1}{2}\right],\end{aligned}$$

where we have taken the first eq.(281) into account. This leads to

$$\begin{aligned}\frac{d}{dt}a_j &= \frac{1}{2}\omega_j\left[2(\mathrm{Im}a_j) - 2\left(\mathrm{Re}a_j\right)i\right] = -i\omega_j a_j \\ &\Rightarrow a_j\left(t\right) = a_j\left(0\right)\exp\left(-i\omega_j t\right)\end{aligned} \tag{283}$$

Another example is the down-conversion process in a nonlinear crystal, to be studied in detail in the next section.

I emphasize again that the WW formalism provides an alternative formulation of quantum optics, physically equivalent to the more common HS. But it suggests a picture of the optical phenomena quite different from the latter, where photons are the fundamental concept. Indeed, the WW picture may provide a realistic interpretation in terms of random variables and stochastic processes, as presented in the following.

The mentioned properties of the WW formalism are sufficient for the interpretation of experiments involving pure radiation fields interacting with macroscopic bodies, defined by their bulk electric properties like the refractive index or the nonlinear electrical susceptibility. Within the WW formalism the interaction between the fields (either signals or vacuum fields) and macroscopic bodies may be treated as in classical electrodynamics. This is for instance the case for the action of a laser on the crystal with nonlinear susceptibility, to be discussed in the following [**14**], [**15**].

6.4. Spontaneous parametric down conversion (SPDC)

6.4.1. Production of correlated signal and idler fields. SPDC has been the main source of 'entangled photon pairs' during the last forty years. In the following I provide a theoretical interpretation of SPDC experiments within the WW formalism in the Heisenberg picture, that was initiated in the nineties of the past century [**3**]–[**13**]. In many of those early studies the approach was heuristic and here I shall provide a more formal foundation. The WW formalism suggests an intuitive picture for photon entanglement and the interpretation of SPDC experiments in terms of random variables and stochastic processes without any reference to photons. Nevertheless, I will use sometimes the language common in the HS formalism.

Entangled photons are produced when a pumping laser enters a crystal possessing a nonlinear electric susceptibility. Avoiding a detailed study of the physics inside the crystal [**14**], [**15**] we might describe the phenomenon with a model Hamiltonian [**16**]. I shall start in the standard HS formalism and we will pass to the WW formalism later on via the Weyl transform. The Hamiltonian is

$$\widehat{H}_I = A\hat{a}_s^\dagger \hat{a}_i^\dagger \exp\left(-i\omega_P t\right) + A^* \hat{a}_s \hat{a}_i \exp\left(i\omega_P t\right), \tag{284}$$

when the laser is treated as a classically prescribed, undepleted and spatially uniform field of frequency ω_P. The parameter A is proportional to the pump amplitude and the nonlinear susceptibility. In the WW formalism this Hamiltonian becomes (see eqs.(278))

$$H_I = Aa_s^* a_i^* \exp\left(-i\omega_P t\right) + A^* a_s a_i \exp\left(i\omega_P t\right),$$

whence, taking eqs.(282) and (283) into account, we get

$$\begin{aligned} \frac{d}{dt}a_s &= -i\omega_s a_s - iAa_i^* \exp\left(-i\omega_P t\right), \\ \frac{d}{dt}a_i &= -i\omega_i a_i - iAa_s^* \exp\left(-i\omega_P t\right). \end{aligned} \tag{285}$$

In eqs.(284) to (285) the space dependence has been omitted whence the propagation direction of the fields is not explicit. It is the case that the field with amplitude Aa_i^* produced by the interaction of the pumping laser with the incoming vacuum mode, a_s, travels superposed to the incoming vacuum field a_i, while the field Aa_s^* produced by the vacuum mode a_i travels superposed to the incoming vacuum field a_s. That is, the wave-vectors associated to a_s and a_i^* are parallel and form a finite angle with the wave-vectors associated to a_i and a_s^*, that are also parallel to each other.

We shall assume that the vacuum field a_s evolves as in eq.(283) before entering the crystal and then according to eqs.(285) inside the crystal, that is, during the time T needed to cross it. In order to get the radiation intensity to second order in $AT \equiv C$ (see section 6.2.4 below) we must solve the coupled eqs.(285) also to second order. After some algebra this leads to

(286)

$$\begin{aligned} a_s(t) &= \left(1+\frac{1}{2}\left|C\right|^2\right) a_s(0)\exp\left(-i\omega_s t\right) - iCa_i^*\left(0\right)\exp\left[i\left(\omega_i-\omega_P\right)t\right] \\ &= \left[\left(1+\frac{1}{2}\left|C\right|^2\right) a_s(0) - iCa_i^*\left(0\right)\right]\exp\left(-i\omega_s t\right), \end{aligned}$$

where the last equality is usually interpreted as if the signal and idler photons, with energies $\hbar\omega_s$ and $\hbar\omega_i$, were the result of the division of one laser photon with energy $\hbar\omega_P$, that is, viewed as an 'energy conservation'. However, in the WW formalism that relation is a condition of frequency matching, $\omega_P = \omega_s + \omega_i$, induced by the nonlinear susceptibility with no reference to photons.

Eq.(286) gives the time dependence of the relevant mode of the signal after crossing the crystal, but we should take into account the field dependence on position including a factor $\exp\left(i\mathbf{k}_s\cdot\mathbf{r}\right)$, that is, a phase depending on the path length. Therefore, the correct form of eq.(286) would be, modulo a global phase,

$$a_s(\mathbf{r},t) = \left[\left(1+\frac{1}{2}\left|C\right|^2\right) a_s(0) - iCa_i^*\left(0\right)\right]\exp\left(i\mathbf{k}_s\cdot\mathbf{r} - i\omega_s t\right). \tag{287}$$

A similar result is obtained for $a_i\left(t\right)$, that is,

$$a_i(\mathbf{r},t) = \left[\left(1+\frac{1}{2}\left|C\right|^2\right) a_i(0) - iCa_s^*\left(0\right)\right]\exp\left(i\mathbf{k}_i\cdot\mathbf{r} - i\omega_i t\right). \tag{288}$$

Eq.(287) may be interpreted saying that the vacuum signal is modified by the addition of an amplification of the vacuum idler, but it travels in the same direction of the incoming vacuum signal, and therefore it makes sense to add the initial vacuum signal plus the amplification of the idler. And similarly for a_i, eq.(288).

We may perform a change from C to the new parameter $D = -i\left(1+\frac{1}{2}|C|^2\right)^{-1} C$, whence eqs.(280) become, to order $O\left(|D|^2\right)$,

$$E_s^+ = \left(1+\frac{1}{2}|D|^2\right)\left[a_s + Da_i^*\right]\exp\left(i\mathbf{k}_s\cdot\mathbf{r} - i\omega_s t\right), \tag{289}$$

$$E_i^+ = \left(1+\frac{1}{2}|D|^2\right)\left[a_i + Da_s^*\right]\exp\left(i\mathbf{k}_i\cdot\mathbf{r} - i\omega_i t\right), \quad |D| \ll 1.$$

In the following I will ignore the constant global factor $\left(1+\frac{1}{2}|D|^2\right) \sim 1$ and write, ignoring also the spacetime dependence,

$$E_s^+ = a_s + Da_i^*, E_i^+ = a_i + Da_s^*. \tag{290}$$

Eqs.(287) and (288), although good enough for calculations, are bad representations of the physics. In fact, a physical beam corresponds to a superposition of the amplitudes, $a_{\mathbf{k}}^*$, of many modes with frequencies and wave-vectors close to ω_s and $\mathbf{k}_s$, respectively. For instance, we may represent the positive frequency part of the idler beam created in the crystal, to first order in D, as follows

$$E_i^{(+)}(\mathbf{r},t) = D\int f_i(\mathbf{k})\, d^3\mathbf{k} a_{\mathbf{k}}^* \exp\left[i\left(\mathbf{k}-\mathbf{k}_s\right)\cdot\mathbf{r} - i\left(\omega-\omega_s\right)t\right] + E_{ZPF}^{(+)}, \tag{291}$$

where $\omega = \omega(\mathbf{k})$ and $f_i(\mathbf{k})$ is an appropriate function, with domain in a region of $\mathbf{k}$ around $\mathbf{k}_s$. The field $E_{ZPF}^{(+)}$ is the sum of amplitudes of all vacuum modes, including the one represented by a_s in eq.(289) . (We have neglected a term of order $|D|^2$ so that $E_i^{(+)}$ is correct to order $|D|$). These vacuum modes have fluctuating amplitudes with a probability distribution given by the vacuum Wigner function eq.(277) . Therefore, it might appear that the amplitude a_s is lost 'as a needle in the haystack' within the background of many radiation modes, but it is actually relevant in photon correlation experiments, as shown below. In fact, the vacuum amplitude a_s in the first eq.(289) is fluctuating and the same fluctuations appear also in the signal amplitude a_s^* of the second eq.(289). Therefore, coincidence counts will be favoured when large positive fluctuations of the fields a_s and a_i arrive simultaneously to Alice and Bob detectors. We shall see that this corresponds to what is named 'entanglement between a signal and the vacuum' [**19**] in the HS formalism. Thus in the WW formalism of quantum optics *entanglement appears as a correlation between fluctuating fields in distant places, including the vacuum fields.*

6.4.2. Detection rates. The example of signal-idler cross-correlation. A quantitative theory of detection in the WW formalism may be obtained translating, via the Weyl transform, the standard detection theory in the HS formalism. Here I will use this method, and later on we shall try to provide a realistic stochastic interpretation of the detection probabilities so derived, taking the assumptions of section 6.2.6 into account. I will study only detection events in the form of photocounts ignoring other detection methods, like measuring the total energy arriving at the detector in a time interval, that are less relevant for experiments with entangled photons. Then the measured detection rate is proportional to the probability of a count in one time window. I shall not attempt to present detailed models of the experiments, which should involve many radiation modes in order to represent the signals as (narrow) beams, see eq.(291), but use just a few vacuum radiation modes entering the nonlinear crystal, whence the signals would involve also a few modes.

I will start with the simple case of an experiment measuring the cross-correlation between signal and idler fields and then generalize to more complex experiments. In the HS formalism the field operators corresponding to eqs.(290) are

$$\begin{aligned} \widehat{E}_s^+ &= \left[\hat{a}_s + D\hat{a}_i^\dagger\right] \exp\left(i\mathbf{k}_s{\cdot}\mathbf{r} - i\omega_s t\right), \\ \widehat{E}_i^+ &= \left[\hat{a}_i + D\hat{a}_s^\dagger\right] \exp\left(i\mathbf{k}_i{\cdot}\mathbf{r} - i\omega_i t\right). \end{aligned} \tag{292}$$

These field operators may be obtained either from the Hamiltonian eq.(284) or taking eq.(278) into account, just putting 'hats' in the field amplitudes obtained in the WW formalism. In the HS formalism the Alice single rate is given by the following vacuum expectation to order $O\left(|D|^2\right)$:

$$R_A = \left\langle 0 \left| \widehat{E}_s^- \widehat{E}_s^+ \right| 0 \right\rangle = \left\langle 0 \left| \left[\hat{a}_s^\dagger + D^* \hat{a}_i\right] \left[\hat{a}_s + D\hat{a}_i^\dagger\right] \right| 0 \right\rangle = |D|^2, \tag{293}$$

and similarly for Bob. It is easy to prove that the spacetime factors of eq.(292) cancel. The coincidence rate is given by

$$R_{AB} = \frac{1}{2} \left\langle 0 \left| \widehat{E}_s^- \widehat{E}_i^- \widehat{E}_i^+ \widehat{E}_s^+ \right| 0 \right\rangle + \frac{1}{2} \left\langle 0 \left| \widehat{E}_i^- \widehat{E}_s^- \widehat{E}_s^+ \widehat{E}_i^+ \right| 0 \right\rangle.$$

Taking into account that $\widehat{E}_i^+$ and $\widehat{E}_s^+$ commute, both terms are equal and we have

$$\begin{aligned} R_{AB} &= \left\langle 0 \left| \widehat{E}_s^- \widehat{E}_i^- \widehat{E}_i^+ \widehat{E}_s^+ \right| 0 \right\rangle \\ &= \left\langle 0 \left| \left[\hat{a}_s^\dagger + D^* \hat{a}_i\right] \left[\hat{a}_i^\dagger + D^* \hat{a}_s\right] \left[\hat{a}_i + D\hat{a}_s^\dagger\right] \left[\hat{a}_s + D\hat{a}_i^\dagger\right] \right| 0 \right\rangle \\ &= |D|^2 \left\langle 0 \left| \hat{a}_i \hat{a}_i^\dagger \hat{a}_i \hat{a}_i^\dagger \right| 0 \right\rangle + O\left(|D|^4\right) = |D|^2 + O\left(|D|^4\right). \end{aligned} \tag{294}$$

In the following we will study more complex experiments involving 'entangled photon pairs' produced via SPDC, but restricted to cases where there are only two detectors, say belonging to Alice and Bob. In general there are a number of devices between the nonlinear crystal, or crystals, and the detectors, like lens systems, beam splitters, etc. In any case, the field operator corresponding to the radiation arriving at the detectors may be written in the form

$$Alice : \widehat{E}_A^+ = \widehat{E}_{A0}^+ + \widehat{E}_{A1}^+, \qquad Bob : \widehat{E}_B^+ = \widehat{E}_{B0}^+ + \widehat{E}_{B1}^+,$$

where for every detector the first term is the field operator that would appear in case there was no pumping on the crystals, whilst the second one derives from the fields produced by the action of the pumping. In the simple experiment studied in eqs.(292) to (294) we have

$$\widehat{E}_{A0}^+ = \hat{a}_s, \widehat{E}_{A1}^+ = D\hat{a}_i^\dagger, \qquad \widehat{E}_{B0}^+ = \hat{a}_i, \widehat{E}_{B1}^+ = D\hat{a}_s^\dagger \tag{295}$$

but in other experiments the field operators are more involved. However, there are two features that are general, namely $\widehat{E}_{A1}^+$ is small and consists of a linear combination of *creation* operators, whilst $\widehat{E}_{A0}^+$ is large and consists of a linear combination of *annihilation* operators. That is, we may write

$$\widehat{E}_{A0}^+ = \sum_j c_j \hat{a}_j, \widehat{E}_{A1}^+ = \sum_k d_k \hat{a}_k^\dagger, \tag{296}$$

where by analogy with eqs.(295) I shall say that $\widehat{E}_{A0}^+$ is of the order of unity and $\widehat{E}_{A1}^+$ of order $|D| \ll 1$. And similarly for $\widehat{E}_{B0}^+$ and $\widehat{E}_{B1}^+$.

From these field operators the single detection rate by Alice may be obtained via the following standard rule in the HS formalism

$$\begin{aligned} R_A &= \langle 0 \mid \widehat{E}_A^- \widehat{E}_A^+ \mid 0 \rangle = \langle 0 \mid \left(\widehat{E}_{A0}^- + \widehat{E}_{A1}^-\right) \left(\widehat{E}_{A0}^+ + \widehat{E}_{A1}^+\right) \mid 0 \rangle \\ &= \langle 0 \mid \widehat{E}_{A0}^- \widehat{E}_{A0}^+ \mid 0 \rangle + \langle 0 \mid \widehat{E}_{A1}^- \widehat{E}_{A1}^+ \mid 0 \rangle = \langle 0 \mid \widehat{E}_{A1}^- \widehat{E}_{A1}^+ \mid 0 \rangle, \end{aligned} \tag{297}$$

where we take into account that

$$\widehat{E}_{A0}^+ \mid 0 \rangle = \langle 0 \mid \widehat{E}_{A0}^- = 0. \tag{298}$$

In particular, in the simple case of measuring the detection rate of the signal field, eq.(295) gives

$$R_A = |D|^2 \langle 0 \mid \hat{a}_i \hat{a}_i^\dagger \mid 0 \rangle = |D|^2 , \tag{299}$$

and identical result for Bob's rate.

The HS rule for the coincidence rate is

$$R_{AB} = \frac{1}{2}\langle 0 \mid \widehat{E}_A^- \widehat{E}_B^- \widehat{E}_B^+ \widehat{E}_A^+ \mid 0 \rangle + \frac{1}{2}\langle 0 \mid \widehat{E}_B^- \widehat{E}_A^- \widehat{E}_A^+ \widehat{E}_B^+ \mid 0 \rangle. \tag{300}$$

If $\widehat{E}_A^+$ and $\widehat{E}_B^+$ commute then both terms would be equal, but I shall work out the general case when this is not true. The first term may be calculated as follows

$$\begin{aligned} \langle 0 \mid \widehat{E}_A^- \widehat{E}_B^- \widehat{E}_B^+ \widehat{E}_A^+ \mid 0 \rangle &= \langle 0 \mid \widehat{E}_{A1}^- \widehat{E}_B^- \widehat{E}_B^+ \widehat{E}_{A1}^+ \mid 0 \rangle \\ &\simeq \langle 0 \mid \widehat{E}_{A1}^- \widehat{E}_{B0}^- \widehat{E}_{B0}^+ \widehat{E}_{A1}^+ \mid 0 \rangle \\ &= \langle 0 \mid \widehat{E}_{A1}^- \widehat{E}_{B0}^- \mid 0 \rangle \langle 0 \mid \widehat{E}_{B0}^+ \widehat{E}_{A1}^+ \mid 0 \rangle = \left| \langle 0 \mid \widehat{E}_{B0}^+ \widehat{E}_{A1}^+ \mid 0 \rangle \right|^2 , \end{aligned} \tag{301}$$

where the first equality takes eqs.(298) into account and the second is valid to order $|D|^2$ (terms of order $|D|^3$ are neglected). Finally the third equality is justified as follows. The field $\widehat{E}_{B0}^+$ $\left(\widehat{E}_{A1}^+\right)$ consists of a linear combination of annihilation (creation) operators whence $\widehat{E}_{B0}^+ \widehat{E}_{A1}^+ \mid 0 \rangle$ consists of a sum of the form

$$\widehat{E}_{B0}^+ \widehat{E}_{A1}^+ \mid 0 \rangle = \sum_j \sum_k c_j d_k \hat{a}_j \hat{a}_k^\dagger \mid 0 \rangle, \tag{302}$$

where c_j and d_k are numerical parameters. Terms with $j \neq k$ give no contribution because the operator $\hat{a}_j$ cannot annihilate a photon that had not been created. Therefore, eq.(302) reduces to the form

$$\widehat{E}_{B0}^+ \widehat{E}_{A1}^+ \mid 0 \rangle = \sum_j c_j d_j \hat{a}_j \hat{a}_j^\dagger \mid 0 \rangle = \sum_j c_j d_j \mid 0 \rangle,$$

that is, the vacuum state vector times a numerical coefficient. This justifies the third equality eq.(301) because the inclusion of the operator $\mid 0 \rangle\langle 0 \mid$ does not modify the result. A similar treatment may be made for the other term of eq.(300) and we finally get

$$R_{AB} = \frac{1}{2} \left| \langle 0 \mid \widehat{E}_{B0}^+ \widehat{E}_{A1}^+ \mid 0 \rangle \right|^2 + \frac{1}{2} \left| \langle 0 \mid \widehat{E}_{A0}^+ \widehat{E}_{B1}^+ \mid 0 \rangle \right|^2 . \tag{303}$$

For instance, in the simple case of the cross-correlation between signal and idler we get, taking eq.(295) into account,

$$R_{AB} = \frac{1}{2}\left|D\right|^2 \langle 0 \mid \hat{a}_i \hat{a}_i^\dagger \mid 0\rangle + \frac{1}{2}\left|D\right|^2 \langle 0 \mid \hat{a}_s \hat{a}_s^\dagger \mid 0\rangle = \left|D\right|^2 . \tag{304}$$

Thus the theory predicts a strong positive correlation. In fact, any count by Alice (Bob) will coincide with a count by Bob (Alice) within the same detection window.

Now we are in a position to derive the WW detection rules via a Weyl transform of the stated rules in HS. The following transforms may be easily derived from eqs.(278)

$$\begin{aligned} \langle 0 \mid \hat{a}_i \hat{a}_i^\dagger \mid 0\rangle \quad &\rightarrow \quad \langle |a_i|^2 + \frac{1}{2}\rangle = 2\langle |a_i|^2\rangle, \\ \langle 0 \mid \hat{a}_i^\dagger \hat{a}_i \mid 0\rangle \quad &\rightarrow \quad \langle |a_i|^2 - \frac{1}{2}\rangle = 0. \end{aligned} \tag{305}$$

We must perform the Weyl transform of the general Hilbert-space rule eq.(297) . If we use the last equality of eq.(297) we realize that $\widehat{E}_{A1}^-$ consists of a sum of annihilation operators and $\widehat{E}_{A1}^+$ a sum of creation operators. Hence the first eq.(305) leads to

$$R_A = \langle 0 \mid \widehat{E}_{A1}^- \widehat{E}_{A1}^+ \mid 0\rangle \rightarrow R_A = 2\langle E_{A1}^- E_{A1}^+\rangle, \quad R_B = 2\langle E_{B1}^- E_{B1}^+\rangle. \tag{306}$$

I stress that the Weyl transform of the first equality in eq.(297) leads to the same result, that is,

$$R_A = \langle 0 \mid \widehat{E}_{A0}^- \widehat{E}_{A0}^+ \mid 0\rangle + \langle 0 \mid \widehat{E}_{A1}^- \widehat{E}_{A1}^+ \mid 0\rangle \rightarrow 0 + 2\langle E_{B1}^- E_{B1}^+\rangle.$$

With similar arguments, the HS rule for the coincidence rate, eq.(303) , leads to the WW rule

$$R_{AB} = 2\left|\langle E_{B0}^+ E_{A1}^+\rangle\right|^2 + 2\left|\langle E_{A0}^+ E_{B1}^+\rangle\right|^2 . \tag{307}$$

Eq.(307) provides the desired WW detection rule, which is rather simple if written in terms of field amplitudes arriving at Alice and Bob respectively. By construction it is obvious that, using the rules eqs.(306) and (307), the WW formalism will give the same predictions than the standard quantum HS formalism for all experiments involving entangled photon pairs produced via SPDC. I will analyze below a few relevant experiments.

6.4.3. Detection rates in terms of intensities. For the realistic interpretation to be discussed later it is interesting to write the detection rules eqs.(306) and (307) in terms of field intensities

rather than amplitudes. With this purpose we define several partial intensities in terms of field components as follows

(308)

$$I_{A0} = E^+_{A0}E^-_{A0}, \quad I_{A1} = E^+_{A0}E^-_{A1} + E^+_{A1}E^-_{A0}, \quad I_{A2} = E^+_{A1}E^-_{A1},$$
$$I_{B0} = E^+_{B0}E^-_{B0}, \quad I_{B1} = E^+_{B0}E^-_{B1} + E^+_{B1}E^-_{B0}, \quad I_{B2} = E^+_{B1}E^-_{B1},$$
$$I_A = E^+_A E^-_A = I_{A0} + I_{A1} + I_{A2}, \quad I_B = E^+_B E^-_B = I_{B0} + I_{B1} + I_{B2}.$$

Actually, I_{A1} and I_{B1} are not true intensities, in particular they are not positive definite. I point out that I_{A0}, I_{A1} and I_{A2} are of order $1, |D|$ and $|D|^2$, respectively, in the small parameter $|D| \ll 1$, and similarly for I_{B0}, I_{B1} and I_{B2}. Eq.(306) may easily be obtained in terms of intensities taking eqs.(308) into account. We get

$$R_A = 2\langle I_A\rangle - 2\langle I_{A0}\rangle = 2\langle I_{A2}\rangle,$$
$$R_B = 2\langle I_B\rangle - 2\langle I_{B0}\rangle = 2\langle I_{B2}\rangle, \tag{309}$$

the terms I_{A1} and I_{B1} not contributing, as may be realized.

In order to put the coincidence detection rate in terms of intensities we start writing $\langle I_{A0}I_{B2}\rangle$ in terms of the fields taking eqs.(308) into account. We get

$$\langle I_{A0}I_{B2}\rangle = \left\langle E^+_{A0}E^-_{A0}E^+_{B1}E^-_{B1}\right\rangle,$$

which may be transformed as follows taking into account that the fields involved are Gaussian. In fact, the average of the product of 4 Gaussian random variables, say A, B, C, D, may be obtained from the averages of every pair, that is,

$$\langle ABCD\rangle = \langle AB\rangle\langle CD\rangle + \langle AC\rangle\langle BD\rangle + \langle AD\rangle\langle BC\rangle. \tag{310}$$

Hence it follows that

$$\begin{aligned}\langle I_{A0}I_{B2}\rangle &= \left\langle E^+_{A0}E^-_{A0}E^+_{B1}E^-_{B1}\right\rangle = \left\langle E^+_{A0}E^-_{A0}\right\rangle\left\langle E^+_{B1}E^-_{B1}\right\rangle \\ &\quad + \left\langle E^+_{A0}E^+_{B1}\right\rangle\left\langle E^-_{A0}E^-_{B1}\right\rangle + \left\langle E^+_{A0}E^-_{B1}\right\rangle\left\langle E^-_{A0}E^+_{B1}\right\rangle \\ &= \langle I_{A0}\rangle\langle I_{B2}\rangle + \left|\left\langle E^+_{A0}E^+_{B1}\right\rangle\right|^2,\end{aligned}$$

the average $\left\langle E^+_{A0}E^-_{B1}\right\rangle$ not contributing as may be realized. A similar method gives $\left|\left\langle E^+_{B0}E^+_{A1}\right\rangle\right|^2$ in terms of intensities whence eq.(307) leads to

$$\begin{aligned}R_{AB} &= 2\left[\langle I_{A0}I_{B2}\rangle + \langle I_{A2}I_{B0}\rangle - \langle I_{A0}\rangle\langle I_{B2}\rangle - \langle I_{A2}\rangle\langle I_{B0}\rangle\right] \\ &= 2\left\langle (I_{A0} - \langle I_{A0}\rangle)I_{B2}\right\rangle + 2\left\langle (I_{B0} - \langle I_{B0}\rangle)I_{A2}\right\rangle. \end{aligned} \tag{311}$$

The conclusion of this section is that the WW formalism is a useful tool for the study of SPDC. In particular, the predictions for the experiments may be obtained as easily as in the standard HS using for the single rates either eq.(306) or eq.(309) and for the coincidence rates either eq.(307) or eq.(311), respectively. In each case the first prediction of rates appears in terms of field amplitudes and the second is in terms of intensities.

6.5. Experiments with entangled photons from SPDC

6.5.1. Experiment of two-photon interference using two crystals. In the following we illustrate the use of the WW formalism for the interpretation of two SPDC experiments. The former consists of two crystals pumped by two coherent beams, obtained via dividing one laser beam by means of a balanced beam splitter [**17**], [**18**], [**19**]. The signal beams from both crystals are sent to the two incoming channels of a beam splitter, BS1, and in front of one of the outgoing channels a detector, say Alice, is placed. Similarly the idler beams of both crystals are mixed via another beam splitter, BS2, and the field from one of the outgoing channels goes to another detector, say Bob. The measured quantity is the coincidence counting rate between A and B as a function of the path length of one of the beams, which may be controlled at will.

The interest of the experiment is that it involves second order field correlation functions, rather than first order correlation as in the measurement of the crosscorrelation of signal and idler from a single crystal, studied in section 6.4.2 above. In a simple quantum treatment we may assume that in the first crystal an 'entangled photon pair' is produced with an amplitude $\psi_1 \exp(i\phi_1)$, the exponential factor introduced in order to take account of the space dependence of the fields (see e.g. eq.(289)). Similarly in the second crystal another pair is produced with amplitude $\psi_2 \exp(i\phi_2)$. If both pairs arrive at Alice and Bob detectors, according to the standard quantum rules the probability of joint detection should be proportional to the total amplitude, obtained adding both amplitudes. This is similar to the sum of amplitudes in the two-slits interference experiment. Therefore, the coincidence rate (or coincidence detection probability within a time window) will be

$$\begin{aligned} R_{AB} &\propto |\psi_1 \exp(i\phi_1) + \psi_2 \exp(i\phi_2)| \\ &= |\psi_1|^2 + |\psi_2|^2 + 2|\psi_1||\psi_2|\cos(\phi_1 - \phi_2). \end{aligned} \tag{312}$$

In the actual experiments the phase difference is controlled via changing at will the path length of the signal coming from the second crystal. If $|\psi_1| = |\psi_2|$ the main prediction is that 100% interference visibility is obtained; that is, the coincidence rate may be written

$$R_{AB} \propto 1 + \cos(\phi_1 - \phi_2).$$

A detailed calculation has been made [**19**] that reproduces this result. The method involves the standard Hilbert-space formalism in the Heisenberg representation. In our approach, that follows, we use the WW formalism that closely parallels the HS treatment starting in eq.(292) for the simple experiment of signal-idler cross-correlation. This allows the experiment to be interpreted in terms of stochastic processes involving the vacuum fields as follows.

The signal and idler fields produced in the first and second crystals are given by eq.(289). That is, ignoring the factor $\left(1 + \frac{1}{2}|D|^2\right) \sim 1$ and the space dependence, we shall write

(313)

$$E^+_{s1} = [a_{s1} + Da^*_{i1}] \exp(-i\omega_s t), \quad E^+_{i1} = [a_{i1} + Da^*_{s1}] \exp(-i\omega_i t),$$
$$E^+_{s2} = [a_{s2} + Da^*_{i2}] \exp(-i\omega_s t), \quad E^+_{i2} = [a_{i2} + Da^*_{s2}] \exp(-i\omega_i t),$$

where the subindices 1 and 2 refer to the first and second crystal respectively, with a notation similar to ref. [**19**].

The amplitude of the light beam arriving at Alice detector will consists of an appropriate superposition of the field amplitudes E^+_{s1} and E^+_{s2} in the beam-splitter BSA in front of Alice's detector, see eq.(267), although here we do not assume that BSA is balanced. We shall take into account the phase difference between the two beams mixed at BSA, due to the path length difference that may be changed at will. We get

$$E^+_A = \{[a_{s1} + Da^*_{i1}]r \exp(i\phi_1) + [a_{s2} + Da^*_{i2}]t\} \exp(-i\omega_s t), \quad (314)$$

where t and r are the transmission and reflection coefficients of BSA, and the imaginary unit in eq.(267) has been absorbed in the (complex) reflection coefficient r. There is a global factor in front of the right side of eq.(314) that is irrelevant for our purposes and we ignore. In this factor we absorb the term in $|D|^2$ of eq.(326), needed to get the intensity to that order. That is, the parameter D of eq.(314) actually corresponds to $D/(1 + |D|^2/2)$ of eqs.(287) and (288).

Similarly, the idler fields of both crystals are sent to another beam splitter BSB and hence to Bob's detector. The positive frequency part

of the field arriving at Bob may be written, assuming that D, r and t are the same for Alice and Bob beams,

$$E_B^+ = \{[a_{i1} + Da_{s1}^*]t + [a_{i2} + Da_{s2}^*]r \exp(i\phi_2)\} \exp(-i\omega_i t). \quad (315)$$

The notation in eqs.(314) and (315) is similar to Ref. [**19**].

The interesting quantity to be compared with experiments is the coincidence counting rate, R_{AB}, of Alice and Bob. In the WW formalism we should calculate it using eq.(307) where E_A^+ and E_B^+ are given by eqs.(314) and (315) and E_{A0}^+, E_{B0}^+ may be written, ignoring the time-dependent factor,

$$E_{A0}^+ = ra_{s1} \exp(i\phi_1) + ta_{s2}, \quad E_{B0}^+ = ta_{i1} + a_{i2} r \exp(i\phi_2). \quad (316)$$

The fields E_{A1}^+, E_{B1}^+ may be obtained subtracting the first and second eqs.(316) from eq.(314) and eq.(315) respectively. After some algebra we get

$$\begin{aligned} R_{AB} &= 2|D|^2 |rt|^2 \left[|\exp(i\phi_1) + \exp(i\phi_2)|^2 \right] \\ &\simeq 4|D|^2 |rt|^2 [1 + \cos(\phi_2 - \phi_1)], \end{aligned} \quad (317)$$

to order $O\left(|D|^2\right)$. This agrees with the result got using the HS formalism [**19**] as it should.

In our approach the amplitudes $a_{s1}, a_{s2}, a_{i1}, a_{i2}$, are random variables representing the amplitudes of vacuum modes arriving at the nonlinear crystals, that we assume to be real stochastic fields.

6.5.2. Induced coherence and indistinguishability in two photon interference. In 1991 Mandel and coworkers performed an experiment [**20**] in which fourth order interference is observed in the superposition of signal photons from two coherently pumped crystals, when the paths of the idler photons are aligned. The experiment consists of two nonlinear crystals, $C1$ and $C2$, pumped by two mutually coherent, classical pump waves. The two pumping fields come from a single laser beam divided via a nonpolarizing balanced beam splitter. The *signal* fields produced in the two crystals are superposed via a beam splitter BS and then send to Alice's detector, put in front of one outgoing channel of BS. The crystals are placed so that an *idler* beam produced in the crystal $C1$ would enter in the crystal $C2$ at a point and with a direction such that it is precisely superposed to an eventual idler produced in $C2$. The resulting idler beam would arrive to Bob's detector and the set-up is such that when a photon is detected Bob cannot know whether it was produced in $C1$ or in $C2$.

Therefore, in a quantum calculation we should add the amplitudes of both possibilities. The difference with the experiment of the previous subsection is that the idler fields are not superposed via a beam splitter, but directly superposed because the two crystals are aligned.

In the calculation using the WW formalism [**4**] the field arriving at Alice will be the same of eq.(314). However, the field arriving at Bob will not be eq.(315) but, to order $|D|$,

$$E_B^+ = \{[a_{i10} + Da_{s10}^*] + [a_{i20} + Da_{s20}^*] \exp(i\phi_2)\} \exp(-i\omega_i t), \tag{318}$$

that is, 1 is substituted for the coefficients t and r. The result is an interference effect in the coincidence detection rate similar to eq.(317) with minor changes.

The interest of the experiment is that fourth order interference disappears when the idlers from $C1$ and $C2$ are misaligned or separated by a beam stop. In this case *we know* that the idler beam arriving at Bob has been produced in $C2$ and therefore *we should add the intensities, not the amplitudes of the fields arriving at Alice from the two crystals, and the interference effect disappears*, because the fields arriving at Alice are no longer coherent. This is in agreement with the popular Feynman's rule for interference in the two-slit experiment: If it is not possible even in principle to know which slit crosses the particle, then we should add the amplitudes of both paths. If it is possible to know which slit, then we should add probabilities. From another point of view when the idlers from $C1$ and $C2$ are misaligned or separated by a beam stop, we should assume that any coincidence count would derive from a signal and an idler both produced in $C2$. In contrast, if the signal is produced in $C1$, the partner idler will not arrive at Bob and no coincidence count will take place.

In any case, the peculiarities of the quantum behaviour, if we interpret the experiment in terms of photons, are such that the experiment has been called 'mind boggling' [**21**]. Indeed, it is difficult to understand how the interference of two signal photons in detector A may depend on whether the idler photon arriving to detector B is surely produced in crystal $C2$ or we cannot know in which crystal, $C1$ or $C2$, it has been produced. In the second case there is interference, in the first one there is not.

The explanation of the experimental result is simple with our realistic interpretation that assumes that the vacuum fields are actual stochastic fields. A vacuum field entering $C1$ produces an idler field E_{i1} that travels towards $C2$ where another idler field E_{i2} is produced

that is superposed to the former. Then the amplitudes (not the intensities!) of both idler fields should be added because E_{i1} and E_{i1} are coherent. On the other hand every signal is coherent with the corresponding idler as shown in section 6.4.2. Hence the fields E_{s1} and E_{s2} become coherent and therefore their superposition in the beam splitter BS is sensitive to the relative phase, that may be controlled at will by changing the path lengths. In contrast, if the two idler fields E_{i1} and E_{i2} are separated by a beam stop they will be incoherent, whence also E_{s1} and E_{s2} become incoherent and no interference can be observed. A quantitative treatment may be made in the WW formalism that reproduces the empirical results but it will not be included here, see ref. [**4**].

6.6. Tests of Bell inequalities

The Bell inequalities have had a deep impact in the foundations of quantum physics, as mentioned in chapter 3. In the last two decades most tests of the inequalities have used entangled photon pairs produced in SPDC. In this section we shall analyze a representative test similar to those claiming to provide a loophole-free violation of a Bell inequality [**22**], [**23**]. Their empirical violation is usually interpreted as a refutation of local realism, but in the following it is shown that the experiments we are discussing may be interpreted as local and realistic even if a Bell inequality is violated.

6.6.1. Photons entangled in polarization. Many experimental tests of Bell inequalities using SPDC entangled photon pairs have been performed. Some of the early tests involved entanglement in phase and momentum, or energy and time. However, the most relevant experiments used polarization entanglement. In the following I will apply the WW formalism to the interpretation of these experiments. The loophole-free tests performed [**22**], [**23**] did not use maximally entangled photons for reasons explained in chapter 3 section 3.4.3. However, here we will study maximally entangled photons because it is simpler and our purpose is not to analyze the actual experiments but to provide a counterexample that may be interpreted as local realistic in spite of violating a Bell inequality.

The electromagnetic radiation is a vector field with two possible polarizations. Therefore, I should take into account this fact, for instance including vectors in the description. Thus, after sending the signal and idler beams to the incoming channels of a polarizer beam

splitter, we shall have in the outgoing channels,

$$\begin{aligned}\mathbf{E}_A^+ &= [(a_s + Da_i^*)\exp(-i\omega_s t)\,\mathbf{v} + i\,(a_i + Da_s^*)\exp(-i\omega_i t)\,\mathbf{h}],\\ \mathbf{E}_B^+ &= [(a_s + Da_i^*)\exp(-i\omega_s t)\,\mathbf{h} - i\,(a_i + Da_s^*)\exp(-i\omega_i t)\,\mathbf{v}],\end{aligned}$$

where $\mathbf{h}$ is a horizontal unit vector and $\mathbf{v}$ a vertical one. We have not written explicitly the dependence on position, that could be restored without difficulty. Furthermore, from now on I will ignore all space-time dependence, that usually contributes phase factors irrelevant for our argument. However, that dependence should be taken into account in more realistic calculations involving many radiation modes, see eq.(291) . The complex conjugate of these fields will be represented as follows

$$(\mathbf{E}_A^+)^* \equiv \mathbf{E}_A^-, \qquad (\mathbf{E}_B^+)^* \equiv \mathbf{E}_B^-$$

These equations represent 'two photons entangled in polarization' as seen in the Weyl-Wigner formalism. The beams will arrive at the Alice and Bob polarization analyzers put at angles θ and ϕ with the vertical respectively. Hence the fields emerging from them may be written as

$$\begin{aligned}E_A^+ &= [(a_s + Da_i^*)\cos\theta + i\,(a_i + Da_s^*)\sin\theta],\\ E_B^+ &= [(a_s + Da_i^*)\sin\phi - i\,(a_i + Da_s^*)\cos\phi],\end{aligned} \tag{319}$$

and they are polarized at angles θ and ϕ with the vertical, respectively. E_A^+ $\left(E_B^+\right)$ gives one of the two components of the vector field arriving at Alice (Bob). The other component corresponds to a vacuum field that will no longer enter the calculation. I define also the fields corresponding to the case when the pumping is off, that is,

$$E_{A0}^+ = a_s\cos\theta + ia_i\sin\theta, E_{B0}^+ = a_s\sin\phi - ia_i\cos\phi, \tag{320}$$

The difference with the total fields eqs.(319) is

$$E_{A1}^+ = Da_i^*\cos\theta + iDa_s^*\sin\theta, \quad E_{B1}^+ = Da_i^*\sin\phi + iDa_s^*\cos\phi. \tag{321}$$

In order to get the single detection rate we shall use eq.(306). Alice's single rate becomes

$$\begin{aligned}R_A &= 2\left\langle \left|E_{A1}^+\right|^2\right\rangle\\ &= 2\,|D|^2\left(\left\langle |a_i|^2\right\rangle\cos^2\theta + \left\langle |a_s|^2\right\rangle\sin^2\theta\right)\\ &= |D|^2 = R_B,\end{aligned} \tag{322}$$

Bob's detection rate being similar. For the coincidence detection rate we shall use the first eq.(307), leading to

(323)

$$\begin{aligned} R_{AB} &= |\langle (a_s \sin\phi - ia_i \cos\phi)\,(Da_i^* \cos\theta + iDa_s^* \sin\theta)\rangle|^2 \\ &+ |\langle (a_s \cos\theta + ia_i \sin\theta)\,(Da_i^* \sin\phi + iDa_s^* \cos\phi)\rangle|^2 \\ &= |D|^2 \left[\frac{1}{2}\,(\sin\phi \sin\theta + \cos\phi \cos\theta)\right]^2 \times 2 = \frac{1}{2}\,|D|^2 \cos^2(\theta - \phi). \end{aligned}$$

Eqs.(322) and (323) reproduce the well known quantum result of the Hilbert-space formalism. The advantage of the WW calculation is that it suggests a physical interpretation.

Our aim now is to try finding a realistic interpretation for the WW description of the SPDC experiments resting on the assumption that the quantum fields are stochastic fields. I will start searching for an intuitive picture of the detection rules eqs.(309) and (311) via an elaboration of the ideas put forward in section 6.4.3 above. For the picture it is convenient to deal with detection probabilities per time window, say P_A, P_B for single counts and P_{AB} for coincidences, rather than detection rates that we have labeled R_A, R_B and R_{AB}. They are related via $R = NP$, where N is the number of windows per unit time.

6.6.2. Realistic interpretation of entangled photons. The picture of the quantum optical phenomena suggested by the WW formalism is quite different from the picture in terms of photons suggested by the HS [**24**]. In the latter some photons of the (usually pulsed) laser beam are assumed to split by the interaction with the nonlinear crystal giving two photons. This produces outgoing radiation whence two beams may be selected, named signal and idler, that emerge from the crystal in different directions. Each beam consists of a flow of photons that are paired amongst the two beams, thus producing a set of entangled photon pairs, one member of the pair in each beam. After crossing appropriate devices the photons may arrive at Alice and Bob respectively. The probability of producing an entangled photon pair by splitting of one laser photon is of order $|D|^2 \ll 1$ whence the simultaneous arrival of entangled photons at Alice and Bob happens for a small fraction of the laser pulses. The detection, or not, of the photons determines the correlation, which eventually may violate a Bell inequality. In summary, it is assumed

that the pumping laser produces entangled photon pairs in the crystal with a *small* probability and then there is a detection probability of the order of *unity*, conditional to the photon production. The latter is defined as detection efficiency.

In the WW approach the probability of photocounts by Alice or Bob does not factorize that way. Furthermore, the concept of photon does not appear at all, but there are *continuous fluctuating fields including a real ZPF* arriving at the detectors that are activated when the radiation intensity is big enough. As already said I will speak of single probabilities, P_A and P_B, and coincidence probability, P_{AB}, of photocounts within a time window. They are proportional to the single and coincidence detection rates. Every time window may agree with one laser pulse.

It is interesting how the 'quantum correlation' appears in the WW formalism, which is labelled strange from a classical point of view because it is a consequence of the phenomenon of entanglement. The origin is the correlation between the signal I_{B2} produced in the crystal and the part I_{A0} that comes from the ZPF, see eq.(311) . This correlation derives from the fact that the same normal modes enter in both fields, E^+_{A0} and E^+_{B1} see eqs.(320) and (321), that go to Alice and Bob respectively. And similarly for E^+_{A1} and E^+_{B0}. However, eq.(311) may look strange because it consists of two terms contributing to the coincidence detection: The first gives some detection probability by Alice that is $I_{A0} - \langle I_{A0}\rangle$, and a different detection probability by Bob, I_{B2}, and the second gives I_{A2} and $I_{B0} - \langle I_{B0}\rangle$ to Alice and Bob, respectively. Also the (Alice) ensemble average of $I_{A0} - \langle I_{A0}\rangle$ is zero, meaning that *only the fluctuations* are involved, and similarly for Bob term $I_{B0} - \langle I_{B0}\rangle$. Furthermore, it seems that for each party one of the terms comes from a signal (that is I_{A2} and I_{B2}) but the other part comes from the ZPF (that is $I_{A0} - \langle I_{A0}\rangle$ and $I_{B0} - \langle I_{B0}\rangle$). This leads to a physical interpretation of entanglement: *it is a correlation between fluctuations involving the vacuum fields*. Usually entanglement is defined as a mathematical property derived from the HS formulation of quantum theory, without any physical interpretation behind (see chapter 2). Actually, the picture offered by eq.(311) is not strange if we take into account the vacuum ZPF field, as we show in the following.

In order to get a plausible picture in terms of stochastic processes several conditions should hold. Firstly we must assume that a detector cannot distinguish the different partial field intensities defined

in eq.(308), it responds to the radiation flux arriving at the sensitive area of the detector within an activation time. The flux corresponds to the time integral of the component of the Poynting vector in the direction perpendicular to the active area of the detector. The net flux will consists of the flux coming from the signal minus the flux of the ZPF that comes from rear, see section 6.2.6.

The picture offered by our interpretation for the single detection probability eq.(309) is clear. In agreement with the above hypotheses we shall assume that the detection probabilities per time window, T, by Alice and Bob respectively, are

$$\begin{aligned} P_A &= \langle M_A \rangle \,, \qquad P_B = \langle M_A \rangle \,, \\ M_A &\equiv T^{-1} \int_0^T \vec{n}_A \cdot \vec{I}^A_{total} \left(\mathbf{r}_A, t \right) dt, \\ M_B &\equiv T^{-1} \int_0^T \vec{n}_B \cdot \vec{I}^B_{total} \left(\mathbf{r}_B, t \right) dt, \end{aligned} \tag{324}$$

where $\vec{I}^A_{total}$ is the Poynting vector of the incoming radiation and $\vec{n}$ a unit vector in the direction of the incoming signals, that I assume perpendicular to the active area of the detector. I use units such that both the intensities and the detection rates are dimensionless, the latter because they are defined as probabilities within a time window T, this being equal or greater than the photocounter activation time.

The total ZPF arriving at any point has the property of isotropy on the average, therefore producing nil mean flux in the detector (modulo fluctuations that might contribute to the dark rate). Hence the Poynting vector of the radiation arriving at the active area of the detectors may be written

$$\begin{aligned} Alice &: \vec{I}^A_{total} \left(t \right) = \vec{I}^A_{ZPF} \left(t \right) + \vec{I}_A \left(t \right) , \\ Bob &: \vec{I}^B_{total} \left(t \right) = \vec{I}^B_{ZPF} \left(t \right) + \vec{I}_B \left(t \right) . \end{aligned} \tag{325}$$

$\vec{I}_A$, $\vec{I}_B$, are due to the fields emerging from the non-linear crystal after they have been transformed by lens systems, apertures, beam splitters, etc. These vectors have the direction of $\vec{n}_A$ and $\vec{n}_B$ respectively, see eqs.(324), whence their moduli would be equal to the arriving signal field intensities. $\vec{I}^A_{ZPF}$ and $\vec{I}^B_{ZPF}$ correspond to radiation belonging to the vacuum field, ZPF. Therefore, eqs.(324) and (325) become,

written in terms of intensities,

$$M_A = I^A_{ZPF} + I_A, I^A_{ZPF} \equiv T^{-1} \int_0^T \vec{n}_A \cdot \vec{I}^A_{ZPF}\left(\mathbf{r}_A,t\right) dt \tag{326}$$

and similarly for M_B. The signal intensities I_A and I_B should also be time averaged, but I do not display it explicitly in eq.(326) . From now on we will use the labels I_A and I_B for the time averages of the signal intensities. Also the partial intensities defined in eq.(308), this is, $I_{A0}, I_{A1}, I_{A2}, I_{B0}, I_{B1}, I_{B2}$, will be assumed to represent time averages.

In order to get the Alice single detection rate we need the average of M_A, that we will evaluate by comparison with the case when the pumping on the nonlinear crystal is off, everything else being identical. In this case I_A becomes I_{A0}, that is, radiation coming from the nonlinear crystal but not modified by the action of the pumping field. It is therefore a part of the ZPF arriving at the detector. Then the Poynting vector of all vacuum fields arriving at Alice, namely $\vec{I}^A_{ZPF}\left(t\right) + \vec{I}_{A0}\left(t\right)$, should have nil average due to the isotropy of the total ZPF. And the same for Bob. As a consequence the intensities arriving at the detectors fulfil the following equalities

$$\left\langle I^A_{ZPF} + I_{A0} \right\rangle = \left\langle I^B_{ZPF} + I_{B0} \right\rangle = 0. \tag{327}$$

Of course, there are other components of the ZPF that are balanced on the average and they will not enter in our calculation. Taking eqs.(326) and (327) into account we get the single detection rate eq.(309), that is,

$$P_A = \langle M_A \rangle = \langle I^A_{ZPF} + I_A \rangle = \langle I_A - \langle I_{A0} \rangle \rangle = \langle I_A \rangle - \langle I_{A0} \rangle \,, \tag{328}$$

and similarly for P_B. There is a factor 2 in comparison with the single detection rate R_A eq.(309) that derives from different conventions used (compare eq.(299) with eq.(309)).

The coincidence detection probability for a given time window will be the product of probabilities, whence the detection rate (or mean detection probability per window) should be written as follows

$$P_{AB} = \langle M_A M_B \rangle = \left\langle \left(I_A + I^A_{ZPF}\right) \left(I_B + I^B_{ZPF}\right) \right\rangle , \tag{329}$$

that should be equivalent to eq.(311) except for a factor 2, that derives also from different conventions. As in the derivation of eq.(328), the mean detection probability P_{AB} should be zero when the pumping is off, whence we get

$$\left\langle \left(I_{A0} + I^A_{ZPF}\right) \left(I_{B0} + I^B_{ZPF}\right) \right\rangle = 0. \tag{330}$$

In order to proceed it is convenient to write I_A and I_B in terms of the partial intensities as defined in eqs.(308) and we obtain

$$P_{AB} = \langle \left(I_{A0} + I^A_{ZPF} + I_{A1} + I_{A2}\right) \left(I_{B0} + I^B_{ZPF} + I_{B1} + I_{B2}\right) \rangle. \tag{331}$$

This expression may be simplified as follows. We shall write it as a sum of products whence, taking eq.(330) into account, we get

$$\begin{aligned} P_{AB} &= \langle \left(I_{A0} + I^A_{ZPF}\right) \left(I_{B1} + I_{B2}\right) \rangle + \langle \left(I_{A1} + I_{A2}\right) \left(I_{B0} + I^B_{ZPF}\right) \rangle \\ &\quad + \langle \left(I_{A1} + I_{A2}\right) \left(I_{B1} + I_{B2}\right) \rangle. \end{aligned}$$

We shall show below that the average of I_{A1} with everything else is zero, and the same is true for I_{B1}, whence these terms may be removed and we obtain

$$P_{AB} = \langle \left(I_{A0} + I^A_{ZPF}\right) I_{B2} \rangle + \langle I_{A2} (I_{B0} + I^B_{ZPF}) \rangle + \langle I_{A2} I_{B2} \rangle.$$

The average $\langle I_{A2} I_{B2} \rangle$ gives a contribution of order $|D|^4$ that we may neglect. Also, we shall take into account that the two ZPF contributions arriving at Alice and Bob are uncorrelated with the signals; whence we get

$$\begin{aligned} P_{AB} &= \langle I_{A0} I_{B2} \rangle + \left\langle I^A_{ZPF} \right\rangle \langle I_{B2} \rangle + \langle I_{A2} I_{B0} \rangle + \langle I_{A2} \rangle \left\langle I^B_{ZPF} \right\rangle \\ &= \langle I_{A0} I_{B2} \rangle - \langle I_{A0} \rangle \langle I_{B2} \rangle + \langle I_{A2} I_{B0} \rangle - \langle I_{A2} \rangle \langle I_{B0} \rangle , \end{aligned} \tag{332}$$

where we have taken eqs.(327) into account in the last equality. This agrees with the quantum prediction obtained in the WW formalism eq.(311), modulo the global factor 2 mentioned above.

The proof that the terms I_{A1} and I_{B1} do not enter in the final result follows. The two terms $\langle I_{A1} \left(I_{B0} - \langle I_{B0} \rangle\right) \rangle$ and $\langle I_{A1} I_{B2} \rangle$ are nil. In fact, the partial 'intensities' I_{A1} and I_{B1} eqs.(308) consist of sums of terms having either the form $a_j a_l$ or $a^*_i a^*_k$, that is, products of two amplitudes with 'star' or two without 'star'. Therefore, the products involving I_{A1} times $(I_{B0} - \langle I_{B0} \rangle)$ and I_{A1} times I_{B2} will give no contribution to the average values because the products will contain a different number of amplitudes with and without 'star'. Similarly for the products of I_{B1} times $(I_{A0} - \langle I_{A0} \rangle)$ and I_{B1} times I_{A2}.

A more extended discussion is needed in order to prove that the average $\langle I_{A1} I_{B1} \rangle$ does not contribute. In fact, writing it in terms of fields we have

$$\langle I_{A1} I_{B1} \rangle = \left\langle \left(E^+_{A0} E^-_{A1} + E^+_{A1} E^-_{A0}\right) \left(E^+_{B0} E^-_{B1} + E^+_{B1} E^-_{B0}\right) \right\rangle . \tag{333}$$

Performing the product there are 4 terms that we may calculate using again the property of the product of four Gaussian variables. The

first term leads to a null result, namely

$$\left\langle E^+_{A0}E^-_{A1}E^+_{B0}E^-_{B1}\right\rangle = \left\langle E^+_{A0}E^-_{A1}\right\rangle\left\langle E^+_{B0}E^-_{B1}\right\rangle + \left\langle E^+_{A0}E^+_{B0}\right\rangle\left\langle E^-_{A1}E^-_{B1}\right\rangle + \left\langle E^+_{A0}E^-_{B1}\right\rangle\left\langle E^-_{A1}E^+_{B0}\right\rangle = 0,$$

because all the averages are zero, as proved taking eqs.(296) into account. Similarly $\left\langle E^+_{A1}E^-_{A0}E^+_{B1}E^-_{B0}\right\rangle = 0$. The two remaining terms coming from eq.(333) are complex conjugate of each other whence we may write

$$\langle I_{A1}I_{B1}\rangle = 2Re\left\langle E^+_{A0}E^-_{A1}E^+_{B1}E^-_{B0}\right\rangle. \tag{334}$$

If we evaluated this expectation using the Gaussian property as previously, we would get

$$\left\langle E^+_{A0}E^-_{A1}E^+_{B1}E^-_{B0}\right\rangle = \left\langle E^+_{A0}E^-_{A1}\right\rangle\left\langle E^+_{B1}E^-_{B0}\right\rangle + \left\langle E^+_{A0}E^+_{B1}\right\rangle\left\langle E^-_{A1}E^-_{B0}\right\rangle + \left\langle E^+_{A0}E^-_{B0}\right\rangle\left\langle E^-_{A1}E^+_{B1}\right\rangle = \left\langle E^+_{A0}E^+_{B1}\right\rangle\left\langle E^-_{A1}E^-_{B0}\right\rangle + \left\langle E^+_{A0}E^-_{B0}\right\rangle\left\langle E^-_{A1}E^+_{B1}\right\rangle,$$

the first term being null. The last two terms consist of a product of two averages each. Every average consists of a field amplitude arriving at Alice times another amplitude arriving at Bob. In these conditions we cannot ignore the phases of the spacetime factors like $\exp\left[i\mathbf{k}\cdot\mathbf{r} - i\omega t\right]$, see eq.(291), that would be different in the Alice and Bob beams, and uncorrelated. Therefore, it is plausible that the average over these phases should result in a null value for the expectation whence the average eq.(334) will be zero, thus finishing the proof that the term $\langle I_{A1}I_{B1}\rangle$ of eq.(329) does not contribute. I point out that the phase correlation condition is quite different in the other terms of eq.(329) because they involve absolute values, making the spacetime phases irrelevant. This is the case for instance in the average $\langle I_{A0}I_{B2}\rangle = \left|\left\langle E^+_{A0}E^-_{B1}\right\rangle\right|^2$.

6.6.3. Local realism weaker than Bell's. The interpretation of SPDC experiments in terms of stochastic processes, including the vacuum fields, allows local models violating a Bell inequality. This contradicts the wide consensus that Bell's is the unique local realistic formalism appropriate for experiments measuring correlations between distant parties.

There are several possible criticisms to our approach. Firstly our final eq.(332) (or (311)) may seem strange because apparently it does not fulfil the condition that the detection depends on the total radiation arriving at the detector. However, the physical basis is eq.(331) that obviously does fulfil the condition. Another possible problem is

the lack of positivity of the functions M_A and M_B, with obvious interpretation as probabilities in a single time window. Indeed, in order to obtain an average probability we should sum only (positive) probabilities. The possible lack of positivity is caused by the fluctuations, but we may assume that fluctuations will contribute only slightly if the integration time T in eq.(326) is large enough. This fact would be clear using the threshold detection rule. This rule is more plausible as said in section 6.2.6, but it would give a cumbersome calculation. Therefore, we may assume that M_A and M_B are positive with fairly good approximation.

Another objection to our approach is that, if M_A and M_B are positive and we use eqs.(328) and (329) in order to obtain the detection probabilities, our starting point would look identical to Bell's proposal, see chapter 3 section 3.3.3. If this was the case our model should fulfil the Bell inequalities and could not agree with quantum predictions. I believe that the introduction of the vacuum field effects produces a substantial difference. In fact, in Bell proposal the correlations derived from the source, but in our approach the vacuum ZPF contribution is essential. In any case, a more accurate treatment involving many radiation modes, not just two, is worthwhile.

As a conclusion, our approach suggests that the WW formalism might provide an interpretation of experimental tests of Bell inequalities that agrees with both the standard quantum predictions and a local realistic view of nature. But carrying on the program presents difficulties and further work is needed in order to clarify the matter.

Bibliography

[1] P. Grangier, G. Roger, A. Aspect: Experimental evidence for a photon anticorrelation effect on a beam splitter: a new light on single-photon interferences'. *Europhys. Lett.* **1**, 173 (1986).

[2] T. W. Marshall, E. Santos: Stochastic optics: A reaffirmation of the wave nature of light'. *Found. Phys.* **18**, 185-223 (1988).

[3] A. Casado, T. W. Marshall, E. Santos: Parametric downconversion experiments in the Wigner representation. *J. Opt. Soc. Amer. B* **14**, 494-502 (1997).

[4] A. Casado, A. Fernández-Rueda, T. W. Marshall, R. Risco-Delgado, E. Santos: Fourth order interference in the Wigner representation for parametric downconversion experiments. *Phys. Rev. A* **55**, 3879-3890 (1997).

[5] A. Casado, A. Fernández-Rueda, T. W. Marshall, R. Risco-Delgado, E. Santos: Dispersion cancellation and quantum eraser experiments analyzed in the Wigner function formalism. *Phys. Rev. A* **56**, 2477-2480 (1997).

[6] A. Casado, T. W. Marshall, E. Santos: Type II parametric downconversion in the Wigner formalism. Entanglement and Bell's inequalities. *J. Opt. Soc. Amer. B* **15**, 1572-1577 (1998).

[7] A. Casado, T. W. Marshall, E. Santos: Parametric down and up conversion in the Wigner representation of quantum optics. *J. Mod. Opt.* **47**, 1273-1287 (1999).

[8] A. Casado, T. W. Marshall, R. Risco-Delgado, E. Santos: Spectrum of the parametric down converted radiation calculated in the Wigner function formalism. *Eur. Phys. J. D* **13**, 109-119 (2001).

[9] A. Casado, S. Guerra, J. Plácido: Wigner representation for experiments on quantum cryptography using two-photon polarization entanglement produced in parametric down-conversion. *J. Phys. B: At. Mol. Opt. Phys.* **41**, 045501 (2008).

[10] A. Casado, S. Guerra, J. Plácido: Partial Bell-state analysis with parametric down conversion in the Wigner function formalism. *Adv. Math. Phys.* **2010**, 501521 (2010).

[11] A. Casado, S. Guerra, J. Plácido: Wigner representation for polarization momentum hyperentanglement generated in parametric down-conversion, and its application to complete Bell-state measurement. *Eur. Phys. J. D* **68**, 338-348 (2014).

[12] A. Casado, S. Guerra, J. Plácido: Wigner representation for entanglement swapping using parametric down conversion: the role of vacuum fluctuations in teleportation. *J. Mod. Opt.* **62**, 377-386 (2015).

[13] A. Casado, S. Guerra, J. Plácido: Rome teleportation experiment analysed in the Wigner representation: the role of the zero-point fluctuations in complete one-photon polarization-momentum Bell-state analysis. *J. Mod. Opt.* **65**, 1960-1974 (2018).

[14] K. Dechoum, T. W. Marshall, E. Santos: Parametric up and down conversion in the Wigner representation of quantum optics. *J. Mod. Opt.* **47**, 1273-1287 (2000).

[15] A. Casado, A. Fernández-Rueda, T. W. Marshall, J. Martínez, R. Risco-Delgado, E. Santos: Dependence on crystal parameters of the correlation time between signal and idler beams in parametric down conversion calculated in the Wigner representation. *Eur. Phys. J. D* **11**, 465-472 (2000).

[16] Z. Y. Ou, L. J. Wang, L. Mandel: Evidence for phase memory in two-photon down conversion through entanglement with the vacuum. *Phys. Rev. A* **41**, 566 (1990).

[17] A. Heuer, R. Menzel, P. W. Milonni: Cmplementarity in biphoton generation with stimulated or induced coherence. *Phys. Rev. A* **92**, 033834 (2015).

[18] A. Heuer, R. Menzel, P. W. Milonni: Induced coherence, vacuum fields and complementarity in biphoton generation. *Phys. Rev. Lett.* **114**, 053601 (2015).

[19] R. Menzel, A. Heuer, P. W. Milonni: Entanglement, complementarity, and vacuum fields in spontaneous parametric down-conversion. *Atoms* **7**, 27 (2019).

[20] X. Y. Zou, L. J. Wang and L. Mandel: Induced coherence and indistinguishability in optical interference. *Phys. Rev. Lett.* **67**, 318 (1991).

[21] D. M. Greenberger, M. A. Horne and A. Zeilinger: Multiparticle interferometry and the superposition principle. *Physics Today* **46**, 8, 22 (1993).

[22] L. K. Shalm et al.: A strong loophole-free test of local realism. *Phys. Rev. Lett.* **115**, 250402 (2015).
[23] M. Giustina et al.: A significant loophole-free test of Bell's theorem with entangled photons. *Phys. Rev. Lett.* **115**, 250401 (2015).
[24] E. Santos: Local model of entangled photon experiments compatible with quantum predictions based on the reality of the vacuum fields. *Found. Phys.* **50**, 1587-1607.

CHAPTER 7

Quantum effects in astrophysics and cosmology

7.1. Introduction

Quantum effects are relevant in several areas of astrophysics and cosmology. This chapter is devoted to study some of them, which may be helpful for the interpretation of quantum theory.

Our understanding of astrophysics and cosmology rests on general relativity. It is commonly believed that a good treatment of phenomena whose interpretation involves also quantum mechanics requires a quantum gravity theory. Thus, section 7.2 is devoted to present a personal view about the possible approach to a unified theory of general relativity and quantum mechanics.

As said in chapter 1 our interpretation of quantum theory involves the hypothesis that the vacuum fields are real stochastic fields. Therefore, most of the present chapter will be devoted to the influence that I believe vacuum fields have in astrophysics. As an introduction, the vacuum fields of quantum electrodynamics are briefly revisited in section 7.3.

The possible effects of the quantum vacuum will be studied in 3 areas. In section 7.4 it is proposed that 'dark energy' is actually an effect of the quantum vacuum fluctuations. In section 7.5 we argue that the vacuum fields may be also the origin of the phenomena attributed to 'dark matter', in particular the observed flat rotation curves of stars and gas in the haloes of galaxies.

Compact stars, like white dwarves, supermassive stars and neutron stars, may have a mass-to-radius ratio so large that, according to the standard opinion, they would eventually collapse. In contrast, ordinary stars, like the Sun, and planets are far from gravitational instability and quantum theory plays a minor role for them, reduced to explaining their equations of state. Therefore, our main interest is the possible quantum effects on the eventual collapse of compact stars. Section 7.6 is devoted to that matter.

7.2. General relativity

7.2.1. The theory of curved spacetime. General relativity (GR) proposes that spacetime is a manifold with intrinsic curvature. In the following I present a brief summary of its fundamental concepts; see any book on GR, such as [**1**] or [**2**], for details.

The analogy with a surface is useful. A sphere has intrinsic curvature, but the curvature of a cylinder is not intrinsic, it makes reference to the three-dimensional space. We may make a cylinder by folding a

sheet of paper, which is not possible for a sphere, as is well known from maps of large regions in Earth. The intrinsic curvature of a sphere is said positive and it is measured by a single number, the Gauss curvature, which is the inverse of its squared radius, $1/R^2$. There are surfaces with negative intrinsic curvature, the typical example being a saddle. For any surface we may consider a small region around a point and identify it with a region of either a sphere or a saddle; that would define the intrinsic curvature at that point.

The generalization to varieties of N dimensions was done by Riemann. Defining the curvature of a variety requires many numbers, 14 for the spacetime of 4 dimensions. Once a system of coordinates $\left\{x^1, x^2, x^3, x^4\right\}$ has been chosen, curvature may be derived from the metric

$$ds^2 = \sum_\mu \sum_\nu g_{\mu\nu}\left(x^1, x^2, x^3, x^4\right) dx^\mu dx^\nu, \tag{335}$$

that allows calculating a kind of distance ds between two near points. In eq. (335) I have included the $\sum$ symbols for clarity, but it is standard practice to omit them. The sum over repeated indices, without including the $\sum$ symbol, is named Einstein convention and I shall use it in the following, with few exceptions. Combining elements of the metric tensor $g_{\mu\nu}$, and their first and second derivatives with respect to the coordinates, it is possible to get the Riemann curvature tensor $R_{\mu\nu\lambda\sigma}$, which has $4^4 = 256$ components. However, there are a number of relations amongst them so that only 14 are independent, as said above.

Curvature is related to the basic properties of matter like energy, momentum and angular momentum, and the relation is given by the Einstein equation. Or, rather, we might say that Einstein's equation is just a 'translator' from the physical language (velocity, energy, etc.) to the geometrical one (world line, curvature). Einstein equation of general relativity is a relation between tensors that may be written (with the notation of the book of Weinberg [**1**]) as

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \equiv G_{\mu\nu} = -8\pi G T^{matt}_{\mu\nu}, \tag{336}$$

The left side $G_{\mu\nu}$, named Einstein tensor, combines the Ricci tensor $R_{\mu\nu}$ and the curvature scalar R, that may be got from the Riemann tensor, that is,

$$R_{\mu\nu} = \sum g^{\lambda\sigma} R_{\lambda\mu\sigma\nu}, \qquad R = \sum g^{\mu\nu} R_{\mu\nu}. \tag{337}$$

The right side of eq. (336) is the matter contents, $T_{\mu\nu}^{matt}$ being the stress-energy tensor. If there is no curvature, i.e. in Minkowski space, both $R_{\mu\nu}$ and R are zero, which would imply $T_{\mu\nu}^{matt} = 0$ according to eq. (336). That is, matter 'cannot live' in Minkowski space.

The relation between the geometry of spacetime and matter determines the fundamental laws of physics like conservation of energy, momentum and angular momentum, and it governs the motion of particles, that are compelled to travel along geodetic lines in 4 dimensions if free, and with appropriate motion if forces (e.g. electromagnetic) act on them. I stress again that in my view (shared with many people including Einstein, I believe) gravity is not a force but an effect of the curvature of spacetime on the motion of bodies.

In practice, solving eq.(336) in order to determine the metric tensor $g_{\mu\nu}\left(x^1, x^2, x^3, x^4\right)$ requires knowledge of the stress-energy tensor $T_{\mu\nu}^{matt}$ as a function of the coordinates or, more usually, some relations between its components (e.g. between energy density and pressure, what is called *equation of state*). Finding the solution of eq. (336) is rather difficult and only a few simple cases may be solved exactly. On the other hand, approximations are problematic due to the nonlinear character of the equations. For a small region all physics may be studied locally as in flat, Minkowski space, locally meaning an infinitesimal region around a point of spacetime. In practice this is the case in laboratory physics where typical sizes are small in comparison with astronomical distances. Of course, passing from an infinitesimal to a finite region is not straightforward. This is the case, for instance, in the study of quantum fields in curved space.

The belief that GR is fundamentally a theory of spacetime is standard, although some authors seem reluctant to accept it. I go further and believe that *GR is exclusively a theory of spacetime and its relation with matter*. It is not a theory of gravity because gravity is not a force. As an example let us consider the motion of Earth around the Sun. People living on Earth do not experience any force of attraction from the Sun because that force is balanced by the centrifugal force (this is correct in the center of Earth, and approximately true on the surface). Of course, any physics teacher would instruct students that there is an important difference, the Sun attraction is a true force whilst centrifugal force is fictitious, it looks like a force because we are observing from Earth, whose motion is not inertial. However, according to GR it is the case that the motion of the Earth *is inertial*, it corresponds to travelling along a geodetic in a curved spacetime.

Indeed, formulating physics in a curved spacetime is similar to studying physics in an accelerated reference frame. It is so similar that it was the starting point of GR, that Einstein proposed in 1907 as a '*principle of equivalence*'.

Gravitational force is fictitious like the centrifugal one. In other words, there are not 4 fundamental forces in nature, just 3: electromagnetic, weak and strong (or two if we consider the unification of the former two). Thus, quantization of gravity is quantization of spacetime, something that for a not educated observer might look bizarre because quantum phenomena do happen in spacetime.

There is one reason that has lead to treat gravity as the fourth force. It is the fact that in the absence of matter Einstein eq. (336) becomes

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \equiv G_{\mu\nu} = 0, \tag{338}$$

and this equation may have solutions aside the trivial one without curvature (i.e. Minkowski space). The most obvious example is the space curvature outside massive bodies like the Earth. This curvature may produce acceleration on particles and it is named gravitational field in the Newtonian approximation.

Other solutions of eq. (338) are the gravity waves. These waves are produced by motion of matter and travel through space in a form similar to electromagnetic waves produced by electric charges in motion; that is, they travel in the absence of charges. In summary, we cannot have matter without spacetime curvature, but we may have curvature without matter. In fact, flat (Minkowski) space implies $G_{\mu\nu} = 0$ whence Einstein eq. (336) gives $T_{\mu\nu} = 0$, but $G_{\mu\nu} = 0$ does not imply a nil Riemann tensor, that is, absence of curvature. Also, solutions of eq. (338) might exist in the form of fluctuations of the metric, that is, fluctuations of the elements $g_{\mu\nu}$ of eq. (335). I propose that such fluctuations may be a fundamental ingredient for the quantum behaviour [**3**], but this question will not be studied here further.

7.2.2. Is general relativity a field theory? The study of gravitational and electrostatic interactions started from very similar laws due to Newton and Coulomb. Ignoring their vector character, irrelevant here, and including the acceleration produced by the force

we may write, respectively

(339)
$$F_{Newton} = -G\frac{m_1 m_2}{r^2} = -m_2\frac{d^2x_2}{dt^2}, \quad F_{Coulomb} = \frac{q_1 q_2}{r^2} = m_2\frac{d^2x_2}{dt^2},$$

where m_j are masses, q_j are charges, and x_2 is the position of the accelerated particle.

But there are two differences between the two theories: firstly, gravity is always attractive, the mass being positive, but the electrostatic interaction may be either attractive or repulsive because there are charges q_j of either sign. The second difference is more subtle but very important. Newton law requires the introduction of a constant, G, but in Coulomb law the charges might be defined so that no constant is necessary. This is because the electric charge just plays the role of source of the interaction. However, in the first eq. (339) mass has two (apparently) quite different roles: it is both the source of gravity on the left side and a measure of inertia on the right side.

We may go further and ask what happens if we remove the constant also in front of the gravity equation putting $G = 1$. The result is that we are compelled to define mass as having the dimensions of length (if we measure times in units of length taking the velocity of light as unit, consistently with special relativity). This might be interpreted saying that mass is a property of space and GR tells us how to relate it with spacetime curvature.

Eqs. (339) suggest defining a gravitational field, $\mathbf{g}$, and an electric field, $\mathbf{E}$, whence it is possible to define a gravitational and an electrostatic energy, E, in a volume V, namely

$$E_{grav} = -\frac{G}{2}\int_V |\mathbf{g}|^2\, d^3r, \qquad E_{elect} = \frac{1}{2}\int_V |\mathbf{E}|^2\, d^3r.$$

Thus the energy may be believed as distributed in space, the material particles being just the sources of the fields. The concept of field was introduced by Faraday, extended by Maxwell to the whole electromagnetic theory and later applied to all known interactions. Maxwell's was a relativistic field theory, or rather it was the origin of relativity. Once special relativity was clearly formulated, by Einstein in 1905, it became standard belief that all fundamental fields of nature should be compatible with relativity theory. This fact suggested the need of a relativistic theory of gravity, and a big effort was devoted to search for it in the first few years of the 20th Century. But the dual role of mass

above mentioned made the new theory much more than just a relativistic field theory of gravity. The appropriate theory was found by Einstein, not as a field theory but as a theory of the intimate relation between spacetime and matter.

The question is whether Einstein's general relativity is really a field theory like Maxwell's electromagnetism. In my opinion it is not.

Actually, it is possible to formulate gravity as a field theory able to reproduce the predictions of Einstein's theory. The standard method is to rewrite the Einstein eq. (336) in a form that looks like a field equation of gravity (see e.g. the book of Weinberg [**1**], section 7.6). This approach may be useful for calculations, but it is artificial and it darkens the physics, in my opinion.

A different approach has been proposed by Logunov [**4, 5**] with the name 'relativistic theory of gravity'. He took the Einstein eq. (336) as starting point, so that most tested predictions of general relativity are reproduced. However, Logunov changes the interpretation: the tensor $g_{\mu\nu}\left(x^1, x^2, x^3, x^4\right)$ is not a metric tensor, but a tensor field of gravity, in analogy with Maxwell theory in terms of a vector field in 4 dimensions. Then both Logunov's gravitational field and Maxwell's field 'live' in a flat (Minkowski) spacetime with a metric

$$ds^2 = \sum h_{\mu\nu} dx_\mu dx_\nu,$$

where $h_{\mu\nu}$ is a tensor such that, in appropriate (Cartesian) coordinates, the metric may be written

$$ds^2 = dx_1 + dx_2 + dx_3 - dx_4.$$

Quantizing a field theory of gravity may be similar to quantizing the electromagnetic field or, say, the electron-positron field; but quantizing Einstein's 'general relativity' is quite different, as commented in more detail in the following.

7.2.3. General relativity and quantum mechanics. The unification of GR with QM is currently viewed as the most important open problem in theoretical physics. It is also common opinion that the solution would be to quantize GR. I may agree with the former sentence, but not fully with the latter. In particular, I do not believe that QM is more fundamental than GR. Of course, QM has had a spectacular success in the prediction of empirical facts, specially after the development of quantum electrodynamics. GR has had also relevant successes in explaining many observations, but it is well below QM in this regard and, more importantly, it has far smaller impact in

the technological development. I think that this is why most physicists believe that QM is the fundamental theory of nature, whilst GR is seen as a theory of gravity, this considered just one of the four fundamental forces of nature.

In contrast, I believe that GR is more fundamental than QM because it establishes the structure of spacetime, which is the framework for the whole physics. On the other hand, as explained in the previous chapters of this book, I support the view that quantization is just a mathematical formalism for the statistical description of the behaviour of matter. Indeed, I believe that the material world is so complex and our observational possibilities so limited that we should use a probabilistic approach. It is the case that the statistical properties of the quantum world are peculiar and demand an appropriate formalism to characterize them. That formalism seems to be Hilbert space (or C^*algebras for systems with infinitely many degrees of freedom).

The view of nature offered by general relativity is as follows. The material world consists of a four-dimensional manifold with intrinsic curvature, that we name spacetime. We believe, with Einstein, that physics does not deal with the evolution of a three-dimensional object, but with a four-dimensional object. That is, time is a coordinate with a role not too different from the space coordinates. The truly special role of time for us, human beings, is an anthropic prejudice, not a fundamental physical fact, as discussed in more detail in chapter 3, section 3.6. Therefore, I will treat time and space in the same footing, using natural units $c = 1$ and speaking about length with the meaning of either spacial length or time interval.

Attached to points of spacetime there are fields, e.g. scalar, vector or spinor. Field theory provides differential equations that relate the values of the fields in different points. Thus, constructing a field theory requires a previously given spacetime. In the most simple case we assume Minkowski space and this has been the common choice for the development of (quantum) electrodynamics first and the whole relativistic (quantum) field theory later on.

General relativity allows connecting the fields with the spacetime curvature via three steps. Firstly there are equations that provide the mass-energy contents of the fields in the form of a stress-energy tensor, although they are usually formulated in Minkowski space, which is an approximation. Then Einstein equation of GR relates that tensor with

the Einstein tensor. Finally the mathematical theory of Riemann relates the Einstein tensor with the metric tensor. In summary, general relativity plus field theory allows connecting field amplitudes with the metric tensor via 3 steps.

I believe that we make conceptual progress if we reduce the steps from 3 to 2. We cannot remove the equations of the first and the last steps, both being rather involved. However, it is very easy to remove the second step because Einstein equation is conceptually very simple, it is just an equality of two tensors modulo the Newton constant G. It is enough to take the Einstein equation as a kind of 'dictionary' that just translates an alleged physical property (stress-energy tensor) to geometrical language (Einstein tensor). A dimensional (Newton) constant is needed because the units used in either metric or Einstein tensors (i.e. combinations of lengths) are different from the units used for the stress-energy tensor (involving mass-energy). With this reduction from 3 to 2 steps we arrive at interpreting mass-energy, momentum and angular momentum as geometrical properties of spacetime. Indeed, the equations that relate the values of the fields at different points are strongly constrained by the need that the Einstein tensor derived from them represents a possible curvature, that could be described by a metric tensor. The constraints lead in particular to the conservation laws of energy, momentum and angular momentum.

Up to here everything has been conceptually simple because we have related classical physics with GR, that may be labeled classical too. The problem appears if the fields are quantum, because then the two theories (QM and GR) do not match. If the equations of quantum field theory were interpreted as stochastic, the solution of the problem would be straightforward. We should consider a probability distribution of Einstein tensors corresponding to the distribution of stress-energy tensors. The big difficulty is that quantum distributions are hidden behind a formalism, say Hilbert spaces, that seem appropriate for the peculiar stochastic processes involved. However, deciphering the stochastic contents of the quantum formalism appears as extremely difficult to the point that many (or most) physicists believe that it is not possible, that quantum theory is not a stochastic theory. This book is precisely devoted to provide some ideas in support of the opposite belief.

7.2.4. How to quantize general relativity. I stress again that in my view GR is not a field theory of gravity because gravity is a

fictitious force, as the centrifugal force. Quantization of GR might be just a procedure to allow characterizing a stochastic spacetime. The problem is that we do not know how to do that. We only know the rules for the quantization of linear fields, and this quantization is made in Minkowski space, which is necessarily an approximation because the existence of fields is not compatible with flat spacetime. I believe that only after we understand quantization as a stochastic treatment of the fields we could go beyond the standard purely formal quantization rules.

Shelving the task of interpreting quantum theory as a stochastic theory we might start quantizing GR by considering that the elements of the Einstein tensor are operators similar to the operators of the stress-energy tensor, whence also the elements of the metric tensor might be got as quantum operators. The problem is that the relation between the Einstein tensor and the metric tensor, provided by Riemann geometry, is not linear; whence a difficulty arises with the ordering of the metric operators. An additional obstacle is the existence of solutions of eq. (338) (without matter contents), whence quantization of spacetime is not just a consequence of a possible quantum Einstein eq. (336) after quantizing the stress-energy tensor.

The difficulty of getting appropriate quantum operators for the metric tensor elements is one of the difficulties to find a 'quantum gravity theory', but it is not the only one. The main problem derives from the fact that Einstein eq. (336) is not linear in the metric tensor components. This precludes a straightforward quantization, achieved e.g. expanding a field in plane waves and promoting the amplitudes of the waves to (creation and annihilation) operators, a procedure pioneered by Dirac for the electromagnetic field in 1927. (As an illustration of the quantization procedure see section 7.3.1 below, dealing with the quantum electromagnetic field).

The early Dirac quantization method evolved to the canonical method via commutation or anticommutation relations of conjugate operators, or via path integral quantization. In the former approach one starts with a variational formulation of the evolution in terms of a Lagrangian density involving field amplitudes and their derivatives with respect to the coordinates. Hence, canonical momentum densities may be derived and then promoted to operators (see any book of quantum field theory). In this approach the integral of the Lagrangian density becomes the action in the path integral, whence the propagation of the fields may be obtained (for a nonrelativistic example see

chapter 4). In GR the action appears in the Hilbert variational formulation. The problem is that the quantized theory is not renormalizable; it gives rise to divergences that cannot be eliminated so that no finite predictions may be obtained. Many attempts have been made to devise modifications that might produce a renormalizable gravity theory, without too much success till now. In my opinion, this failure is a hint that general relativity cannot be quantized in the sense commonly understood. Indeed, according to the interpretation supported in this book quantization is just the substitution of random variables or stochastic processes for classical quantities. More specifically, the evolution might be deterministic in principle, but we cannot know the states of the fields with accuracy, whence we must treat them as random, or stochastic, fields. Quantization is the characterization of these stochastic fields, that should be made consistently for all fields. Indeed, the basic idea of the present book is that quantization has an epistemological, rather than ontological, meaning. It is impossible for us to have complete information about phenomena, specially in the microworld, and our approach would require a probabilistic, stochastic, treatment. That is quantization: the substitution of appropriate random variables for deterministic variables in the study of the physical world.

7.2.5. Extended gravity theories. General relativity has passed all observational tests so far, but the real theory of gravity may well differ from it in some special cases. In fact, the difficulties in quantizing Einstein's theory and astrophysical observations not yet explained in a satisfactory way have led to the study of extended gravity theory. Some of those theories have been proposed as an explanation for the observable effects attributed to the hypothetical dark energy and dark matter or as a possible mechanism to avoid the star collapse leading to a singularity (i.e. to a black hole). I shall deal with these three applications in sections 7.4, 7.5 and 7.6 respectively.

Extended gravity theories [**6**] usually rest upon the same fundamental idea of GR, namely that gravity is, or is associated to, curvature of spacetime. Thus the modification of GR consists of changing the left side of Einstein eq. (336). The change should preserve the symmetry properties of the theory, that is, covariance under arbitrary changes of coordinates. The standard method to achieve this is to formulate GR in terms of the Hilbert action in the form

$$S = \frac{1}{16\pi G}\int d^4x\sqrt{-g}R + S_{matter}, \tag{340}$$

where g is the determinant of the metric tensor, that for a diagonal tensor would be just the product of the four nonzero elements. The integral extends over a region of spacetime. It may be shown that equating to zero the variation of the action leads to Einstein equation [**1**], [**2**]. In extended gravity theories a function f of the 3 possible scalars that may be obtained from the Riemann tensor via contractions is added to R. That is,

$$f = f\left(R, R_{\mu\nu}R^{\mu\nu}, R_{\lambda\mu\sigma\nu}R^{\lambda\mu\sigma\nu}\right). \tag{341}$$

Two important particular cases are $f(R)$-gravity, where f is only a function of R [**12**], and the most general quadratic Lagrangian, that may be written without loss of generality (with the summation of repeated indices assumed) as

$$f = aR^2 + bR_{\mu\nu}R^{\mu\nu}. \tag{342}$$

Here Riemann square does not appear because it may be eliminated using the Gauss-Bonnett combination

$$R_{GB} \equiv R^2 - 4R_{\mu\nu}R^{\mu\nu} + R_{\lambda\mu\sigma\nu}R^{\lambda\mu\sigma\nu},$$

whose presence in the Hilbert action would not contribute to the field equations. The extended gravity theory should behave appropriately in the low gravity limit so that all terrestrial and Solar System observations are not contradicted, which is a strong constraint. For instance, in the quadratic Lagrangian, eq. (342), the parameters a and b should be at most a few millimeters (but see section 7.6.3).

Up to here we have interpreted the function f eq. (341), added to R in the Hilbert action eq. (340), as a modification of GR; that is, as a new theory resting on a curvature tensor $G_{\mu\nu}$ different from the tensor in GR, which is the left side of eq. (336). However, we may pass the function f eq. (341) to the right side with changed sign and interpret it as a contribution to the stress-energy tensor, possibly associated to the vacuum fields. With this interpretation Einstein eq. (336) should be written

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R \equiv G_{\mu\nu} = -8\pi G(T^{matt}_{\mu\nu} + T^{vac}_{\mu\nu}). \tag{343}$$

The vacuum stress-energy tensor could be obtained from the function f via the variational eq. (340). It is a function of the elements $g_{\mu\nu}$ of the metric tensor and their derivatives with respect to the coordinates. This means that in this approach we are assuming that *the vacuum stress-energy tensor depends on the curvature* of spacetime via the metric, see eq. (335). The argument may be also stated as

follows. Let us assume that the quantum vacuum contributes a stress-energy tensor that depends on the curvature of spacetime. Then if we want that such dependence respects the symmetries of the said curved space, then it is appropriate to derive it via the Hilbert eq. (340).

The calculation of $T^{vac}_{\mu\nu}$ eq. (343) with f eq. (341) substituted for R in eq. (340) is straightforward but lenghty. For instance, see the result for the quadratic Lagrangian eq. (342) in Ref. [**13**].

7.3. The vacuum fields of quantum electrodynamics

This section is devoted to an illustrative calculation of the energies and pressures of the fields involved in quantum electrodynamics (QED), that is, the free electromagnetic and positron-electron fields. A sketch of the interaction of both is included for completeness. The purpose is to introduce some properties of the vacuum fields that will be useful in sections 7.4 and 7.5.

7.3.1. Zeropoint energy, divergences and regularizations. As a result of the quantization of fields a zero-point energy appears, as mentioned in chapter 1. If the field is expanded in normal modes (plane waves in in the case of free space) the zero-point field (ZPF) contributes an energy $\frac{1}{2}\hbar\omega$ per normal mode, ω being the frequency. The problem is that this leads to big (quartic) ultraviolet divergence of the ZPF energy. For instance, if we perform the sum for all normal modes of the electromagnetic radiation quantum field, we would get an infinite total energy in any finite volume, that is, an energy density as follows

$$\rho_{EM} = \frac{1}{V}\sum_{\mathbf{k},s}\frac{1}{2}\hbar\omega \quad \rightarrow \quad \frac{1}{8\pi^3}\int_0^{k_{max}} \hbar\omega d^3k = \frac{\hbar\omega^4_{max}}{8\pi^2c^2}, \tag{344}$$

where $\mathbf{k}$ is the wave-vector of a plane wave, $|\mathbf{k}| = k = \omega/c$, we sum over 2 polarizations, and ω_{max} is a cut-off frequency.

In most calculations of quantum mechanics the problem of the divergence is solved in practice with the 'normal ordering rule', that is, putting the annihilation operators to the right as described in chapter 1. The rule amounts to subtracting effectively the ZPF or, equivalently, to fixing the zero of energies at the level of the vacuum. Indeed, usually only energy differences have a physical meaning and the zero may be fixed arbitrarily, *except* if gravity is involved. Also the ZPF is modified by the presence of matter (e.g., an atom) producing radiative corrections (e.g., the Lamb shift).

If we take the electromagnetic vacuum energy eq. (344) as real then a natural cutoff appears at the Planck scale, due to gravitational effects, defined in terms of the universal constants $c, \hbar, G$. Thus, if we choose

$$E_{max} = \frac{1}{2}\hbar\omega_{max} = E_{Planck} = \sqrt{\frac{2\hbar}{G}}c^2 = 2.0 \times 10^6 joule,$$

in eq. (344) we would get an energy (or mass) density

$$\rho_{EM} = \frac{1}{8\pi^2}\rho_{Planck}, \qquad \rho_{Planck} = \frac{c^5}{G^2\hbar} \simeq 10^{97} kg/m^3, \tag{345}$$

a huge value indeed. The electromagnetic zero-point pressure is 1/3 the density in units $c = 1$, that is,

$$\rho_{EM} = \frac{E^4_{max}}{8\pi^2\hbar^3}, \qquad P_{EM} = \frac{E^4_{max}}{24\pi^2\hbar^3}. \tag{346}$$

Similarly, we may calculate the energy (or mass) density ρ and the pressure P of other fields, e.g. for the free electron-positron (Dirac) field we have

$$\rho_{D0} = -\frac{1}{V}\sum_{\mathbf{k},\varepsilon} E \quad \to \tag{347}$$

$$-(2\pi)^{-3}\sum_s \int_0^{p_{\max}} E d^3p = -\pi^{-2}\int_0^{p_{\max}} \sqrt{m^2+p^2}\,p^2 dp$$

$$= -\frac{p^4_{\max}}{4\pi^2} - \frac{p^2_{\max}m^2}{4\pi^2} + \frac{m^4}{8\pi^2}\ln\left(\frac{p_{\max}}{m}\right) + O\left(p^{-2}_{\max}\right),$$

$$P_{D0} = -\frac{1}{V}\sum_{\mathbf{k},\varepsilon}\frac{p^2}{3E} \quad \to \tag{348}$$

$$-\frac{1}{3\pi^2}\int_0^{p_{\max}} \left(m^2+p^2\right)^{-1/2} p^4 dp$$

$$= -\frac{p^4_{\max}}{12\pi^2} + \frac{p^2_{\max}m^2}{12\pi^2} - \frac{m^4}{8\pi^2}\ln\left(\frac{p_{\max}}{m}\right) + O\left(p^{-2}_{\max}\right),$$

where s is the polarization and I have used units $\hbar = c = 1$, as will be made in the rest of this section, although I shall include these parameters sometimes for clarity. In the subindices, D stands for Dirac and 0 for free field. In contrast with the positive sign of the density and the pressure of the electromagnetic field, eqs. (344) and (345), that were a consequence of the commutation relations of creation and annihilation operators (for Bose fields), now the sign is negative due

to the fact that the field operators anticommute (for Fermi fields). For all massive free fields the ZPF energy and pressure are similar to eqs. (347) and (348) except that the sign is opposite for Bose fields. In all cases there is a quartic divergence, another quadratic one, and finally a logarithmic one. In addition to the free fields, the contributions of the field interactions should be taken into account. With a cutoff at the Planck scale, the energies and pressures of the vacuum fields are of the order of eq.(345).

A popular procedure to deal with the divergences in the calculation of the vacuum energy is the regularization of the infinities (see e.g. [**7**]). Actually, the cutoffs leading to eqs. (344) to (348) are not covariant, with the consequence that they do not provide a Lorentz invariant 'vacuum equation of state', that would correspond to density and pressure related by $P = -\rho$. For instance, for the free electromagnetic field one obtains a 'radiation equation of state' where pressure and density are related by $P = 1/3\rho$, which is also the case for the leading term of the electron-positron field. In contrast, Lorentz covariant regularizations, in addition to providing a vacuum equation of state, may lead to a decrease of the divergence from quartic in the particle energies to logarithmic. More specifically, in a covariant regularization the terms with quartic and quadratic divergence of eqs. (347) and (348) usually do not appear and the leading terms are the logaritmically divergent ones. For instance, in the free electron-positron field

$$\rho_{DO} = \frac{1}{8\pi^2} m^4 \ln\left(\frac{p_{\max}}{m}\right), \qquad P_{DO} = -\frac{1}{8\pi^2} m^4 \ln\left(\frac{p_{\max}}{m}\right),$$

that correspond to a vacuum equation of state $P = -\rho$. With a cutoff at the Planck scale, the energy density and the pressure become of order $10^{20} kg/m^3$, still very large but much smaller than eq. (345). In Minkowski space the vacuum equation of state $P = -\rho$ is satisfactory because it is Lorentz invariant. But see the comment at the beginning of section 7.4.5.

Covariant regularizations gives rise to some difficulties, in particular the ambiguity derived from the fact that different regularizations may produce different results, even opposite sign for the regularized quantity [**7**], [**8**]. Another strange result is that regularization gives nil energy density for the radiation vacuum contribution, in sharp contrast with the fact that the integrals involved in the calculation of the energy density (similar to those used to get eq. (344)) have a positive definite integrand. In conclusion, covariant regularizations

are mathematical procedures useful for some purposes, in particular renormalization in QED. But I do not think that they provide an appropriate physical picture of the vacuum fields. Of course, a correct picture should involve *finite* energies only, which is not available at the moment. In any case, the most plausible assumption is that the mean energy and pressure of all vacuum fields taken together are both zero, as discussed in the next section 7.4.

In summary, the divergencies of the ZPF do not put any difficulty for calculations not involving gravity, that give finite results using renormalization techniques, but they present a big unsolved problem for calculations of gravitational effects. The standard view is that the problem could not be solved until we have a 'quantum gravity theory' unifying general relativity with quantum mechanics, see section 7.2 above. In the meantime we may assume that the contributions to the vacuum energy and pressure cancel. A hint for this possibility is that the most divergent terms in the vacuum free fields have opposite signs for bosons and fermions. Of course, there are also contributions due to field interactions.

Here I will not comment any more on the problem of divergencies. In the next three subsections I recall some aspects of the calculation of the energy density and pressure in the QED fields, that is, electromagnetic and electron-positron fields. These calculations are standard and well known, but I reproduce them as a reference for sections 7.4 and 7.5 respectively.

7.3.2. The free electromagnetic field. We will work in Minkowski space, although it is not a good approximation when the spacetime curvature is relevant. I will start calculating the stress-energy tensor of the free electromagnetic field, ignoring all other vacuum fields. Assuming homogeneity and isotropy of space, the stress-energy tensor is defined by just two parameters, energy density ρ and pressure P.

The quantization of a classical field involves promoting the amplitudes of the plane waves expansion to (creation or annihilation) operators. For instance, the classical free electromagnetic field, in the Coulomb gauge, may be represented by an expansion in plane waves

of the vector potential, $\mathbf{A}(\mathbf{r},t)$, that is,

$$\mathbf{A}(\mathbf{r},t) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k},\boldsymbol{\varepsilon}} \left(\frac{2\pi\hbar c}{k}\right)^{1/2} \times$$

$$\times \left[\alpha_{\mathbf{k},\boldsymbol{\varepsilon}} \boldsymbol{\varepsilon} \exp\left(i\mathbf{k}\cdot\mathbf{r} - ikt\right) + \alpha^*_{\mathbf{k},\boldsymbol{\varepsilon}} \boldsymbol{\varepsilon} \exp\left(-i\mathbf{k}\cdot\mathbf{r} + ikt\right)\right], \tag{349}$$

where $k = |\mathbf{k}|$. In the rest of this section I will use the notation of the book of Sakurai [**9**] and units $\hbar = c = 1$. From eq. (349) it is easy to get the electric field, $\mathbf{E} = -\partial\mathbf{A}/\partial t$, and the magnetic field, $\mathbf{B} = \bigtriangledown \times \mathbf{A}$. The polarization vector $\boldsymbol{\varepsilon}$ depends on $\mathbf{k}$ (in fact $\mathbf{k}\cdot\boldsymbol{\varepsilon} = 0$) whence there are two possible values so that we should write $\boldsymbol{\varepsilon}_j(\mathbf{k}), j = 1, 2$, but I will use a simplified notation except when some confusion might arise.

In the quantized field an annihilation operator $\hat{\alpha}_{\mathbf{k},\boldsymbol{\varepsilon}}$ is substituted for the amplitude $\alpha_{\mathbf{k},\boldsymbol{\varepsilon}}$, and a creation operator $\hat{\alpha}^\dagger_{\mathbf{k},\boldsymbol{\varepsilon}}$ for the amplitude $\alpha^*_{\mathbf{k},\boldsymbol{\varepsilon}}$; whence the electric and magnetic fields, $\widehat{\mathbf{E}}$ and $\widehat{\mathbf{B}}$, become vector operators. They may be written via an expansion in plane waves taking a quantized counterpart of eq. (349) into account. From these expansions, that I do not write explicitly, it is straightforward to obtain the energy density operator. We get

$$\hat{\rho}_{EM}(\mathbf{r},t) \equiv \frac{1}{2}\left(\widehat{\mathbf{E}}^2 + \widehat{\mathbf{B}}^2\right) = \frac{1}{4V} \sum_{\mathbf{k},\boldsymbol{\varepsilon}} \sum_{\mathbf{k}',\boldsymbol{\varepsilon}'} K_1 (\hat{\alpha}^\dagger_{\mathbf{k}',\boldsymbol{\varepsilon}'} \hat{\alpha}_{\mathbf{k},\boldsymbol{\varepsilon}} + \hat{\alpha}_{\mathbf{k},\boldsymbol{\varepsilon}} \hat{\alpha}^\dagger_{\mathbf{k}',\boldsymbol{\varepsilon}'})$$

$$+ \frac{1}{4V} \sum_{\mathbf{k},\boldsymbol{\varepsilon}} \sum_{\mathbf{k}',\boldsymbol{\varepsilon}'} \left(K_2 \hat{\alpha}_{\mathbf{k},\boldsymbol{\varepsilon}} \hat{\alpha}_{\mathbf{k}',\boldsymbol{\varepsilon}'} + K_2^* \hat{\alpha}^\dagger_{\mathbf{k},\boldsymbol{\varepsilon}} \hat{\alpha}^\dagger_{\mathbf{k}',\boldsymbol{\varepsilon}'}\right),$$

where the functions K_1 and K_2 are numbers (not operators) given by

$$K_1 = \sqrt{kk'}\,\boldsymbol{\varepsilon} * \boldsymbol{\varepsilon}' \exp\left[i\left(\mathbf{k} - \mathbf{k}'\right)\cdot\mathbf{r} - i\left(k - k'\right)t\right] \tag{350}$$

$$K_2 = \sqrt{kk'}\,\boldsymbol{\varepsilon} * \boldsymbol{\varepsilon}' \exp\left[i\left(\mathbf{k} + \mathbf{k}'\right)\cdot\mathbf{r} - i\left(k + k'\right)t\right], \tag{351}$$

For notational simplicity I have introduced the following 'star product'

$$\boldsymbol{\varepsilon} * \boldsymbol{\varepsilon}' \equiv \boldsymbol{\varepsilon}\cdot\boldsymbol{\varepsilon}' + \frac{1}{kk'}\left[(\mathbf{k}\times\boldsymbol{\varepsilon})\cdot(\mathbf{k}'\times\boldsymbol{\varepsilon}')\right], \tag{352}$$

which satisfies the properties

$$\sum_j (\boldsymbol{\varepsilon}_i * \boldsymbol{\varepsilon}'_j)(\boldsymbol{\varepsilon}'_j * \boldsymbol{\varepsilon}''_l) = \boldsymbol{\varepsilon}_i * \boldsymbol{\varepsilon}''_l, \qquad \sum_{ij} \boldsymbol{\varepsilon}_i * \boldsymbol{\varepsilon}_j = 4$$

$$\sum_{ij} (\boldsymbol{\varepsilon}_i * \boldsymbol{\varepsilon}'_j)^2 = 2\left[1 + \frac{\mathbf{k}\cdot\mathbf{k}'}{kk'}\right]^2. \tag{353}$$

It is convenient to write the terms of $\hat{\rho}_{EM}$ in normal order, that is, with the annihilation (creation) operators to the right (left). Taking the commutation rules of the field operators into account we get

$$\hat{\rho}_{EM} = \rho_{EM0} + \hat{\rho}_{EM1} + \hat{\rho}_{EM2}, \tag{354}$$

$$\rho_{EM0} = \frac{1}{2V} \sum_{\mathbf{k},\boldsymbol{\varepsilon}} k.$$

$$\hat{\rho}_{EM1} = \frac{1}{2V} \sum_{\mathbf{k},\boldsymbol{\varepsilon}} \sum_{\mathbf{k}',\boldsymbol{\varepsilon}'} K_1 \hat{\alpha}^{\dagger}_{\mathbf{k}',\boldsymbol{\varepsilon}'} \hat{\alpha}_{\mathbf{k},\boldsymbol{\varepsilon}},$$

$$\hat{\rho}_{EM2} = \frac{1}{4V} \sum_{\mathbf{k},\boldsymbol{\varepsilon}} \sum_{\mathbf{k}',\boldsymbol{\varepsilon}'} \left(K_2 \hat{\alpha}_{\mathbf{k},\boldsymbol{\varepsilon}} \hat{\alpha}_{\mathbf{k}',\boldsymbol{\varepsilon}'} + K_2^* \hat{\alpha}^{\dagger}_{\mathbf{k},\boldsymbol{\varepsilon}} \hat{\alpha}^{\dagger}_{\mathbf{k}',\boldsymbol{\varepsilon}'} \right),$$

where ρ_{EM0} is a numerical constant (times the unit operator).

The Hamiltonian is obtained by performing a space integral of the energy density eq. (354), that is,

$$\widehat{H}_{EM} = \lim_{V\to\infty} \int_V \hat{\rho}_{EM}\,(\mathbf{r})\, d^3\mathbf{r} = \sum_{\mathbf{k}\boldsymbol{\varepsilon}} k(\hat{\alpha}^{\dagger}_{\mathbf{k},\boldsymbol{\varepsilon}} \hat{\alpha}_{\mathbf{k},\boldsymbol{\varepsilon}} + \frac{1}{2}). \tag{355}$$

The integral removes the spacetime dependence and cancels the denominator V. It is interesting that the space integral also leads to $\mathbf{k}' = -\mathbf{k}$ in K_2 eq. (351); whence the term $\hat{\rho}_{EM2}$ does not contribute to the Hamiltonian nor, therefore, to the average energy density. It contributes only to the fluctuations above or below the mean, a relevant fact for our calculations in sections 7.4 and 7.5. In quantum language these fluctuations consist of the creation or annihilation of two virtual photons with zero total momentum, as shown by the field operators involved.

For the free electromagnetic field, the vacuum state $|0\rangle$ may be defined as the state with the minimal energy, that is, as the eigenvector of the operator eq. (355) with the smallest eigenvalue. It is a state with zero photons and it has the properties

$$\alpha_{\mathbf{k},\boldsymbol{\varepsilon}} \mid 0\rangle = 0, \qquad \langle 0 \mid \alpha^{\dagger}_{\mathbf{k},\boldsymbol{\varepsilon}} = 0.$$

The state $|0\rangle$ is different from the actual QED vacuum $|vac\rangle$, which takes the electron-positron field and the interaction into account. For the purely electromagnetic vacuum state $\mid 0\rangle$ the expectation of the energy density is

$$\langle 0\,|\hat{\rho}_{EM}|\,0\rangle = \rho_{EM0} = \frac{1}{V} \sum_{\mathbf{k},\boldsymbol{\varepsilon}} \frac{1}{2} \hbar k = \frac{1}{V} \sum_{\mathbf{k}} \hbar k, \tag{356}$$

where the last equality derives from the two possible polarizations. We have taken into account that the $\hat{\rho}_{EM1}$ term does not contribute to the expectation value because both the annihilation operator placed on the right and the creation operator on the left give zero when acting on the vacuum state. In the limit $V \to \infty$ we obtain eq. (344). I will not give a detailed calculation of the pressure; for the result see eq. (346).

In addition to density and pressure we might find the two-point correlations of vacuum energy density, the most relevant quantity for the study of the quantum fluctuations. I will define it at equal times after subtracting the mean density, that is

$$C(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{2} \langle 0 | \hat{\rho}(\mathbf{r}_1) \hat{\rho}(\mathbf{r}_2) + \hat{\rho}(\mathbf{r}_2) \hat{\rho}(\mathbf{r}_1) | 0 \rangle - \rho_{vac}^2, \tag{357}$$

where the fact that $\hat{\rho}(\mathbf{r}_1)$ does not commute with $\hat{\rho}(\mathbf{r}_2)$ has been taken into account. It is easy to see that ρ_{EM0} eq. (354) does not give a contribution to $C(\mathbf{r}_1, \mathbf{r}_2)$ because there is a cancelation in eq. (357). Similarly $\hat{\rho}_{EM1}$ gives zero in both terms. Therefore, the whole contribution to the correlation derives from the term $\hat{\rho}_{EM2}$. We get

$$C(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{16V^2} \left\langle 0 \left| \sum_{\mathbf{k},\boldsymbol{\varepsilon}} \sum_{\mathbf{k}',\boldsymbol{\varepsilon}'} K_2 \hat{\alpha}_{\mathbf{k},\boldsymbol{\varepsilon}} \hat{\alpha}_{\mathbf{k}',\boldsymbol{\varepsilon}'} \sum_{\mathbf{k}'',\boldsymbol{\varepsilon}''} \sum_{\mathbf{k}''',\boldsymbol{\varepsilon}'''} K_2^* \hat{\alpha}^\dagger_{\mathbf{k}'',\boldsymbol{\varepsilon}''} \hat{\alpha}^\dagger_{\mathbf{k}''',\boldsymbol{\varepsilon}'''} \right| 0 \right\rangle,$$

whence, putting the operators in normal order via the commutation rules, we obtain after some algebra

$$C_{EM0}(r) = \frac{1}{4V^2} \sum_{\mathbf{k},\boldsymbol{\varepsilon}} \sum_{\mathbf{k}',\boldsymbol{\varepsilon}'} k k' \left[1 + \frac{\mathbf{k} \cdot \mathbf{k}'}{k k'} \right]^2 \exp\left[i (\mathbf{k} + \mathbf{k}') \cdot \mathbf{r} \right],$$

where eqs. (351) and (353) have been taken into account. Substituting integrals for the sums this leads to

$$C_{EM0}(r) = \frac{1}{4(2\pi)^6} \int k d^3\mathbf{k} \int k' d^3\mathbf{k}' \left[1 + \frac{\mathbf{k} \cdot \mathbf{k}'}{k k'} \right]^2 \exp\left[i (\mathbf{k} + \mathbf{k}') \cdot \mathbf{r} \right]. \tag{358}$$

The integrals are convergent for any finite r, after a regularization for $k, k' \to \infty$, and we get

$$C_{EM0}(r) = \frac{3\hbar^2}{\pi^4 c^2 r^8}. \tag{359}$$

The correlation strongly diverges as r^{-8} for $r \to 0$, which is consistent with the quartic divergence of the mean energy density eq. (344). However, this behaviour is only appropriate for long distances, corresponding to low energies. In fact, for that range the most relevant vacuum field is the electromagnetic one, whose excitations (photons) have zero mass. In contrast, at short distances the vacuum fluctuations should involve many vacuum fields simultaneously; that is, the fluctuations of all interacting fields together are correlated making them relatively weak. We may take this fact into account including a cutoff in eq. (358) by means of the exponential factor $\exp[-(k+k')/m]$, whence we get

$$C_{EM_0}(r) = \frac{3m^8c^6}{2\pi^4\hbar^6}\frac{(2x^2-4x+3)}{(x+1)^6}, \qquad x \equiv (mrc/\hbar)^2. \tag{360}$$

7.3.3. The electron-positron vacuum field. We will get the stress-energy tensor of the Dirac field, that is, the energy density ρ and the pressure P. In order to avoid any confusion caused by the different notations used in the literature, I will write ρ and P in terms of the original Dirac matrices α_k and β, rather than the gamma matrices that have been defined in several different forms by different authors. Thus the corresponding operators may be written

$$\hat{\rho}_D = \frac{i}{2}\left(\widehat{\psi}^\dagger \frac{d\widehat{\psi}}{dt} - \frac{d\widehat{\psi}^\dagger}{dt}\widehat{\psi}\right), \tag{361}$$

and

$$\widehat{P}_D^{(k} = \frac{i}{2}\left(\widehat{\psi}^\dagger \alpha_k \frac{\partial\widehat{\psi}}{\partial x^k} - \frac{\partial\widehat{\psi}^\dagger}{\partial x^k}\alpha_k\widehat{\psi}\right), \tag{362}$$

where $\widehat{\psi}$ and $\widehat{\psi}^\dagger$ are quantized fields and k may be either 1, 2 or 3, the resulting pressure being the same in the three cases due to the assumed isotropy of the vacuum. (Note that in (362) there is no summation of repeated indices). Expanding $\widehat{\psi}$ and $\widehat{\psi}^\dagger$ in plane waves we get

$$\begin{aligned}\widehat{\psi}(\mathbf{r},t) &= \sqrt{\frac{1}{V}}\sum_{\mathbf{p},s}\sqrt{\frac{m}{E}}\hat{b}_{\mathbf{p}s}u_{\mathbf{p}s}\exp(i\mathbf{p}\cdot\mathbf{r} - iEt) \\ &\quad + \sqrt{\frac{1}{V}}\sum_{\mathbf{p},s}\sqrt{\frac{m}{E}}\hat{d}^\dagger_{\mathbf{p}s}v_{\mathbf{p}s}\exp(-i\mathbf{p}\cdot\mathbf{r} + iEt),\end{aligned}$$

$$\begin{aligned}\widehat{\psi}^{\dagger}(\mathbf{r},t) &= \sqrt{\frac{1}{V}}\sum_{\mathbf{p},s}\sqrt{\frac{m}{E}}\hat{b}^{\dagger}_{\mathbf{p}s}u^{\dagger}_{\mathbf{p}s}\exp\left(-i\mathbf{p}\cdot\mathbf{r}+iEt\right)\\ &\quad+\sqrt{\frac{1}{V}}\sum_{\mathbf{p},s}\sqrt{\frac{m}{E}}\hat{d}_{\mathbf{p}s}v^{\dagger}_{\mathbf{p}s}\exp\left(i\mathbf{p}\cdot\mathbf{r}-iEt\right),\end{aligned}$$

where $\hat{b}_{\mathbf{p},s}\left(\hat{d}_{\mathbf{p},s}\right)$ is the annihilation operator of an electron (positron) with momentum $\mathbf{p}$ and spin $s(=1,2)$, $\hat{b}^{\dagger}_{\mathbf{p},s}\left(\hat{d}^{\dagger}_{\mathbf{p},s}\right)$ is the corresponding creation operator and $u, u^{\dagger}, v, v^{\dagger}$ are appropriate spinors. (Our notation follows the book of Sakurai [**9**]). m is the electron (or positron) mass, $\mathbf{p}$ the momentum and E the energy of either an electron or a positron. Inserting these expressions in eq. (361) we get $\hat{\rho}_D$ as a sum of four terms with a product of 2 operators each. We want all the terms with the creation and annihilation operators in normal order so that, taking the anticommutation rules into account, we should make a replacement as follows

$$\begin{aligned}&\hat{d}_{\mathbf{p}s}\exp\left(i\mathbf{p}\cdot\mathbf{r}-iEt\right)\hat{d}^{\dagger}_{\mathbf{p}'s'}\exp\left(-i\mathbf{p}'\cdot\mathbf{r}+iE't\right)\\ =\;&\delta_{\mathbf{p}\mathbf{p}'}\delta_{ss'}-\hat{d}^{\dagger}_{\mathbf{p}'s'}\exp\left(-i\mathbf{p}'\cdot\mathbf{r}+iE't\right)\hat{d}_{\mathbf{p}s}\exp\left(i\mathbf{p}\cdot\mathbf{r}-iEt\right),\end{aligned}$$

where $\delta_{\mathbf{p}\mathbf{p}'}$ and $\delta_{ss'}$ are Kronecker deltas. Thus we obtain 5 terms, all in normal order, that we shall label as follows

$$\hat{\rho}_D(\mathbf{r},t)=\rho_{D0}+\hat{\rho}_b(\mathbf{r},t)+\hat{\rho}_d(\mathbf{r},t)+\hat{\rho}_{bd}(\mathbf{r},t)+\hat{\rho}^{\dagger}_{bd}(\mathbf{r},t). \tag{363}$$

As in the electromagnetic field, the first term, ρ_{D0}, is a number not an operator (strictly speaking, it is proportional to the unit operator), that is

$$\rho_{D0}=-\frac{1}{V}\sum_{p,s}\sqrt{m^2+p^2}. \tag{364}$$

The term $\hat{\rho}_b$ (resp. $\hat{\rho}_d$) may create or destroy one electron (resp. positron). These two terms contribute neither to the mean energy density nor to the fluctuations. The term $\hat{\rho}^{\dagger}_{bd}$ may create an electron-positron pair and the term $\hat{\rho}_{bd}$ may destroy a pair. These two terms do not contribute to the mean energy density, but they do contribute to the fluctuations.

Integration of eq. (363) with respect to $\mathbf{r}$ gives the Hamiltonian of the free electron-positron field, which is

$$H_D=\sum_{p,s}\sqrt{m^2+p^2}\left(b^{\dagger}_{\mathbf{p},s}b_{\mathbf{p},s}+d^{\dagger}_{\mathbf{p},s}d_{\mathbf{p},s}-1\right). \tag{365}$$

The vacuum state, $|0\rangle$, of the free field corresponds to the eigenvector of the Hamiltonian eq. (365) with the smallest eigenvalue. It consists of zero electrons and zero positrons. Therefore, from now on we may define $\mid 0\rangle$ to be the QED unperturbed vacuum state, which is a simultaneous eigenvector of both Hamiltonians eqs. (355) and (365). Thus the state $|0\rangle$ is defined as having zero photons, electrons and positrons. It should be distinguished from the physical vacuum state, $\mid vac\rangle$, which is an eigenvalue of the total QED Hamiltonian, including the interactions. From eq. (365) the vacuum energy density eq. (347) is obtained.

The stresses may be calculated as vacuum expectations of the operators $\widehat{P}_D^{(k}$, eq. (362). As in the case of the energy density, eq. (363), we may write the operator $\widehat{P}_D^{(k}$ as a sum of terms with the creation and annihilation operators in normal order, that is,

$$\widehat{P}_D^{(k}(\mathbf{r},t) = P_{D0}^{(k} + \widehat{P}_b^{(k}(\mathbf{r},t) + \widehat{P}_d^{(k}(\mathbf{r},t) + \widehat{P}_{bd}^{(k}(\mathbf{r},t) + \widehat{P}_{bd}^{(k\dagger}(\mathbf{r},t)\,. \tag{366}$$

The last four terms are similar to those of eq. (363) but I will not give their explicit expressions. They do not contribute to the mean pressure. The first term of eq. (366) does contribute and, after some algebra, it becomes

$$\begin{aligned} P_{D0}^{(k} &= -\frac{m}{V}\sum_{\mathbf{p}s}\frac{p_k}{E}v^\dagger_{\mathbf{p}s}\alpha_k v_{\mathbf{p}s} = -\frac{2}{V}\sum_{\mathbf{p}}\frac{p_k^2}{E} \\ &\to -\frac{1}{\pi^2}\int_0^{p_{\max}}\frac{p_k^2}{E}p^2dp = -\frac{1}{3\pi^2}\int_0^{p_{\max}}\frac{p^4}{E}dp, \end{aligned}$$

where the isotropy has been taken into account in the last equality. That is, the stresses along three orthogonal axes are equal and every one corresponds to the pressure. The result of the momentum, p, integration was given in eq. (348).

7.3.4. The interaction between electron-positron and electromagnetic fields. The physical vacuum of QED, $|vac\rangle$, is different from the zeroth order vacuum, $|0\rangle$, studied above. The latter is the eigenvector, with the smallest eigenvalue, of the unperturbed Hamiltonian $\widehat{H}_0 = \widehat{H}_{EM} + \widehat{H}_D$, see eqs. (355) and (365). The former is the eigenvector of $\widehat{H} = \widehat{H}_{EM} + \widehat{H}_D + \widehat{H}_{int}$ with the smallest eigenvalue, that is, taking the interactions into account. Finding $\mid vac\rangle$ as an exact eigenvector of $\widehat{H}$ is not possible and we should use a perturbation method. That is, we should calculate it as an expansion in powers of the coupling constant, the positron charge, e. Only even powers of e

would appear and the result becomes an expansion in powers of the fine structure constant $\alpha \equiv e^2/(4\pi\hbar c) \simeq 1/137$.

The interaction Hamiltonian (or energy) operator may be written, in the Coulomb gauge,

$$\hat{\rho}_{int}\left(\mathbf{r},t\right) = -e\widehat{\psi}^{\dagger}\boldsymbol{\alpha}\widehat{\psi}\cdot\widehat{\mathbf{A}}. \tag{367}$$

The operators $\widehat{\psi}$, $\widehat{\psi}^{\dagger}$ and $\widehat{\mathbf{A}}$ contain two terms each when expanded in plane waves, every term corresponding to an infinite sum. One of these terms has creation operators and the other one annihilation operators (similar to eq. (349) for the electromagnetic potential vector). This gives rise to 8 terms for $\hat{\rho}_{int}$, eq. (367). Integrating with respect to $\mathbf{r}$ inside the volume V leads to the interaction Hamiltonian. Only two terms survive and we get

$$\begin{aligned}\widehat{H}_{int} &= -e\sum_{\mathbf{p},\mathbf{q},\mathbf{k},s,s',\boldsymbol{\varepsilon}}\frac{m}{V^{3/2}\sqrt{2kEE'}}u_s^{\dagger}\left(\mathbf{p}\right)\boldsymbol{\alpha}\cdot\boldsymbol{\varepsilon}\,v_{s'}\left(\mathbf{q}\right)\\ &\quad\times\delta_{\mathbf{p}+\mathbf{q},\mathbf{k}}\widehat{a}_{\mathbf{k},\boldsymbol{\varepsilon}}^{\dagger}b_{\mathbf{p},s}^{\dagger}d_{\mathbf{q},s'}^{\dagger}+h.c,\end{aligned} \tag{368}$$

where $\delta_{\mathbf{p}+\mathbf{q},\mathbf{k}}$ is a Kronecker's delta, $h.c.$ means Hermitian conjugate and

$$E \equiv \sqrt{p^2+m^2}, \qquad E' \equiv \sqrt{q^2+m^2}.$$

One of the terms of the Hamiltonian may create triples electron-positron-photon and the other term may annihilate triples. The integral in $\mathbf{r}$ causes the Hamiltonian eq. (368) to be invariant under translations and rotations, whence it can couple only states with the same total momentum and total angular momentum.

I will not detail the calculation of the interaction energy, which is involved and will not be needed in the forthcoming sections of this chapter.

7.4. Vacuum fluctuations an alternative to dark energy

7.4.1. The standard cosmological model. Astronomical observations, in particular the study of type Ia supernovae, anisotropies in the cosmic background radiation and matter power spectra inferred from large galaxy surveys, have improved our knowledge of the universe giving rise to a precision cosmology. The data are compatible with the universe having a Friedmann–Robertson–Walker metric with flat spatial slices [**17**] of the form

$$ds^2 = -dt^2 + a(t)^2\left(dr^2 + r^2 d\Omega\right). \tag{369}$$

The parameter $a(t)$ is related, at present time t_0, to the measurable Hubble constant, H_0, and decceleration parameter, q_0, via

$$\left[\frac{\dot{a}}{a}\right]_{t_0} = H_0, \qquad \left[\frac{\ddot{a}}{a}\right]_{t_0} = -H_0^2 q_0. \tag{370}$$

The observations also provide information about the evolution of $a(t)$ and the distribution of matter in the past.

The available knowledge is summarized in the ΛCDM model. In it, the baryonic matter density ρ_B represents about 4.6% of the matter content of the universe while two hypothetical ingredients named cold dark matter (CDM) and dark energy (DE) contribute with densities $\rho_{DM} \sim 23\%$ and $\rho_{DE} \sim 73\%$ respectively [**10**]. These densities are related to the metric eq. (369) via the Friedmann equations, derived from general relativity, giving the following relations [**11**]

$$\begin{aligned} \left[\frac{\dot{a}}{a}\right]^2 &= \frac{8\pi G}{3}\left(\rho_B(t) + \rho_{DM}(t) + \rho_{DE}\right), \\ \frac{\ddot{a}}{a} &= \frac{8\pi G}{3}\left(\frac{1}{2}\left[\rho_B(t) + \rho_{DM}(t)\right] - \rho_{DE}\right), \end{aligned} \tag{371}$$

where I neglect small effects of radiation and matter pressure. The baryonic density ρ_B is well known from the measured abundances of light chemical elements, which allows calculating ρ_{DE} and ρ_{DM} from the empirical quantities H_0 and q_0 via comparing eqs. (371) and (370). The values obtained by this method agree with data from other observations. For instance, cold dark matter, in an amount compatible with ρ_{DM}, is needed in order to explain the observed (almost flat) rotation curves in galaxies. However, the nature of dark matter and dark energy remain open problems. The ΛCDM model rests upon the assumption that general relativity (GR) is indeed the correct theory of gravity. However, it is conceivable that both cosmic speed-up and dark matter represent signals of a breakdown of GR.

A common assumption is that, in addition to dark energy density, there is pressure with a value $P_{DE} = -\rho_{DE}$, corresponding to a cosmological constant, as will be discussed in the next subsection. The hypothesis agrees with observations [**10**] and it has been used in the derivation of eq. (371). In any case the problem is the origin of that stress-energy. A possibility studied in subsections 7.4.3 to 7.4.5 below is that the dark energy is actually an effect of the quantum vacuum fluctuations.

7.4.2. The cosmological constant problem. Many proposals have been made for the origin of dark energy (for a review see Copeland [**14**]). As said above it is usual to identify it with a cosmological constant or, what is equivalent in practice (see below), to assume that it derives from the quantum vacuum. In fact, Einstein equation of general relativity may be written (with the notation of the book of Weinberg [**1**])

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - \Lambda g_{\mu\nu} = -8\pi G T^{matt}_{\mu\nu}, \tag{372}$$

where Λ is the cosmological constant and G the Newton constant. In the original Einstein formulation of 1915 the term with Λ did not appear (see eq. (336) in the introduction section). It was later included by Einstein in order to get a stationary (although not stable) universe, but he proposed to remove it after the discovery of the expansion of the universe [**15**]. However, the cosmological constant was a recurrent possibility for the next 70 years, but without too much empirical support. That opinion changed radically in 1999 because it was discovered that the expansion of the universe was accelerating [**16**], [**17**] and the cosmological constant provided an explanation, although the less committed name 'dark energy' was introduced.

The assumption that the quantum vacuum is the origin of the accelerating expansion derives from the fact that eq. (372) may be written passing the term Λ to the right side with opposite sign, that is

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = -8\pi G \left(T^{matt}_{\mu\nu} + T^{vac}_{\mu\nu} \right), \tag{373}$$

where $T^{vac}_{\mu\nu}$ may be interpreted as the stress-energy tensor of the vacuum fields. In order to have Lorentz invariance it is generally assumed that the vacuum stress-energy should be proportional to the metric tensor, that is, $T^{vac}_{\mu\nu} \propto g_{\mu\nu}$, whence the identification

$$8\pi G T^{vac}_{\mu\nu} = -\Lambda g_{\mu\nu},$$

leads from eq. (373) to eq. (372).

The hypothesis that dark energy is a gravitational effect of the quantum vacuum poses a well known problem [**18**]. In fact, it is plausible that the vacuum energy density, ρ_{DE}, should correspond to a combination of the universal constants c, $\hbar$, G. There is a unique combination with dimensions of density, given in eq. (345). That is,

$$\rho_{Planck} = \frac{c^5}{G^2 \hbar} \simeq 10^{97} kg/m^3, \tag{374}$$

a value about 123 orders greater than the required dark energy density, with current empirical value [**19**]

$$\rho_{DE} = -P_{DE} \simeq (6.0 \pm 0.2) \times 10^{-27} kg/m^3, \tag{375}$$

$P_{DE} < 0$ being the pressure.

The large discrepancy between eq. (374) and the observed value eq. (375) is named the 'cosmological constant problem' [**18**], [**20**], [**7**]. Many solutions have been proposed [**18**] that will not be discussed here. I just mention a recent one assuming that the observed cancelation at large scale derives from an averaging of rapid fluctuations of energy and pressure in regions at the Planck scale [**21**]. Another assumption is that the vacuum fields have positive and negative contributions that cancel with each other. A complete cancelation might appear possible, even plausible, but a partial cancelation giving the result eq. (375), more than hundred orders smaller than Planck's density, looks conspiratorial [**18**].

In the following I propose a possible solution to the cosmological constant problem resting on the following assumptions:

1. The energy density and the pressure in the vacuum averaged over large spacetime regions (free of matter) are both null, as was suggested in section 7.3.

2. The fluctuations of energy and pressure give rise to an effective tensor fulfilling eq. (375). This is studied in subsections 7.4.3 to 7.4.5 below.

7.4.3. Vacuum fluctuations as the origin of dark energy. Any theory aimed at explaining dark energy would lead to a value for that energy as a combination of some parameters. That value should agree with eq. (375). We may use the universal constants c, $\hbar, G$, but we need another parameter in the form of a mass, length, time or a combination of them whence a parameter, say μ, may be found such that the empirical density eq. (375) is obtained in the form

$$\rho_{DE} = \rho_{Planck} f\left(c, \hbar, G, \mu\right)$$

where f is a dimensionless function to be derived from the (yet unknown) theory. Without loss of generality we may choose μ to be a mass, m, whence dimensional analysis leads to

$$\rho_{DE} = \frac{c^5}{G^2\hbar} f\left(N\right), \qquad N = \frac{\hbar c}{Gm^2}, \tag{376}$$

N being the most simple dimensionless quantity that may be got combining the universal constants with a mass. The introduction of a parameter different from mass does not give anything new. For instance, choosing length, λ, or time, τ, we may get the same result provided the new parameters are related to mass via

$$\lambda = \frac{\hbar}{mc}, \qquad \tau = \frac{\hbar}{mc^2}. \tag{377}$$

It is the case that without any theory we may guess the function f. It is plausible that m is of order the mass of a fundamental particle. If we consider a charged particle then N eq. (377) is of order the ratio between the electrostatic repulsion and the gravitational attraction between two of these particles, that is,

$$\frac{e^2}{Gm^2} = \frac{e^2}{\hbar c} \times \frac{\hbar c}{Gm^2} = \alpha N, \quad \alpha \simeq \frac{1}{137}, \tag{378}$$

where α is the fine structure constant. The (dimensionless) number N may be recognized as one of Dirac's large numbers of order 10^{40}. Therefore, we may guess that the function $f(N)$ should be of order N^{-3} because as said above the ratio between eqs. (374) and (375) is of order 10^{123}. This leads to

$$\rho_{DE} \sim \frac{c^5}{G^2\hbar}\left(\frac{\hbar c}{Gm^2}\right)^{-3} = \frac{Gm^6c^2}{\hbar^4}, \tag{379}$$

that provides a very good agreement with eq. (375) if we choose m to be 80 times the electron mass (or about the mass of the pion if we write $h = 2\pi\hbar$ rather than $\hbar$ in eq. (379)).

An energy density proportional to the Newton constant G, eq. (379), suggests that ρ_{DE} is a gravitational energy density. Therefore, the theory of dark energy, that we are searching for, should explain what is the matter that originates that gravitational energy. Eq. (379) was proposed by Zeldovich [**22**] as a possible cosmological constant more than 40 years ago, well before the discovery of the accelerating expansion of the universe. Zeldovich also provided an interpretation writing eq. (379) in terms of a mass m and its associated 'Compton wavelength' λ eq. (377), that is,

$$\rho_{DE} \sim -\frac{Gm^2}{\lambda c^2} \times \frac{1}{\lambda^3}, \quad \lambda \equiv \frac{\hbar}{mc}. \tag{380}$$

Thus eq. (380) is the mass density (rather than energy density, hence the division by c^2) corresponding to the (Newtonian) gravitational energy of two particles of mass m placed at a distance λ, assuming

that such an energy appears in every volume λ^3. A problem is that in eq. (380) the gravitational energy ρ_g is negative if both masses are positive. Zeldovich interpreted that the 'particles' were actually vacuum fluctuations, that may involve positive or negative energy density departures from the mean. Hence the hypothesis that the dark energy derives from the fluctuations of the quantum vacuum, an assumption further studied elsewhere [**23**], [**24**], [**25**]. A summary is presented in the following.

7.4.4. Newtonian treatment of the gravity of density correlations. For the sake of clarity I shall begin with arguments involving Newtonian gravity (rather than general relativity). I will assume that the vacuum fields fluctuate. That is, at a given time there are regions with energy density above the mean (positive energy fluctuations) and other regions with density below it (negative fluctuations). Thus, assuming that there is a complete cancelation of the vacuum energy so that its average is zero, the energy density may be represented by a stochastic field, $\rho(\mathbf{r},t)$ with zero mean. I will assume that both the ensemble average $\langle\rho(\mathbf{r},t)\rangle$ and the spacetime average over a large region are zero. This fits in our assumption in section 7.4.2 that the vacuum stress-energy tensor is zero.

We will also assume, as in special relativity, that energy gravitates, because energy is essentially the same as mass. Then we should accept that Newtonian gravity associates negative (positive) gravitational potential to positive (negative) mass-energy. (Of course, Newtonian gravity is not consistent with special relativity, but for a simplified argument we may combine both). Thus, using Newtonian theory, the gravitational energy, E_G, of the vacuum mass density in the volume V would be

$$E_G = -\frac{G}{2}\int_V \rho(\mathbf{r}_1,t)\, d^3\mathbf{r}_1 \int_V \frac{\rho(\mathbf{r}_2,t)}{|\mathbf{r}_2-\mathbf{r}_1|} d^3\mathbf{r}_2, \tag{381}$$

where G is the Newton constant and V a volume large in comparison with the range of the fluctuations. Hence we may define the ensemble average of the gravitational energy in the volume V as

$$\bar{\rho}_G = \frac{\langle E_G\rangle}{V} = -\frac{G}{2V}\int_V d^3\mathbf{r}_1 \int_V |\mathbf{r}_2-\mathbf{r}_1|^{-1} \langle\rho(\mathbf{r}_1,t)\,\rho(\mathbf{r}_2,t)\rangle\, d^3\mathbf{r}_2. \tag{382}$$

If the self-correlation of fluctuations (at a gven time) depends only on the distance $r = |\mathbf{r}_2-\mathbf{r}_1|$, consistent with the assumed homogeneity

and isotropy of the vacuum, then changing from the variables $\{\mathbf{r}_1, \mathbf{r}_2\}$ to the new ones $\{\mathbf{r}_0, \mathbf{r}\}$ via

$$\mathbf{r}_0 = \frac{\mathbf{r}_1 + \mathbf{r}_2}{2}, \qquad \mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1,$$

we may define the two-point correlation, $C(\mathbf{r}_0, \mathbf{r})$ of the energy density around the point $\mathbf{r}_0$ by

$$C(\mathbf{r}_0, \mathbf{r}) \equiv \langle \rho(\mathbf{r}_0 + \mathbf{r}/2, t)\, \rho(\mathbf{r}_0 - \mathbf{r}/2, t) \rangle . \tag{383}$$

If there is isotropy and homogeneity the correlation depends only on r, that is

$$C(r) \equiv \langle \rho(\mathbf{r}_0 + \mathbf{r}/2, t)\, \rho(\mathbf{r}_0 - \mathbf{r}/2, t) \rangle , \quad r \equiv |\mathbf{r}| . \tag{384}$$

In this case the mean gravitational energy density will be

$$\bar{\rho}_G = -\frac{G}{2} \int \frac{1}{r} C(r)\, d^3\mathbf{r} = -2\pi G \int_0^\infty C(r)\, r dr. \tag{385}$$

In order that the integral does not diverge at the origin the correlation should fulfil

$$|C(r)| < r^{-2} \text{ for } r \to 0. \tag{386}$$

This gives a plausible limit on the strength of the vacuum fluctuations at short distances. Of course, it is also plausible that the correlation of fluctuations decreases at infinity so rapidly that no divergence would appear for $r \to \infty$.

The density $\bar{\rho}_G$ in eq. (385) may be positive or negative depending on the form of the function $C(r)$. It is plausible that $C(r)$ is positive for small r whence the short range part will give a negative contribution to the gravitational energy. At large distances $C(r)$ should go to zero, but the behaviour at intermediate distances is unknown. It is plausible that within Newtonian gravity the total gravitational energy is negative because the $1/r$ dependence makes the short range behaviour most relevant. In any case, the question is purely academic because a treatment with general relativity is required, that will be made in the next subsection.

Up to here the approach has been classical (including special relativity) but a quantum treatment is straightforward. It is enough to substitute operators $\hat{\rho}(\mathbf{r}, t)$ for the classical densities $\rho(\mathbf{r}, t)$ and vacuum expectation values for ensemble averages. Indeed, in this book we support that quantum mechanics is a (peculiar) stochastic theory,

although we are not always able to find the appropriate probability distributions of the random variables involved. In particular, the quantum counterpart of the definition eq. (384) should be

(387)
$$C(r) \equiv \lim_{V\to\infty}\left[\frac{1}{2V}\int_V \langle vac|\hat{\rho}(\mathbf{r}_1)\,\hat{\rho}(\mathbf{r}_2)+\hat{\rho}(\mathbf{r}_2)\,\hat{\rho}(\mathbf{r}_1)|\,vac\rangle\, d^3\mathbf{r}_0\right],$$

where we have made the plausible assumption that the symmetrical order is appropriate for the operators that may not commute.

The conclusion from our Newtonian analysis is that, in addition to the mean energy density of the vacuum, if any, there will be a *negative energy* caused by the gravitational interaction of density fluctuations. However, according to eq. (375) we need a *positive energy and a negative pressure*, so that the Newtonian approach fails. In the following I show that a treatment within GR may provide plausible results.

7.4.5. Vacuum fluctuations in general relativity. A simplified model. The quantum vacuum is incompatible with Minkowski spacetime even in the absence of any form of matter, due to the fluctuations. In fact, according to general relativity Minkowski space implies a null Einstein tensor $G_{\mu\nu}$ whence the stress energy tensor $T_{\mu\nu}$ should be also zero, see eq. (336). However, $T_{\mu\nu}$ is different from zero if there are quantum vacuum fluctuations.

The argument may be presented also in the form of a paradox as follows. In Minkowski space we should expect that the two-point density correlation would be Lorentz invariant, whence the correlation function eq. (387) would depend only on the relativistic invariant

(388)
$$\sigma^2 \equiv (t_2-t_1)^2-|\mathbf{r}_1-\mathbf{r}_2|^2,$$

that reduces to $r=|\mathbf{r}_1-\mathbf{r}_2|$ for $t_2=t_1$. Of course, the correlation might depend on whether the interval is timelike or spacelike (in other words, on whether σ as defined above is real or imaginary). In the following I show that in any case the said invariance prevents the existence of density fluctuations. In fact, let us consider a Minkowski space and two arbitrary points, A and B, with coordinates $(\mathbf{r}_A,t_A)$ and $(\mathbf{r}_B,t_B)$, respectively. It is always possible to find another point C with coordinates $(\mathbf{r}_C,t_C)$ which is lightlike separated from both A and B. Indeed, any point in the intersection of the light cones of A and B fulfils that condition. Then eq. (388) implies that the vacuum energy density in C will be the same as in A and the same as in B. But the points A and B being arbitrary we conclude that the density

will be the same in all points of Minkowski space. There would be no fluctuations at all! The solution of the paradox is that, as said above, if there are fluctuations the spacetime cannot be Minkowskian. Thus we are compelled to study the deviation from Minkowski space due to vacuum fluctuations.

We might believe that the modification of Minkowski space caused by the quantum vacuum fluctuations would consist of just localized time-dependent fluctuations of the metric, that is, a small region of curved space around every fluctuation. However, this is not the case. In fact, in the following we show that, with plausible approximations, the quantum fluctuations lead to a global change of spacetime that may be understood as due to the long-range nature of gravity combined with the nonlinear character of GR. That is, masses localized in a small region give rise to space curvature over large regions.

In order to study the effects of vacuum fluctuations on space curvature I start recalling some well known facts. For a small enough region of the universe around us, but large in comparison with typical distances between galaxies, the spacetime metric eq. (369) may be rewritten, near present time, using spherical coordinates as follows [**11**]

$$ds^2 \simeq g_{rr}dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) - g_{tt}dt^2, \tag{389}$$
$$g_{rr} = \left[1 + \left[\frac{\dot{a}}{a}\right]_0^2 r^2\right], \quad g_{tt} = \left[1 + \left[\frac{\ddot{a}}{a}\right]_0 r^2\right],$$

where terms of order $O\left(r^3\right)$ have been neglected and it is assumed that the (slow) change of the metric coefficients with time may be ignored. The metric elements g_{rr} and g_{tt} of eq. (389) may be got from the known contents of the universe, which is cold matter consisting of baryonic and dark matter with densities ρ_B and ρ_{DM} respectively and negligible pressure, plus a hypothetical 'dark energy' with density ρ_{DE} and pressure $-\rho_{DE}$. Then Einstein eq. (336) leads to the following components of the metric tensor from eq. (389)

$$g_{rr} = 1 + \frac{8\pi G}{3}\left[\rho_B(t) + \rho_{DM}(t) + \rho_{DE}\right] r^2 + O\left(r^3\right),$$
$$g_{tt} = 1 + \frac{8\pi G}{3}\left[\frac{1}{2}\left[\rho_B(t) + \rho_{DM}(t)\right] - \rho_{DE}\right] r^2 + O\left(r^3\right). \tag{390}$$

This metric corresponds to an expanding universe due to baryonic and dark matter, with the expansion accelerating due to the 'dark energy' term ρ_{DE}. In the following I will show that the term ρ_{DE} of the metric

eq. (390) may be a consequence of the vacuum fluctuations without the need of assuming any dark energy. Of course, the metric eq. (389) possesses spherical symmetry only after spreading homogeneously in space both matter and, in some sense to be explained below, the vacuum fluctuations. Spreading matter, both baryonic and dark, is standard in cosmology, in spite of the fact that the distribution is highly inhomogeneous, the mass being concentrated mainly in stars, grouped in galaxies and clusters.

In order to obtain the effect of the density and pressure due to vacuum fluctuations in a classical approach we should consider a general metric, eq.(335), and solve Einstein eq. (336) for a suitable stress-energy tensor associated to matter plus vacuum fluctuations. After that we should perform an ensemble average with respect to the possible samples of the metric, every metric associated to a possible sample of vacuum fluctuations. The calculation is lengthy and it has been performed with several simplifying assumptions [**25**]. Here I use an alternative approximate treatment that, being more simple, may help getting an intuitive understanding of the physics. We shall assume that a spacetime average of the fluctuations is equivalent to the ensemble average, a kind of ergodic property.

For simplicity I will ignore matter, both baryonic and dark, in the following calculation. If our hypothesis that 'dark energy effects' actually derive from vacuum fluctuations is correct, then the metric might be written in the form

$$ds^2 = (1+\Lambda r^2)dr^2+r^2(d\theta^2+\sin^2\theta d\phi^2)-(1-\Lambda r^2)dt^2+O\left(r^3\right), \tag{391}$$

where Λ may be seen as a cosmological constant related to the hypothetical dark energy ρ_{DE} by

$$\Lambda \equiv \frac{8\pi G}{3}\rho_{DE}. \tag{392}$$

We will assume that spacetime is globally modified by the vacuum fluctuations, but only slightly. Indeed, the Newton constant G is small in the sense that large concentrations of matter are needed to produce a relatively small curvature of space, as is the case for instance in the solar system. Also, masses and energies at the atomic or macroscopic laboratory scale produce a negligible curvature (gravitational field). Thus I propose studying a vacuum with density and pressure fluctuations using a metric appropriate for spherical symmetry, that is,

$$ds^2 = A\,(r,t)\,dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) - B\,(r,t)\,dt^2, \tag{393}$$

ignoring the small *local* departures due to the vacuum fluctuations. I will just study the long range modifications with respect to Minkowski.

The elements of the Einstein tensor in terms of the components A and B of the metric eq. (393) contain derivatives with respect to the r coordinate, but not to time, whence t appears just as a parameter that I will ignore for notational simplicity. It is easy to invert those relations, getting the metric in terms of density $\rho(r)$ and pressure $P(r)$, that is,

$$A(r) = \left(1 - \frac{2Gm(r)}{r}\right)^{-1}, \qquad m(r) \equiv 4\pi \int_{\mathbf{0}}^{r} \rho(z)\, z^2 dz,$$

$$B(r) = \exp\gamma, \qquad \gamma = 2G \int_0^r \frac{m(x) + 4\pi x^3 P(x)}{x^2 - 2Gm(x)\, x} dx. \tag{394}$$

As pointed out Newton constant may be taken as very small whence it is appropriate to approximate eqs. (394) in powers of G. We must go up to terms of order G^2 because the terms of order G do not contribute as we will see. The approximation gives

$$A(r) = 1 + \frac{2Gm(r)}{r} + \frac{4G^2 m(r)^2}{r^2} + O\left(G^3\right), \tag{395}$$

and

$$\begin{aligned} B(r) &= 1 + 2G \int_0^r \left[x^{-2} m(x) + 4\pi x \bar{P}(x)\right] dx \\ &\quad + 4G^2 \int_0^r \left[x^{-3} m(x)^2 + 4\pi m(x)\, \bar{P}(x)\right] dx \\ &\quad + 2G^2 \left\{\int_0^r \left[x^{-2} m(x) + 4\pi x \bar{P}(x)\right] dx\right\}^2 + O\left(G^3\right), \end{aligned} \tag{396}$$

where I take $\bar{P}(x)$ as an angular average of the pressure, that is,

$$\bar{P}(x) \to \frac{1}{4\pi} \int_0^{\pi} \sin\theta d\theta \int_0^{2\pi} d\phi P(\mathbf{x}) = \frac{1}{4\pi x^2} \int P(\mathbf{z})\, d^3 z \delta(x - z). \tag{397}$$

Here $\delta()$ is Dirac delta and the $\mathbf{z}$ integral may be extended to the whole space. We take into account that both $\rho(\mathbf{x})$ and $P(\mathbf{x})$ are fluctuating quantities (i.e. stochastic fields) with zero mean changing over typical distances and time intervals of atomic size. We will ignore the details of these changes but take them into account using averages that involve the two-point density correlation function $C(\mathbf{r}_0, \mathbf{r})$, eq. (383), a similar correlation for P and the cross-correlation between ρ and P.

Now we proceed to the calculation of the metric elements A and B via averaging eqs. (395) and (396). In eq.(395) the term of order G gives zero because the average of $m(r)$ becomes an average of ρ, which is zero. Indeed, we are assuming that the density fluctuations may be positive (energy greater than the mean) or negative (smaller than the mean) and also assume that the mean is zero. For the term of order G^2 we should average m^2 and we get

$$\left\langle m\left(r\right)^2\right\rangle = \left\langle \left[\int_{|\mathbf{z}|\leq r} \rho\left(\mathbf{x}\right) d^3x\right]^2\right\rangle = \int_{|\mathbf{x}|\leq r} d^3x \int_{|\mathbf{y}|\leq r} d^3y \left\langle \rho\left(\mathbf{x}\right)\rho\left(\mathbf{y}\right)\right\rangle$$

$$= \int_{|\mathbf{x}|\leq r} d^3x \int_{|\mathbf{y}|\leq r} d^3y C\left(|\mathbf{x}-\mathbf{y}|\right), \tag{398}$$

where $C(r)$ was defined in eq. (384). Hence we should obtain the metric coefficient eq. (395) to order $O\left(G^2\right)$, that is,

$$\begin{aligned} A &= 1+\frac{4G^2}{r^2}\left\langle m\left(r\right)^2\right\rangle \\ &= 1+\frac{4G^2}{r^2}\int_{|\mathbf{x}|\leq r} d^3x \int_{|\mathbf{y}|\leq r} d^3y C\left(|\mathbf{x}-\mathbf{y}|\right). \end{aligned} \tag{399}$$

In order to perform the integrals we need the function $C\left(|\mathbf{x}-\mathbf{y}|\right)$ that was estimated in (360). It is plausible that the function is bigger than the estimation at short distances and I propose to use the following simple function

$$C(r) = \frac{K_{\rho\rho}}{r^2}, \tag{400}$$

where $K_{\rho\rho}$ is a constant. A reason for choosing (400) is that dimensional arguments show that a correct dependence on r in eqs. (391) is achieved with that choice; that is, the leading term after unity in the two relevant metric elements is proportional to r^2.

Eq. (398) gives, taking eq. (400) into account,

$$\begin{aligned} \left\langle m\left(r\right)^2\right\rangle &= K_{\rho\rho}\int_{|\mathbf{x}|\leq r} d^3x \int_{|\mathbf{y}|\leq r} d^3y \left|\mathbf{x}-\mathbf{y}\right|^{-1} \\ &= 8\pi^2 K_{\rho\rho}\int_0^r xdx\int_0^r ydy \ln\frac{x+y}{|x-y|} = 2\pi^2 K_{\rho\rho} r^4 \end{aligned}$$

where we have performed an angular integration in the first equality. Hence eq. (399) leads to

(401)
$$A = 1 + 8\pi^2 G^2 K_{\rho\rho} r^2 \quad \Rightarrow \quad \Lambda = 8\pi^2 G^2 K_{\rho\rho} \quad \Rightarrow \quad \rho_{DE} = 3\pi G K_{\rho\rho}.$$

Thus our model allows relating the two-point vacuum density correlation with the measured value of the dark energy. For instance, if we assume eq. (400) for the correlation, we may get $K_{\rho\rho}$ from the last eq. (401). We may also relate it with Zeldovich eq. (379) and we get

$$K_{\rho\rho} \sim \frac{m^6 c^2}{\hbar^4} = \left(\frac{\hbar}{\lambda^3 c^2}\right)^2, \tag{402}$$

where m is about the pion mass and λ its Compton wavelength.

The calculation of the metric element B eq. (396) is more involved. We shall start defining an expression similar to eq. (400) for the correlations of density and pressure, needed for the averages of the integrals eqs. (396), that is,

$$\begin{aligned} C_{PP}\left(|\mathbf{x}-\mathbf{y}|\right) &\equiv \left\langle P\left(\mathbf{x}\right) P\left(\mathbf{y}\right)\right\rangle = \frac{K_{PP}}{|\mathbf{x}-\mathbf{y}|^2}, \\ C_{\rho P}\left(|\mathbf{x}-\mathbf{y}|\right) &\equiv \left\langle \rho\left(\mathbf{x}\right) P\left(\mathbf{y}\right)\right\rangle = \frac{K_{\rho P}}{|\mathbf{x}-\mathbf{y}|^2}, \end{aligned}$$

where we neglect terms of order $O\left(|\mathbf{x}-\mathbf{y}|^{-1}\right)$ as proposed above. Then the term of order G has zero average and we should evaluate 5 averages involved in the terms of order G^2 of eq. (396). We get

$$\begin{aligned} B_1 &= 4G^2 \int_0^r x^{-3} \left\langle m\left(x\right)^2 \right\rangle dx = 8\pi^2 G^2 K_{\rho\rho} \\ B_2 &= 4G^2 \int_0^r 4\pi \left\langle m\left(x\right) \bar{P}\left(x\right)\right\rangle dx = 16\pi^2 G^2 K_{\rho P}, \\ B_3 &= 2G^2 \int_0^r x^{-2} dx \int_0^r y^{-2} \left\langle m\left(x\right) m\left(y\right)\right\rangle dy = 16\pi^2 G^2 K_{\rho\rho}, \\ B_4 &= 2G^2 \int_0^r 4\pi x dx \int_0^r 4\pi y \left\langle \bar{P}\left(x\right) \bar{P}\left(y\right)\right\rangle dy \\ &= 16 \ln 2 \pi^2 G^2 K_{PP} \simeq 11.09\pi^2 G^2 K_{PP}, \end{aligned}$$

$$\begin{aligned} B_5 &= 4G^2 \int_0^r x^{-2} dx \int_0^r 4\pi y \left\langle m\left(x\right) \bar{P}\left(y\right) \right\rangle dy \\ &= \left(\frac{128}{3} \ln 2 - \frac{1}{3} \right) \pi^2 G^2 K_{\rho P} \simeq 18.91 \pi^2 G^2 K_{\rho P}. \end{aligned}$$

The result of the calculation is

$$B = 1 + G^2 \pi^2 r^2 \left[32 K_{\rho\rho} + 11.09 K_{PP} + 34.91 K_{\rho P} \right], \tag{403}$$

while the desired result eq. (391) requires

$$32 K_{\rho\rho} + 11.09 K_{PP} + 34.91 K_{\rho P} = -\Lambda G^{-2} \pi^{-2} = -8 K_{\rho\rho}. \tag{404}$$

Plausibly the quantities $K_{\rho\rho}$ and K_{PP} should be positive, but $K_{\rho P}$ negative if we assume that in the quantum vacuum fluctuations the pressure acts with a sign opposite to the pressure, in agreement with the Lorentz invariant vacuum equation of state $P = -\rho$. This also suggests identifying $K_{\rho\rho} = K_{PP}$. Then in order to fulfil eq. (404) we choose

$$K_{\rho\rho} = K_{PP}, \qquad K_{\rho P} \simeq -1.46 K_{\rho\rho} \simeq -\frac{3}{2} K_{\rho\rho}.$$

Up to now the calculation has been classical, but we may define the two-point correlations as in eq. (387). This makes the treatment become quantum theoretical, in the semiclassical approximation of general relativity. That is, we identify the (classical) Einstein tensor with the (quantum) vacuum expectation of the two-point correlation.

In conclusion, I propose as a plausible possibility that ρ_{DE} is not an actual (dark) energy density, but a parameter taking into account the effect of the quantum vacuum fluctuations on spacetime curvature. It is satisfactory that the metric elements eqs. (401) and (403) depart from unity by a term of order $\mathrm{O}\left(G^2\right)$, which is required in order that the 'dark energy' is proportional to G as in Zeldovich formula eq. (379). Furthermore, if the theory is correct, and the approximations made in the calculation are fair, then eq. (404) allows getting the behaviour of the two-point correlations of vacuum fluctuations at short distances from the measurable value of the dark energy density eq. (375). The agreement is a prediction of the theory that could be supported if an independent calculation of the two-point correlation of vacuum fluctuations is made from the theory of fundamental fields.

7.5. Dark matter or effect of the vacuum fields?

The existence of dark matter of unknown origin is the current solution proposed for two different problems which may be called astrophysical and cosmological, respectively [**27**]. The astrophysical problem is the more or less flat rotation curves of stars or gas in the haloes of galaxies and clusters of galaxies. Flat means that the velocity of stars and gas around the center of the galaxy is fairly the same for all approximately circular paths, independently of the radius. This contrasts with the predicted velocity decreasing outwards if there is only baryonic matter. In order to account for the flat rotation curves of spiral galaxies, dark matter should dominate the total mass of the galaxy and should be concentrated in the outer baryonic regions of galactic disks, as well as in the surrounding haloes. The cosmological problem is the need of a large amount of non-baryonic cold matter in the universe, at the recombination time, in order to explain the formation of structures. The amount of non-baryonic matter is fairly the same in both cases, which has lead to the popular assumption that both may be solved with the same hypothesis, namely dark matter. As mentioned in the previous section the standard ΛCDM cosmological model predicts that about 70% of the mass-energy budget of the Universe is composed of dark energy and about 30% is the mass, dominated with more than 5/6 by dark matter, whilst ordinary baryonic matter constitutes less than 1/6 of the total mass [**19**].

7.5.1. Alternative explanations for the anomalous gravity in galaxies. The existence of dark matter has been discussed for more than a century [**26**], the anomaly named initially 'the missing mass problem'. In the 1970s, astronomers and cosmologists have began to build what is generally assumed today as a compelling body of evidence for this elusive component of the universe, based on a variety of observations that include temperature anisotropies of the Cosmic Microwave Background, baryonic acoustic oscillations, type Ia supernovae, gravitational lensing of galaxy clusters, and rotation curves of galaxies. However, the Standard Model of particle physics contains no suitable particle to explain these observations. Proposed candidates for dark matter span 90 orders of magnitude in mass, ranging from ultra-light bosons, often referred to as fuzzy dark matter, to massive primordial black holes.

The class of dark matter candidates that has attracted the most attention over the past four decades is weakly interacting massive

particles (WIMPs). They appeared for a long time as a perfect dark matter candidate, as new particles at the weak scale may be produced in the early universe. Despite much effort, no particle other than the Higgs boson has been convincingly detected at the weak scale so far [**27**].

The fact that none of the known particles is a good candidate to form the dark matter and the failure to discover new particles with the required properties, in spite of the big effort made at the observational level, has led to alternatives consisting of a modification of either current gravity theories (i.e. general relativity or Newtonian gravity) or dynamics. There are also proposals resting on the quantum vacuum. Actually, these modifications attempt mainly to solve the astrophysical problem, with less implication in the cosmological one. In the following I comment briefly on modified gravity or dynamics and later we study in more detail the possible alternative derived from vacuum effects.

Modified dynamics. The most popular alternative during several years was the 'modified Newtonian dynamics' (MOND) [**28**], for a review see [**29**]. MOND proposes, for galaxies approximately spherical, that the flat (i.e. independent of the radius) rotation velocities, v, in galactic haloes are roughly given by

$$v^2 \simeq \sqrt{a_0 G M_B}, \tag{405}$$

where G is the Newton constant, M_B the baryonic mass of the galaxy, and a_0 a new universal constant, that is,

$$a_0 = 2 \times 10^{-10}\ \mathrm{m/s^2}. \tag{406}$$

The dependence on the square root of the total baryon mass of the galaxy fits in the empirical Tully-Fisher law, which is $v^4 \propto M_B$. Also MOND is able to successfully explain the dynamics of galaxies outside clusters. Early in its history, it was appreciated that MOND was capable of explaining the observed dynamics of many spiral and elliptical galaxies. While MOND does reduce the need for additional mass in clusters, significant quantities of dark matter are still required [**26**].

Modified gravity. Modifications of general relativity are the theory of $f(R)$ gravity or its generalizations (see section 7.2.5). The former provides sufficient freedom to accommodate also dark matter. In particular, it allows good fits to the rotation curves in galaxies [**30**] which therefore might be explained as a curvature effect. In particular, for $f(R) \propto R^{3/2}$, the MOND acceleration regime is recovered. Also

$f(R)$ theories of gravity are able to explain the baryonic Tully-Fisher relation of gas-rich galaxies in a natural way. As discussed in section 7.2.5, modifications of general relativity might be also seen as effects of the quantum vacuum fields. However, there are effects of vacuum fields that cannot be seen as modifications of general relativity, for instance the effect studied in the next section.

7.5.2. Vacuum polarization as an alternative to dark matter. Another hypothesis proposed to explain the flat rotation curves of galaxies is gravitational vacuum polarization. In particular, Hajdukovic [**31**]–[**35**] assumed a gravitational repulsion between particles and antiparticles which would give rise to a gravitationally polarizable vacuum. The author attributes the repulsion to a negative gravitational mass (but positive inertial mass) of antiparticles. Then dipoles would exist in vacuum consisting of virtual particle-antiparticle pairs, producing a gravitational polarization similar to the electrical or magnetic polarization.

The most serious problem of Hajdukovic's approach is the assumption that antiparticles have negative gravitational mass but positive inertial mass. This assumption breaks the most cherished principle of general relativity, namely the equivalence principle, that requires identity between gravitational and inertial mass. Also no empirical evidence exists for the negative mass of antiparticles.

I propose two mechanisms that might give rise to a gravitational polarization of the vacuum without assuming Hajdukovic negative mass of antiparticles. Firstly, the Bose fields contribute positive energy to the vacuum, but Fermi fields contribute negative, as illustrated in section 7.3, see also [**8**].

If we assume that the vacuum fields are real, as supported all throughout this book, the observed null (or very small) value of the vacuum energy compels us to believe that the Bose and Fermi field energies plus the field interactions cancel on the average. Then it is plausible that an additional gravitational field, say from a galaxy, attracts Bose fields and repels Fermi fields, giving rise to a polarized vacuum. An attempt for a quantitative approach to this effect has been made [**36**] involving just the electromagnetic and the Dirac electron-positron fields. It failed because most of the calculated quantities are divergent, which may be an indication that the whole set of fields should be taken into account for any realistic quantitative

treatment (another example is the two-point correlation of the vacuum electromagnetic field alone that leads to the strongly divergent result eq. (359), see section 7.3.2 above). An alternative mechanism of gravitational vacuum polarization might be the modification of either the size or the distribution of the quantum vacuum fluctuations in the presence of gravity. In fact, according to our treatment in section 7.4, the quantum vacuum density $\rho(\mathbf{r},t)$ may be treated as a stochastic field whose fluctuations may have positive or negative energy with respect to the average value, that we should assume zero. Actually, both hypotheses, Bose-Fermi displacement and modified vacuum fluctuations, are not contradictory, whence both may contribute.

If we assume a gravitational vacuum polarization, via the mechanisms just mentioned, we may get results similar to those derived by Hajdukovic [**31**]–[**35**]. They are revisited in the following, with some elaboration. We will work within Newtonian gravity, which is a good enough approximation for galaxies. In order to have a quantitative approach we shall define a polarization vector $\vec{P}_g$ at every point in space. This vector makes sense as an average over regions that are large with respect to the typical sizes of fluctuations, of order atomic dimensions, although small compared with astrophysical objects. The quantity of interest will be the polarization mass density ρ_{DM}, related to the polarization vector $\vec{P}_g$ via

$$\rho_{DM} = -\nabla \cdot \vec{P}_g, \qquad \rho_{DM} = \frac{1}{r^2}\frac{d}{dr}\left(r^2 \left|\vec{P}_g\right|\right). \tag{407}$$

where ρ_{DM} would be a mass density alternative to 'dark matter'. The second eq. (407) corresponds to the particular instance of spherical symmetry. We must point out that polarization does not create net mass or energy, it would just effectively displace some mass to regions with great negative gravitational potential from other regions. In particular, to the neighbourhood of a galaxy from far away regions. We might assume that the displaced mass would contribute to either the dark energy effects or cancel with a small positive homogeneous mass density, a kind of 'dark matter'. I will not discuss this point further on.

As argued above, it is plausible that the polarization vector is a function of the gravitational field $\mathbf{g}$, that is,

$$\vec{P}_g = f(|\mathbf{g}|)\,\mathbf{g}.$$

Without a more detailed theory we cannot determine the function f, but it is plausible that it is a constant for low gravitational field, making the polarization proportional to the gravitational field. However, it should either not increase indefinitely, or the increase should be slowing down for large fields. Thus for large $|\mathbf{g}|$ we may assume that there is a saturation value of $\vec{P}_g$. In fact, a saturation is plausible due to resistance to depart from the equilibrium state. If the polarization derives from the change of vacuum fluctuations we may estimate the saturation value from our study in the previous section. Indeed, the parameter $K_{\rho\rho}$ in eq. (400), which characterizes the two-point correlation function, has dimensions of the square of a polarization vector. Thus, it is plausible to estimate the maximum polarization to be of order the square root of $K_{\rho\rho}$. That is,

$$P_{gmax} \sim \sqrt{K_{\rho\rho}} = \sqrt{\frac{\rho_{DE}}{3\pi G}}, \tag{408}$$

where ρ_{DE} is the dark energy density and we have taken eq. (401) into account. This gives for the associated critical field

$$g_{cr} = GP_{gmax} \sim \sqrt{\frac{G\,\rho_{DE}}{3\pi}}, \tag{409}$$

where G is the Newton constant. Hajdukovic has shown that the estimate eq.(409) roughly fits with observed properties in galaxies, which strongly suggests a connection between dark energy ρ_{DE} and dark matter ρ_{DM}. The former may be related with the pion mass via Zeldovic formula eq. (379), introducing appropriate numerical constants of order unity [**33**].

In fact, Hajdukovic proposed the critical field g_{cr} to be of order the fundamental acceleration a_s conjectured by MOND, see eq. (405), that is,

$$g_{cr} = a_s, \tag{410}$$

which is not far from the estimate eq. (409). Then, taking eqs. (402) and (408) into account, the saturation polarization P_{gmax} and the critical field g_{cr} may be written in terms of the (non-reduced) Compton wavelength λ associated to the pion mass. That is,

$$P_{gmax} = A\frac{\hbar}{\lambda^2 c^2} \quad \Rightarrow \quad g_{cr} = A\frac{G\hbar}{\lambda^2 c^2}, \tag{411}$$

where $A \simeq 0.29$. Eqs. (411) have been the starting point of Hajdukovic for the derivation of several results in fair agreement with

observations [**34**]. In the following we summarize the most relevant ones.

Let us consider a spherical body with mass M_b. We will study three spherical regions around it separated by two radii R_0 and R_1. In the first region $r < R_0$ the gravitational field is sufficiently strong for the polarization to achieve the maximum value eq. (411), which taking eqs. (407) into account leads to

$$\rho_{DM} \propto 1/r \quad \Rightarrow \quad \frac{dM\,(r)}{dr} \propto r, \tag{412}$$

that is, a constant gravitational field directed towards the body center. This result has been related to the Pioneer anomaly, a small, constant sunward deceleration of spacecrafts Pioneer 10 and Pioneer 11 not predicted by the standard Newtonian treatment of the solar system [**35**]. The anomalous acceleration of the Pioneer happens to be of the order the MOND acceleration eq. (406), but the agreement might be just an accident.

In the outer region $r > R_1$ the gravitational field will be weak, thus proportional to the field g, that is,

$$\left|\vec{P}_g\right| \propto g = \frac{GM\,(r)}{r^2}.$$

Taking the last eq. (407) into account, this is consistent with ρ_{DM} being small and the mass $M(r)$ roughly constant, increasing with r but slightly.

In the intermediate region $R_0 < r < R_1$, the dependence of $|\vec{P}_g|$ on r should pass from $|\vec{P}_g| \propto r^0$ to $|\vec{P}_g| \propto r^{-2}$. Thus I will study the intermediate case with the simple approximation

$$\left|\vec{P}_g\right| \propto 1/r \quad \Rightarrow \quad \rho_{DM} \propto 1/r^2 \quad \Rightarrow \quad M\,(r) = M\,(R_0) + \mu\,(r - R_0)\,,$$

where μ is some constant. Hence the gravitational field $g\,(r)$ would be

$$g\,(r) = \frac{G\,[M\,(R_0) - \mu R_0]}{r^2} + \frac{G\mu}{r}. \tag{413}$$

If the first term is small in comparison with the second one, the rotation velocity of bodies around a galaxy will be approximately flat, which is the most relevant observation to be explained by the dark matter hypothesis. Now if the first term is small we shall have

$$M\,(R_0) \simeq \mu R_0 \quad \Rightarrow \quad \frac{dM\,(r)}{dr} = \mu \simeq \frac{M\,(R_0)}{R_0},$$

where eq. (413) has been taken into account. If we identify $M(R_0)$ with the baryonic mass M_b and choose R_0 so that the acceleration at this radius equals the acceleration a_s of MOND we have, taking eqs. (409) and (411) into account,

$$\frac{dM_{dm}(r)}{dr} = \frac{B}{\lambda}\sqrt{mM_b}, \tag{414}$$

where M_b is the baryonic mass, the parameter $B \simeq 0.58$, and m and λ are the pion mass and wavelength, respectively. This reproduces the main features in the rotation curves of galaxies eq. (405) and it is compatible with solar system and terrestrial observations. In particular, if M_b is the mass of the Sun, the polarization corresponds to saturation up to large distances. That is, eq. (412) holds and the only effect predicted is a small sunwards constant force, as said above.

Many other results in fair agreement with observations have been derived by Hajdukovic that may be seen in the quoted articles [**31**]–[**35**]. In particular, this author has studied four benchmark measurements: the universality of the central surface density of galaxy dark matter haloes, the cored dark matter haloes in dwarf spheroidal galaxies, the non-existence of dark disks in spiral galaxies, and distribution of dark matter after collision of clusters of galaxies (the Bullet cluster is a famous example). Hajdukovic has claimed that only some of these phenomena can in principle be explained by dark matter and by theories of modified gravity, but the framework of the gravitational polarization of the quantum vacuum allows the understanding of the totality of them.

As a conclusion, the gravitational behaviour in galaxies and clusters might be explained from the assumption of a gravitational polarization of the vacuum. It is remarkable that the vacuum fluctuations are also able to explain phenomena attributed to dark energy. For unknown reasons the pion mass plays an important role in both cases, although it is plausible that it appears just as an appropriate average of the masses in the Standard Model of fundamental particles.

7.5.3. Summary: Dark energy and dark matter as quantum effects. The main empirical data of both dark energy (DE) and dark matter (DM) may be interpreted almost independently of any theory. The most relevant feature of DE is its cosmological character, that is, the fact that it appears everywhere in the universe with the same value. Thus we might attempt to explain its value as a combination of the universal constants. However, it is the case that the

unique combination leading to an energy density is Planck's, which is about 10^{123} times too big. This huge value strongly suggests that this quotient is a ratio between macro and microphysical quantities. Indeed, we know that the ratio between the electrostatic and the gravitational interaction energies of two charged fundamental particles is of order 10^{40}, one of the Dirac large numbers. This number is also of order $Gm^2/\hbar c$, whence it is straightforward to derive Zeldovich formula eq. (379), as shown in section 7.4.2. A relevant fact is that the mass m involved in the formula is of order the pion mass.

The best known effects of dark matter are the rotation curves in the haloes of galaxies and clusters. The most simple explanation of this fact is the existence of some non-baryonic mass fulfilling the condition that the derivative, $dM_{DM}(r)/dr$, with respect to the radial coordinate is roughly a constant. On the other hand the squared rotation velocity in different galaxies is known to be proportional to the square root of their baryonic mass M_B, which is the empirical Tully-Fisher law. Therefore, in natural units $c = \hbar = 1$ we should write

$$\frac{dM_{DM}}{dr} \sim m^{3/2}\sqrt{M_B}, \tag{415}$$

where m is a parameter required by dimensional considerations, that may be chosen a mass without loss of generality. With this choice the equation, written in usual units, becomes eq. (414) and the mass parameter m should be of the order of the pion mass to match observations. Eq.(415) fits in MOND (405) with $m^{3/2}$ substituted for $\sqrt{Ga_0}$.

I believe that it is not an accident that the interpretation of both DE and DM via eqs. (379) and (414), respectively, requires the same mass parameter. This suggests that both DE and DM should derive from similar mechanisms. Eq. (379) makes almost compelling to attribute dark energy to the quantum vacuum fluctuations, as proposed by Zeldovich long ago [**22**], and our section 7.4.5 makes this still more plausible. Then the relation of DM with DE, via the same (pion) mass parameter, suggests to attribute the dark matter effects also to vacuum fluctuations. Certainly, this hypothesis for DM is less compelling than the assumption for DE and it presents some problems. In particular, the approach of section 7.5.2 rests on the assumption that vacuum fluctuations may lead to a vacuum polarization, which may explain the effects at the galaxy scale, e.g. the flat rotation curves. However, gravitational vacuum polarization means that some mass is

transferred from distant regions to the neighbourhood of the galaxies. It seems implausible that this mechanism could solve the cosmological problem of the large amount of non-baryonic cold matter needed in the universe, at the recombination time, in order to explain the formation of structures. We conclude that further work is worthwhile.

7.6. Compact stars

7.6.1. From ordinary stars to black holes. Ordinary stars, like the Sun, appear in the evolution of clouds of hydrogen with small amounts of helium, called protostars. By the action of gravity the clouds contract, increasing their temperature and pressure until nuclear reactions start, mainly the fusion of hydrogen, provided that the protostar mass is large enough. As a consequence the gas becomes a plasma of protons, helium nuclei and electrons. Temperature and pressure increase again but the contraction stops, giving rise to an almost stationary state that may last for a hundred million years. When a large fraction of hydrogen is burn out the evolution is complicated but the star may end as a white dwarf, a compact star with a radius similar to the Earth whose pressure comes mainly from the electrons behaving as a Fermi gas (i.e. the pressure is a consequence of the Pauli principle). After cooling by radiation, the white dwarf may contract again to a neutron star, a compact object with radius of a few kilometers. Finally, neutron stars with mass greater than about two solar masses are assumed to collapse again producing black holes. Also in the center of most (or all) galaxies there were initially supermassive stars with mass of order equal to 10^6 to 10^{10} solar masses, that after cooling are assumed to collapse giving supermassive black holes. Details about the formation, evolution and death of stars may be seen in astrophysics books. Here I will treat mainly compact stars and their eventual collapse [**37**], [**38**].

7.6.2. Singularities, a stain for Einstein. The actual collapse to black holes, presenting a spacetime singularity, is still questioned by some people. In order to understand why, it is illustrative to revisit the historical evolution of the ideas about star collapse [**39**], summarized in the following. In 1916, soon after Einstein presented general relativity (GR in the following), Karl Schwartzschild calculated terrestrial gravity with the Earth modelled as a sphere of homogeneous density. He also found that for spheres with homogeneous density no stationary regular solution of the Einstein equation exists

if the radius and mass fulfilled the inequality (with units $c = 1$)

$$R \leq 2GM, \tag{416}$$

(actually the constraint is stronger in stars with homogeneous density, see section 7.6.3). In fact, the solution regular at infinity presents a *Schwartzschild singularity* at the star center. At that moment no astrophysical object was known fulfilling the inequality and the result remained purely academic. For the Earth, the left side of eq. (416) is about 10^9 times the right side.

The actual possibility of stellar collapse arose in the early 1930's, the initial impulse coming when Chandraskhar, then a young student, proved that white dwarves are not stable if their mass surpasses some limit. The possibility of collapse to a Schwartzschild singularity was strongly dismissed by Arthur Eddington, then a respected astrophysicists, whose belief may be summarized as follows: *Some unknown physical mechanism should exist preventing collapse towards a singularity.*

The reason for that belief was that GR rests upon the idea that spacetime is a continuum whose curvature is governed by differential equations. Therefore, a singularity is a kind of inconsistency or at least a stain in GR, "the most beautiful theory of physics" according to Lev Landau. Surely Einstein supported Eddington's view when he published an article in 1939 [**40**] claiming to prove that collapse is impossible due to the relativistic limit on the velocity of particles (that is, the speed of light). Einstein studied a model of star consisting of non-interacting particles moving in circles all with the same center, and such that the ensemble of particles contributes a mass density possessing spherical symmetry. The stars in this model may reach equilibrium with different mass distributions without in any case arriving at the instability condition eq. (416).

What Chandrashekar calculated around 1930 was that, in the process of cooling, the radius of a white dwarf decreases and if the star is massive enough the pressure becomes unable to support the gravitational force and it collapses. However, it was later shown that the strong pressure favours a nuclear reaction inverse to neutron β decay, that is, protons and electrons combine to produce neutrons and neutrinos, the latter leaving the star. Thus, massive white dwarves indeed collapse, but giving rise to a neutron star possibly without singularity. Such collapses have been observed as supernovas, processes

emitting so huge amounts of energy that may be observed even in stars belonging to distant galaxies.

The next step was to see whether neutron stars also would collapse to a singularity. A few weeks after Einstein star model appeared in press [**15**], Robert Oppenheimer published a relevant article on the subject in collaboration with G. M. Volkoff [**41**]. They calculated the structure of stars consisting of free neutrons and showed that no stable star would exist with mass greater than about one solar mass. In a subsequent paper, in collaboration with H. Snider [**42**] a calculation was reported about the collapse of a cloud of free point particles. This is today seen as an idealized model of collapse to a Schwartzschild singularity.

Oppenheimer and Volkoff (OV) started with the following equation of hydrostatic equilibrium for spherical symmetry:

(417)
$$\frac{dP}{dr} = -g\,(\rho + P)\,, \quad g = \frac{G\left(m + 4\pi r^3 P\right)}{r^2 - 2Gmr}, \quad m \equiv \int_0^r 4\pi r^2 \rho dr.$$

Here $\rho\,(r)$ is the mass density and $P(r)$ the pressure of matter in the star, both related by an equation of state (EoS) that gives P as a function of ρ. The quantity g plays the role of the gravitational field intensity in Newtonian gravity. Indeed, for $P \ll \rho$ and $2Gm \ll r$ the GR eq. (417) becomes Newton's.

The seeming contradiction between the conclusions of Einstein and OV may be understood as follows. In spherical symmetry the stress tensor at any point might not be isotropic but possess two principal pressures, radial and transversal, that I will label p and q respectively. When $p \neq q$ the GR equilibrium equation is no longer eq. (417) but the following

(418)
$$\frac{dp}{dr} + 2\frac{p-q}{r} = -g\,(\rho + p)\,, \qquad g \equiv \frac{G\left(m + 4\pi r^3 p\right)}{r^2 - 2Gmr}.$$

In (special) relativity the pressure on a surface perpendicular to the Z axis corresponds to

(419)
$$P_z = \frac{1}{\Delta V}\sum_j \frac{p_{zj}^2}{E_j},$$

where p_{zj}^2 is the Z component of the linear momentum of the particle j, and E_j its energy, the sum extended to all particles present in the volume ΔV at a given time. In Einstein's model the particle motions

have no radial component, whence there is no radial pressure p, only transverse q. Therefore, the equilibrium eq. (418) becomes

$$p = \frac{1}{2}\frac{Gm}{r-2Gm}\rho, \qquad m \equiv \int_0^r 4\pi r^2 \rho dr,$$

whence the condition that the pressure cannot surpass the density, which is easily derived from eq. (419), leads to

$$r \geq \frac{5}{2}Gm.$$

This puts a lower limit to the radius that a star in equilibrium may possess for a given mass, but collapse leading to a singularity is not possible, see eq. (416).

Einstein was actually flawed when he claimed that his proof for the impossibility of collapse should be valid for any star. The argument is correct for his model, but the conclusion cannot be extrapolated to all stars, in particular those with local isotropy $p = q$. Furthermore, it is plausible that local equilibrium at every point in the interior of the star would led to local isotropy, except if there existed a mechanism not yet known that could produce local anisotropy. Thus, although equilibrium solutions with local anisotropy (fulfilling eq. (418)) exist, Einstein model being an example, those stars are not locally stable because matter in local equilibrium is plausibly isotropic.

In the 1930's there was an important development in nuclear physics, that had been initiated after the study of radioactivity by Becquerel, and Pierre and Marie Curie. This development and the beginning of World War II made the interest of physicists turn to nuclear physics in the 1940's. E.g., R. Oppenheimer became the scientific leader of the Manhattan Project aimed at the production of nuclear weapons. Interest in star collapse arose again in the 1950's after it was discovered that there is a strong repulsion between neutrons at short distances (usually named a 'hard core'), so that realistic equations of state for interacting neutrons could support far greater pressure than free neutrons. However, when J. A. Wheeler and other people studied in detail equilibrium of stars using realistic neutron equations of state, the conclusion was that an upper mass limit does exist, although greater than OV's, at about 2 solar masses [**2**]. Wheeler claimed that as no alternative stopping mechanism had been yet found then probably none existed. Therefore, he concluded that collapse of massive enough neutron stars should be unavoidable. Since cooling these stars

would end their lives as 'black holes', a name that he proposed for collapsed objects possessing a Schwartzschild singularity. After that time the existence of black holes has become a paradigm in theoretical astrophysics [**38**]. Other stars different from neutron's are assumed to collapse towards black holes, the limiting mass depending on the equation of state of matter. In particular, for an electron-proton plasma at a finite temperature it is several million solar masses, so that supermassive stars would collapse when they cool down.

Even in Newtonian gravity a gas of free particles would collapse to a point if it cools down indefinitely. However, this possibility remains of purely academic interest because no particles strictly noninteracting exist and a pressure would arise preventing collapse to a point. In contrast, in GR the pressure may be unable to stop collapse. The argument is that what prevents collapse of bodies like the Earth or the Sun is the pressure that balances the gravitational attraction, but in GR the pressure contributes also to gravity, whence in massive enough bodies, where the pressure may be very large, nothing could stop collapse. More quantitatively, the GR equilibrium eq. (417) contains two terms that, for given mass and radius, increase with increasing pressure; namely $\left(m + 4\pi r^3 P\right)$ and $(\rho + P)$, both contributing to the contraction of the star. However, this is not the case if the pressure is not isotropic, as Einstein model shows. Therefore, a mechanism stopping collapse might exists involving anisotropic pressure. Also, the observations could barely distinguish a black hole from a star having $2GM/R$ very close to unity [**43**], which justifies the search for collapse not leading to a singularity.

In spite of my opinion against the actual existence of black holes (BH) I remind some features about them in the following. When $R \to 2GM$ the effective field intensity g goes to infinity producing a number of peculiar, I would say even bizarre, phenomena. Firstly, for a body of mass M when its radius R approaches $2GM$ the gravitational attraction, g eq. (418), diverges so that nothing could escape from the body, not even light, hence the name 'black'. (However, BH are very luminous objects because particles in the surroundings are accelerated by the strong gravity, with emission of radiation). After all matter of the star has crossed inwards the surface defined by

$$R = 2GM, \tag{420}$$

named event horizon, no particles may escape. In crossing the surface some elements of the metric tensor, determining the curvature of

spacetime, change sign and the radial (temporal) coordinate becomes temporal (radial). An observer falling in the BH would arrive at the surface in a finite time, but due to the relativity of time in GR, for an external observer that time would be infinite. Hence we, external observers, should not say that *there are* BH, but that *there will be*. That is, for a collapsing star the information that the process has finished would require an infinite time to reach us. Eventually, all matter would arrive at a point, the center of the sphere, which is the Schwartzschild singularity. It is actually assumed that due to quantum effects BH emit radiation (Hawking radiation) so that BH are not strictly black, but dark. As a consequence, BH would evaporate in times that are short for small BH but greater than the age of the universe for massive ones.

Since about 1965 it is a common view that black holes exist. Indeed, it is assumed that there are, or have been, from small BH created in the early universe, to BH with a few solar masses or supermassive BH with million or billion solar masses. Hundreds of articles and many books have been devoted to the theoretical study of BH. And there are also many observations of phenomena attributed to BH, for instance quasar luminosity [**38**] or gravitational waves, recently observed, that are assumed to be created in the collision of two BH.

7.6.3. Attempts to stop collapse to a singularity.

Local anisotropy. The belief in the existence of BH rests upon observational facts, but it also *requires a theoretical argument,* namely that no mechanism exists able to stop the collapse of massive neutron stars or supermassive stars. Of course, the argument is questionable from an epistemological point of view because the future discoveries in physical research are not known. Also, the observable predictions of phenomena near, but outside, a BH would not be too different from those taking place near compact objects with a very large but finite gravitational field intensity $|g|$. In view of these facts and the unpleasant presence of a mathematical singularity in BH, the search for mechanisms that could prevent star collapse makes sense. In the following I comment on some of these efforts.

The possibility that collapse is prevented by local anisotropy goes back to Einstein's model mentioned above [**40**]. Local anisotropy [**44**] has been applied to the study of compact stars in order to estimate its possible influence on equilibrium. White dwarves with local anisotropy have been studied and configurations of equilibrium with

mass above the Chandrasekhar limit have been found [**45**], as well as supermassive stars [**46**] and neutron stars [**47**], [**48**]. In all cases, however, spontaneous local anisotropy does not prevent collapse. In fact, it has been shown that if there exists a locally isotropic stable star configuration with mass M and baryon number N, then all locally anisotropic equilibrium configurations having the same N, in the neighbourhood of that configuration, possess a mass not smaller than M. Therefore, it appears that local anisotropy might solve the problem only if there is some mechanism that induces it. This might be the case by the influence of the quantum vacuum or due to modifications of general relativity known as 'extended gravity theories'. Actually, both vacuum effects and modifications of gravity theory may be equivalent in practice as discussed above in section 7.2.5.

Modified gravity. Many studies applying extended gravity theories (see subsection 7.2.5) to star equilibrium have been made [**6**], [**51**], [**52**]. Most of them correspond to $f(R)$ gravity, where the function $f(R)$ of the Ricci scalar R is added to R in the Hilbert action, which gives rise to a modification of general relativity. The conclusion of such studies is that, although they tend to stabilize the stars, they do not prevent collapse.

A more drastic modification is to substitute for the Ricci scalar the most general scalar that may be obtained from the Riemann (curvature) tensor [**52**], namely an arbitrary function of the scalars R and $\sum_{\mu\nu} R_{\mu\nu}R^{\mu\nu}$. An example will be reviewed in section 7.6.3.

Gravastars. An alternative to black holes, rather than a mechanism preventing collapse, is the gravastar [**49**], a hypothetical object consisting of a dark energy condensate surrounded by a strongly correlated thin shell of anisotropic matter. More specifically, gravastars have 3 regions with the following equations of state:

$$\begin{aligned} \text{interior:} &\quad 0 \le r < r_1, \quad \rho = -P, \\ \text{shell:} &\quad r_1 < r < r_2, \quad \rho = +P, \\ \text{exterior:} &\quad r_2 < r, \quad \rho = P = 0. \end{aligned}$$

For the exterior region the metric is Schwartzschild. The matching is made assuming continuity of the metric coefficients at r_1 and r_2. A large negative pressure in the internal region prevents collapse as may be seen in the equilibrium hydrostatic eq. (418). Indeed, there we have

$$m + 4\pi r^3 P \simeq \frac{4\pi}{3} r^3 \rho + 4\pi r^3 P = \frac{4\pi}{3} r^3 \left(\rho + 3P\right) = -\frac{8\pi}{3} r^3 \rho < 0,$$

whence the effective gravitational field would be repulsive. The model has been studied by several authors, in particular the comparison of gravastars with black holes [**50**].

7.6.4. Change of vacuum equation of state in strongly curved space. In the following I propose a possible mechanism able to avoid collapse, resting on the effect of the quantum vacuum fields. As in sections 7.4 and 7.5 we assume that the vacuum field properties may be different in diverse conditions. Indeed, in section 7.4 the effect was studied at cosmic scale with the result that Minkowski space remains a good approximation for scales as large as galactic. In section 7.5 we considered spacetime with a curvature so small that it may be studied within Newtonian gravity. Now I consider the possibility of a big change of the vacuum stress-energy when the curvature is very large, as may be the case in massive compact objects. I propose that in these conditions the equation of state of the vacuum might depart dramatically from the standard one, $\rho = -P$, appropriate for Minkowski space. In fact, it is common opinion that in the collapse to black hole the ratio GM/R is achieved, which implies that the effective gravitational field intensity $g \sim GM/(R^2 - 2GMR)$ may reach values indefinitely large. In particular, the effective gravitational field g, eq. (418) might become as large as the inverse of the atomic, or even nuclear, radius (in natural units $c = 1$). In my opinion it is implausible that the quantum vacuum properties are only slightly modified in fields so strong. I believe that very big changes should occur so that the equation of state is far from the free space one $P = -\rho$.

A simple model. In the following I present a simple model that illustrates how those changes might prevent collapse to black hole. We will recall a more elaborate model that presents a similar behaviour in section 7.6.4. In the simple model I assume that in a curved space with spherical symmetry the radial, p_{vac}, and the transverse, q_{vac}, stresses (or pressures) of the vacuum fields may be different from each other. In this case a more general equation of state that maintains the mean pressure opposite to the density should fulfil

$$\rho_{vac} = -\frac{1}{3}p_{vac} - \frac{2}{3}q_{vac}. \tag{421}$$

In order to have an equation of state another relation is needed and the assumption is made that the ratio p_{vac}/q_{vac} is related to the increase of lengths in the radial direction, that is, to the metric coefficient

$1/\sqrt{1-2m/r}$. In particular, I propose the following relation

$$q_{vac} = p_{vac}\sqrt{1-2Gm/r}. \tag{422}$$

We may remove q_{vac} amongst eqs. (421) and (422) and we get

$$\rho_{vac} = -\frac{1}{3}p_{vac}\left(1+2\sqrt{1-2Gm/r}\right), \tag{423}$$

this being the equation of state to be used in the following.

In the study of equilibrium of the spherical body (that I will name star in the following) we shall assume that the quantum vacuum and the matter behave as two independent fluids. Then the hydrostatic eq. (418) should hold for each fluid, that is,

$$\frac{dp_j}{dr} + 2\frac{p_j - q_j}{r} = -g\left(\rho_j + p_j\right).$$

In our case we have the vacuum for $j = 1$ and matter for $j = 2$, whence

$$\frac{d\ln|p_{vac}|}{dr} = -2\left(\frac{g}{3}+\frac{1}{r}\right)\left(1-\sqrt{1-2Gm/r}\right),$$

$$\frac{dp_{mat}}{dr} = -g\left(\rho_{mat}+p_{mat}\right), \tag{424}$$

where we have taken eqs. (422) and (423) into account in the first, vacuum, equation and assumed that the pressure is isotropic in the last, matter, equation. The quantities m and g depend on both vacuum and matter, that is,

$$m = \int_0^r 4\pi r^2\left(\rho_{mat}+\rho_{vac}\right)dr,$$

$$g = \frac{G\left[m+4\pi r^3\left(p_{vac}+p_{mat}\right)\right]}{r^2-2Gmr}. \tag{425}$$

In our model of compact stars I will also assume that the equation of state of matter fulfils $\rho_{mat} > p_{mat}$. In the following I will make a qualitative study of eqs. (424) and (425).

We are interested in knowing whether there are equilibrium configurations for stars having a definite surface at a radius R and fulfilling the following conditions

$$2Gm/R = 1-\varepsilon, \qquad 0 < \varepsilon \ll 1. \tag{426}$$

In order that the external region of the star is similar to that of a black hole I assume that ε is very small.

In the *external region* $r > R$ we have $\rho_{mat} = p_{mat} = 0$ and, for large values of r, $2Gm \ll r$. Then the first eq. (424) becomes in the asymptotic region

$$\frac{d \ln |p_{vac}|}{dr} \sim 0 \quad \Rightarrow \quad p_{vac} \sim -K,$$

and the integration constant K should be identified with the standard vacuum energy density (the 'dark energy'). That is, eqs. (422) and (423) go to the standard vacuum equation of state $\rho_{vac} = -p_{vac} = -q_{vac}$.

Als,o in both the external region and the part of the internal region not too close to the star center, eq. (425) leads to $g > 0$, whence $|p_{vac}|$ increases inwards. For large enough $|p_{vac}|$ the effective gravitational field g will change sign, say at $r = r_0$. Therefore, in the region $r < r_0$ gravitation would be repulsive, thus stoping collapse. However, $|p_{vac}|$ cannot increase indefinitely because for very large $|p_{vac}|$ the quantity g will become greater than $1/3r$; whence, taking eq. (424) into account, $|p_{vac}|$ would start decreasing inwards. Therefore, the vacuum radial pressure and density (see eq. (423)) will remain always finite, and the same will hold for the metric coefficients.

Solving the coupled first-order differential eqs. (424), taking eqs. (425) into account, may allow the study of particular stars in equilibrium once the equation of state of matter is known. In principle, we expect nonsingular solutions having the asymptotic behaviour for any central matter density $\rho_{mat}(0)$. Every choice of $\rho_{mat}(0)$ would require an appropriate value of the central vacuum pressure $p_{vac}(0)$, in order to get the asymptotic behaviour for large $r \gg R$. The solution of the differential equations will not be studied here in more detail.

The relation between baryon number and mass. An interesting feature of any compact star, in particular our model, is the nontrivial relation between baryon number N and mass M of the star. In a Newtonian star, that is, when Newtonian gravity is a good approximation, M is roughly proportional to N. For instance, in ordinary stars like the Sun the ratio $M/N = \mu$ is the mass of a hydrogen atom (actually slightly larger due to the small amount of helium). We should take into account also the interaction energy, the thermal energy and the gravitational energy, but the contribution of these effects is less than 0.1%. In relativistic stars, however, the gravitational energy may be quite large and the said approximation fails. The baryon number should be calculated taking into account the volume element that,

using the standard metric for spherical symmetry eq. (393), reads

$$dV = 4\pi r^2 g_{rr} dr = 4\pi r^2 \left(1 - 2Gm/r\right)^{-1} dr.$$

Hence the baryon number is given by

$$N = \int_0^R \frac{n(r)}{1 - 2Gm/r} 4\pi r^2 dr. \tag{427}$$

where $n(r)$ is the baryon number density. The mass however is not given in a similar manner, but it is

$$\begin{aligned} M &= \int_0^R \rho(r) 4\pi r^2 dr \\ &= \int_0^R \rho(r) \left[\frac{1}{1 - 2Gm/r} - \frac{2Gm/r}{1 - 2Gm/r}\right] 4\pi r^2 dr, \end{aligned} \tag{428}$$

where $\rho(r)$ is the mass density and the second term may be interpreted as the (negative) gravitational energy. The point is that, in contrast with Newtonian stars where $2Gm/r \ll 1$, in our model we assume that $2Gm(r)/r$ may be close to unity. Hence, taking eqs. (428) and (427) into account, we may have

$$N \gg M/\mu. \tag{429}$$

That is, for a family of stars with increasing baryon number N it may be that the mass M does not always increase; it may decrease for large N. This suggests that there might be an upper mass limit for stars, with no limit for the baryon number.

If there was a mechanism preventing collapse to black hole that would allow equilibrium stars with the ratio $2GM/R$ very close to unity, then a hot very massive neutron star should have a decreasing mass when it cools, but never developing a Schwartzschild singularity. Rather, the mass would decrease by radiation of energy without any decrease of the baryon number until it arrives at a cold state of equilibrium with $M < N\mu$. Therefore, the observed lack of cold neutron stars with large mass might not be a confirmation of the Oppenheimer-Volkoff limit, see section 7.6.1, but a consequence of the inequality eq.(429).

As a simple example, let us consider a star with both homogeneous mass density σ and homogeneous mass per baryon μ. Taking eqs. (428) and (427) into account the total mass and baryon number

will be, respectively,

$$\begin{aligned} M =& \frac{4}{3}\pi\sigma R^3, \\ N =& \frac{4\pi\sigma}{\mu}\int_0^R \frac{r^2 dr}{\sqrt{1-8\pi G\sigma r^2/3}} = \frac{4\pi\sigma R^3}{\mu\alpha^3}\int_0^\alpha \frac{x^2 dx}{\sqrt{1-x^2}} \\ =& \frac{2\pi\sigma R^3}{\alpha^3\mu}\left[\arcsin(\alpha) - \alpha\sqrt{1-\alpha^2}\right], \\ \alpha^2 =& 8\pi G R^2\sigma/3 = \frac{2GM}{R}. \end{aligned}$$

For $2GM/R \equiv \alpha^2 \ll 1$ we have $M/N \simeq \mu$, but for greater values of α the ratio M/N decreases, and for $2GM/R \to 1$ we get $M/N \to 4\mu/3\pi$. Actually, for stars with homogeneous density and local isotropy (i.e. fulfilling eq. (417)) there is the constraint $2GM/R < 8/9$ [**1**]. However, if we allow for stars with local anisotropy (i.e. radial pressure different from tangential pressure fulfilling eq. (418)) the constraint would be $2GM/R < 1$.

A similar behaviour appears in the calculation that we report in the following.

A more realistic model. The main objection to the simple model presented above is that our starting eq. (423) has no fundamental motivation. It might be seen as an *ad hoc* assumption with the aim of proving that collapse might be stopped. A better way to derive a vacuum effect is to use an extended gravity theory, which is equivalent as discussed in section 7.2.5.

In the following I briefly review a calculation using the quadratic Lagrangian,

$$f = aR^2 + b\sum_{\mu\nu} R_{\mu\nu}R^{\mu\nu} \tag{430}$$

(see eq. (342) in section 7.2.4), that does prevent collapse of neutron stars. The differential equations of the model are involved and the solution has been obtained numerically [**52**]. The result of the calculation is that the star mass increases with increasing central density until about 1 solar mass, and then it decreases. However, the baryon number increases monotonically.

The shortcoming of the model is that it requires values of the parameters a and b that are too large to be compatible with terrestrial and solar system observations. However, both preventing collapse and fitting in terrestrial observations would be possible if the actual

Lagrangian is approximately quadratic at the high densities of neutron stars, but very small at the low densities of terrestrial observations. This would be the case for instance if the following Lagrangian is substituted for eq. (430):

$$F = f - c\ln\left(1 + f/c\right), \tag{431}$$

where c is a constant such that f/c is dimensionless. For an appropriate choice of the parameters a, b and c the Lagrangian F fulfils the desired properties. Indeed, for low density, i.e. $f/c \ll 1$, expanding the logarithm in powers of f/c and retaining the lowest nonzero term we get

$$F \simeq f - c\left(\frac{f}{c} - \frac{f^2}{2c^2}\right) = \frac{f^2}{2c} \ll f,$$

while for large density $f/c \gg 1$, the logarithmic term may be neglected:

$$F = f - c\ln(1 + f/c) \simeq f - c\ln(f/c) \simeq f.$$

This makes the model somewhat contrived, but the result shows that there are possibilities of preventing collapse to black hole in extended gravity theories.

These arguments suggest that, if we assume that the vacuum becomes locally anisotropic, with a radial pressure different from the transverse pressure as in eq. (422), then star collapse may be stopped, contrary to the standard view.

Bibliography

[1] S. Weinberg: *Gravitation and Cosmology*. John Wiley and Sons, New York, 1972.

[2] C. Misner, K. Thorne, J. Wheeler: *Gravitation*. Freeman, San Francisco. 1973.

[3] E. Santos: On a heuristic point of view concerning the motion of matter. From random metric to Schrödinger equation. *Phys. Lett. A* **352**, 49-54 (2006).

[4] A. A. Logunov, M. A. Mestvirishvili: Relativistic theory of gravitation. *Found. Phys.* **16**, 1-26 (1986).

[5] A. A. Logunov: *The theory of gravity* (transl. by G. Pontecorvo). Nauka, Moscow, 2001.

[6] S. Capozziello, M. De Laurentis: Extended theories of gravity. *Phys. Rep.* **509**, 167 (2011).

[7] J. Martin: Everything you always wanted to know about the cosmological constant problem (but were afraid to ask). *Comptes Rendus Physique* **13**, 566 (2012).

[8] G. Ossola and A. Sirlin: Considerations concerning the contributions of fundamental particles to the vacuum energy density. *Eur. Phys. J.* **C31**, 165 (2003)

[9] J. J. Sakurai: *Advanced quantum mechanics*. Addison-Wesley, Reading (Massachusetts), 1967.
[10] C. L. Bennett et al.: Nine-year Wilkinson microwave anisotropy observation (WMAP). Final maps and results. *Astrophys. J. Suppl. Ser.* **192**, 17 (2011).
[11] J. Rich: *Fundamentals of Cosmology*. Springer-Verlag, Berlin, 2001.
[12] T. P. Sotiriou, V. Faraoni: f(R) theory of gravity. *Rev. Mod. Phys.* **82**, 451 (2010).
[13] E. Santos: Quantum vacuum effects as generalized f(R) gravity. Application to stars. *Phys. Rev.* **D 81**, 064030 (2010)
[14] E. J. Copeland, M. Sami, S. Tsujikawa: Dynamics of dark energy, *Int. J. Mod. Phys. D* **15**, 1753–1936 (2006).
[15] A. Pais: *"Subtle is the Lord..."*. *The science and life of Albert Einstein*. Oxford University Press, 1984.
[16] P. J. E. Peebles, Bharat Ratra: The cosmological constant and dark energy. *Rev. Mod. Phys.* **75**, 559-606 (2003).
[17] Varun Sahni: Dark matter and dark energy. In *The physics of the early universe. Proceedings, 2nd Aegean School, Ermoupolis, Greece, September 22-30, 2003*. Lect. Notes Phys. **653**, 141-180 (2004).
[18] S. Weinberg: The cosmological constant problem. *Rev. Mod. Phys.* **61**,1-23 (1989)
[19] Planck Collaboration, P. A. R. Ade, N. Aghanim, et al.: Planck 2015 results - XIII. Cosmological parameters. *Astronomy & Astrophys.* **594**, A13 (2016).
[20] T. Padmanabhan: Cosmological constant. The weight of the vacuum. *Phys. Rept.* **380**, 235 (2003)
[21] Quingdi Wang, Zhen Zhu, William G. Unruh: How the huge energy of the quantum vacuum gravitates to drive the slow accelerating expansion of the universe. *Phys. Rev.* **D95**, 103504 (2017).
[22] B.Y. Zeldovich: Vacuum theory: a possible solution to the singularity problem of cosmology. *Sov. Phys. Usp.* **24**(3), 216 (1981).
[23] E. Santos: Quantum vacuum fluctuations and dark energy. *Astrophys. Space Sci.* **326**, 7-10 (2010)
[24] E. Santos: Space-time curvature due to quantum vacuum fluctuations. An alternative to dark energy? *Phys. Lett.* **A 374**, 709 (2010).
[25] E. Santos: Dark energy as a curvature of space-time induced by quantum vacuum fluctuations. *Astrophys. Space Sci.* **332**, 423-435 (2011)
[26] G. Bertone, D. Hooper: A history of dark matter. *Rev. Mod. Phys.* **90**, 45002 (2018).
[27] G. Bertone, T.M.P. Tait: A new era in the quest for dark matter. *Nature* **562**, 51 (2018).
[28] M. Milgrom: A modification of Newtonian dynamics as a possible alternative to the hidden mass hypothesis. *Astrophys. J.* **270**, 365 (1983).
[29] B. Famaey, S. S. McGaugh: Modified Newtonian dynamics(MOND). *Living Rev. Relativity* **15**,10 (2012).
[30] S. Capozziello, P. Jovanović, V. Borka Jovanović, D. Borka: Addressing the missing matter problem through a new fundamental gravitional radius. *J. Cosmol. Astropart. Phys.* **1706**, 044 (2017).
[31] D. S. Hajdukovic: Quantum vacuum and dark matter. *Astrophys. Space Sci.* **337**, 9-14 (2012).

[32] D. S. Hajdukovic: On the gravitational field of a point-like body immersed in quantum vacuum. *MNRAS* **491**, 4816 (2020).
[33] D. S. Hajdukovic: A few provoking relations between dark energy, dark matter and pions. *Astrophys. Space Sci.* **326**, 3-5 (2010).
[34] D. S. Hajdukovic: Is dark matter an illusion created by the gravitational polarization of the quantum vacuum? *Astrophys. Space Sci.* **334**, 215-218 (2011).
[35] D. S. Hajdukovic: On the vacuum fluctuations, Pioneer anomaly and modified Newtonian dynamics. *Astrophys. Space Sci.* **330**, 207-209 (2010).
[36] E. Santos: Dark matter as an effect of the quantum vacuum. *Astrophys. Space Sci.* **363**, 74 (2018).
[37] S. L. Shapiro, S. A. Teukolsky: *Black holes, white dwarfs and neutron stars: The physics of compact objects.* John Wiley, New York, 1983.
[38] M. Camenzind: *Compact objects in astrophysics: White dwarfs, neutron stars and black holes.* Springer-Verlag, Berlin, 2007.
[39] K. S. Thorne: *Black holes and time warps. Einstein's strangeous legacy.* Norton and Co. New York, 1994.
[40] A. Einstein: On a stationary system with spherical symmetry consisting of many gravitating masses *Ann. Math.* **40**, 922-936 (1939).
[41] J. R. Oppenheimer and G. M. Volkoff: On massive neutron cores. *Phys. Rev.* **55**, 374 (1939).
[42] J. R. Oppenheimer and H. Snider: On continued gravitational contraction. *Phys. Rev.* **56**, 445 (1939).
[43] M. A. Abramowicz, W. Kuzniak, J. P. Lasota: No observational proof of the black-hole event-horizon, *Astronomy Astrophys.* **396**, L32 (2002).
[44] L. Herrera, L. O. Santos: Local anisotropy in self-gravitating systems, *Phys. Rep.* **286**, 53-130 (1997).
[45] E. Santos-Corchero: White darfs with anisotropic pressure in post-Newtonian gravity. *Astrophys. Space Sci.* **259**, 32 (1998).
[46] E. Santos-Corchero: Equilibrium of spheres with local anisotropy in post-Newtonian gravity. Aplication to supermasive stars. *Class. Quantum Grav.* **15**, 3645-3654 (1998).
[47] E. Santos-Corchero: Quantum approach to neutron stars leading to configurations with local anisotropy and mass above the Oppenheimer- Volkoff limit. *Astrophys. Space Sci.* **275**, 259-274 (2001).
[48] E. Santos-Corchero: Quantum ground state of stars consisting of noninteracting fermions. *Class. Quantum Grav.* **19**, 417-427 (2002).
[49] P. Mazur, E. Mottola: Gravitational condensate stars. An alternative to black holes. *Proc. Nat. Acad. Sci. USA* **101**, 9545 (2004).
[50] T. Harko, Z. Kovács, F. S. N. Lobo: Can accretion disk properties distinguish gravastars from black holes? *Class. Quant. Grav.* **26**: 215006 (2009).
[51] E. Santos: On the stability of slightly relativistic stars in f(R) gravity. *Phys. Letters A*, **376**, 2321 (2012).
[52] E. Santos: Neutron stars in generalized $f(R)$ gravity. *Astrophys. Space Sci.* **341**, 411-416 (2012).